THE UNIVERSITY
of EDINBURGH

Tumour associated macrophages in Diffuse Large B cell lymphoma

Tamasin Naomi Doig

For the degree of Doctor of Philosophy

University of Edinburgh

2015

Abstract

Tumour associated macrophages (TAMs) have been associated with prognosis in a wide variety of tumours with most studies showing a high number of macrophages equating with poor prognosis. This is postulated to be due to TAMs providing support to the tumour through a wide variety of mechanisms including suppression of the immune response, promotion of angiogenesis and provision of growth supporting signals. Previous work within the group has characterised some of the mechanisms by which Burkitt lymphoma cells attract macrophages to the tumour and some of the mechanisms by which these macrophages support tumour cell growth. This thesis extends some of the work carried out in Burkitt lymphoma to Diffuse Large B cell lymphoma (DLBCL) and examines TAMs in this tumour type.

Diffuse Large B cell lymphoma is the commonest high grade lymphoma in the Western world. Like Burkitt lymphoma it is characterised by diffuse sheets of lymphoid blasts. In contrast to Burkitt lymphoma, it represents a less well-defined entity that encompasses tumours with variable morphology, genetic abnormalities and outcome. Rates of proliferation and apoptosis vary between individual tumours, and unlike Burkitt lymphoma not all cases are characterised by a prominent macrophage infiltrate. Previous work within the group has shown a relationship in Burkitt lymphoma between apoptosis, macrophage infiltration and proliferation suggesting that apoptosis recruits macrophages to provide support to the tumour cells. This relationship was studied here in a large cohort of patients with DLBCL and the same relationship shown to exist in this tumour also.

Following this observation, a bioinformatic approach was taken to define a gene expression signature of the TAM in DLBCL in situ in an unbiased way. Using large publicly available human tumour gene expression datasets, a graph clustering approach using the tool Biolayout Express 3D was used to explore the transcriptome of DLCBL and other human tumours. Signatures of immune cells and stromal cells, functional pathways and tumour specific signatures were defined from individual

tumour type transcriptomes by study of clusters of co-expressed genes. Further work used a novel graph clustering approach based on mean Pearson correlations to define a 'core' transcriptome signature shared across many unrelated tumour types and in which elements of the tumour stroma were prominent. To validate the TAM signature derived from the DLBCL dataset, protein expression of selected elements of the signature were analysed at the protein level by immunohistochemistry in an unrelated cohort of DLBCL.

Selected markers from the DLBCL TAM signature were then assessed for relationship to outcome in a cohort of patients treated with CHOP chemotherapy. Of the proteins studied, a significant difference in outcome was demonstrated only for leukocyte associated immunoglobulin receptor 1 (LAIR1) expression by TAMs, where low intensity staining for LAIR1 in TAMs was associated with better overall survival.

LAIR1 is a collagen-binding inhibitory receptor expressed only in cells of haemopoetic lineage whose role is little studied in macrophages. The final results chapter presents some preliminary data from co-culture experiments in which the expression of LAIR-1 on the 'macrophage-like' cell line THP-1 is studied in various polarisation states and the ability of these cells to support or constrain tumour cell growth studied in the presence or absence of collagen.

Lay summary

Diffuse Large B cell lymphoma is the commonest high grade lymphoma in the western world. It affects a wide age range from children to the very elderly, although the disease is commonest in adults. With current therapies the disease is curable in approximately 60% of case, leaving a significant burden of incurable disease.

In this tumour, as in many others, there is a significant infiltrate of inflammatory cells in to the tumour, including macrophages, known in this context as TAMs. Macrophages are normally important in fighting infection and, in the aftermath of infection or inflammation, in helping tissues repair and wounds to heal. Some of the functions that are used in healing and repair can be 'borrowed' by tumour cells to help the tumour grow. If we can better understand the ways in which tumours interact with TAMs, then these macrophages may represent another facet of the tumour susceptible to attack by drugs or other therapies.

Others within the group have looked at the ways in which apoptosis (a particular form of programmed cell death) of tumour cells can attract macrophages into the tumour in Burkitt lymphoma, another form of high grade lymphoma. This thesis starts by looking at the relationship of apoptosis, macrophage infiltration and tumour cell proliferation in a large group of tumour samples from patients diagnosed with Diffuse Large B cell lymphoma, and shows that there is a relationship in this tumour type also between the number of tumour cells undergoing programmed cell death and the number of macrophages in the tumour, suggesting that the dying cells may be recruiting the macrophages to the tumour.

To better understand the macrophages, a gene expression signature of the cells was then sought. The gene expression signature is simply a list of genes being expressed by the macrophages and gives an idea of what functions the cells are undertaking rather than studying genes of interest one by one. Macrophages are very changeable

cells and isolating them from the tissue in which they reside is likely to make them alter the genes they are expressing. In an effort to overcome this, a signature was sought of the cells in the tissue. To do this a computational method was used, whereby gene signatures of whole tumours were studied. Diffuse Large B cell lymphoma was studied, as were several other malignancies including breast, colorectal, and ovarian carcinomas, testicular and primary brain tumours to ensure that the computational method was widely applicable. The computational method relies on the fact that every tumour is different, so that tumour A might have more macrophages in it than tumour B, which would be different again from tumour C. Because the macrophage genes will alter with the number of macrophages present in the tumour they should all show the same or similar pattern of expression across the samples. The computational method used allows genes with the same profiles to be identified as clusters. This approach was applied across multiple datasets representing data from more than one thousand patients to generate signatures of macrophages, as well as many other cell types and cellular processes.

Having identified a macrophage signature by this approach, this was verified by looking for the proteins encoded by a subset of genes using immunohistochemistry to ensure that the proteins were expressed in macrophages and not tumour cells. Some of these selected genes were studied further to assess whether expression by the macrophages equated with a good or a poor outcome for the patient. Only one of the markers studied (LAIR1) showed a relationship with outcome with patients with a low level of staining for the protein in their TAMs being associated with a better outcome. LAIR1 is a receptor which binds collagen and sends an inhibitory signal into the cell, preventing it being activated. This suggested that patients who had lower levels of staining for LAIR1 might have easier to activate macrophages, leading to a better outcome. The final part of the thesis looked at expression of LAIR1 on macrophages, how it was altered in different settings and how these affected the way tumour cells grow in culture. This very preliminary work hinted that LAIR1 signalling might affect the way tumour cells and macrophages interacted.

Declaration

I certify:

(a) that the thesis has been composed by me, and

(b) either that the work is my own, or, where I have been a member of a research group, that I have made a substantial contribution to the work, such contribution being clearly indicated, and

(c) that the work has not been submitted for any other degree of professional qualification except as specified.

Tamasin Doig

Acknowledgements

My thanks firstly to my supervisors Professor Chris Gregory and Dr John Goodlad for their advice, support and guidance.

Thanks to all the members of the Inflammation and Cancer Group, and especially John Pound, Catriona Ford and Lynsey Melville for the practical help, and guidance they provided, as well as their company which made my years working with the group such a pleasant experience.

I'm especially grateful to Professor Tom Freeman who provided the inspiration for the bioinformatic analysis in chapter 4 and who provided endless enthusiasm for the project, created new code specifically for the analysis and taught me a huge amount about the use of bioinformatics in analysis of complex data.

I'm grateful for the help of both the Tayside Tissue Bank and Edinburgh Experimental Cancer Medicine Centre who provided human tumour samples and created the tissue microarrays. Clinical outcome data was provided by Michel Sieniawski who manages the Scotland and Newcastle Lymphoma Group database. I thank him for his help, and all the members of the SNLG who contributed the clinical data to the database.

And finally, I thank friends and family, especially Stephen, for their support (and forbearance) during the writing of this thesis.

Table of contents

List of abbreviations used.

ABC	Activated B- cell
ADAM	A disintegrin and metalloproteinase
AIDS	Acquired immunodeficiency syndrome
AIF1	Allograft inflammatory factor
AML	Acute myeloid leukaemia
ATP	Adenosine triphosphate
BAFF	B cell activating factor
BCL	B cell lymphoma gene
BCR	B cell receptor
BL	Burkitt lymphoma
BL2	Burkitt lymphoma 2 cell line
BP	Biological process
BSA	Bovine serum albumin
CC	Cellular component
CCL	Chemokine (C-C motif) ligand
CCR	Chemokine (C-C motif) receptor
CD	Cluster of differentiation
CEL	CEL file – Affymetrix specific data file format
CHOP	chemotherapeutic regimen consisting of Cyclophosphamide, doxorubicin (otherwise known as Hydroxydaunomycin), vincristine (previously called Oncovin) and Prednisolone
CXCL	Chemokine (C-X-C motif) ligand
DAB	Diaminobenzidine
DAVID	Database for Annotation, Visualisation and Integrated Discovery
DDR	Discodin domain receptor
DLBCL	Diffuse Large B cell lymphoma

DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphates
EBV	Epstein Barr Virus
ECL	Enhanced chemiluninescence
ECM	Extracellular matrix
ECOG	Eastern cooperative oncology group
EDTA	Ethylene diamine tetra acetic acid
FACS	Fluorescence activated cell sorting
FCS	Foetal calf serum
FFPE	Formalin-fixed, paraffin embedded.
GBM	Glioblastoma multiforme
GCB	Germinal centre B-cell
GEO	Gene Expression Omnibus
GM-CSF	Granulocyte-macrophage colony stimulating factor
GO	Gene ontology
GSEA	Gene set enrichment analysis
HBSS	Hank's buffered salt solution
HEPES	N-2-hydroxyethylpiperazine- N-2-ethanesulfonic acid
HER2	Human epidermal growth factor
HIF	Hypoxia inducible factor
HIV	Human immunodeficiency virus
HPF	High-powered field
HRP	Horseradish peroxidase
IFN	Interferon
IgG	Immunoglobulin gamma

IgH	Immunoglobulin heavy chain
IL	Interleukin
IMS	Industrial methylated spirit
ITIM	Immunoreceptor tyrosine-based inhibitory receptor
KEGG	Kyoto encylopedia of genes and genomes
LAIR	Leukocyte associated immunoglobulin-like receptor
LDH	Lactate dehydrogenase
LDL	Low density lipoprotein
LGALS3BP	Lectin, galactosidase-binding, soluble, 3 binding protein
LMP	Low malignant potential
LPS	Lipopolysaccharide
MCL	Markov clustering algorithm
MF	Molecular function
MFGE8	Milk fat globule epidermal growth factor 8
MHC	Major histocompatability complex
MMP	Matrix metalloproteinase
MSig DB	Molecular signatures database
MSI	Microsatellite instability
MSS	Microsatellite stable
NFκB	Nuclear factor of kapp ligh polypeptide gene enhancer in B cells
NK	Natural killer
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PI	Propidium iodide
PID	Pathway interaction database
PMA	Phorbol 12-myristate 13-acetate

PMBL	Primary mediastinal B cell lymphoma
QC	Quality control
RMA	Robust multiarray average
RNA	Ribonucleic acid.
RPMI	Roswell Park Memorial Institute medium
RT-PCR	Reverse transcription polymerase chain reaction
SDS	Sodium dodecyl sulfate
SNOMed	Systemised nomenclature of medicine
SP-PIR	Swiss protein protein information resource
TAM	Tumour-associated macrophage
TBS	Tris buffered saline
TGF	Transforming growth factor
THP-1	THP-1 monocytic cell line
TIMP	Tissue inhibitor of metalloproteinase
TMA	Tissue microarray
TNF	Tumour necrosis factor
TYMP	Thymidine phosphorylase
UTP	Uridine triphosphate
VEGF	Vascular endothelial growth factor
WHO	World Health Organisation
WT	Wilms tumour

Chapter 1

Introduction

1.1 Diffuse large B cell lymphoma

Diffuse large B cell lymphoma (DLBCL) as currently defined by in the WHO in Classification of Tumours of Haemopoetic and Lymphoid tissues, 4th edition is the commonest high grade lymphoma in the Western world (Swerdlow et al 2008). As such it still represents a relatively rare malignancy, with all non-Hodgkin lymphoma combined ranking 6th in the UK in terms of incidence and DLBCL accounting for about half of those cases giving a crude incidence rate of 8 or 9 per 100,000 of the population in the UK (data from Cancer Research UK). Unlike many commoner malignancies however, diffuse large B cell lymphoma affects a very wide age range stretching from the paediatric age group, through all adult ages to the very elderly. There are currently many related subtypes of DLBCL which are separately classified due to unique clinical and pathological features such as EBV-positive diffuse large B cell lymphoma of the elderly, chronic infection related DLBCL, primary DLBCL of the central nervous system, primary cutaneous DLBCL, leg-type and lymphomatoid granulomatosis which appear to represent unique disease entities and will not be discussed further, this thesis focussing on the major disease classification of DLBCL, not otherwise specified.

The WHO classification of lymphoma draws together histological, clinical and genetic features to create a classification system and it is clear that DLBCL actually represents a heterogeneous collection of diseases defined as a single entity in the absence of our ability to further subdivide these in a clinically meaningful way. Whilst intellectually unsatisfying such a pragmatic approach to classification groups together diseases predicted to behave in a broadly similar way, that are therefore treated in a broadly similar way. The classification encompasses DLBCL arising de novo as well as high grade transformation of low grade lymphomas such as follicular lymphoma, marginal zone lymphoma and small lymphocytic lymphoma/chronic

lymphocytic leukaemia (Swerdlow et al 2008). DLBCL may present as any age, though is commoner in the adult population. It can present at any site in the body, either in a nodal or extranodal location and it tends to behave in an aggressive manner. A variety of genetic abnormalities are described in the classification including mutations in *MYC*, *PIMI*, *RHOH/TTF* and *PAX5* and translocations involving the *BCL6* gene, *BCL2* gene and the *MYC* gene but none are characteristic or necessary to reach the diagnosis (Pasqualucci et al 2001, Bastard et al 1994, Offit et al 1994, Ohno and Fukuhara 1997, Lipford et al 1987, Weiss et al 1987, Yunis et al 1989, Hummel et al 2006). Histologically the disease shows wide morphological variation with three major morphological variants (immunoblastic, centroblastic or anaplastic) and other rare variants. These morphological differences show poor inter- and intra-observer reproducibility and, perhaps because of this, do not predict tumour behaviour (Harris et al 1994). The tumours also differ considerably in the content of other cells such as macrophages and small lymphocytes that are present or the degree of fibrosis present.

Prognosis is predicted by various clinical features many of which are used to calculate the IPI (International Prognostic Index) including the patient age, stage of disease, performance status, presence of extranodal disease and serum levels of lactate dehydrogenase (LDH) (International Non-Hodgkin's Lymphoma Prognostic Factors project, 1993). There is considerable heterogeneity of clinical response within IPI grouping however, reflecting the heterogeneity of the underlying tumours. Poor outcome and particular patterns of tumour spread can also be predicted by features such as the presence of a *MYC* translocation in the tumour or involvement of 'sanctuary sites' such as the testis and treatment tailored accordingly, but currently little information about the abnormalities driving the tumour is used in the planning of treatment, most patients being uniformly treated with CHOP (cyclophosphamide, doxorubicin, vincristine and prednisolone), or a CHOP-like regimen plus Rituximab, as dictated by patient fitness (Barrans et al 2010, Cheah et al 2014, Ferreri et al 2015, Ghose et al 2015). Multiple individual features or combinations of features have been assessed over recent years in an effort to provide further information about prognosis but while many studies have shown associations with outcome, much of

the data have been contradictory and very little of this data has influenced individual patient treatment. (Adida et al 2000, Barrans et al 2004, Berglund et al 2005, Colomo et al 2003, De Paepe et al 2005, Gasgoyne et al 1997, Hans et al 2005, Hermine et al 1997, Iqbal et al 2006, Milleret al 1988, Muris et al 2006, Veelken et al 2007, Yamguchi et al 2002)

Whilst the disease has an aggressive behaviour it is frequently sensitive to chemotherapy and is potentially curable by this approach. Prior to the introduction of the anti-CD20 monoclonal antibody Rituximab the cure rate was in the region of 40-50%. Since the widespread introduction of Rituximab to standard chemotherapeutic regimens in the last decade the cure rate has increased dramatically to about 60-70%, although clearly longer term data on this are limited and captured within this figure there are still differences in outcome between subtypes and prognostic groups (Alizadeh et al 2000, Lenz et al 2008, reviewed in Roschewski et al 2014 and Sehn and Gascoyne 2015). There is clearly however still a large burden of incurable disease, and a greater understanding of the biology of this disease/these diseases is required if the percentage of patients who are cured is to be increased. Currently a significant proportion of the incurable disease cohort represents patients with primary refractory disease or those whose tumours recur very rapidly after cessation of the initial chemotherapy (Martelli et al 2013).

Several large studies of DLBCL in recent years using high throughput gene expression and gene mutational analysis have furthered our understanding of the diversity of the diseases captured under the DLBCL umbrella. A seminal paper by Alizadeh et al in 2000 showed that DLBCL could be subtyped on the basis of their gene expression profile, with two broad classes being described; those whose gene expression signature resembled that of normal germinal centre B cells (GCB-type) and those whose expression signature resembled that of peripheral blood B cells activated in vitro (activated B cell (ABC)-type) (Alizadeh et al 2000). This gene expression signature as well as revealing aspects of the biology of the tumours appeared to provide prognostic information with patients whose tumours had an ABC-type pattern of gene expression having a considerably poorer prognosis when

treated with standard chemotherapy regimens, than those with a GCB-type pattern of gene expression. Tumours with the ABC-profile were characterised by expression of genes such as IRF4, a gene transiently expressed in B cell activation, and anti-apoptotic genes FLIP and BCL-2, while GCB-type tumours showed expression of genes characteristic of germinal centre B cells such as CD10 and BCL-6, as well as genes known to be mutated in other haematological malignancies BCL-7A and LMO2 (Alizadeh et al 2000). Of importance was the observation that there was no single gene whose expression predicted which group the tumours would fall into. Instead it was the signature identified of many genes that allowed stratification of the tumours into the subtypes. A further study from the same group using an array cGH based approach to a large number of tumours identified 272 recurrent chromosomal abnormalities, 30 of which were differentially present in the DLBCL subtypes, suggesting that these do truly represent different disease entities characterised by different oncogenic pathways (Lenz et al 2008). ABC-type tumours were characterised by trisomy 3, deletion of chromosome arm 6q, gain or amplification of an area on 18q, deletion of the tumour suppressor locus *INK4a* on chromosome 9 and gains or amplications of an area of chromosome 19. GCB tumours were characterised by amplifications on chromosomes 2, 12 and 13, as well as a deletion on chromosome 10. Identified targets of these mutations included loss of the *PTEN* tumour suppressor gene and amplification of *REL* (Lenz et al 2008).

Gradually a picture has built up of the underlying genetic abnormalities in these subtypes of DLBCL. The ABC-subtype appears to be characterised by signalling through NFκB pathways (Davis et al 2001). ABC-tumours are characterised by IRF4 expression, a target of NFκB. In the normal setting this is expressed by B cells differentiating towards plasma cells (Shaffer et al 2009). In the malignant B cells of the ABC-DLBCL tumours terminal differentiation to plasma cells is blocked by the inactivation of Blimp-1 by a variety of mutations included those affecting *PRDP1* (encoding Blimp-1), *BCL6* and *SPIB* (Iqbal et al 2007, Lenz et al 2008a, Mandelbaum et al 2010, Pasqualucci et al 2006, Schmidlin 2008, Tam et al 2006). Thus ABC-type tumours appear stuck in a normally transient stage in the differentiation pathway as a plasmablast. Multiple influences on the NFκB pathway

are represented in ABC-type tumours; mutations in CARD11 and CD79 lead to activation through the BCR signalling pathway; mutations in MYD88 seen in 39% of ABC-type tumours lead to signalling through the TLR pathway, as well as driving secretion of IL-6, IL-10 and IFN-β and downstream autocrine signalling (Davis et al 2010, Lenz et al 2008b, Ngo et al 2011). Signalling through the BCR-complex appears important in ABC-tumours which maintain expression of the IgM heavy chain (and hence its functional signalling qualities) in their functional BCR despite the non-productive allele undergoing class switch recombination (Pasqualucci et al 2011a). BCRs in ABC tumours are immobile and clustered on the cell surface, a property seen normally in antigen-stimulated B cells, and inhibition of BCR signalling components results in cell death in vitro (Davis et al 2010). Intriguingly, recent data suggests the antigen recognised by the BCR in some ABC-DLBCL lymphoma cell lines may be a 'self' antigen derived from apoptotic tumour cells (L. Staudt, unpublished data). The underlying mechanisms in GCB-DLBCL are less well characterised. GCB-type tumours are often characterised by translocations involving BCL2, rendering the cells less responsive to apoptosis (Iqbal et al 2004). BCL6 activity is frequently dysregulated in DLBCL, either by translocation, or mutation (reviewed in Pasqualucci et al 2003). BCL6 is a central transcriptional regulator in the normal germinal centre reaction with multiple downstream targets with three broad functions; blocking differentiation to plasma cells, repressing inhibitors of the cell cycle and blocking expression of many genes associated with inflammation (Shaffer et al 2000). In tandem with mutations of BCL2, cells with dysregulation of BCL6 would be predicted to proliferate without differentiating and be protected from apoptosis (Ci et al 2009). Epigenetic phenomena clearly also have a role to play, with mutations in histone modifying enzymes being relatively common. These tend to be seen more in GCB than ABC subtype with 41% of GCB tumours having mutations in *CREBBP* compared to only 17% in ABC type (Pasqualucci et al 2011b). These mutations impact on acetylation of BCL6, representing one of the mechanisms by which this dysregulated.

Other approaches to subclassification of DLBCL by gene expression analysis have been undertaken by groups looking at genes involved in apoptosis, clustering of all

genes on the chip (as opposed to the approach used by Alizadeh which focused on lymphocyte genes only) or clustering based on good or poor prognosis and these approaches have also yielded insights into tumour behaviour (Alizadeh et al 2000, Lenz et al 2008c, Linderoth et al 2008, Monti et al 2005 and Muris et al 2007, Rosenwald et al 2002). The study reported by Rosenwald et al expanded on the Alizadeh et al paper using a customised chip enriched in immune/lymphoid genes but in contrast to the previous study where only genes expressed in lymphocytes were used for clustering, used all genes and derived signatures that as well as representing the germinal centre, also represented the background lymph node, proliferation, and a signature they called MHC class II (Rosewald et al 2002). Of these, the proliferation signature was correlated with poor outcome and the lymph node signature and MHC II signature with good outcome suggesting a role for the host response in determining outcome. Monti et al analysed gene expression using Affymetrix chips and showed that when the whole gene expression pattern of the tumour is considered, DLBCL showed three main patterns that they characterised as 'oxidative phosphorylation', 'B cell receptor/proliferation' and 'host response' (Monti et al 2005). These clusters, unlike those described by Alizadeh et al, did not predict outcome, and was unsurprising given this, did not overlap with the 'cell of origin' signatures. The approach taken in the Muris et al paper was to select genes involved in apoptosis that hence may represent a mechanism of resistance to chemotherapy and cluster based only on expression of these genes (Muris et al 2007). They identified three clusters, a 'cellular cytotoxic response', an 'activated apoptosis cascade' and 'hyperplastic lymphoid tissue' cluster, of which the first two were associated with poor prognosis and the 'activated apoptosis cascade' showed some overlap with the ABC type of DLBCL. The Linderoth et al paper clustered genes on the basis of patient outcome and found a better prognosis associated with genes of the tumour microenvironment (Linderoth et al 2008). Lenz et al clustered DLBCL on the basis of all genes on Affymetrix U133 Plus2 arrays and showed a germinal centre cluster similar to the GCB cluster, which was expressed in the malignant B cells and two stromal clusters, one of good and one of poor prognosis associated with the non-B cell component of the tumour (Lenz et al 2008c). This analysis suggested that a vascular signature predicted poor response to chemotherapy

while a macrophage and extracellular matrix-rich signature predicted good response to chemotherapy. It is of interest that there is very little overlap between signatures from different analyses strategies despite for example, multiple signatures predicting poor prognosis and contradictory findings as to the significance of these clusters. The 'cell of origin' analysis which can at least partly be predicted at the protein level by immunohistochemistry and thus used outwith the research setting where fresh frozen tissue is available, remains the subclassification most used (Choi et al 2009, Hans et al 2004, Muris et al 2006).

Although Rituximab has made an enormous contribution to the treatment of DLBCL (and many other B cell lymphomas) its mechanisms of action have been poorly understood. In vitro the molecule has multiple actions including induction of apoptosis and complement mediated cell death, but the situation in vivo is far less well characterised (Golay et al 2000, Maloney et al 2002). There is limited in vivo data as to the mechanism of action of Rituximab, but what there is points to its activity being due to a combination of activation of complement, direct induction of programmed cell death of tumour cells and antibody dependent recruitment of effector cells, especially cells of monocytic origin with the recruitment of effector cells being believed to be the most important, hinting that the relationship of the TAM with the tumour is likely to be of importance not just in the development and survival of the tumour but also its response to therapy (Illidge et al 2014, Minard-Colin et al 2008). Considerable work has been ongoing for many years to produce additional monoclonal antibodies against B cells that are more effective than Rituximab. Many of these have been designed to have specific functions such as enhanced activation of complement, induction of cell death or delivery of a cytotoxic molecule in an effort to address 'effector-cell exhaustion', one proposed short-coming of Rituximab (Honeychurch et al 2012, Illidge et al 2014, Goede et al 2015, Palanca-Wessels et al 2015). Several of these drugs are currently in clinical trial, but not yet in routine use.

1.2 Tumour associated macrophages

There is extensive data both from spontaneously developing human tumours and experimental models suggesting that macrophages play multiple important roles in tumour formation, survival and spread (Allavena and Mantovani 2012, Qian and Pollard 2010). Tumour associated macrophages (TAMs) have been ascribed roles in the development of tumours, the angiogenic switch allowing tumour growth past a few millimetres in dimension, suppression/evasion of effective immune responses to the tumour, acquisition of the ability to breach basement membranes and invade into blood or lymphatic channels and the ability to extravasate from blood or lymph and proliferate in metastatic sites.

The evidence for the role of macrophages in the development of tumours comes from a combination of epidemiological data from humans and experimental models. The increased risk of malignancy in chronic inflammatory conditions is well recognised. In ulcerative colitis for example, a disease characterised by relapsing and remitting inflammation largely confined to the colonic mucosa, there is an increase in the risk of development of colorectal carcinoma compared to the general population (Burisch and Munkholm 2015). This risk is influenced by the length of exposure to the inflammation, and the extent and severity of the inflammation. Similar increased risk of malignancy is also seen in other chronic inflammatory conditions such as Helicobacter-associated inflammation in the stomach, which increases the risk of both extranodal marginal zone lymphoma and adenocarcinoma, chronic viral infections with Hepatitis B or C which increase the risk of hepatocellular carcinoma, and chronic inflammation related to Schistosome ova which increases the risk of bladder cancer in areas where Schistosomiasis haematobium is endemic (Fried et al 2011, Oh and Weiderpass 2014). This correlative human data is supported by data from animal models where it has been shown for example that ablation of *Stat3* in myeloid cells leads to a chronic colitis and development of carcinoma (Deng et al 2010). Similarly mice deficient in GM-CSF and IFNγ demonstrated inflammatory lesions and an increased rate of spontaneously developing malignancy, both lymphoproliferative disorders and a range of solid tumours (Enzler et al 2003).

TAMs, as well as being implicated in the development of tumours, are also believed to play a role in supporting established tumours. One of the key functions described here is in promoting angiogenesis (Qian and Pollard 2010). Without the ability to develop new blood vessels to provide oxygen and nutrients and carry away waste products, tumours are severely constrained in their growth even if the tumour cells possess many of the other properties of malignancy such as independence from growth signals and resistance to death (Bergers and Benjamin, 2003). Angiogenesis occurs to a limited extent in the adult normally being confined to specific settings such as the endometrium during the menstrual cycle and in wound healing and repair in response to injury and in these normal settings non-inflammatory/reparative macrophages play a role (Okada et al 2014, Novak et al 2013). In order to develop a fully malignant phenotype, tumours appear to need to harness some of these functional abilities of macrophages to drive new vessel formation. Correlative data from many human malignancies looking at microvessel density in established tumours shows a correlation between macrophage number and microvessel density as well as between microvessel density and poor outcome (Ch'ng et al, 2013, Espinonsa et al, 2010, Fujita et al, 2014, Hirayama et al, 2012, Koh et al, 2014, Kawahara et al, 2010, Kurahara et al, 2012, Murri et al, 2008, Romero et al, 2012, Shigeoka et al, 2013, Suyani et al, 2013, Wu et al, 2012). Several drugs are currently in trials or routine use which target the tumour vasculature in malignancies including Bevacizumab, an anti-VEGF monoclonal antibody (Carmeliet and Jain, 2011). The mechanisms by which TAMs promote angiogenesis have been partially elucidated in various animal models and in vitro systems. Various molecules identified in TAMs have been shown to be involved in angiogenesis including semaphorin 4D, thymidine phosphorylase and VEGF (Lewis et al 2000, Mitselou et al 2012, Sierra et al 2008).

Knockdown of macrophages either by genetic depletion of these cells or pharmaceutical depletion using liposomal clondronate results in reduced growth of tumours either arising in situ or transplanted (Lin et al 2006, Zeisberger et al 2006). Macrophages influence on angiogenesis is mediated partly via the expression of

hypoxia inducible factor (HIF), expressed by macrophages and responsive to hypoxia (Fang et al 2009, Rius et al 2008,). This has multiple downstream targets including the canonical angiogenic factor VEGF. VEGF expression has been demonstrated in TAMs in both mouse and human tumours (Ding et al, 2012, Gonzales et al, 2007, Kurahara et al, 2012, Lewis et al, 2000). As well as direct secretion by macrophages, VEGF can also be released by macrophages from sequestration in the extracellular matrix by the action of macrophage-expressed molecules such as MMP-9 (Bergers et al 2000).

A further function of macrophages believed to be subverted by tumours for their own means is that of invasion through tissues (Qian and Pollard 2010). Normal epithelial cells are anchored in an organised and structured tissue and confined by the basement membrane. To invade through tissues tumours need to be able to breach these physical barriers and migrate. As cells which are migratory in the normal situation, and play a role in the degradation and remodelling of stroma in the wound healing and repair context, macrophages possess the ability to degrade extracellular matrix by a variety of mechanisms including secretion of proteolytic enzymes such as matrix metalloproteinases, cathepsins and urokinase plasminogen activator (uPA) (Kessenbrock et al 2010, Liguori et al 2011). Recruitment of macrophages to tumours allows tumour cells to take advantage of these mechanisms (Hagemann et al 2005, Kessenboeck et al 2010). TAMs in human tumours can be demonstrated to express a wide variety of matrix proteases including MMPs 1,7,9,12,14 and 19, uPA, cathepsins B, C, D, L and Z, ADAMs and SPARC (Liguori et al 2011, Meyer et al 2011). The role of these is far more complex than simply degradation/remodelling of the tumour stroma however with cleavage of ECM components releasing signalling molecules that affect other elements of the tumour microenvironment (Liguori et al 2011).

Macrophages are also implicated in formation of metastatic foci with data indicating a direct role for macrophages in aiding malignant cells into the vasculature and into metastatic sites (Kaplan et al 2005, Wyckoff et al 2007). There appears to be a very

close relationship between macrophages and invasion with Ojalvo et al demonstrating a specific sub-population of macrophages associated with invasive tumour cells. (Oljalvo et al, 2010). Recent work has highlighted the role played by CCL2 and CCL3 in recruitment and retention of these metastasis specific macrophages (Qian et al 2011, Kitamura et al 2015)

A further role for tumour associated macrophage appears to be in alteration of the immune response for the benefit of the tumour. At some point in tumour progression there appears to be a switch in the phenotype of the macrophage from the 'inflammatory' type cells seen in the cancer promoting inflammatory conditions to a more reparative/immune suppressive phenotype seen in established tumours (Qian and Pollard 2010). This later stage TAM has some features in common with the 'M2' macrophage described as an in vitro phenomenon on polarisation with IL-4, but to simply view TAMs as 'M2' macrophages is to vastly oversimplify a complex cell type subject to multiple influences (Biswas and Mantovani 2010). The immunosuppressive role of the TAM appears to be mediated in part by the cytokines it secretes with these tending to secrete IL-10 and TGFβ and lacking expression of IL-12 thus tending to recruit Th2 and regulatory T cells rather than Th1 T cells and cytotoxic T cells (Biswas et al 2006, Hagemann et al 2008, Torroella-Kouri et al 2009). In human ovarian tumours secretion of CCL22 by TAMs recruits regulatory T cells (Curiel et al 2004). TAMS may also act directly on T cells to induce apoptosis through expression of B7-family members and via arginase to inhibit the availability of L-arginine (Biswas and Mantovani, 2010).

The role that TAMs play in lymphoma is not yet entirely clear. Some of the functions ascribed to TAMs in solid tumours such as the acquisition of the ability to invade the blood stream and migrate through tissue are likely to be of less relevance in tumours of lymphoid origin as normal lymphoid cells are much more mobile than epithelial cells and at some stages of their life possess the ability to circulate through the body as part of their normal function. Some functions such as suppression of an effective immune response are clearly relevant however. Much of the data in lymphoma comes from human studies looking at TAM numbers and outcome. In

general these show that higher TAM numbers are associated with a poorer prognosis (Andjelic et al 2012, reviewed in Kridel et al 2015, Steidl et al 2010, Tan et al 2012). As these data are all derived from clinical samples, they necessarily capture outcome data in the light of chemotherapy rather than purely reflecting the relationship between TAMs and behaviour. Of interest in this context is that the addition of rituximab to the chemotherapeutic regimen has altered this relationship, with patients who have large number of TAMs who are treated with CHOP chemotherapy plus rituximab having a better outcome (Alizadeh et al 2000, Lenz et al 2008, Kridel et al 2015). This result suggests that the TAMs in these tumours are not 'fixed' in a pro-tumorogenic role but are potentially reprogrammable to a tumour-suppressive phenotype.

Much of the mechanistic data as to the role macrophages play in tumours comes from mouse models and as yet, the extent to which the phenomena seen in inbred populations of mice map to real human disease is unclear. Care is required in extrapolating directly from mice to humans as it is clear that there are major differences in macrophage activation profiles between these organisms (Martinez et al 2013). There is a large body of data looking at classically (M1) or alternatively (M2) activated macrophages and drawing comparisons particularly between M2 macrophages and TAMs which are often said to have an 'M2-like' profile (reviewed in Mantovani and Sica 2010, and Mantovani et al 2013). When mouse and human macrophages are activated in an identical fashion (with IL-4 in this case), the well-described markers of 'M2' activation in mice e.g. *arg*, *Ym1*, *Fizz1*, etc are not seen in human macrophages (Martinez et al 2013). Indeed, many of the best characterised mouse markers of so-called M2 activation have no human homologues suggesting that human macrophages mount a different response to this stimulus (Raes et al 2005, Martinez et al 2013).

TAMs in human tumours are an appealing target however for novel therapies as, unlike the malignant cells they support, these cells do not possess the increased rate of mutagenesis that allows the malignant cells of the tumour to mutate away from the

selective pressure applied by treatment and as the data from lymphoma studies hints, these cells which are present in often large numbers in tumours, are reprogrammable to an anti-tumour role.

1.3 The role of apoptosis in modulating macrophage function in tumours

Apoptosis, a form of programmed cell death important in normal development, is considered to be a 'quiet' way for cells to die in contrast to necrotic cell death where the cellular contents are spilled from the disrupted cell and can act on cells in the microenvironment to induce an inflammatory response (Poon et al 2014). Apoptosis is a tightly regulated programme of a cascade of effector molecules triggered either by an external signal to the cell ('the death receptor' pathway) or a signal derived from inside the cell in response to cellular damage (the intrinsic pathway). The characteristic features of apoptosis with retained membrane integrity, nuclear condensation and formation of pyknotic nuclear fragments were described by Kerr, Wylie and Currie in 1972 (Kerr et al 1972). Apoptotic cells are rapidly cleared from tissues, preventing them becoming secondarily necrotic and in many situations this appears to be a function of the neighbouring cells (Grimsley and Ravichandran 2003). In situations of high apoptosis such as that seen in resolving inflammation or within tumours, professional phagocytes are recruited to deal with these dying cells (Gregory and Pound 2011).

Despite the 'quiet' nature of apoptotic cell death, this is not a neutral event with apoptotic cells shown to influence the activation state of macrophages in a variety of inflammatory settings (Voll et al 1997, Fadok et al 1998, Huyn et al 2002, Duffield et al 2001). Defects in apoptosis are associated in humans with autoimmune

conditions including systemic lupus erythematosus, where there is an uncommon but strong association with deficiencies in the complement component C1q (Savill et al 2002). Mouse models looking at genetic defects in phosphatidylserine-mediated apoptotic cell clearance have demonstrated autoimmune phenomena (Reviewed in Elliot and Ravichandran 2010).

Cells dying by apoptosis appear to release 'find me' signals that recruit macrophages to the tumour including ATP/UTP, fracktalkine, lysophophatidylcholine and monocyte chemotactic protein (Elliot et al 2009, Kobara et al 2008, Lauber et al 2003, Truman et al 2008). Once the macrophage is in the local environment of the apoptotic cell, a variety of molecules on the cell surface allow engulfment by macrophages either directly or via the action of a linker molecule. The 'eat me' signals include phosphatidylserine 'flipped' from the inner to the outer cell membrane, and oxidized low density lipoproteins-like sites (Savill et al 2002). Characterised linker molecules include the complement component C1q, mannose binding lectin, Gas6, and milk fat globule epidermal growth factor 8 (MFGE8) amongst others (Elliott and Ravichandran 2010, Lauber et al 2004, Savill et al 2002). Cell surface receptors on the macrophage involved in apoptotic cell clearance include CD14, CD91, calreticulin, the scavenger receptors CD36, Scavenger receptor A, CD68 and LOX1, β2 integrins, the $\alpha_v\beta_3$ integrin, Mer, Bai1, Tim-2 and Stabilin-2 (Elliott and Ravichandran 2010, Lauber et al 2004, Savill et al 2002) . There appears to be a large amount of redundancy in the system as might be predicted by the ever growing list of molecules involved in phagocytosis of apoptotic cells and it is likely that different combinations of molecules are used in different settings.

Apoptotic cells influence their neighbourhood. They have been demonstrated to result in increased secretion by macrophages of anti-inflammatory molecules such as IL-10 and TGFβ and down-regulation of pro-inflammatory mediators including IL-6, IL-8, IL-12 and TNFα (Fadok et al 1998, Huynh et al 2002, Grimsley and Ravichandran 2003, Voll et al 1997). This anti-inflammatory effect of apoptotic

cells was seen even in the face of previous activation by IFNγ and LPS (Reiter et al 1999). This effect is mediated by contact with apoptotic cells and does not require phagocytosis by the macrophage (Gregory and Pound, 2011). Ingestion of apoptotic cells can also lead directly to VEGF secretion, contributing to angiogenesis (Golpon 2004).

Tumours are microenvironments characterised by high rates of apoptosis, representing as they do a proliferative environment but one in which the cells proliferating have multiple mutations and hence are susceptible to apoptosis. While the rate of proliferation in a tumour clearly outbalances cell loss or the tumour would not grow, and as a whole entity tumours are resistant to appropriate apoptosis, it is clear from histological examination of any tumour that apoptosis is a frequent event. This is therefore one mechanism by which tumours can influence the inflammatory microenvironment in which they exist.

Recent work by our group has demonstrated the importance of apoptosis in tumour cell growth (Ford et al 2015). Prevention of apoptosis by upregulation of anti-apoptotic proteins constrained tumour cell growth in vivo in an animal model, suppressed angiogenesis and inhibited macrophage accumulation in the tumour. Improved tumour cell growth and macrophage accumulation was seen on inoculation of apoptotic tumour cells along with viable tumour cells in a transplant tumour model, and the effect of this was seen to be partly IL-4 dependent, being reduced in IL-4 knockout mice. Previous work has demonstrated enhanced expression of IL-10 and secretion of the B cell growth factor BAFF on phagocytosis of apoptotic cells suggesting mechanisms by which apoptotic cell death might influence tumour growth (Ogden et al 2005).

Tumour supportive effects of macrophages exposed to apoptotic cells clearly have implications for therapy. Many chemotherapeutic drugs act by inducing apoptosis in the target cells and an unintended consequence of this may be enhanced support of tumour cell growth, angiogenesis and immune evasion by the TAMs. This effect clearly requires consideration in planning future chemotherapeutic regimens/immunochemotherapy.

1.4 The influence of the microenvironment in lymphoma

In a recent review Scott and Gascoyne characterised three main patterns of interaction of lymphoma cells with their microenvironment which they called 're-education, recruitment and effacement' (Scott and Gascoyne, 2014). Follicular lymphoma typifies the 're-education' pattern being extremely reliant on the microenvironment for survival, with the tumour 're-educating' the normal lymph node environment to provide trophic support for tumour cell survival and growth. The malignant follicles contain follicular dendritic cells and follicular T helper cells as in the normal lymph node, however there is an increase in number of T_{FH} cells as well as a skew towards CD4 positive T_{regs} and away from T_H17 cells (Ai et al 2009, Yang et al 2007, Yang et al 2009). This appears to be driven by secretion of CCL22 by the tumour cells and expression of co-stimulatory molecules CD70, CD80 and CD86 (Yang et al 2006). Follicular lymphoma when it involves the bone marrow can induce the formation of follicular-like structures containing some of the same cellular elements seen in the lymph node. The dependence on the microenvironment for follicular lymphoma growth and survival is demonstrated by its inability to grow in culture (Tweeddale et al 1987). Intriguing work looking at the BCR expressed by follicular lymphoma has shown that many of these are mannosylated and can therefore interact with the mannose receptor and CD209 expressed by cells in the microenvironment (Coelho et al 2010, Zhu et al 2002) This interaction appears to signal to the lymphoma cell and provide a survival signal. CD40 expression by the T_{FH} cells recruited to the microenvironment and expression of IL-4 have also been demonstrated to play a role (Calvo et al 2008, Johnson et al 1993, Pangault et al 2010). This appears to be a tumour very dependent on its microenvironment.

Classical Hodgkin lymphoma shows what they called the 'recruitment' pattern, with a microenvironment characterised by a diffuse and variable mixture of cells including neutrophils, eosinophils, macrophages, plasma cells and T and B lymphocytes as well as stromal components recruited by the relatively sparse tumour cells (Scott and Gascoyne 2014). This very prominent cellular and stromal infiltrate

appears to be driven by multiple chemokine and cytokines secreted by the Reed-Sternberg cells which draw into the microenvironment of the involved lymph node elements that are not normally present. These include CCL5, CCL17, CCL22, IL-5, CSF-1 and fracktalkine (reviewed in Steidl et al, 2011). The role played by these elements has yet to be unpicked in Hodgkin lymphoma, but given the enormous number of 'recruited' cells to the number of tumour cells, must play some role in the tumour growth and survival. Like follicular lymphoma Hodgkin lymphoma cells are difficult to culture out of their normal microenvironment (Diehl et al 1981). There is some suggestive data from gene expression profiling and immunohistochemical analysis that poor outcomes are predicted by large number of macrophages, NK cells and cytotoxic T cells in the tumour and low number of T_{regs} and B cells (Alvaro et al 2005, Chetaille et al 2009, Kelley et al 2007, Steidl et al 2010).

The 'effacement' pattern is that of Burkitt lymphoma where the microenvironmental component of the tumour is relatively inconspicuous suggesting the normal microenvironment of the lymph node has been effaced by the tumour cells which are associated with few other cell types with only macrophages being appreciable by normal light microscopy (Scott and Gascoyne 2014). However, as work from our group has shown, 'effacement' of the lymph node microenvironment does not equate to no role for the microenvironment in tumour cell growth (Ogden et al 2005, Ford et al 2015).

Diffuse large B cell lymphoma according to this scheme, lies between 're-education' and 'effacement' and appears closer to the situation in Burkitt lymphoma than that in follicular lymphoma being able in some cases to grow in culture but having a less completely effaced microenvironment than Burkitt lymphoma (Tweeddale et al 1987). The large gene expression studies discussed earlier have highlighted several facets of the microenviroment in DLBCL and triggered a wealth of studies, many of them somewhat contradictory in different patient cohorts. The studies by Monti et al, Rosenwald et al, Lenz et al and Linderoth et al all identified signatures of the microenvironment in gene expression studies of whole tumour tissue from DLBCL (Lenz et al 2008c, Linderoth et al 2008, Monti et al 2005, Rosenwald et al 2002).

The possible influences these have on prognosis is relatively unclear with a 'lymph node' signature and a 'microenvironment signature' reported to correlate with good outcome, a 'cellular cytotoxic signature' to correlate with poor outcome and two stromal signatures indentified in the Lenz paper with different predictions of outcome. The signatures identified in this paper are of particular interest suggesting a macrophage signature is associated with a good prognosis while a vascular signature, which might have been predicted to be quite closely aligned with the macrophage signature, predicts poor prognosis (Lenz et al 2008c). Numerous studies have looked at the numbers of TAMS in DLBCL and the relationship this has with prognosis. The majority of these studies have used either CD68 or CD163 or both as markers of TAMS. As reviewed by Kridel et al recently, while there are contradictory findings in some of these studies, overall the pattern appears to be that large numbers of TAMs predict poor outcome in the pre-rituximab era and good outcome when rituximab is included in the treatment regimen (Kridel et al 2015).

Like follicular lymphoma, there appears to be some role for signalling from the microenvironment for tumour cell growth and survival with secretion of APRIL reported by neutrophils (and very rarely by macrophages or other stromal cells) in DLBCL, (Schwaller et al 2007) despite neutrophils apparently not forming a large part of the tumour microenvironment on histological examination.

Tumour cells avoid recognition by the immune cells normally present in their microenvironment by several mechanisms. In some cases there is loss of expression of MHC II by the cells, a feature common to many subtypes of lymphoma (Riemersma et al 2000, Wilkinson et al 2011). In DLBCL growing in immune privileged sites this is frequently due to chromosomal deletions affecting the MHC II genes, whilst the ABC-subtype downregulates expression as part of its abortive progression towards a plasma cell (Riemersma et al 2000, Wilkinson et al 2011). This downregulation of MHC II correlates with poorer outcome (Rimsza et al 2004). Frequent rates of loss of both MHC class I molecules (via mutations in β2-microglobulin) and CD58 together in DLBCL suggest that these cells avoid the attention of both cytotoxic T cells and NK cells by downregulating the surface

molecules required for recognition by these cells (Challa-Malldi et al 2011). Again similarly to follicular lymphoma, DLBCL cells express CD80 in a large percentage of cases, with some expression also seen in T cells providing another potential modulator of immune activation (Dakappagari et al 2002). A recent intriguing study in a mouse model of lymphoma suggests a central role for T cells in the development of lymphoma (Afshar-Sterle et al 2014). In this model mice carried B cells with either inactivation of *Blimp-1* or overexpression of BCL6, both common mutations in DLBCL, but rarely developed spontaneous tumours. These mice crossed onto a T cell deficient background developed overt lymphoma, and the T cell control of the lymphoma was seen to be mediated via Fas-Fas-ligand interactions (Afshar-Sterle et al 2014). In established human tumours numbers of CD3-positive T cells, CD4-positive T cells, FOXP3-positive T_{regs} or numbers of cytotoxic T cells have all been shown to correlate with prognosis although the data within these studies is somewhat contradictory probably reflecting differences in the populations studied and methods used to define the cells of interest (Coutinho et al 2015, Hasselblom et al 2007, Keane et al 2013, Tzankov et al 2008).

A study looking for genes assessable in routine diagnostic practice that predict prognosis identified two genes predictive of outcome, one of which, LMO2, is expressed by tumour cells and the other, TNFRSF9, expressed by cells of the microenvironment, in this case a small population of T cells of both CD4 and CD8 subtypes (Alizadeh et al 2011). The role played by these cells has not yet been elucidated but did not involve expression of the ligand on the tumour cells, suggesting an indirect effect of these activated T cells.

Other components of the microenvironment have also been studied. Mast cells in the tumour have been associated with increased microvessel density, with a relationship also demonstrated of mast cells with T cells (Marinaccio et al 2014, Marinaccio et al 2015. Mast cells have also been associated with an increase in fibrosis (Fukushima et al 2006). In this study, the mast cells were also identified as a source of IL-4. In lymphoplasmacytic lymphoma, a low grade B cell lymphoma, mostly confined to the bone marrow, there is an increase in mast cells associated with the tumour suggesting

active recruitment of these cells by the lymphoma cells (Swerdlow et al 2008). While not as obvious in DLBCL it is possible that these cells are recruited to perform a specific role.

Huang et al demonstrated preferential expression of a molecule, Tim3, in endothelium in lymphomas that had a suppressive effect on T cells (Huang et al 2010). A study by Brandt et al looked at the expression of two of the extracellular matrix genes identified in the 'stromal signature' associated with good prognosis in the gene expression study by Lenz et al, fibronectin and SPARC. (Brandt et al 2013, Lenz et al 2008c). Expression of these in the extracellular matrix and endothelium at the protein level by immunohistochemistry confirmed the good prognosis seen in the earlier paper, but did not provide any insights into the role these molecules play.

It has long been recognised that the tumour microenvironment in haematological malignancies contributes to resistance to drug therapies (reviewed in Shain et al 2015) but much of the interaction of DLBCL with its microenvironment currently remains poorly understood. Questions that are currently unanswered include; the mechanisms by which macrophages and T cells are recruited to the tumour, or re-educated in the tumour; the role of the microenvironment in providing trophic support to the tumour; mechanisms by which macrophages and other immune cells are influenced by the lymphoma cells and influence the lymphoma cells; and the role of the extracellular matrix.

This thesis examines some of the features of macrophages in DLBCL in tissue derived from human tumours in an effort to better understand the tumour associated macrophage in this tumour. The underlying hypothesis is that the tumour associated macrophage in diffuse large B cell lymphoma, is not simply a bystander but plays a role in the pathogenesis of the tumour. The relationship between apoptosis, proliferation and macrophage number is studied to assess the hypothesis that, as in Burkitt lymphoma, apoptosis of tumour cells represents a mechanism by which macrophages are recruited to tumours. Generation of a gene expression signature of the tumour associated macrophage is then addressed using a bioinformatic approach

on a large number of human tumours and elements of that signature validated in a different cohort of patients. Some of the markers of the tumour associated macrophage signature in human DLBCL are examined in relation to outcome in a standardly treated cohort of patients. Finally, some in vitro experiments look at LAIR1, derived from the macrophage signature to assess possible mechanisms by which this molecule may affect outcome in DLBCL.

Chapter 2

Materials and Methods

All work described here was carried out by the candidate unless otherwise stated.

2.1 Studies on formalin-fixed paraffin embedded human tissue

2.1.1 Tissue microarray used in chapter three.

The DLBCL tissue microarray (TMA) used in chapter 3 was created by the Tayside tissue bank in Dundee containing 82 of cases of DLBCL diagnosed in Tayside and Highland region over an 8 year period (http://www.tissuebank.dundee.ac.uk). A minimum of two, but in most cases four, 0.6mm cores, were sampled from each tumour. Areas of tissue for inclusion in the TMA were selected by two pathologists (not the candidate) to ensure cores contained tumour.

The status of the cases with regard to the t(14:18)(q32;q21)translocation, which brings *BCL2* and *IGH* together, was also made available. Cases were classified as GCB or non-GCB-type using the immunohistochemical classification of Hans et al (Hans et al 2004).

All sections from the TMA were cut by Tayside tissue bank at 3 µm onto Superfrost slides (Thermo scientific, Loughborough, UK). Sections were stained by Tayside tissue bank for markers of proliferation, apoptosis and macrophage infiltration using antibodies against ki67, active caspase-3 and CD68 respectively (Clones were; ki67 - MM1, Vector Labs (Peterborough, UK); active caspase-3 - JHM62, Novocastra (Newcastle, UK) and CD68 - PG-M1, DAKO, (Cambridge, UK) with citrate buffer heat antigen retrieval by microwave pressure cooking. These stained slides were made available to the candidate for analysis.

2.1.2 Cell counting

Cells were counted by a single observer using an eyepiece graticule. A representative area to count was established by calculation of a cumulative mean of positive cells until the mean reached stability for each antibody and for cases representing 'high', 'medium' and 'low' levels of positive cells. An area 0.25mm x 0.125mm was counted for each core on the array. To overcome variability in tumour cell size all values were expressed as percentage of total cells. Mean values across cores were calculated for each case. Cases were excluded for analysis if there were less than 2 cores available for analysis.

2.1.3 Statistical analysis

Associations were calculated by Pearson correlation. Differences between groups were assessed by Kruskall-Wallis test.

2.1.4 Creation of a CHOP treated DLBCL TMA used in chapters 5 and 6.

A cohort of cases were identified by a search of the Scotland and Newcastle Lymphoma database to identify cases of DLBCL diagnosed at Western General Hospital, Edinburgh, Edinburgh Royal Infirmary or St John's Hospital, Livingston, who had received CHOP chemotherapy. All cases were reviewed and seventy-three cases identified in which there was sufficient formalin-fixed paraffin embedded (FFPE) material available for inclusion in a tissue-microarray. Cases were excluded if the FFPE-material was missing from the archive or if there was insufficient material to allow sampling for the tissue-microarray while preserving tissue for future use. Representative areas of tumour were marked for sampling.

The tissue microarray was built by the Edinburgh Experimental Cancer Medicine Centre (http://www.ecmcnetwork.org.uk/network-centres/edinburgh/edinburgh). Briefly, 2mm diameter cores of tissue were punched from the donor blocks of FFPE-

material and placed into spaces created in a recipient paraffin block with the exact position of each donor core being recorded on a map of the tissue-microarray. Up to three cores were sampled from each donor block. Once completed, the recipient block was heated gently to fuse the donor paraffin wax with the recipient paraffin wax.

Clinical outcome data and associated statistical analysis for cases examined in chapter 6 was provided from the Scotland and Newcastle Lymphoma database by Michel Sieniawski who manages the database.

2.1.5 Creation of a reactive lymph node TMA used in chapters 5 and 6.

Cases were identified from the archives of Edinburgh Royal Infirmary pathology department by a search using the SNOMed terms 'lymph node' and 'hyperplasia' over the period 1996-2005. Cases were excluded if they were associated with a cancer resection specimen, even if not involved by cancer. Cases were also excluded if there was insufficient material available in the FFPE-block for inclusion in a TMA while preserving material for future use. Cases were also excluded where FFPE-material was missing from the archive. Sixty one cases were identified that were suitable for inclusion in a TMA.

The tissue microarray was built by the Edinburgh Experimental Cancer Medicine Centre. Briefly, 1.5mm diameter cores of tissue were punched from the donor blocks of FFPE-material and placed into spaces created in a recipient paraffin block with the exact position of each donor core being recorded on a map of the tissue-microarray. Up to three cores were sampled from each donor block. Once completed, the recipient block was heated gently to fuse the donor paraffin wax with the recipient paraffin wax.

2.1.6 Single colour immunohistochemistry.

Sequential three micron sections were cut from the TMA onto Superfrost coated slides (Thermo Scientific, Loughborough, UK). Sections were dewaxed in xylene (Sigma, Poole, UK) and rehydrated through a series of dilutions of IMS (Sigma, Poole, UK) before antigen-retrieval and staining as listed in table 2.4 below. All staining was performed using the Sequenza manual immunohistochemical staining system (Thermo Scientific, Loughborough, UK). Endogenous peroxidise activity was blocked by incubating for 5 mins in 3% hydrogen peroxide (Sigma, Poole, UK) in PBS (Sigma, Poole, UK) before washing and incubating with the primary antibody. All antibodies were diluted in DAKO antibody diluent (DAKO, Cambridge, UK). The wash buffer was TBS-Tween in all cases (Sigma, Poole, UK). DAB (DAKO, Cambridge, UK) was used as the chromogen in all cases. Counterstain was haematoxylin (Sigma, Poole, UK) for 30 seconds, then blued in sodium bicarbonate (Sigma, Poole, UK) for 1 minute before dehydration through a series of alcohols to Histoclear II (National diagnostics, USA) then mounting. Unless otherwise stated all reactions occurred at room temperature. Negative controls used primary antibody species and isotype matched IgG at equivalent concentration to the primary antibody of interest, except for CD68 where primary antibody was omitted in the negative control. Details of antigen retrieval, antibody concentrations and incubation periods are given in Table 2.3.

2.1.7 Two colour immunohistochemistry

Sequential three micron sections were cut from the TMA onto Superfrost coated slides (Thermo Scientific, Loughborough, UK). Sections were dewaxed in xylene and rehydrated through a series of dilutions of IMS before antigen-retrieval and staining as listed in table 2.3 below. Endogenous peroxidise activity was blocked by incubating for 5 mins in 3% hydrogen peroxide (Sigma, Poole, UK) in PBS (Sigma, Poole, UK) before washing and incubating with the primary antibody. Both primary antibody incubations were performed simultaneously overnight at 4°C. Secondary detection with a goat anti-rabbit-HRP conjugate was performed first with chromogenic detection using VIP peroxidise substrate (Vector, Peterborough, UK) then a sheep anti-mouse HRP-conjugate was used with chromogenic detection by

DAB (DAKO, Cambridge, UK). All staining was performed using the Sequenza system. All antibodies were diluted in DAKO antibody diluent (DAKO, Cambridge, UK). The wash buffer was TBS-Tween in all cases. There was no counterstain used. Unless otherwise stated reactions occurred at room temperature. Negative controls included rabbit IgG and Mouse IgG with anti-rabbit and anti-mouse HRP conjugates, anti-TYMP rabbit monoclonal antibody with anti-mouse HRP conjugate, anti-CD163 mouse monoclonal with anti-rabbit HRP conjugate, rabbit IgG with anti-rabbit HRP conjugate and mouse IgG with anti-mouse HRP conjugate. No staining was detected in any of the negative control conditions. Appropriate positive controls were included which showed appropriate specific staining without background. Details of antigen retrieval, antibody concentrations and incubation periods are given in Table 2.4.

2.2 Bioinformatic analysis

2.2.1 Selection of Datasets

The datasets used here were selected on the following criteria; analysis of primary human tumour, large study size, availability of data with provision of clinical annotation and genome-wide analysis using the Affymetrix U133 platforms (either U133A+B or U133Plus2.0). These datasets were identified from Gene Expression Omnibus (GEO) or caArray and CEL files downloaded. Details of the 7 datasets used for this study are given in Table 1.

Database reference	*Reference (PMID)*	*Tumour type(s)*	*No of cases analysed (data available)*	*Platform*
GSE11318	**Lenz et al. (18765795)**	**DLBCL and PMBL**	**194 (203)**	**Affymetrix U133Plus2.0**
GSE1456	**Pawitan et al. (16280042)**	**Breast carcinoma**	**134 (159)**	**Affymetrix U133A & B**
GSE9891	**Tothill et al. (18698038)**	**Ovarian (epithelial) carcinoma**	**265 (285)**	**Affymetrix U133Plus2.0**
GSE3218	**Korkola et al. (16424014)**	**Testicular tumours**	**86 (107)**	**Affymetrix U133A & B**
GSE13294	**Jorissen et al. (19088021)**	**Colorectal carcinoma**	**150 (155)**	**Affymetrix U133Plus2.0**
caArray/rembr-00037	**REMBRANT – Repository for Molecular Brain Neoplasia Data**	**Primary CNS tumours**	**253 (271)**	**Affymetrix U133Plus2.0**
GSE7553	**Riker et al. (18442402)**	**Skin tumours**	**77(82)**	**Affymetrix U133Plus2.0**

Table 2.1 Details of datasets used in bioinformatic analysis with database reference, original associated publication, tumour type and platform given. The total number of cases originally available for study and the number of cases passing the quality control step and therefore used in the analysis are given.

The initial analysis used six individual datasets. The breast cancer dataset consisted of samples taken from 159 patients treated in Sweden. These were stratified on the basis of molecular tumour type (basal, HER2 positive, luminal A or B, or normal-like), Ellis-Elston grade, survival outcome and recurrence outcome. The colorectal dataset contained 155 cases of colorectal carcinoma derived from an Australian population divided into microsatellite stable and unstable tumours. No information was available as to grade or stage. The lymphoma dataset consisted of 203 cases of diffuse large B cell lymphoma (DLBCL). These were stratified into germinal centre B cell-like (GCB), activated B cell-like (ABC), primary mediastinal B cell (PMBL) and unclassified, based on gene expression. The data was further organised on the

basis of sex and patient outcome. The glioma dataset, derived from caArray, contained samples of 271 tumours which were stratified on the basis of histological type (astrocytoma, oligodendroglioma, glioblastoma), WHO grade and sex. The ovarian dataset was from 285 cases from several hospitals in Australia and the Netherlands. These were stratified based on whether they were malignant or of low malignant potential, histological type (endometrioid or serous), grade, stage and primary site (ovary, peritoneum or fallopian tube). The testicular dataset was made up of 107 patient samples from the USA. This contained primary germ cell tumours and was stratified first on whether the tumours were pure or mixed histological types and then within each category on the constituent parts of the tumour using the WHO classification (seminoma, teratoma, embryonal carcinoma, yolk sac tumor, choriocarcinoma). A final dataset of 82 samples was derived from various types of skin cancer including basal cell carcinoma, squamous cell carcinoma, primary melanoma, including melanoma in-situ, and melanoma metastatic to subcutaneous tissue, lymph node, brain and adrenal gland.

2.2.2 Quality control

The quality of the raw data from each dataset was reanalysed by Fios genomics using the arrayQualityMetrics package in Bioconductor (http://www.bioconductor.org/) and scored on the basis of 5 metrics, namely maplot, spatial, boxplot, heatmap and rle. Any array failing on more than one metric was removed from the dataset (Table 2.1) and in cases comprising A and B arrays, failure of one chip (A or B) resulted in removal of data from both arrays from the dataset. In cases where the data was derived from A and B arrays, these were merged to create a single file for each sample. Normalisation for each dataset was performed independently using the robust multi-array average (RMA) expression measure. Probesets were then annotated using latest annotation available in Bioconductor (26 June 2009) and samples ordered according to clinical grouping to ease interpretation of the data.

2.2.3 Network Analysis

Each dataset was saved as an '.expression' file. These contain a unique identifier for each row of data (Gene symbol concatenated to probeset ID), followed by columns of gene annotations which can be used as class-sets for the overlay and analysis of information with respect to the graph and finally natural scale normalised data values for each sample (each column of data being derived from a different sample). These files were then loaded into the network analysis tool BioLayout *Express*^{3D}. Pairwise Pearson correlations were calculated for each probeset on the array(s). All Pearson correlations where $r \geq 0.6$ were saved to a '.pearson' file. Based on a user defined threshold designed to include approximately 40% of the data available, undirected network graphs of the data were generated. In this context nodes represent individual probesets (genes/transcripts) and the edges between them Pearson correlation coefficients above the selected threshold. The network was then clustered into groups of genes sharing similar profiles using the MCL algorithm with an MCL inflation value (which controls the granularity of clustering) set to 2.2. Graphs of each dataset were explored extensively in order to understand the significance of the gene clusterings and their relevance to the pathology of the tumours.

2.2.4 Cluster annotation

Gene set enrichment analysis was performed on clusters using DAVID and GSEA MSigDB web-based analysis tools (Huang et al 2009a, Huang et al 2009b, Subramanian et al 2005, Mootha et al 2003). Analysis through DAVID used the 'functional annotation clustering' tool to identify clusters of related hits. This tool compares the submitted gene list with data from a variety of sources including gene ontology data, Swiss protein interaction data and curated pathway data such as is available in KEGG and Biocarta to see whether the submitted gene list is enriched in genes falling into any of these groupings beyond that which would be expected by chance. In this analysis the background from which the tested gene list is derived is considered in analysis ie which genes could be present because of the chip used rather than all genes in the genome. The functional clustering tool pulls together

groups of similar annotation from different sources to highlight functional annotation and reduce redundancy in the generated list of gene sets with significant overlap with the test set. The GSEA MSigDB tool has a similar function allowing determination of whether there are overlaps beyond that expected by chance between entered gene sets and the gene sets that are present in the GSEA database. The available genes sets for comparison include gene ontology sets, oncogenic signatures derived from microarray data deposited in GEO, immunological signatures generated by manual curation of published studies, positional gene sets reflecting individual gene position on chromosomes and curated gene sets from various sources including KEGG, Biocarta and Reactome. Analysis using this tool mostly focussed on curated pathways, gene ontology terms, oncogenic and immune signatures although other gene sets were less frequently used. Unlike DAVID this approach does not take into account the background from which the tested gene set was derived. GSEA analysis generated multiple hits for most clusters and clusters were annotated if hits of high significance showed a common trend as to function. These analyses were supplemented by comparison of the clusters generated here with tissue- and cell-specific clusters derived from network-based analyses of a human tissue atlas and an atlas of purified leukocyte populations, analysis of protein expression of selected targets using the Human Protein Atlas and review of the published literature (Ulhen et al 2010, Uhlen et al 2015).

2.2.5 Comparison of Expression Patterns Across Six Cancers and Validation of Signatures

To allow direct comparison of all datasets, probesets that were not represented in all datasets i.e. were only present on the U133Plus 2.0 array, were removed leaving 44,754 probesets common to all U133 platforms. The Pearson correlation coefficient between data derived from each probe-set in each dataset was calculated using BioLayout *Express*^{3D} and all values written to file. A mean Pearson correlation was then calculated for each probe i.e. the average correlation between probesets across the six datasets. The mean Pearson correlations were then filtered to remove any mean correlation values below a user defined threshold and network graphs

constructed. The majority of the work described here is based on the use of graphs constructed using a mean Pearson threshold of $r \geq 0.6$ and clustered with an MCL inflation value of 2.2.

2.2.6 Validation of 'core' signatures on independent dataset.

A skin cancer dataset, not used in the 6 dataset merge analysis above was treated as described for the other individual datasets. The gene clusters and annotations from the merged graph were mapped onto the dataset and clusters in the skin dataset analysed for enrichment of genes from the 'core' signatures that were greater than would expected by chance, using Fisher's test with an adjustment for multiplicity of testing.

2.3 In vitro cell culture studies

2.3.1. Cell line maintenance

The EBV-negative Group 1 sporadic Burkitt lymphoma cell line BL2 was used in these studies (Rowe et al 1986). These were cultured in suspension in RPMI 1640 (Invitrogen, Paisley, UK) supplemented with 10% foetal calf serum (PAA, USA), 2mM L-glutamine, 100 IU/ml penicillin and 100 μg/ml streptomycin at 37°C in 5% CO^2.

THP-1 cells, a human acute moncytic leukaemia cell line (European collection of authenticated cell cultures, Porton Down, Cat no. 88081201) were maintained in suspension in RPMI (Invitrogen, Paisley, UK), plus 10% foetal calf serum (PAA, USA), 2mM L-glutamine, 100 IU/ml penicillin and 100 μg/ml streptomycin at 37°C in 5% CO^2.

2.3.2 THP-1 differentiation and polarisation

THP-1 cells, were plated out a concentration of 0.2 $x10^5$ cells/ml and PMA in DMSO added at a concentration of 200nM. After three days, cells were washed in Dulbecco PBS (PAA, USA) and incubated in fresh culture medium, with the addition of cytokines and LPS as indicated. Cytokine suppliers and final concentrations given below (Table 2.2). The cells were incubated in the presence of the cytokines or LPS for five days and then washed with Dulbecco PBS before being used in experiments.

Cytokine	Supplier	Concentration
Human IL-10	R&D	5ng/ml
Human IFNγ	R&D	20ng/ml
Human IL-4	R&D	20ng/ml
LPS	Sigma	100ng/ml

Table 2.2. Cytokine concentrations used and suppliers.

2.3.3 Detachment of adherent cells from culture flasks

Cells were washed with HBSS (Thermo-scientific, Loughborough, UK) to remove medium then cold detachment buffer added to the cell. They were then shaken on an orbital shaker at 80rpm at 4°C for 30mins. Cells were then gently lifted with a cell lifter and washed once in HBSS before use.

2.3.4 Culturing THP-1 cells on collagen

Collagen IV from human placenta (Sigma, Poole, UK) was dissolved in 0.25% sterile filtered acetic acid and was spread onto tissue culture dishes or chamber slides (Sigma, Poole, UK). This was UV treated overnight to sterilise and cross-link the collagen. These were stored at room temperature and before use were sprayed with 70% alcohol them washed with sterile water and allowed to dry. Control dishes with no collagen were treated in an identical manner but without collagen being present in the acetic acid diluent. All THP-1 differentiation and polarisation steps as described above then took place on collagen, or control coated plates.

2.3.5 Co-culture of THP-1 cells and BL2 cells

THP-1 cells were grown on collagen-coated or dummy-coated slides as described above. After three days of growth in the presence of PMA and a further 5 days growth in the presence of polarising cytokines, cells were washed and BL2 cells were added in fresh culture medium at concentrations of 0.1 or 0.2 $x10^6$cells per ml.

2.3.6 Cell counting

Cells were counted on a Beckman-Coulter XL flow cytometer by adding 25 µl of a known concentration of fluorescent beads (Beckman-Coulter, High Wycombe, UK) to a fixed volume of BL2 cells from culture and capturing events until 2000 beads

have been counted. BL2 viable and non-viable counts were obtained by gating on the basis of forward and side scatter.

2.3.7 Assessment of LAIR-1 surface expression by flow cytometry

For each FACS tube 0.5 x 10^6 cells per ml was used. Cells were first washed in HBSS then resuspended in 100µl of FACS buffer (Dulbecco PBS with 5% normal goat serum and 10% normal mouse serum) (PAA, USA). Primary antibody or isotype control was added at the concentrations indicated in Table 2.3 and incubated at 4°C for 30mins. Cells were then washed in FACS buffer and resuspended in FACS buffer containing matching $F(ab')^2$ fragment conjugated to Alexa 488 conjugate (Invitrogen, Paisley, UK). Cells were incubated at 4°C for 30 minutes before washing in FACS buffer and resuspension in FACS buffer. Samples were run on a Beckman-Coulter XL flow cytometer and post-measurement analysis performed using 'Flow Jo' (Tree Star). Histograms of cells were drawn only from cells that fell into a 'viable' gate based on forward and side scatter.

2.3.8 Assessment of LAIR1 intracellular expression by flow cytometry

To assess LAIR-1 expression using an antibody directed against the intracellular component of the molecule cells were fixed and permiabilised prior to staining using a commercial kit 'Fix and Perm' (Caltag, Buckingham, UK). For each FACS tube 0.5 x 10^6 cells per ml was used. Cells were first washed in HBSS then resuspended in 100µl of FACS buffer (Dulbecco PBS with 5% normal goat serum and 10% normal mouse serum). An equal volume of Reagent A (Fix) was added and incubated at room temperature for 15 minutes. Cells were washed in FACS buffer and then resuspended in reagent B (Perm) and Primary antibody or isotype control was added at the concentrations indicated in Table 2.3. Cells were then vortexed to mix and then washed in FACS buffer and resuspended in FACS buffer containing matching $F(ab')^2$ fragment conjugated to Alexa 488 conjugate. Cells were kept in the dark at room temperature for 40 minutes before washing in FACS buffer and

resuspension in FACS buffer. Samples were run on a Beckman-Coulter XL flow cytometer and post-measurement analysis performed using 'Flow Jo' (Tree Star). Histograms of cells were drawn only from cells that fell into a 'viable' gate based on forward and side scatter.

2.3.9 Immunoblotting

2.3.9.1 Preparation of whole cell lysates

Cells were washed 3 times in Dulbeco PBS to remove culture medium and then scraped from the culture dishes with a cell scraper. The cells were collected and lysed with freshly made lysis buffer, 12.5µl of lysis buffer per million cells, and then transferred to a fresh eppendorf. Protease inhibitors were added (P8340, Sigma, Poole, UK) (1µl of inhibitor per million cells), before vortexing for 20 seconds and then leaving on ice for at least 30 mins to lyse. Lysed cells were then spun at 18,000g for 10 mins at 4°C to remove cellular debris. Supernatants were transferred to fresh eppendorfs and frozen at -20°C until used in analysis. Protein concentrations in whole cell lysates were measured using the Bradford assay

2.3.9.2 Bradford Assay

Protein concentrations were measured using Bradford dye reagent (Bio-Rad, Hemel Hempstead, UK). Dilutions of whole cell lysate and serial dilutions of IgG (Sigma, Poole, UK) of known concentration were mixed with the dye reagent and colour concentration measured on a plate reader at 595nm wavelength. A standard curve was created from the measurements of known concentration and protein concentration of cell lysates calculated.

2.3.9.3 Electrophoresis and Western blotting

Electrophoretic sample separation and transfer to nitrocellulose membranes were performed using the Invitrogen NuPage system and reagents according to the manufacturers' instructions (Invitrogen, Paisley, UK). Samples were prepared as follows to give a total volume of 10µl in each well.

Reagent	Reduced sample
Sample	x µl
NuPAGE LDS sample buffer(x4)	2.5µl
NuPAGE reducing agent(x10)	1µl
Deionized water	to 6.5µl
Total volume	10µl

Samples were denatured at 70°C for 10 mins before being loaded into the wells of the precast NuPage 4-12% Bis-Tris gel. SeeBlue Plus 2 prestained standard (Invitrogen, Paisley, UK) was loaded as a molecular weight indicator. Protein separation was carried out in NuPage MOPS SDS running buffer at 150-200v for 30-40 mins. Proteins were transferred onto PVDF membranes in NuPaGE transfer buffer at 30V for 1 hour. Membranes were blocked with 0.1% Tween-20/PBS with 5% non-fat milk and incubated at room temperature for 1 hour. Antibodies were diluted to the required concentration (see table 2.3) in 0.1% Tween-20/PBS with 5% non-fat milk and incubated at 4°C overnight. Membranes were then washed thoroughly in 0.1% Tween-20/PBS and sheep anti-mouse HRP conjugated antibody (GE Healthcare, Amersham, UK) in 0.1% Tween-20/PBS with 5% non-fat milk added and incubated at room temperature for 90 minutes. Membranes were again washed and bands were developed using ECL reagents (GE Healthcare, Amersham, UK). Chemiluminescence was captured using photographic film (GE Healthcare, Amersham, UK).

For staining blots with multiple antibodies, membranes were stripped by incubating twice in stripping buffer for 10 minutes then washing twice with PBS and twice with TBS-Tween before proceeding to the blocking stage.

2.3.10 Quantative reverse transcription PCR

2.3.10.1 RNA extraction from whole cells

RNA extraction was performed using the RNeasy mini kit (Qiuagen, Manchester, UK) following the manufacturers' instructions. Cell lysis was performed on the cell culture plates using RLT buffer and RNA extraction performed as per the handbook including an on-column DNA digestion step. RNA concentration was measured using the Nanodrop system (Thermo Scientific, Loughborough, UK).

2.3.10.2 cDNA synthesis

cDNA was generated using the Supersript III First strand synthesis Supermix for qPCR (Invitrogen, Paisley, UK) following the manufacturer's instructions. Each reaction included 10µl of RT reaction mix, 2µl of RT enzyme mix and RNA and water (Invitrogen) to total volume of 20µl. Samples were mixed and incubated at 25°C for 10 minutes, followed by 50°C for 30 minutes. The reaction was terminated at 85°c for 5 minutes then cooled to 4°C. 1µl of RNA-ase H (Invitrogen) was added to each reaction and incubated at 37°c for 20 minutes. cDNA was then stored at -20°C until use.

2.3.10.3 Quantitative PCR

Quantitative PCR was carried out using the SYBR Green system on the FAST 7500 Real Time PCR system (Applied Biosystems, Thermoscientific, Loughborough, UK).

Primers used;

LAIR1 (from Jansen et al 2007) – (synthesised by eurofins genomics, Ebersberg, Germany)

Forward 5' – CCT GAC CTG GCT GTT GAT GTT CT – 3'

Reverse 5' – GCC CGG GCT GTC CTC TGT – 3'

18S - (synthesised by Sigma, Poole, UK)

Forward 5' – CGA AGA CGA TCA GAT ACC GT – 3'

Reverse 5' – GGT CAT GGG AAT AAC GCCG – 3'

Each reaction was set up containing 10µl of SYBR green master mix (Qiagen, Manchester, UK), 2µl of cDNA, primers at a concentration of 0.3µM and Ultrapure water (Invitrogen, Paisley, UK) to 20µl. The reaction was run on the 7500 Fast Real Time PCR system (Applied Biosystems) using the following parameters; 2 minutes at 50°C once, 15 minutes at 95°C once, then 40 cycles of 15 seconds at 95°C followed by 1 minute at 60°C. Melting curve analysis was performed. 18S was used as the endogenous control. Undifferentiated THP-1 cells were used as the calibrator in performing $\Delta\Delta C_T$ analysis.

2.3.11 Induction and evaluation of apoptosis in BL2 cells

BL2 cells from confluent cultures were washed 3 x 5mins in serum-free RPMI (Invitrogen, Paisley, UK) with 2mM L-glutamine, 100 IU/ml penicillin and 100 µg/ml streptomycin, counted and resuspended in the same medium at a concentration of $1x10^6$ cells/ml. Apoptosis was induced by treatment with Stuarosporine (Calbiochem, Millipore, Watford, UK) at 1µM for five minutes. Cells were them washed 3 times in Dulbecco's PBS without Ca and Mg, once in Dulbecco's PBS with added Ca and Mg and a final time in RPMI (Invitrogen, Paisley, UK), plus 10% foetal calf serum (PAA), 2mM L-glutamine, 100 IU/ml penicillin and 100 µg/ml

streptomycin. Induction of apopotosis in the BL2 cells was assessed after 75 minutes by staining for Annexin V (AxV)-FITC (Life technologies, Thermo scientific, Loughborough, UK) and propidium iodide (PI) (Life technologies), using Beckman-Coulter XL flow cytometer. Briefly, 1 x 10^6 cells were placed in a FACS tube, spun down and re-suspended in 100 µl of Annexin V binding buffer (10 mM HEPES/NaOH (pH 7.4); 140 mM NaCl; 2.5 mM CaCl2). 1 µl of Annexin V-FITC was added per tube and incubated for 10 min on ice. Cells were next washed and re-suspended in 500 µl of binding buffer. 10 µl of propidium iodide was added to each tube prior to FACS measurement.

2.3.12 Co-culture of apoptotic BL2 cells with THP-1 cells

THP-1 cells were previously differentiated for 8 days as described above. BL2 cells induced to undergo apoptosis were added at three concentrations of 10,000 cells/ml, 100,000 cells per ml and 1 x10^6cells per ml. Control culture conditions included co-culture with BL2 cells not induced to undergo apoptosis with staurosporine, and culture of THP-1 cells with supernatant derived from the last wash of the staurosporine treated cells to exclude any effect of residual struarosporine on the THP-1 expression of LAIR1. Cells were cultured in 5% CO_2 at 37°C. To assess whether co-culture with apoptotic cells altered expression of LAIR1 surface staining was assessed by flow cytometry using an antibody directed at the extracellular component of the molecule as described above at various times points ranging from 2 to 24 hours.

Table 2.3 Antibodies and dilutions used.

Antibody	**Catalogue number**	**Supplier**	**IHC**	**Flow cytometry**	**Western blotting**
Primary antibodies					
Rabbit polyclonal anti-LAIR-1 (clone HAP011155)	HPA011155	Sigma	**1:300**	**1µg/ml**	**1:5000**
Mouse monoclonal anti-human LAIR-1 (clone DX26) (IgG1)	550810	BD Pharmingen	-	**10µg/ml**	**1:1000**
Mouse monoclonal anti-human LAIR-1 (clone NKTA255) (IgG1)	NB100-66540	Novus biologicals	-	**10µg/ml**	**1:1000**
Mouse monoclonal anti-CD68 (clone PG-M1) (IgG3)	Ab783	abcam	**1:50**	-	-
Rabbit polyclonal anti-LGALS3BP	HPA000554	Sigma	**1:400**	-	-
Rabbit polyclonal anti-ECGF1 (TYMP)	HPA001072	Sigma	**1:100**	-	-
Goat polyclonal anti-Iba1 (AIF1)	Ab5076	abcam	**1:100**	-	-
Mouse monoclonal anti-CD163 (IgG1)	NCL-CD163	Leica	**1:100**	-	-
β-actin	A5441	Sigma	-	-	**1:1000**
Negative controls					
Mouse IgG1	MCA928 NCG01 X0931	AbD Serotec Abcam DAKO	**Equivalent concentration to primary antibody**	**Equivalent concentration to primary antibody**	
Goat IgG	I 9140	Sigma	**Equivalent concentration to primary antibody**		
Rabbit Immunoglobulin	X0936	Dako	**Equivalent concentration to primary antibody**	**2µg/ml**	
Secondary conjugated antibodies/F(ab')$_2$ fragments					

Goat anti-rabbit F(ab')$_2$ fragment, Alexa 488 conjugated.	A-11077	Invitrogen	-	**1:100**	
Rabbit anti-mouse F(ab')$_2$ fragment, Alexa 488 conjugated.	A-21204	Invitrogen	-	**1:100**	
Goat polyclonal anti-rabbit immunoglobulin/HRP conjugated	P0488	DAKO	**1:100**		**1:2000**
Rabbit polyclonal anti-goat immunoglobulin/HRP conjugated	P0449	DAKO	**1:100**		
Anti mouse immunoglobulin/HRP conjugated.		DAKO	**1:100**		
Sheep anti-mouse IgG peroxidise linked	NA931	ECL GE Healthcare	**1:100**		**1:1000**

Table 2.4 Methods of antigen retrieval, blocking and incubation times for immunohistochemistry. All incubations are at room temperature unless otherwise stated.

Target	Antigen retrieval	Blocking agent	Primary antibody	Primary antibody incubation time (mins)	Secondary antibody	Secondary antibody incubation time (mins)
LGALS3BP	None	3% H_2O_2 in PBS	Anti-LGALS3BP (Sigma HPA 554)	105	DAKO goat anti-rabbit – HRP	60
LAIR1	Sodium citrate buffer with microwave heat retrieval	DAKO peroxidise block	Anti-LAIR-1 (Sigma HPA011155)	120	DAKO goat anti-rabbit - HRP	30
TYMP – single colour IHC	Vector antigen unmasking solution with microwave heat retrieval	3% H_2O_2 in PBS	Anti-ECGF1 antibody (Sigma HPA 001072	16 hours at 4°C	DAKO goat anti-rabbit - HRP	60
TMYP - two colour IHC	Vector antigen unmasking solution with microwave heat retrieval	3% H_2O_2 in PBS	Anti-ECGF1 antibody (Sigma HPA 001072	16 hours at 4°C	DAKO goat anti-rabbit - HRP	60
CD163	Vector antigen unmasking solution with microwave heat retrieval	3% H_2O_2 in PBS	Anti-CD163 (Leica Novocastra NCL-CD163	60	DAKO anti mouse HRP	60
CD163 – two colour IHC	Vector antigen unmasking solution with microwave heat retrieval	3% H_2O_2 in PBS	Novocastra NCL-CD163	16 hours at 4°C	GE Healthcare sheep anti-mouse-HRP	60
CD68	Tris-EDTA buffer with microwave heat retrieval	3% H_2O_2 in PBS	Mouse anti human CD68 (PGM1) – Abcam (ab783	120	Sheep anti-mouse HRP (– GE healthcare NXA931)	30
AIF1	Sodium citrate buffer with microwave heat retrieval	DAKO peroxidise block	Anti –lba1 Abcam ab5076	60	Rabbit anti – goat HRP conjugate (DAKo P0449)	60

2. 4 Reagent list

Acetic acid	Sigma-Aldrich
Antibody diluent	DAKO
Antigen unmasking solution (citrate)	Vector
Annexin V-FITC	Life technologies
BSA (low endotoxin)	PAA
Bradford dye	Bio-Rad
Collagen IV (human)	Sigma-Aldrich
DAB (3, 3'-diaminoebenzidine)	Vector
DAKO antibody diluent	DAKO
Dulbecco's PBS	PAA
DMSO	Sigma-Aldrich
dNTP mix	Sigma-Aldrich
EDTA	Sigma-Aldrich
ECL chemoluminescence reagents	GE Healthcare
$F(ab)^2$ Goat anti-mouse IgG (H+L) TRI-COLOUR conjugated	Invitrogen
FCS	PAA
Flow count fluorespheres	Beckman Coulter
Fix and Perm cell permiabilisation reagent.	Caltag Laboratories
Harris haematoxylin	Sigma-Aldrich
HEPES	Sigma-Aldrich
HBSS	Thermo Scientific
Histoclear II	National Diagnostics, USA
Hydrogen peroxide	Sigma-Aldrich
Hydrochloric acid	Sigma-Aldrich
Industrial methlylated spirit (IMS)	Sigma

LPS	Sigma
L-glutamine	PAA
MicroAmp Fast optical 96 well plate	Applied Biosciences
MicroAmp Optical adhesive film	Applied Biosciences
NuPAGE LDS sample buffer	Invitrogen
NuPAGE reducing agent	Invitrogen
NuPAGE 4-12% Bis –Tris precast gel	Invitrogen
NuPAGE MPOS SDS running buffer	Invitrogen
NuPAGE transfer buffer	Invitrogen
PBS	Sigma-Aldrich
Penicillin/streptomycin	Invitrogen
Peroxidase block	DAKO
PMA	Sigma
Propidium iodide	Life technologies
Protease inhibitor (P8340)	Sigma
QuantiTect SYBR green PCR kit	Quiagen
RPMI 1640	Invitrogen
RNeasy Mini kit	Quiagen
RNase free DNAase set	Quiagen
SDS	Sigma-Aldrich
SeeBlue plus 2 prestained standard	Invitrogen
Sodium bicarbonate	Sigma-Aldrich
Staurosporine	Calbiochem
Superscript III First strand synthesis mix for qPCR	Invitrogen
SYBR green master mix	Quiagen
Tris-base	Sigma-Aldrich
Tri-sodium citrate	Sigma-Aldrich

Triton –x-100	Sigma-Aldrich
Tween-20	Sigma -Aldrich
UltraPure™ DNase/RNase-Free Distilled Water	Invitrogen
VIP peroxidise substrate	Vector
Xylene	Sigma-Aldrich

Antibodies

IgG1 (mouse) MCA928	Serotec
IgG1 (mouse) NCG01	Abcam
IgG (goat)	Sigma-Aldrich
Immunoglobulin fraction (goat)	DAKO
Anti -LAIR-1 (HPA011155)	Sigma
Anti –LAIR-1 (DX26)	BD Bioscience
Anti-LAIR1 (NKTA255)	Novus Biologicals
Anti-CD68 (PG-M1)	abcam
Anti-LGALS3BP (HPA000554)	Sigma
Anti-ECGF1 (HPA001072)	Sigma
Anti-lba1 (ab5076)	abcam
Anti-CD163 (NCL-CD163)	Leica

Cytokines

Human IL-10	R&D
Human IFNγ	R&D
Human IL-4	R&D

Solutions

Lysis buffer

10mM HEPES
1mM EDTA
1% Triton – X-100
H_20

Membrane stripping buffer

15g Glycine
1g SDS
10ml Tween-20
Made up to 1L with distilled water.
pH adjusted to 2.2 by addition of hydrochloric acid

Detachment buffer

HBSS (with no Mg+ or Ca+)
5mm EDTA
0.2% BSA

FACS buffer

Dulbecco's PBS (with no Mg+ or Ca+)
5% normal goat serum
10% normal mouse serum

Annexin V binding buffer

10mM HEPES
140mM NaCl
2.5mM CaCl
Adjust pH to 7.4 with NaOH

Sodium citrate buffer

10mM sodium citrate
0.05% Tween -20
Adjust pH to 6.0 with hydrochloric acid

Tris-EDTA buffer

10mM Tris base
1mM EDTA
0.05% Tween-20
Adjust pH to 9.0

Chapter 3

Relationship of proliferation, apoptosis and macrophage infiltration in Diffuse Large B cell lymphoma.

3.1 Introduction

Previous work within the group (Ford et al 2015) has demonstrated a significant linear relationship between rates of proliferation, apoptosis and macrophage infiltration in Burkitt lymphoma. Burkitt lymphoma as currently classified in the WHO Classification of Tumours of Haemopoetic and Lymphoid tissue, 2008 is a relatively tightly defined entity that must meet defined clinical, histological and genetic parameters to be diagnosed as such (Swerdlow et al 2008). The tumour is defined in part by the characteristic (although not invariable) presence of a *myc* translocation resulting in high rates of proliferation (often approaching 100%), and consequent high rates of apoptosis. Morphologically, the tumour is characterised by a dense macrophage infiltrate, these macrophages containing abundant apoptotic debris, reflecting the incredibly high rate of cell loss in these tumours. In these tumours therefore, there is a constant presence of tumour-associated macrophages, which are visibly involved in clearance of apoptotic cells and as such are likely to be functionally modified by these apoptotic cells. Much work within the group has focussed on how these apoptotic cells may modulate the tumour microenvironment to provide a growth advantage to the tumour. Various mechanisms including release of lactoferrin and fractalkine from apoptotic cells and production of BAFF from TAMs have been demonstrated (Bournazou 2009, Truman et al 2008, Ogden et al 2005).

This chapter seeks to explore whether a similar relationship exists between apoptosis, proliferation and macrophage infiltration in DLBCL, with the resultant influences by apoptotic cells on macrophage function, to confirm that the phenomenon is not unique to Burkitt lymphoma. In contrast to Burkitt lymphoma, DLBCL as currently defined is a much more heterogeneous group of diseases than Burkitt lymphoma, lacking a common clinical presentation, or genetic abnormality and containing tumours from a wider morphological spectrum. Like Burkitt lymphoma, these tumours tend to be characterised by aggressive behaviour. They show a much wider distribution of rates of proliferation, from relatively low to close to 100%. In some cases, and often in cases with very high proliferation, there is evidence of a myc translocation, but this is not a universal finding and many other genetic abnormalities are also described. The relationship between proliferation and apoptosis would therefore be predicted to be somewhat more complex than in Burkitt lymphoma, but if apoptosis is important in recruitment of macrophages to the tumour site to help support tumour cell growth, then the same relationship would be predicted to occur.

To address this question, a cohort of 82 cases of DLBCL, representing all cases suitable for inclusion in a tissue microarray from two regional pathology laboratories (Tayside and Highland) over a 10 year period were examined for expression of activated caspase-3 as a marker of apoptosis, CD68 as a marker of macrophages and ki67 as a marker of proliferation. The cohort of cases was obtained from the Tayside tissue bank and built into a TMA in which each case was represented at least twice and up to 4 times. Cases had been divided into either 'germinal centre' type or 'non-germinal centre' type (roughly equating with the activated B cell type defined by gene expression analysis) using the immunohistochemical algorithm of Hans (Hans et al, 2004). There was also information available for all cases as to whether there was a t(14:18) translocation in the tumour bringing bcl-2 under the control of the immunoglobulin heavy chain promoter, and predicted to render the tumour cells more resistant to apoptosis. For each characteristic studied, an index was calculated based on number of positive cells expressed as a percentage of total cells.

3.2.1 Rates of proliferation but not apoptosis or macrophage recruitment differ between subtypes of DLBCL

Rates of proliferation, apoptosis and macrophage infiltration were calculated for all tumours in the microarray. There was a wide variation in the rates of all of these between individual tumours reflecting the heterogeneous nature of the individual tumours and their hosts (see figure 3.1).

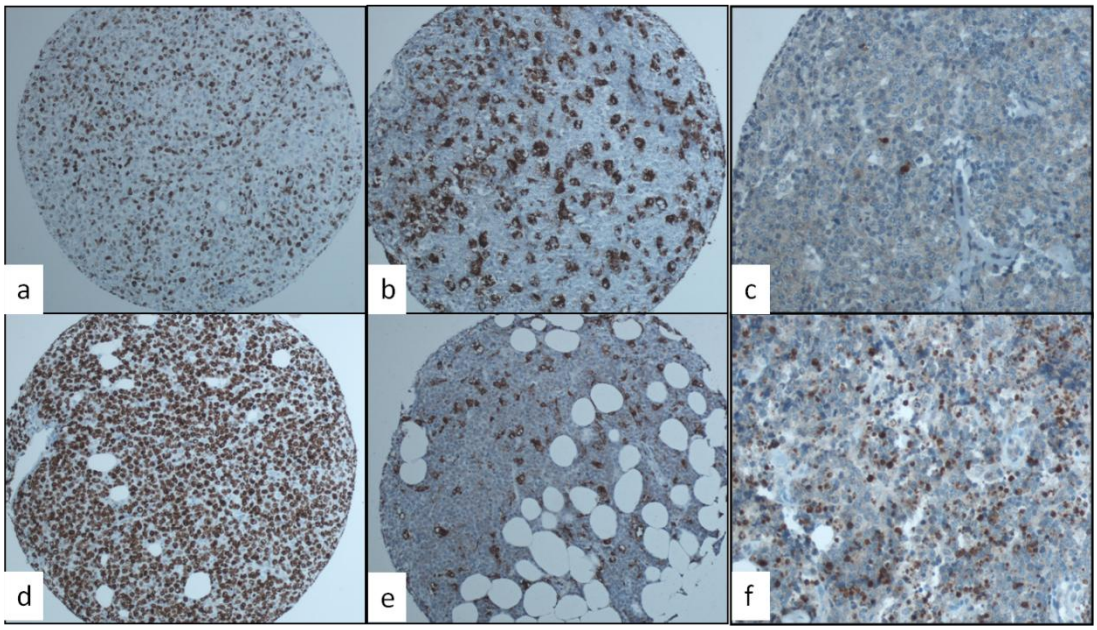


Figure 3.1. Representative photomicrographs to show examples of rates of proliferation [a) low proliferative index, d) high proliferative index], macrophage infiltration [b) high macrophage infiltration, e) low macrophage infiltration] and apoptosis [c) low apoptotic index, f) high apoptotic index]. All images represent different tumours. Magnification in images a-d = x40 and images e &f = x100.

Analysis of medians across the groups demonstrated the germinal centre-type of DLBCL without bcl-2 translocation to show significantly higher rates of proliferation than germinal centre-type with the translocation or non-germinal centre type (Kruskall-Wallis test for differences in rates of proliferation. $p = 0.026$), but there were no significant differences in rates of apoptosis or macrophage infiltration (see figure 3.2 and table 3.1 on page 68). This is somewhat at variance from what might have been predicted based on the known genetic mutations of these tumours. Whilst

high rates of proliferation in GC-type cases without the bcl-2 translocation is not unexpected, it would have been predicted that cases with the translocation would show a lower rate of apoptosis and this was not shown to be the case. Rates of apoptosis varied much more markedly in GC tumours lacking the bcl-2 translocation than others, but there were no significant differences between groups. In part, this may reflect the generally low rate of apoptotic cells seen, with a median of 1.4% of all cells being positive for activated caspase-3. This does not necessarily imply a low rate of apoptosis as clearance of apoptotic cells occurs very rapidly and potentially differences in rates of apoptosis between tumour types may be masked by differences in efficiency in macrophage clearance (reviewed in Gregory and Pound, 2011). An alternative explanation would be that despite the role of bcl-2 in protecting cells from apoptosis, expression of this by DLBCL does not alter the rate at which apoptosis occurs in vivo. The results contrast with those of Bai et al who found higher rates of apoptosis in germinal-centre type tumours but no differences in proliferation between groups in study of 79 cases of nodal and extranodal DLBCL subdivided into GC and non-GC type (Bai et al, 2004). In this study no information on translocations was included.

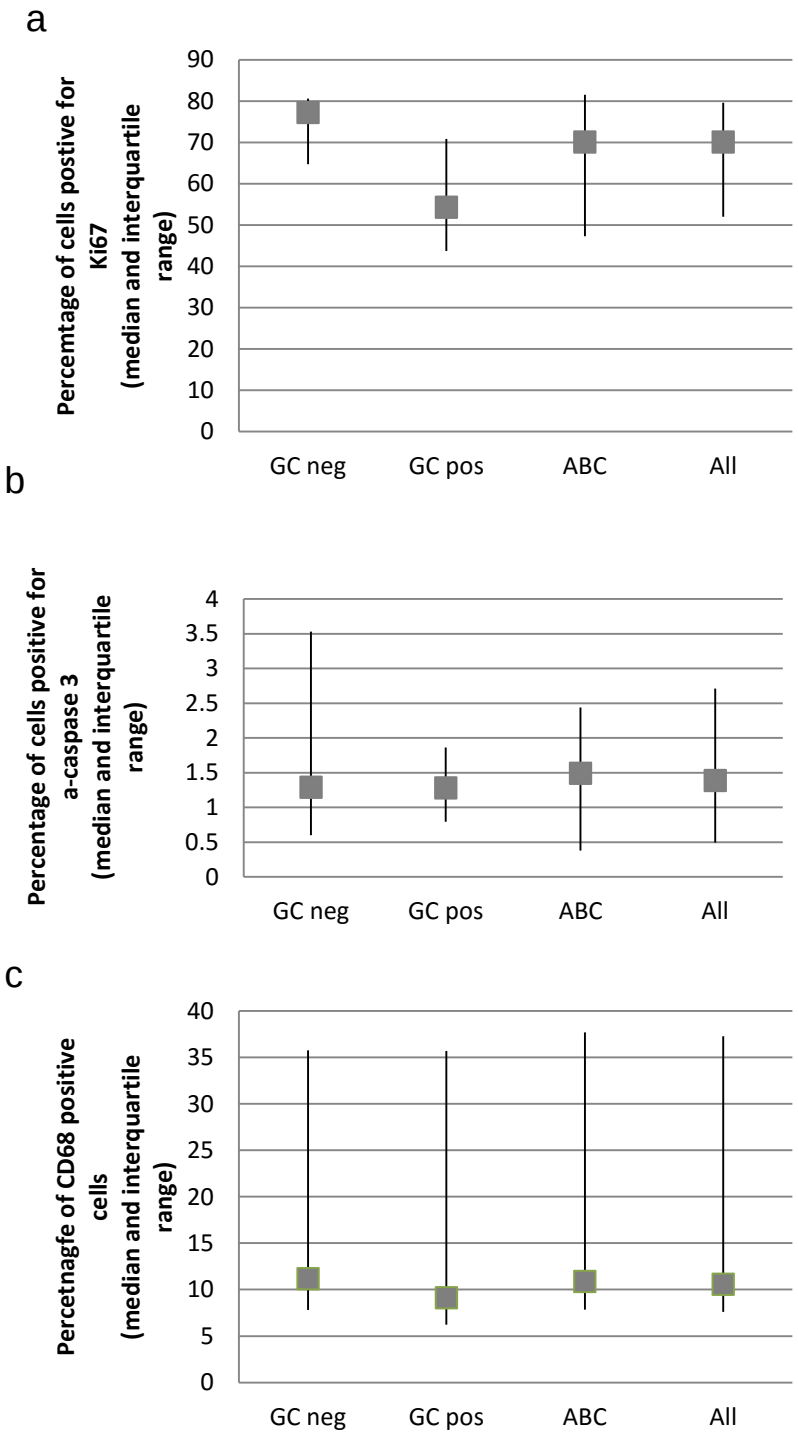


Figure 3.2 a) Median proliferative index (plus interquartile range) for all cases of DLBCL and subtypes of DLBCL. Proliferative indices varied significantly between groups on Kruskall Wallis test (p=0.026). b) Median apoptotic index (plus interquartile range) for all cases of DLBCL and subtypes of DLBCL. There were no significant differences between groups. c) Median macrophage index (plus interquartile range) for all cases of DLBCL and subtypes of DLBCL. There were a wide range of values between tumours but the median for all groups was low. There were no significant differences between groups). GC-neg (n=25) = germinal centre-type cases without the t(14;18) translocation, GC-pos (n=13) = germinal centre-type cases with the t(14:18) translocation, and ABC (n=44) = activated B cell-type or non-germinal centre type.

	Proliferative index Median (IQR)	**Apoptotic index** Median (IQR)	**Macrophage index** Median (IQR)
All cases *n=82*	70 (52-80)	1.4 (0.5-2.7)	11 (8-15)
Non-GCB *n=44*	70 (47-81)	1.5 (0.4-2.4)	11 (8-13)
GCB t(14:18) negative *n=25*	77 (64-81)	1.3 (0.6-3.5)	11 (8-15)
GCB t(14:18) positive *n=13*	54 (44-71)	1.3 (0.8-1.9)	9 (6-16)
Kruskall-Wallis test for differences between groups. (p-value)	0.026	NS	NS

Table 3.1: Summary of variables measured across all tumours assessed. All values are number of positive cells as a percentage of total cells. Differences between groups was assessed by Kruskall-Wallis test. p-values <0.05 were considered significant. (GCB= Germinal Centre B-cell subtype. NS = Not significant).

3.2.2 There is a significant correlation between rates of apoptosis and proliferation, and apoptosis and macrophage infiltration in DLBCL

The relationship between these variables was studied in the cohort of DLBCL cases. Scatter plots in Figure 3.3 show the distribution of variables between individual cases. Associations between these variables were analysed using Pearson correlations. When proliferation and apoptosis was analysed there was a significant correlation (Correlation coefficient r=0.408, p-value <0.001) between proliferation and apoptosis when studied as a whole cohort. This association remained, when cases were subdivided into smaller cohorts of GC-type or non GC-type, but was lost

if the cases were further subdivided into the smaller groups of GC-type divided by presence or absence of bcl-2 translocation. (Table 3.2). Similarly a correlation existed between apoptosis and macrophage infiltration (Correlation coefficient $r=0.238$, p-value <0.05). No such correlation existed between macrophage infiltration and proliferation.

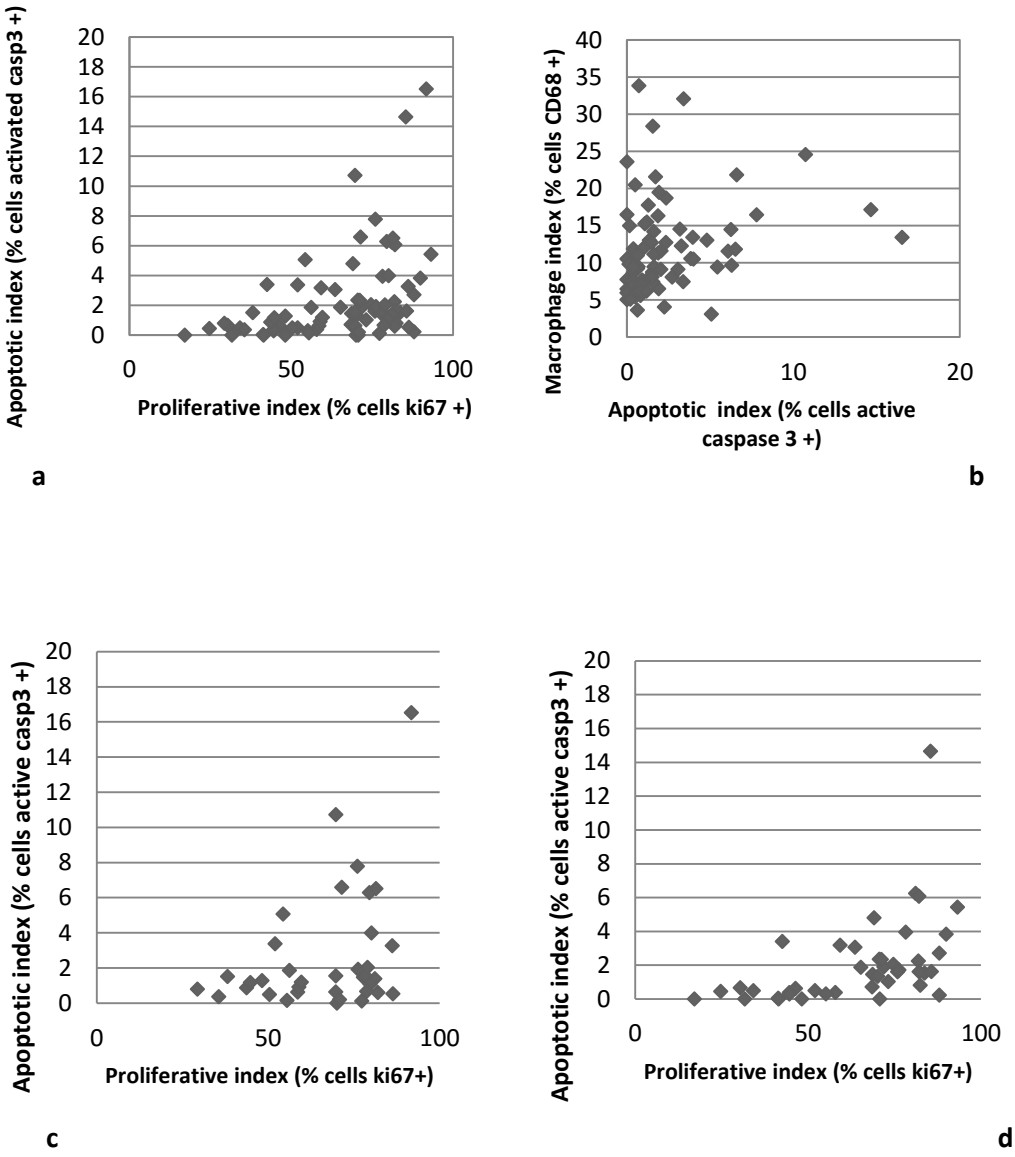


Figure 3.3: Scatterplots demonstrate relationships between apoptosis, proliferation and macrophage infiltration. Each dot represents the mean of two or more cores per tumour present on the array. a) All cases – Proliferative index vs Apoptotic index, b) All cases – Apoptotic index vs Macrophage index, c) GCB cases only – Proliferative index vs apoptotic index, d) Non-GCB cases only – Proliferative index vs Apoptotic index.

This data would support a relationship between proliferation and apoptosis, regardless of the underlying genetic abnormalities of the tumour. Such an

association would be predicted given the known propensity of cells in cell cycle to undergo apoptosis, and would argue that in DLBCL, the cell cycle is not deregulated to the extent that this property is lost (Evan et al, 1992). Similarly, although the association is less strong, there is a clear relationship between apoptosis and macrophage recruitment/retention to the tumour. Again, based on the well characterised role of macrophages as professional phagocytes, and their attraction to apoptotic cells, this relationship was predicted to occur. The data in relation to the association of macrophage infiltration is more difficult to interpret. If, as hypothesised, macrophages recruited and programmed by apoptotic cells support cell proliferation, then an association between these variables would be predicted. The absence of such an association however, does not necessarily imply that this supportive relationship does not in fact exist, but any such support may be only one of multiple factors affecting proliferation in DLBCL. The relative complexity and heterogeneity of the tumours that fall within the classification of DLBCL would argue that there are many influences at play in these tumours. To detect any such association may require study of much larger cohort of cases, which would of necessity require a multi-centre study to collect sufficient cases of this relatively rare tumour type.

	All tumours (n=82) Pearson correlation (r)	*p-value*	Non-GC (n=44) Pearson correlation (r)	*p-value*	GC (n=38) Pearson correlation (r)	*p-value*	GC t(14:18) negative (n=25) Pearson correlation (r)	*p-value*	GC t(14;18) positive (n=13) Pearson correlation (r)	*p-value*
Proliferative Index vs Apoptotic Index	0.408	<0.001	0.477	<0.01	0.351	<0.05	0.364	NS	0.338	NS
Proliferative Index vs Macrophage Index	0.087	NS	0.049	NS	0.148	NS	-0.096	NS	0.315	NS
Apoptotic Index vs Macrophage Index	0.237	<0.05	0.191	NS	0.295	NS	0.368	NS	0.138	NS

Table 3.2: The relationship between proliferation, apoptosis and macrophage infiltration were assessed for the whole cohort and for sub-groups within the cohort by Pearson correlation coefficients (r). Significant correlations were shown to exist between proliferation, apoptosis and macrophage infiltration. P-values of <0.05 were considered to be significant. (NS = Not significant)

Summary

This chapter examined proliferation, apoptosis and macrophage infiltration in a large cohort of cases of human DLBCL. It demonstrated variable rates of all factors in this heterogeneous collection of cases.

- There were no significant differences between these variables in subtypes of tumours with the exception of proliferation which varied significantly between germinal centre-type tumours with the t(14;18) translocation, germinal centre-type tumours without the t(14;18) translocation and the non-germinal centre cell type of tumour.
- There was a significant correlation between rates of proliferation and apoptosis, and rates of apoptosis and macrophage infiltration in these tumours providing support for the hypothesis, that as in Burkitt lymphoma, one of the roles of apoptotic tumours cells is to recruit and potentially modulate tumour-associated macrophages for the benefit of the tumour.

Chapter 4

A bioinformatic approach to deriving gene expression signatures from human tumour data.

Aims of the chapter

Having established that there was a relationship between proliferation, apoptosis and macrophage infiltration in DLBCL as in Burkitt lymphoma, this study then focussed on the role of the tumour associated macrophage in the tumour microenvironment. The role of the tumour associated macrophage is clearly an important one in many types of tumour, but it is a very difficult cell to study due to its inherent plasticity as any analysis in which the macrophage is removed from its microenvironment will inevitably alter the cell. For this purpose, a gene expression signature of the TAM in situ in human tumours was sought. Work by another member of the group focussed on microdissection of TAMs from a mouse model of lymphoma to generate unaltered TAMs (Ford et al 2015). This approach was not suitable for use on human lymphoma tissue which could not be handled in the stringent way required to extract mRNA of sufficient quality to make the analysis possible, and would be derived from a genetically unrelated and heterogeneous population requiring large numbers of specimens to be available for meaningful analysis. An alternative approach was sought and in an effort to generate a tumour associated macrophage signature in an unbiased way, without either disrupting the tissue to remove the tumour associated macrophage or creating an artificial microenvironment *in vitro*, a bioinformatic approach was used. A novel tool, Biolayout *Express* 3D, being developed by Professor Tom Freeman's group at the Roslin Institute, was utilised for this analysis and some additional functions created to allow the analysis of core signatures (Freeman et al 2007, Theocharidis et al 2009). This approach, which defines clusters of genes based on their patterns of co-expression, was used to generate

signatures of co-expressed genes within cancer datasets including signatures of elements of the cancer microenvironment, including the tumour associated macrophage. This chapter presents the bioinformatic analysis from a broad spectrum of human tumours as well as common 'core signatures' derived from this data. The rationale behind the analysis strategy is presented and then individual analyses from individual tumour types presented.

The biolayout graphs presented in this chapter are available in an interactive Biolayout-ready format at the website www.Oncograph.org and selected gene lists for all datasets analysed are included in Appendix 1 as a CD. A paper arising from this work is included as Appendix 2 (Doig et al, 2013).

4.1 Introduction

There is an enormous wealth of gene expression data that has been made publically available covering a wide range of human tumours and including samples from huge numbers of patients, now running to many thousands. The challenge has been to analyse this data which runs to many billions of individual datapoints, and is inherently complex and heterogeneous. Multiple strategies have been derived, each with their individual strengths and weaknesses, and hence comparison of bioinformatic analyses of similar data by different methods often yields very few similarities (Iwomata and Psztai 2010). One of the problems in analysis of the data is the seemingly simple task of visualisation of the data in the way that allows the human brain to see and make sense of patterns in an incredibly complex set of data. Increasingly graphical representation of the data is being pursued although this also presents challenges due to the sheer volume of the data. Biolayout 3D is a 3-dimensional graphing tool to allow visualisation of complex data that is being developed by Professor Tom Freeman's group in which it is possible to create graphs large enough to explore an entire gene expression experiment.

The primary human tumour datasets that are available, in contrast to published gene expression experiments from in vitro experiments, contain gene expression

signatures derived from fragments of real tumours and as such are not pure populations of tumour cells but contain all the elements of the tumour microenvironment such as extracellular matrix, blood vessels and lymphatics, stromal cells and inflammatory cells, and frequently also contain contamination from adjacent normal tissues. Given the large numbers of gene expression studies of intact human tumour tissue that have been undertaken to date, it is clear that there must exist in these datasets a gene expression signature of the tumour-associated macrophage; the challenge is to dissect it from the surrounding signatures. Using the ability of Biolayout to visualise very large complex datasets, and cluster these into highly related cliques of genes, signatures with a high degree of co-expression were isolated from human tumour datasets. These signatures were extensively explored using tools to assess gene set enrichment analysis, comparison with known signatures from isolated cell types and review of the literature. Signatures unique to individual tissues and tumour types and common to all tumours were identified including signatures related to the immune response and extracellular matrix.

4.1.1 Correlation as a measure of relatedness of gene expression

In any given individual tissue or tumour there will be influences that drive the expression of various gene 'signatures' or collections of genes. The driving influences behind these signatures are broad, ranging from collections of genes under the control of certain extracellular signals to genes stably expressed by a particular cell type. Regardless of the driving nature of the signature, if these genes are expressed in a co-ordinated fashion, then they would be predicted to rise and fall in association with each other. Thus a signature driven by a signalling molecule would be predicted to rise in the presence of that activating signal or a cell signature be predicted to reflect the relative abundance of that cell type (Figure 4.1). This co-ordinate expression of genes, if viewed across a large dataset would be expected to rise or fall in a coordinated pattern. The degree of relatedness of the pattern of expression of individual genes can be calculated as a measure of their co-expression. A range of variables have been used as a measure of this relatedness including Spearman Rank Correlation, and Pearson correlation coefficient. In the data

presented below co-expression is measured using the Pearson correlation coefficient, a measure that takes into account only the shape of the data considered and not the amplitude of the signal and should therefore capture low abundance signal that correlates with other gene expression as well as genes expressed at a high level in any given cell/pathway/signature. To create the 3-D graph networks used in Biolayout, a Pearson correlation is created for every single probe on a gene expression array versus every other probe on the array so that there is a measure of relatedness for every single gene covered by the analysis with every single other gene in the analysis.

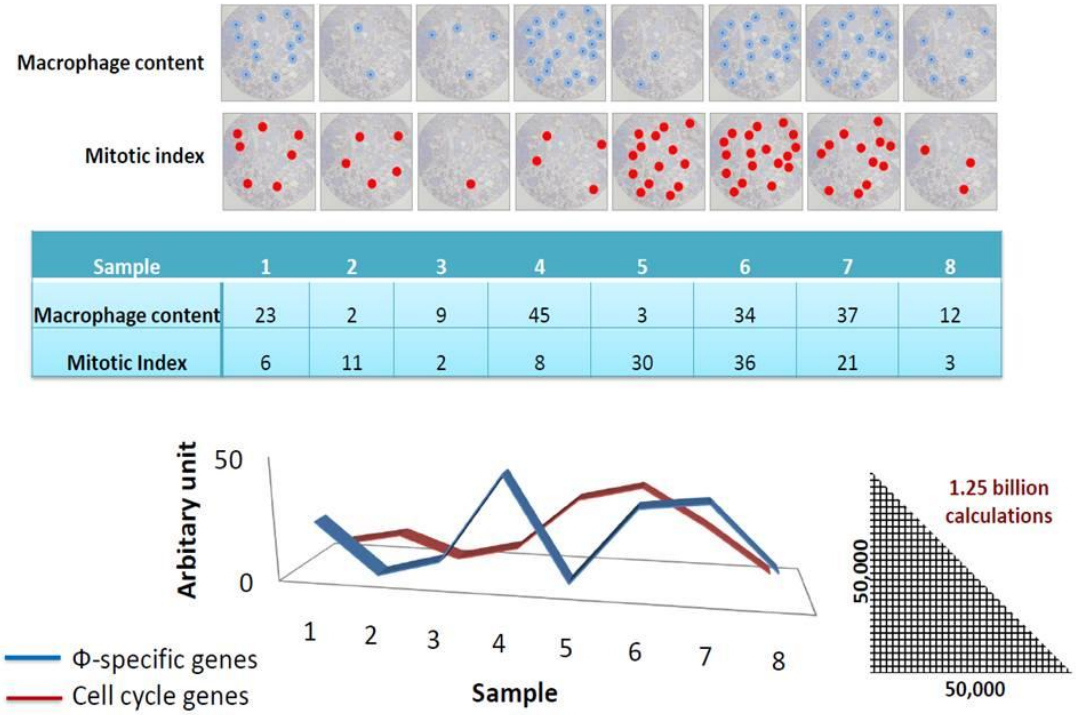


Figure 4.1. Rationale behind the study. The relative number of a specific cell type or activity of certain pathways will vary across a collection of individual tumours. For example, the macrophage content (Φ) will differ in every tumour and so therefore will the mRNA level of macrophage specific genes (in blue). Similarly in every tumour at the point it is sampled the number of cells in mitosis (the mitotic index) will differ and this will be reflected in different levels of expression of cell cycle genes (in red). As a result the expression level of genes specifically expressed by those cells or associated specifically with the pathways will vary accordingly. By calculating the correlation coefficient between every gene on the array and every other gene on the array it is possible to calculate a correlation matrix that includes all these correlation coefficients. Graphs are then used to visualise relationships above a given correlation threshold and clustering used identifying groups of co-expressed genes.

4.1.2 Graphical representation of complex data

Graphical representation of complex data is increasing used in an effort to render large and complex datasets in a form in which the human eye and brain are able to discern and interpret patterns (Barbaris et al 2011, Huber et al 2007, Pavlopoulos et al 2008). Biolayout analysis uses a 3-dimension graphing approaching to visualisation and analysis of data by creating a 3-dimensional graph in which genes with similar patterns of expression will lie close to each other in space. Briefly, it uses a modified Fruchterman-Rheingold algorithm in 3-dimensional space in which nodes representing genes/transcripts are connected by weighted, undirected edges representing correlations above the selected threshold. Each node finds its place in the graph dependent on which other genes it is highly correlated with, these correlations forming the edges that connect the nodes. The degree of correlation that is used as the threshold is dependent on the nature of the data and question posed. Setting a very low correlation threshold will ensure that no data is lost, but will increase the risk of spurious associations, increase the complexity of the graph and obscure the relationships within the graph as well as rendering the size of the graph unmanageable. Use of a correlation threshold that is very high will result in a much smaller graph, with clearly co-expressed genes forming tight clusters but will result in the exclusion of much of the data. Such a threshold would be appropriate in analysis of experimental data from a controlled experiment where there should be relatively little variation in gene expression, but would lose much of interest in the setting presented here where the data is derived from genetically diverse individuals with genetically diverse tumours. The numbers of nodes and edges generated at each correlation threshold above 0.5 for the datasets used here is given in figure 4.2 to demonstrate how these relationships vary in individual datasets. The use of a 3-D graph network allows a global view of the data where all the relationships between genes/transcripts can be visualised and explored. In addition, 'noise' in the data is excluded by the nature of the graph; random, non-biologically relevant associations between genes will be unlikely to form structures in the graph as it is implausible that if a spurious correlation exists between a non-relevant gene (x) and one which forms

part of the graph network (*y*) that the non-relevant gene will also be highly correlated with the genes that surround *y* and as such will form part of the graph network. For this reason, it is not necessary, as it is in many other data analysis approaches to gene expression data, to apply a fluorescence amplitude threshold above which gene expression data is considered to be real rather than representing an artefact of the experiment, as low-signal 'noise' will be naturally excluded from the graph while true low amplitude gene expression should be retained.

In all the graphs presented below the majority of the data is represented in one complex network that occupies one corner of the space available for the graph layout with multiple unconnected clusters mostly of small numbers of highly connected genes/probesets occupying the rest of the space. These smaller clusters represent probesets/genes that show a high degree of correlation with the other probesets/genes in that cluster but little correlation with all the other probesets/genes in the analysis and hence do not form part of the main graph. As such they mostly do not inform the analysis of the overall gene expression patterns in the tumour and most of the data presented here will focus on the large graphs in each dataset.

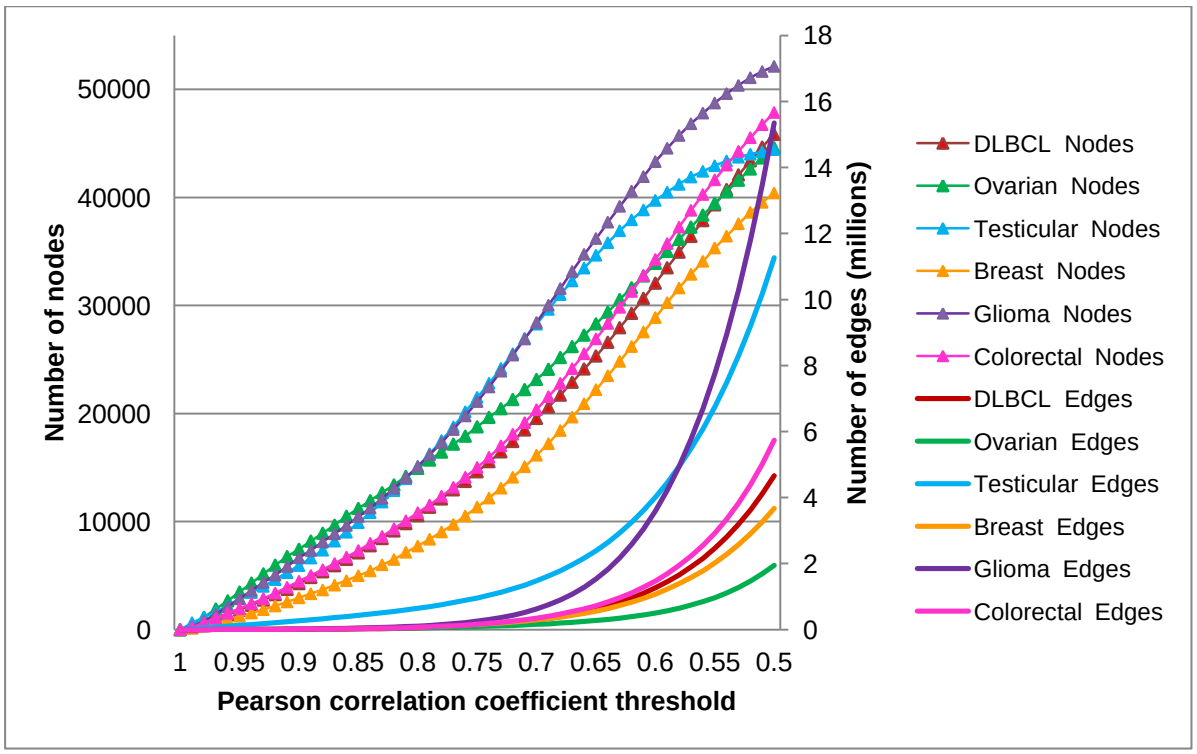


Figure 4.2. Effect of correlation co-efficient on graph size. As the correlation coefficient at which data is included in the graph is decreased (x-axis), the number of nodes and edges increases. The effect differs in every individual dataset dependent on various factors inherent to each dataset. The six datasets used in the main analysis in this chapter are presented here (see table 4.1 for dataset details). It can be seen that for both the testicular and glioma datasets there is a sharp increase in the number of edges formed under a correlation coefficient cut-off of 0.65, while the ovarian tumour dataset adds smaller number of edges as the correlation coefficient is decreased.

4.1.3 Markov clustering algorithm

The Markov clustering algorithm is employed by Biolayout to divide the complex 3-dimensional graphs into cliques or clusters of genes that have a high degree of interconnectivity, with all the elements within the cluster having a similar expression pattern and thus being correlated with many or all of the other elements within the cluster. The algorithm simulates multiple iterations of random flow through the graph to define areas of high connectivity which make up the clusters. There are no predefined number of clusters to be derived from the data, with the number of

clusters being determined by the interconnectivity of the data. The only user defined feature of the clustering is the granularity which determines whether the data divides into many smaller clusters or fewer larger clusters. In all the data presented below an MCL inflation value (which determines the granularity) of 2.2 was used, which has been empirically determined in previous analysis of large datasets to generate meaningful data. Not all nodes within the graph network are assigned to a cluster. Those which lie outwith the areas of tight connectivity that define the clusters are not allocated to a cluster. These nodes often lie within the complex graph network and adjacent to clusters implying that there is a high degree of correlation of expression pattern with the genes within the cluster, but insufficient relationship with the other genes within the cluster to form the multiple connections required to lie within the cluster. There is still scope to analyse these genes and their relationship to the genes that lie close to them within the Biolayout programme where the 'nearest neighbours' of individual genes or groups of genes can be visualised.

4.1.4 Use of Biolayout to extract cell specific signatures

A hypothesis was made; that there must be a signature or signatures that define macrophages to make them the unique cell type they are; that this signature must represent a group of co-expressed genes and that is therefore extractable by clustering groups of genes that are co-expressed. In this analysis of large human cancer datasets the variability and complexity of the data becomes an advantage rather than a problem as the variation in tissue composition, relative abundance of tumour cells to other cells etc, that confound gene expression analysis focussed on the tumour cells provides the patterns in the data that powers the analysis. Large datasets were selected for analysis meaning the data presented in this chapter encompasses many genetically unique individuals totalling more than eleven hundred patients and also providing sufficient variability and complexity in each dataset to allow differentiation of patterns of co-expression. The six large datasets used in the initial analysis were all analysed individually (Figure 4.3). In each case the data was quality controlled, normalised and log-transformed before annotation using

Bioconductor. The samples in each dataset were ordered using the clinical data provided to allow easier visualisation of the data and then loaded into Biolayout. Pearson correlations were calculated for all probes on the array and all correlations over 0.6 stored. Depending on the size of the dataset and the inherent variation of samples, datasets produced graphs of varying sizes at a given correlation threshold value (see figure 4.2). Selected correlation thresholds for individual datasets were designed to include approximately 40% of the available data. Graphs were then laid out and clustered using the Markov clustering algorithm. Graphs of each dataset were then explored to understand the significance of the gene clusters and their relevance to the pathology of the tumours.

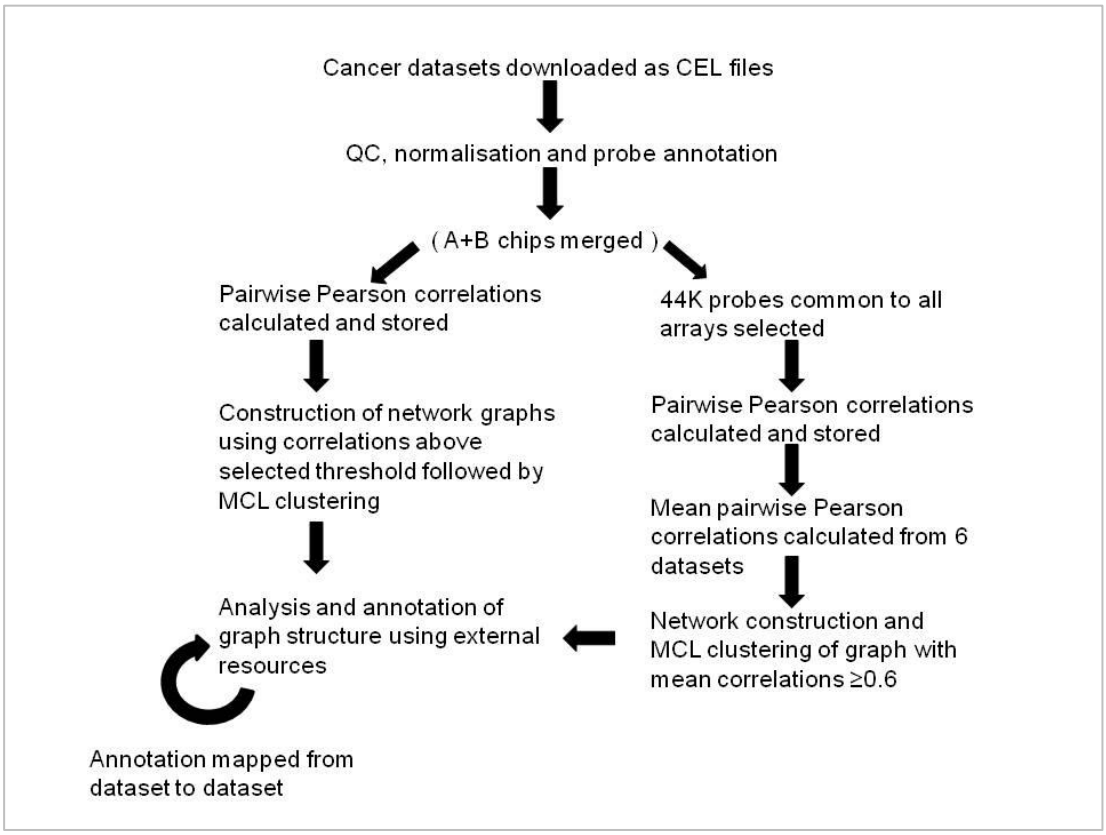


Figure 4.3 Workflow scheme. For each individual cancer dataset the left-hand pathway was followed. All datasets were treated individually and following QC and normalisation steps were laid out in network graphs and clustered. Clusters were explored using external sources of information inclusing GSEA, DAVID, literature review and the Human Protein Atlas. Annotation was mapped from graph to graph in an iterative approach. To create the merged 'core' graph, the right-hand pathway was followed. All pairwise Pearson correlations from 6 datasets were stored and mean correlations calculated for each pairing. Network construction and clustering were then performed using these mean correlations.

Clusters are identified by the known function of genes within them. Indeed clusters can only be meaningfully identified if the function of some of the genes within them are already characterised. Intriguingly one of the most stable and frequently present clusters identified in the data in this chapter and in many other datasets (Freeman, unpublished data) is formed of genes about which very little is known and about which therefore very little can be inferred. It can be speculated however, that this represents a fundamental part of the cell machinery such that is does not alter significantly between cell types, be they normal or malignant and as such has never become the subject of study.

The requirement that to be identified, a cluster must contain genes of known function does not imply that the analysis contributes nothing new to the understanding of the signature of clusters so identified. Rather the presence of known and well-characterised genes forming part of a biologically meaningful cluster allows other genes within that cluster to be assigned the same biological function on a 'guilt-by-association' principle; that is, if all the other genes that can be identified in a cluster are recognised as playing a role in the cell cycle, or belonging to a particular cell type then it can be inferred that other genes that share a similar expression profile share the same or a very closely related function.

The significance of the clusters identified were analysed using several strategies; web-based tools GSEA MSigDB and DAVID were used to look for functionally enriched clusters of genes (Subramanian et al 2005, Huang et al 2009 a&b). This approach proved useful in determining functional signatures, but due to the nature of the available genes sets in these tools proved less successful at identifying individual cell type signatures. Signatures derived from isolated cell types were overlain on the data to help identify these. Literature reviews were undertaken for genes from clusters of interest, and the Human Protein Atlas was used to help confirm the protein localisation of genes of interest (Uhlen et al 2010, Uhlen et al 2015). Clusters were annotated if hits of high significance in GSEA MSigDB and DAVID showed a common trend as to function and/or the genes appear cell specific either

due to comparison with known cell signatures, literature review or protein expression pattern in the Human Protein Atlas. It was not possible to assign putative functions to all clusters based on current knowledge.

Finally, in a novel approach to gene expression signature analysis, a ‘core’ signature of gene clusters common to a wide variety of tumours was calculated (Figure 4.3). All pair-wise Pearson correlations from six human cancer datasets were stored. A mean value was calculated for each pair-wise correlation across those six datasets. A biolayout graph was then created using these mean values to create a graph in which edges were formed only where mean correlations above a set threshold were used. In this way signatures unique to each tissue type or tumour were excluded as while modules of co-expressed genes would show very high levels of correlation in an individual tumour, unless that module also existed in the other datasets then the mean level of correlation between those genes in that module would be low and hence below the threshold at which the graph was drawn. Only those genes forming modules or clusters that existed in all datasets would be expected to be preserved across a wide range of tumour types. The core signature derived from this data was then overlain on further independent datasets to ensure the signatures generated from the ‘core’ signature’ also existed in independent datasets.

Analysis occurred in an iterative approach with identification of clusters from one dataset being used to inform analysis of other datasets and the ‘core’ annotation from the merged datasets mapped back to the original datasets to inform analysis. As a consequence of this approach, datasets were repeatedly laid out and clustered to allow the new annotation to be visualised. Thus although the stability of the clusters was not formally assessed, these were observed to be very stable varying very little between independent clusterings when carried out using the same parameters.

4.2 Selection of data sets analysed

Publically available datasets containing samples from large numbers of human tumours were used in the analysis. Datasets were identified from Gene Expression Omnibus or caArray. All datasets selected used the Affymetrix platform to allow direct mapping of probesets from one dataset to another in the subsequent analysis. This allowed for direct mapping of annotations from one analysis to the next, as all probe set identities remained the same between datasets and the clusters and the genes within them were annotated with functional information as the analysis proceeded. Use of the Affymetrix platform also provided absolute values for fluorescence for each probe set, rather than values relative to normal tissue or a different disease state. Datasets were also selected on the basis of containing large numbers of cases to provide sufficient power to the analysis, ranging in size from 77 cases to 265 cases. The use of large datasets decreases the likelihood that observed correlations between genes across the dataset are random events, rather than reflecting a true co-expressed 'module' of genes with a common function. Finally, only datasets where there was clinical information provided were considered for inclusion to allow the data to be ordered in an informative manner for the subsequent analysis. The datasets used are listed in Table 4.1.

Database reference	Reference (PMID)	Tumor type(s)	Cases analysed	Affymetrix U133 Platform(s)	Graph size (nodes)	Graph size (edges)
GSE11318	Lenz et al. (18765795)	DLBCL	194	Plus 2.0	19,850	614,273
GSE1456	Pawitan et al. (16280042)	Breast carcinoma	134	A & B	19,246	559,761
GSE9891	Tothill et al. (18698038)	Ovarian (epithelial) carcinoma	265	Plus 2.0	19,415	268,471
GSE3218	Korkola et al. (16424014)	Testicular germ cell tumours	86	A & B	18,934	954,082
GSE13294	Jorissen et al. (19088021)	Colorectal carcinoma	150	Plus 2.0	22,687	725,467
caArray/rembr-00037	REMBRANT – Repository for Molecular Brain Neoplasia Data	Primary CNS tumours	253	Plus 2.0	23,015	623,591
GSE7553	Riker et al. (18442402)	Skin tumours	77	A & B	19,623	600,143

Table 4.1. The seven primary human tumour datasets used in the analysis. All used the Affymetrix platform with a mixture of U133A and B and U133 plus 2. To allow direct comparison between datasets only those probesets present on the U133A and U133B chips were used for analysis of data from U133plus 2 chips.

Each dataset had a unique graph structure reflecting the individual gene expression patterns contained within that dataset. All graphs were laid out at a correlation coefficient such that roughly 40% of the probesets were represented on the graph as nodes. As seen in table 4.1 the number of edges contributing to these graphs at that size varied considerably between datasets ranging from only 268,471 edges in the

very large ovarian dataset to 954,082 edges in the much smaller testicular tumour dataset which also contained a much wider variety of individual tumours and hence showed a tendency to form very tight clusters with large numbers of edges relating to each individual tumour type. Each dataset had its own idiosyncrasies owing to the high degree of inherent variation in gene expression data derived from cancer samples. This variation is due the samples' derivation from an outbred population, heterogeneous cellular composition and the inherent variability associated with cancer. Further variation in the data is also down to the type of cancer from which it was derived, the number of subtypes of cancer included in the dataset and the number of samples analysed. All of these factors ultimately determine the overall topology of the network derived from the data.

4.3 Internal validation of the clustering approach

Before presenting the analyses from individual datasets, it is worth discussing some of the types of clusters that were present in all datasets and some of which, while not necessarily informing the analysis with regards to the individual tumour types, provided some internal validation for the graph layout and clustering method.

In all datasets there were biologically explicable clusters, many of which clearly correlated with clinical or histological features of the tumours from which they were derived and provided support to the data analysis being able to accurately graph and partition co-expressed genes into clusters. Identification of tumour specific clusters and relationship to tumour type and outcome was not the focus of this analysis which aimed to identify mainly the common features of tumour datasets and such 'cancer signatures' are only discussed in as much as they provide support for the approach used. Unknown clusters that appear from their expression profiles to provide information about tumour biology were not extensively explored, although there is clearly a wealth of untapped data in the network graphs presented below, far beyond the scope of this thesis to explore. In addition to these more biologically relevant clusters, there were other less interesting clusters that none the less provided

evidence that probesets with similar expression patterns were being clustered together; genes that were represented on the Affymetrix array by multiple probe sets tended to form clusters in which all the members were probes directed to that gene; internal Affymetrix controls formed a single cluster; genes known to be co-regulated were seen to form clusters, eg probe sets to Haemoglobin A and Haemoglobin B genes formed a cluster (Figure 4.4); genes expressed in males or females only were mutually exclusive (figure 4.5); and there were clusters present that represented a 'spike' in the data formed by only a few tumours (Figure 4.6). These 'spikes' frequently appeared to represent contamination of the sampled tumour by tissues that were otherwise absent in the dataset resulting in gene expression profile across the dataset that was very different from the profiles seen in genes that were more commonly expressed across the dataset. As demonstrated in figure 4.6, these contaminants generate very distinctive gene expression profiles and demonstrate one of the weaknesses of the use of Pearson correlation co-efficients as the measure of relatedness of genes; i.e. the measure is very sensitive to outliers which will have very high correlation coefficients. In these very large and complex datasets, this effect however appears not of great importance with the clusters formed of such spikes mostly consisting of only small numbers of genes. Other 'spikes', such as those seen in the teratoma samples of the testicular cancer dataset, could be readily explained by the nature of the tumour with expression profiles of mature non-testicular tissues (such as skin, gut, thyroid, pancreas etc.) representing the differentiated tissues present in these tumours. However, other of these clusters with a 'spiked' expression pattern were more difficult to explain in terms of grouped functions of genes. They frequently however, included multiple transcripts of the same gene (which would argue against these observations being due to hybridisation artefacts where one might expect a truly random set of transcripts to be affected), various subunits of a multi-unit structure or multiple members of a gene family, all of which argue that these are real effects rather than due to technical error. On analysis through GSEA using the 'positional gene sets' collection, it is apparent that at least some of these spikes are enriched in genes from a given area of a single chromosome, suggesting individual tumour-specific dysregulation of genes in that location either through localised chromosomal rearrangements or epigenetic factors.

The fact that these spikes are unique to a very small number of tumours argues that any such translocation or localised dysregulation of gene expression is not essential to development of the tumour or it would be expected to be present in a larger number of samples.

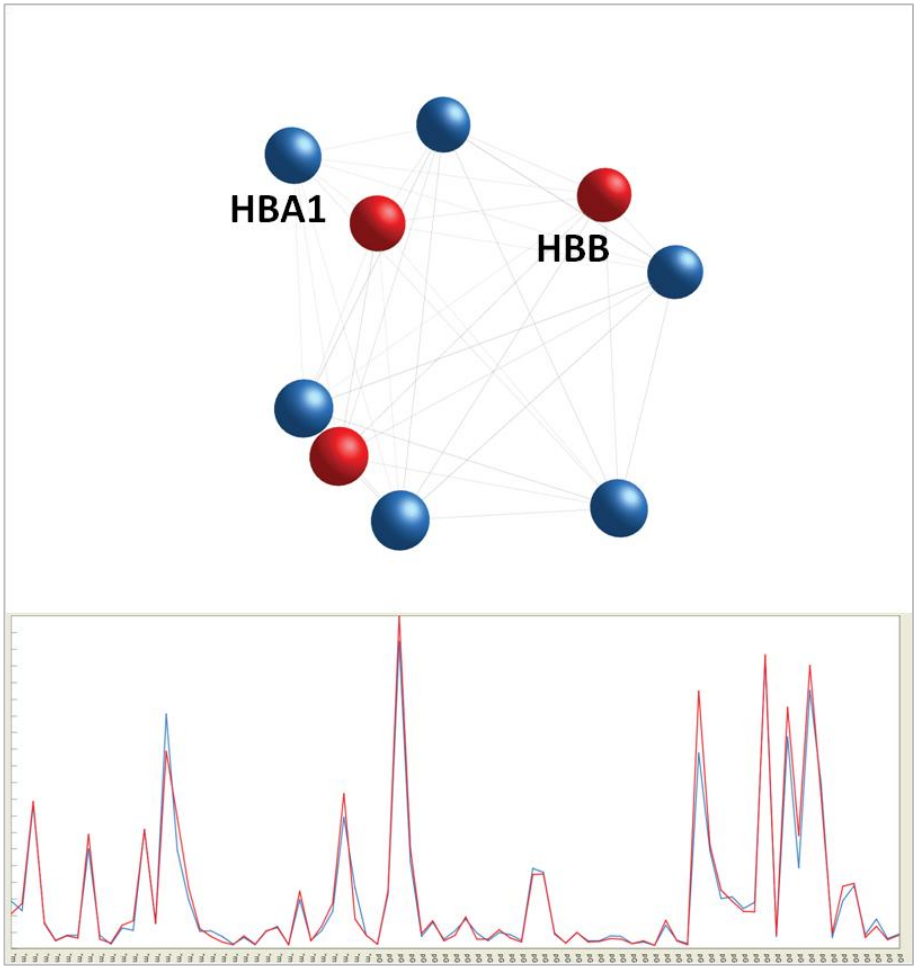


Figure 4.4. Haemogloblin cluster. The probeset directed against the genes for haemoglobin A and B frequently clustered together. Shown above are these genes forming a cluster in the testicular tumours dataset. The top panel of the figure shows the haemoglobin cluster expanded and the rest of the graph hidden to allow visualisation of the multiple connections between these probesets that result in cluster formation. The two genes are represented in different colours with probesets to HBA1 in blue and HBB in red. The bottom half of the figure shows the mean expression pattern of the probesets across the tumour dataset to show how closely the expression of these genes mirrors each other. Again HBA1 is represented in blue and HBB in red. Individual tumours are represented on the x-axis.

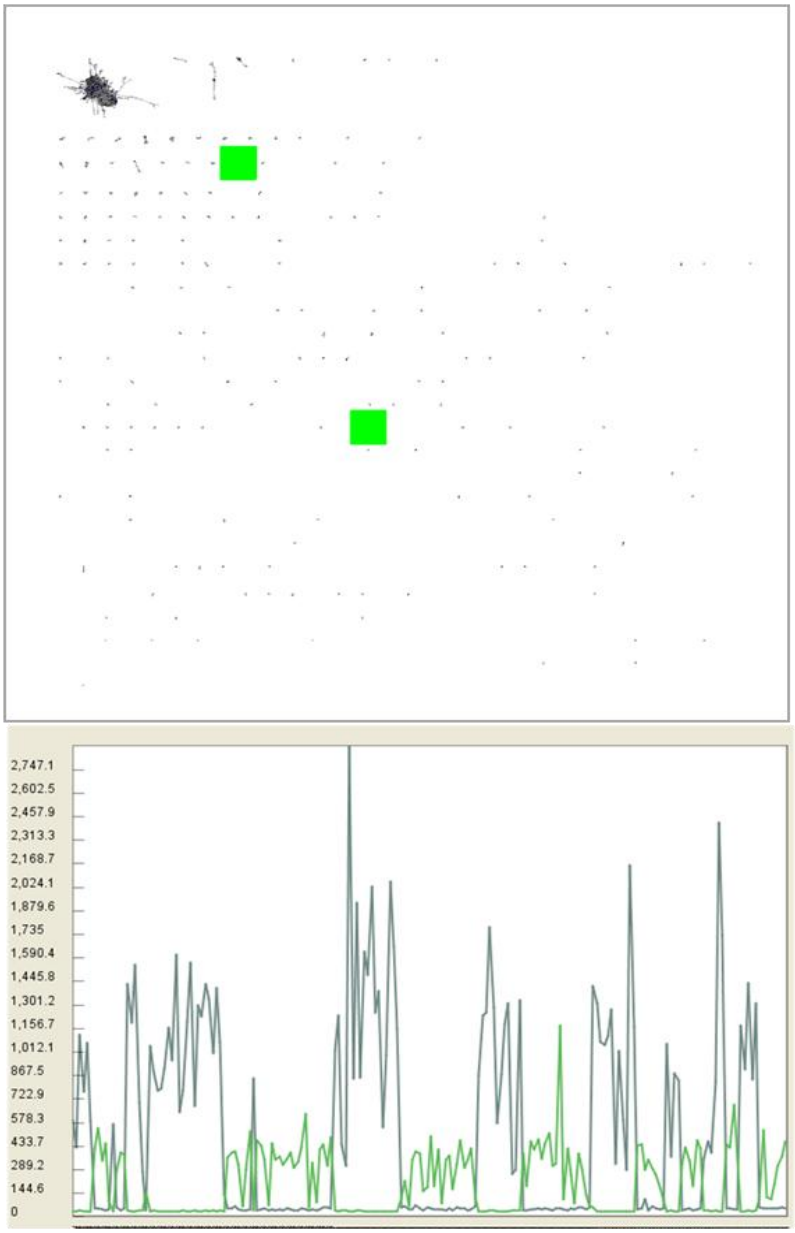
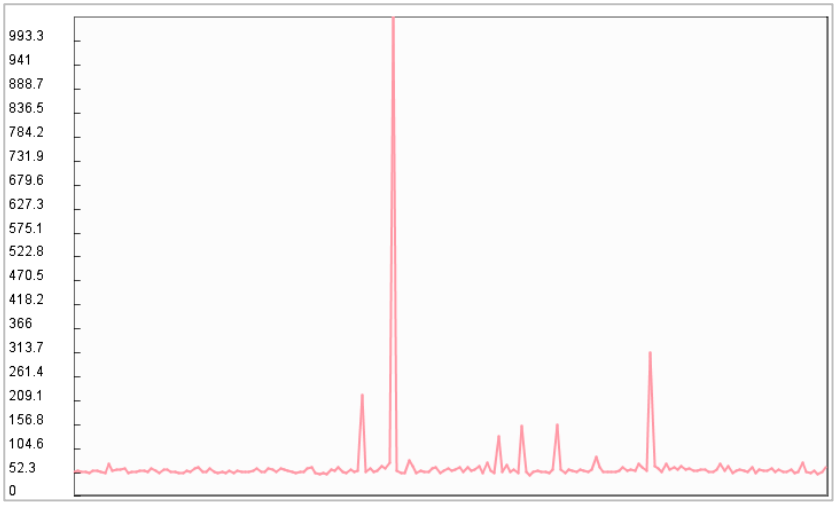


Figure 4.5. Male and female clusters. The DLBCL graph laid out at p≥0.6 and MCL inflation value 2.2 contains mutually exclusive clusters that display contrasting patterns in males and females. The top panel shows a 'scanning power' view of the DLBCL graph network. The majority of the structure lies in the top left hand corner of the graph. Small, highly separate clusters lie in the remainder of the space. High-lighted by green squares are two clusters that represent genes expressed in only males or females. The expression patterns of these genes are very different from much of the rest of the data and form widely separate clusters that do not join the main graph. The lower half of the figure shows the average gene expression values for these clusters (male cluster in grey, female cluster in green). The x-axis represents individual tumours. The tumours have been stratified first by tumour type, then outcome, then sex so that the group 'male patients' lies in multiple discrete areas along the X-axis and the female patients in the intervening areas. See table 4.2 for gene lists.

Gene name;Affymetric probe ID	Description	Cluster annotation
CYorf15A;232618_at	chromosome Y open reading frame 15A	Cluster0075*Y-chromosome-male
CYorf15A;236694_at	chromosome Y open reading frame 15A	Cluster0075*Y-chromosome-male
CYorf15B;214131_at	chromosome Y open reading frame 15B	Cluster0075*Y-chromosome-male
CYorf15B;223645_s_at	chromosome Y open reading frame 15B	Cluster0075*Y-chromosome-male
CYorf15B;223646_s_at	chromosome Y open reading frame 15B	Cluster0075*Y-chromosome-male
DDX3Y;205000_at	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, Y-linked	Cluster0075*Y-chromosome-male
DDX3Y;205001_s_at	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, Y-linked	Cluster0075*Y-chromosome-male
EIF1AY;204409_s_at	eukaryotic translation initiation factor 1A, Y-linked	Cluster0075*Y-chromosome-male
EIF1AY;204410_at	eukaryotic translation initiation factor 1A, Y-linked	Cluster0075*Y-chromosome-male
JARID1D;206700_s_at	jumonji, AT rich interactive domain 1D	Cluster0075*Y-chromosome-male
NA;1560800_at	NA	Cluster0075*Y-chromosome-male
NA;239677_at	NA	Cluster0075*Y-chromosome-male
NA;244482_at	NA	Cluster0075*Y-chromosome-male
NLGN4Y;207703_at	neuroligin 4, Y-linked	Cluster0075*Y-chromosome-male
PRKY;206279_at	protein kinase, Y-linked	Cluster0075*Y-chromosome-male
RPS4Y1;201909_at	ribosomal protein S4, Y-linked 1	Cluster0075*Y-chromosome-male
TTTY15;214983_at	testis-specific transcript, Y-linked 15	Cluster0075*Y-chromosome-male
USP9Y;206624_at	ubiquitin specific peptidase 9, Y-linked (fat facets-like, Drosophila)	Cluster0075*Y-chromosome-male
USP9Y;228492_at	hypothetical protein LOC100130216	Cluster0075*Y-chromosome-male
UTY;208067_x_at	ubiquitously transcribed tetratricopeptide repeat gene, Y-linked	Cluster0075*Y-chromosome-male
UTY;210322_x_at	ubiquitously transcribed tetratricopeptide repeat gene, Y-linked	Cluster0075*Y-chromosome-male
UTY;211149_at	ubiquitously transcribed tetratricopeptide repeat gene, Y-linked	Cluster0075*Y-chromosome-male
ZFY;230760_at	zinc finger protein, Y-linked	Cluster0075*Y-chromosome-male
NA;229315_at	NA	Cluster0205*Female
XIST;214218_s_at	X (inactive)-specific transcript (non-protein coding)	Cluster0205*Female
XIST;221728_x_at	X (inactive)-specific transcript (non-protein coding)	Cluster0205*Female
XIST;224588_at	X (inactive)-specific transcript (non-protein coding)	Cluster0205*Female
XIST;224589_at	X (inactive)-specific transcript (non-protein coding)	Cluster0205*Female
XIST;224590_at	X (inactive)-specific transcript (non-protein coding)	Cluster0205*Female
XIST;227671_at	X (inactive)-specific transcript (non-protein coding)	Cluster0205*Female
XIST;231592_at	X (inactive)-specific transcript (non-protein coding)	Cluster0205*Female
ZFX;229022_at	zinc finger protein, X-linked	Cluster0205*Female

Table 4.2. Gene cluster lists from the DLBCL data in figure 4.4. There are two mutually exclusive clusters representing male and female specific genes.



Selected genes from cluster	Selected functional annotation from DAVID with P values for modified Fisher exact test
Actin , alpha 1, skeletal muscle	GO Term: contractile fibre - P value = 1.3^{e-72}
Creatine kinase, muscle	SP Keywords: muscle protein - P value = 1.6^{e-49}
Desmin	GO Term: Z disc - P value =1.1^{e-24}
Myoglobin	GO Term: myofibril - P value = 1.7^{e-70}
Myosin, heavy and light chains	GO Term: sarcomere - P value = 8.4^{e-65}
Troponins I and T	GO Term: actin cytoskeleton – P value = 4.2^{e-26}

Figure 4.6 Skeletal muscle cluster. The DLBCL graph contains a 'spiked' cluster formed of genes that are only expressed to an appreciable level in 7 cases out of 194 being absent in the remainder of the cases. Analysis of this cluster indicates this represents 'contamination' of the tumour by skeletal muscle i.e. biopsies taken from where the tumour is infiltrating skeletal muscle. The upper half of the figure represents the mean expression profile of the genes in this cluster across the entire dataset demonstrating the small number of cases where there is expression of these genes. The lower part of the figure highlights selected genes from within the cluster typical of skeletal muscle and selected functional annotation from DAVID analysis. The complete gene list forming the cluster is given in appendix 1.

4.4 Gene signatures from DLBCL

The lymphoma dataset consisted of 203 cases of diffuse large B cell lymphoma, 194 of which passed the QC step and were included in the analysis. These were stratified into Germinal centre B cell-like (GCB), activated B cell-like (ABC), primary mediastinal B cell (PMBL) and unclassified, based on gene expression profiles as determined in the paper by Alizadeh et al (Alizadeh et al 2000). The data was further organised on the basis of sex and patient outcome. The DLBCL dataset (figure 4.7) when laid at out at Pearson correlations $r \geq 0.65$ gives a graph of 19,850 nodes joined by 614,273 edges.

The graph breaks into three main areas, with immune related clusters and extracellular matrix lying in one area which appears to broadly represent the tumour stroma, cell cycle genes in another area, and a group of poorly characterised clusters lying in the third main area (Figures 4.7 and 4.8). These poorly characterised clusters tend to show little variation between samples and are often expressed at relatively low level. For ease of discussion these are referred to as housekeeping or HK clusters. Numerous small clusters not forming part of the main graph lie in the remainder of the space. Many of these small clusters are formed of multiple probe sets to the same gene, while others are formed of small numbers of genes whose expression in very closely related to each other but have little relationship to the genes that make up the bulk of the graph. The immune area contains clusters of macrophage, T-cell and interferon related genes and lies close to but distinct from the extracellular matrix clusters. Within the cell cycle cluster there are several distinct but related clusters with genes that are expressed in the S/M phase forming a cluster and other cell cycle genes lying adjacent to this.

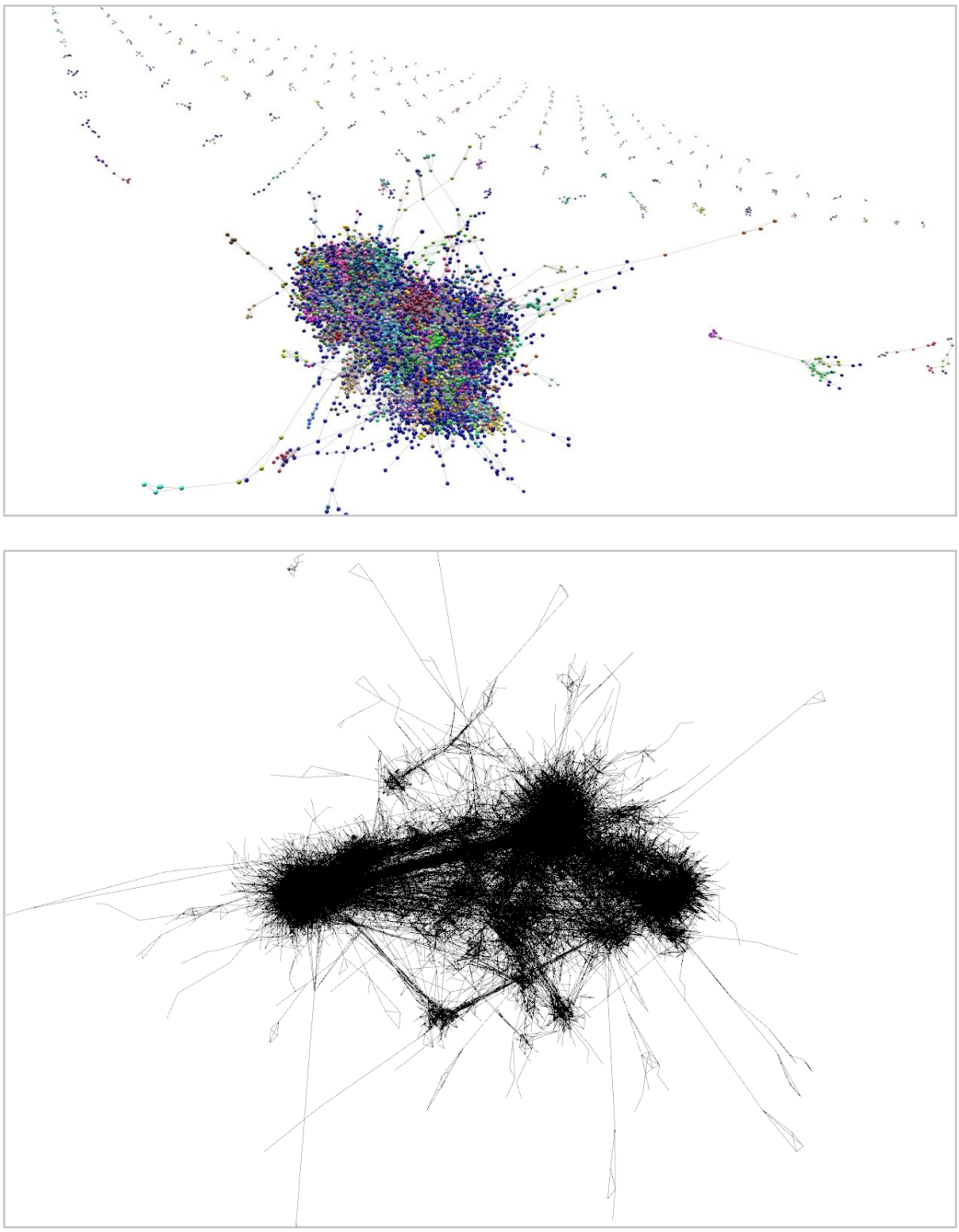


Figure 4.7 Two views of the DLBCL graph. In the upper panel the whole graph is visualised. The main component of the graph lies in front and centre formed of the majority of the probes that correlate with other probes. The genes of interest for the analysis lie mostly within this area. Clusters as defined by the MCL clustering algorithm are represented by colours. Due to the complexity of the graph little of the overall structure can be appreciated at this magnification. Extending off into the distance are a series of small clusters which do not connect with another clusters. These are of genes/probesets that have a high degree of internal correlation but which show very little correlation with the other probesets on the array. In the lower panel the main part of the graph is represented just as edges with the nodes hidden to better visualise the structure of the graph.

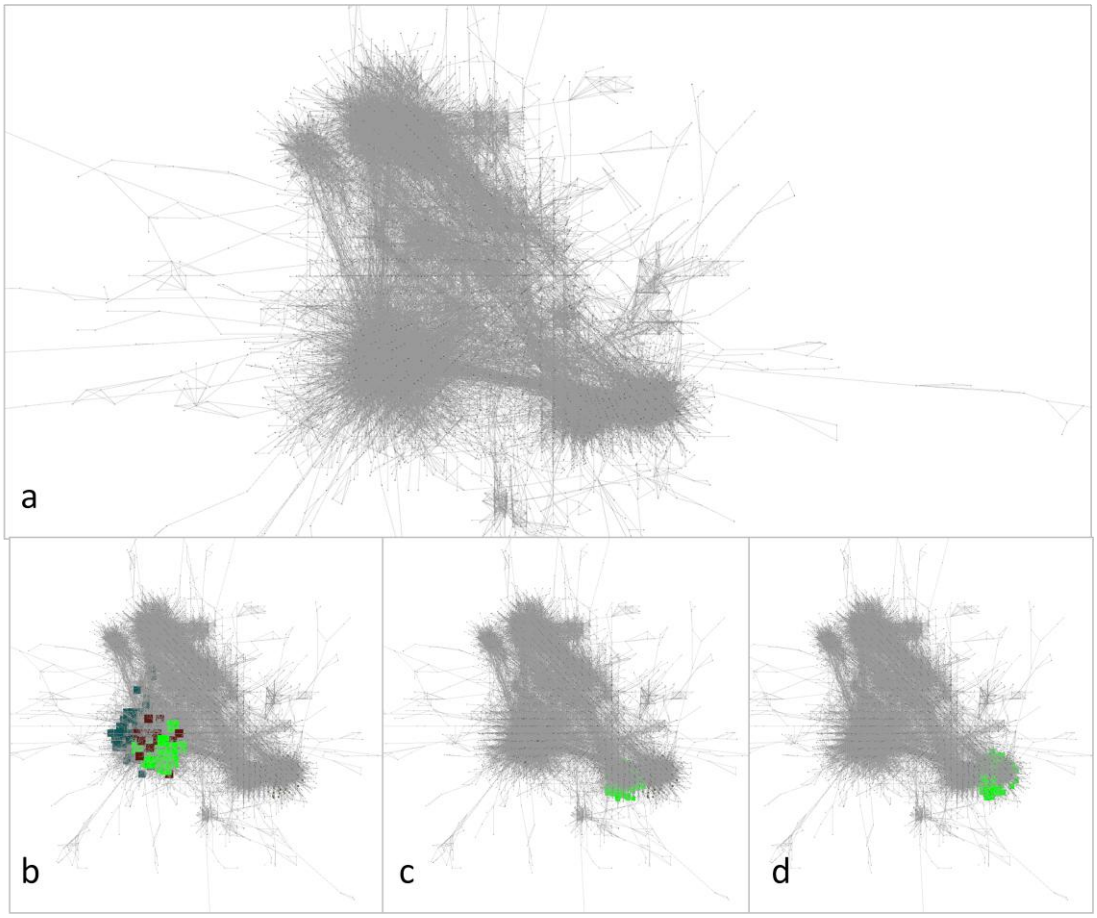


Figure 4.8 Different areas of the overall graph contain different main clusters. These relationships are best appreciated in 3-D within the Biolayout programme where the graph can be interacted with, but a simplified view of the DLBCL graph is presented here to illustrate some of the relationships. a) Overall graph structure with nodes minimised, b) the cell cycle related clusters are superimposed on the graph network with the S/M phase cluster in green and other cell cycle related clusters in blue and brown, c) the ECM clusters occupy a different space in the graph, d) Immune clusters, here highlighting the macrophage cluster, occupy one pole of the large cluster

4.4.1 Stromal signatures

The macrophage cluster contains 769 probe sets representing 497 genes and contains well-characterised markers of macrophage/monocyte differentiation including CD14, CD68, CD163, colony stimulating factor receptors (CSF receptor 1, CFS receptor 2β and CSF receptor 3), and Fc receptors (Fc receptors for IgA, IgE and IgG; FCRG1A, FCRG1B, FCRG2A, FCRG3B). In addition to these there are also multiple

chemokines and cytokines (CCL4, CCL5, CCL18, CXCL9, CXCL10, CXCL11, CXCL12, CXCL16, IL15, IFNG), chemokine and interleukin receptors (CCR1, CCR5, IL15RA, IL17RA, IL31RA), and toll-like receptors (TLR1, TLR2, TLR4, TLR5, TRL8). There are multiple apolipoproteins (APOE, APOL1, APOL2, APOL3, APOL4, APOL6), metallothioneins (MT1E, TM1F, MT1G, MT1M, MT1X, MT2A) and genes encoding lysosomal proteins (CD63, CD68, GM2A, NPL, NPC1, NPC2, ACP2, ACP5, CTSA, CTSB, CTSD, CTSL1, CTS, CTSZ, DNASE2, GALNS, GLAC, GNS, HEXA, HEXB, IGF2R, IFI30, LIPA, LAMP1, LAMP2, M6PR, MANBA, PPT1, PSEN1, PRCP, PSAP, SCARB2, SCL11A2, SCL29A3). The macrophage scavenger receptors are represented (MARCO, SCARB2). A list of selected genes found in the macrophage cluster, omitting less informative genes characterised only as open reading frames and numbered loci without additional annotation is given in table 4.4. Analysis of the gene signature through the DAVID functional annotation clustering software and GSEA MolSigDB software shows the signature to be significantly enriched in genes associated with defence response, myeloid cells, lysosomal genes, pattern recognition receptors and many others as would be expected reading the list of genes identified. These analysis methods when applied to the cluster discussed pull out hundreds of points of annotation, many of which are similar (Table 4.3). Only a few selected annotations are provided in each case for illustration.

Looking at the expression profile of the macrophage signature across the dataset, there is obviously marked variability in expression between individual tumours (Figure 4.9). There is no obvious pattern of expression separating the tumour types or outcomes. This is in contrast to the paper by Lenz et al (Lenz et al 2008) who demonstrated differences in outcome dependent on stromal signatures. Their approach differed from that used here in that they first isolated genes associated with poor outcome and then clustered those using a hierarchical clustering method. Their stromal signatures are derived from these groups of genes that define outcome rather than as here, reflecting simply the nature of the expression network in the tumour.

It is important to recognise that the “macrophage cluster” does not contain obvious markers of other stromal or tumour cell type or process (e.g. cell cycle), so there appears to be no obligate association between macrophage content and any other biological activity.

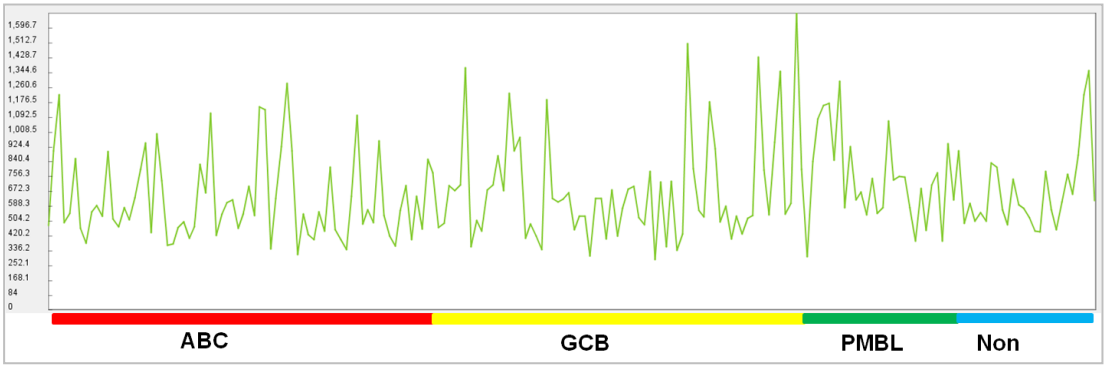


Figure 4.9 Mean expression intensity of the genes present in the ‘macrophage signature’ across the dataset. There is considerable variation between individual tumours but no clear pattern between tumour types or with differences in outcome. Coloured bars denote tumour subtypes. For each tumour subtype patients with good outcome lie in the left part of the bar and poor in the right.

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Defense response	2.6×10^{-23}
DAVID	KEGG pathway	Lysosome	2.7×10^{-17}
DAVID	GOTERM_BP_FAT	Response to molecule of bacterial origin	1.5×10^{-9}
DAVID	GOTERM_MF_FAT	Immunoglobulin binding	4.9×10^{-8}
DAVID	GOTERM_BP_FAT	Chemotaxis	1.4×10^{-11}
DAVID	SP_PIR_KEYWORDS	Immune response	5.5×10^{-16}
DAVID	GOTERM_MF_FAT	Pattern recognition receptor activity	8.7×10^{-7}
GSEA	GSE10325	Genes downregulated in healthy B cells vs healthy myeloid cells	1.12×10^{-95}
GSEA	Stanford tumour modules	Module 84 – Immune and inflammatory response	2.39×10^{-101}

Table 4.3 Functional annotation for the gene list from the 'macrophage signature'. Selected annotation derived from DAVID and GSEA gene expression analysis tools. The macrophage gene list was uploaded to DAVID and analysed through the functional annotation clustering and in GSEA using the molecular signatures database. Only selected annotation is shown here. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA, hypergeometric distribution p-values.

Table 4.4 Macrophage signature from DLBCL dataset laid out at Pearson correlation ≥ 0.6 with MCL inflation values of 2.2. Only probesets that represent genes with official gene symbols are listed, with open reading frames, hypothetical proteins and unannotated probesets excluded. A complete list is given in appendix 1 .

Symbol	Description
AADACL1	arylacetamide deacetylase-like 1
ABCA1	ATP-binding cassette, sub-family A (ABC1), member 1
ABCD2	ATP-binding cassette, sub-family D (ALD), member 2
ABHD12	abhydrolase domain containing 12
ABHD2	abhydrolase domain containing 2
ABHD2	abhydrolase domain containing 2
ACO1	aconitase 1, soluble
ACP2	acid phosphatase 2, lysosomal
ACP5	acid phosphatase 5, tartrate resistant
ACSL1	acyl-CoA synthetase long-chain family member 1
ACSS2	acyl-CoA synthetase short-chain family member 2
ACVR1B	activin A receptor, type IB
ADAMDEC1	ADAM-like, decysin 1
AGTRAP	angiotensin II receptor-associated protein
AIF1	allograft inflammatory factor 1
ALAS1	aminolevulinate, delta-, synthase 1
ALDH1A1	aldehyde dehydrogenase 1 family, member A1
ALDH3B1	aldehyde dehydrogenase 3 family, member B1
ANKRD22	ankyrin repeat domain 22
AOAH	acyloxyacyl hydrolase (neutrophil)
APOC1	apolipoprotein C-I
APOE	apolipoprotein E
APOE	apolipoprotein E
APOL1	apolipoprotein L, 1
APOL2	apolipoprotein L, 2
APOL3	apolipoprotein L, 3
APOL4	apolipoprotein L, 4
APOL6	apolipoprotein L, 6
AQP9	aquaporin 9
ARHGEF10L	Rho guanine nucleotide exchange factor (GEF) 10-like
ARHGEF11	Rho guanine nucleotide exchange factor (GEF) 11
ARRB1	arrestin, beta 1
ASAH1	N-acylsphingosine amidohydrolase (acid ceramidase) 1
ASCL2	achaete-scute complex homolog 2 (Drosophila)
ASGR1	asialoglycoprotein receptor 1
ATG7	ATG7 autophagy related 7 homolog (S. cerevisiae)
ATP13A2	ATPase type 13A2
ATP6AP1	ATPase, H+ transporting, lysosomal accessory protein 1
ATP6V0C	ATPase, H+ transporting, lysosomal 16kDa, V0 subunit c
ATP6V1A	ATPase, H+ transporting, lysosomal 70kDa, V1 subunit A
ATP6V1B2	ATPase, H+ transporting, lysosomal 56/58kDa, V1 subunit B2
B2M	beta-2-microglobulin
BATF2	basic leucine zipper transcription factor, ATF-like 2
BLVRB	biliverdin reductase B (flavin reductase (NADPH))
BLZF1	basic leucine zipper nuclear factor 1
BRI3	brain protein I3
BST1	bone marrow stromal cell antigen 1

BTBD14A	BTB (POZ) domain containing 14A
C1QA	complement component 1, q subcomponent, A chain
C1QB	complement component 1, q subcomponent, B chain
C1QC	complement component 1, q subcomponent, C chain
C1RL	complement component 1, r subcomponent-like
C2	complement component 2
C3AR1	complement component 3a receptor 1
CA11	carbonic anhydrase XI
CALCOCO2	calcium binding and coiled-coil domain 2
CAPG	capping protein (actin filament), gelsolin-like
CARD9	caspase recruitment domain family, member 9
CASP5	caspase 5, apoptosis-related cysteine peptidase
CCL18	chemokine (C-C motif) ligand 18 (pulmonary and activation-regulated)
CCL4	chemokine (C-C motif) ligand 4
CCL5	chemokine (C-C motif) ligand 5
CCL8	chemokine (C-C motif) ligand 8
CCR1	chemokine (C-C motif) receptor 1
CCR5	chemokine (C-C motif) receptor 5
CCRL2	chemokine (C-C motif) receptor-like 2
CD14	CD14 molecule
CD163	CD163 molecule
CD244	CD244 molecule, natural killer cell receptor 2B4
CD300LF	CD300 molecule-like family member f
CD302	CD302 molecule
CD33	CD33 molecule
CD4	CD4 molecule
CD63	CD63 molecule
CD68	CD68 molecule
CD81	CD81 molecule
CD97	CD97 molecule
CDA	cytidine deaminase
CDS2	CDP-diacylglycerol synthase (phosphatidate cytidylyltransferase) 2
CEBPB	CCAAT/enhancer binding protein (C/EBP), beta
CEBPD	CCAAT/enhancer binding protein (C/EBP), delta
CECR1	cat eye syndrome chromosome region, candidate 1
CENTG2	centaurin, gamma 2
CENTG3	centaurin, gamma 3
CHI3L1	chitinase 3-like 1 (cartilage glycoprotein-39)
CHST11	carbohydrate (chondroitin 4) sulfotransferase 11
CLEC10A	C-type lectin domain family 10, member A
CLEC12A	C-type lectin domain family 12, member A
CLEC2B	C-type lectin domain family 2, member B
CLEC4D	C-type lectin domain family 4, member D
CLEC4E	C-type lectin domain family 4, member E
CLEC7A	C-type lectin domain family 7, member A
CMKLR1	chemokine-like receptor 1
CNIH4	cornichon homolog 4 (Drosophila)
CPD	carboxypeptidase D
CPVL	carboxypeptidase, vitellogenic-like
CRAT	carnitine acetyltransferase
CREBL2	cAMP responsive element binding protein-like 2
CREG1	cellular repressor of E1A-stimulated genes 1

CRIM1	cysteine rich transmembrane BMP regulator 1 (chordin-like)
CSF1	colony stimulating factor 1 (macrophage)
CSF1R	colony stimulating factor 1 receptor
CSF2RB	colony stimulating factor 2 receptor, beta, low-affinity (granulocyte-macrophage)
CSF3R	colony stimulating factor 3 receptor (granulocyte)
CST3	cystatin C
CTNND2	catenin (cadherin-associated protein), delta 2 (neural plakophilin-related arm-repeat protein)
CTSA	cathepsin A
CTSB	cathepsin B
CTSD	cathepsin D
CTSL1	cathepsin L1
CTSS	cathepsin S
CTSZ	cathepsin Z
CXCL10	chemokine (C-X-C motif) ligand 10
CXCL11	chemokine (C-X-C motif) ligand 11
CXCL12	chemokine (C-X-C motif) ligand 12 (stromal cell-derived factor 1)
CXCL16	chemokine (C-X-C motif) ligand 16
CXCL9	chemokine (C-X-C motif) ligand 9
CYBB	cytochrome b-245, beta polypeptide
CYFIP1	cytoplasmic FMR1 interacting protein 1
CYP1B1	cytochrome P450, family 1, subfamily B, polypeptide 1
DAB2	disabled homolog 2, mitogen-responsive phosphoprotein (Drosophila)
DAPK1	death-associated protein kinase 1
DENND5A	DENN/MADD domain containing 5A
DIRC2	disrupted in renal carcinoma 2
DLL1	delta-like 1 (Drosophila)
DMXL2	Dmx-like 2
DNASE2	deoxyribonuclease II, lysosomal
DOCK4	dedicator of cytokinesis 4
DOCK5	dedicator of cytokinesis 5
DOK2	docking protein 2, 56kDa
DPYD	dihydropyrimidine dehydrogenase
DRAM	damage-regulated autophagy modulator
DUSP3	dual specificity phosphatase 3
DYNLT1	dynein, light chain, Tctex-type 1
DYSF	dysferlin, limb girdle muscular dystrophy 2B (autosomal recessive)
EIF4E3	eukaryotic translation initiation factor 4E family member 3
EMR2	egf-like module containing, mucin-like, hormone receptor-like 2
EPB41L3	erythrocyte membrane protein band 4.1-like 3
EPHX1	epoxide hydrolase 1, microsomal (xenobiotic)
ERLIN1	ER lipid raft associated 1
F2R	coagulation factor II (thrombin) receptor
FAIM2	Fas apoptotic inhibitory molecule 2
FAM105A	family with sequence similarity 105, member A
FAM125B	family with sequence similarity 125, member B
FAM20A	family with sequence similarity 20, member A
FAM26F	family with sequence similarity 26, member F
FASLG	Fas ligand (TNF superfamily, member 6)

FCAR	Fc fragment of IgA, receptor for
FCER1G	Fc fragment of IgE, high affinity I, receptor for; gamma polypeptide
FCGR1A	Fc fragment of IgG, high affinity Ia, receptor (CD64)
FCGR1B	Fc fragment of IgG, high affinity Ib, receptor (CD64)
FCGR2A	Fc fragment of IgG, low affinity IIa, receptor (CD32)
FCGR3B	Fc fragment of IgG, low affinity IIIb, receptor (CD16b)
FCGRT	Fc fragment of IgG, receptor, transporter, alpha
FCHO2	FCH domain only 2
FER1L3	fer-1-like 3, myoferlin (C. elegans)
FFAR2	free fatty acid receptor 2
FGL2	fibrinogen-like 2
FNIP2	folliculin interacting protein 2
FPR1	formyl peptide receptor 1
FPR2	formyl peptide receptor 2
FRMD4B	FERM domain containing 4B
FTH1	ferritin, heavy polypeptide 1
FTHP1	ferritin, heavy polypeptide pseudogene 1
FTL	ferritin, light polypeptide
FUT4	fucosyltransferase 4 (alpha (1,3) fucosyltransferase, myeloid-specific)
FZD2	frizzled homolog 2 (Drosophila)
FZD5	frizzled homolog 5 (Drosophila)
GAB3	GRB2-associated binding protein 3
GADD45G	growth arrest and DNA-damage-inducible, gamma
GALC	galactosylceramidase
GALM	galactose mutarotase (aldose 1-epimerase)
GALNS	galactosamine (N-acetyl)-6-sulfate sulfatase
GALNT11	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 11 (GalNAc-T11)
GBP1	guanylate binding protein 1, interferon-inducible, 67kDa
GBP2	guanylate binding protein 2, interferon-inducible
GBP3	guanylate binding protein 3
GBP4	guanylate binding protein 4
GBP5	guanylate binding protein 5
GCH1	GTP cyclohydrolase 1
GGTA1	glycoprotein, alpha-galactosyltransferase 1
GJD3	gap junction protein, delta 3, 31.9kDa
GK	glycerol kinase
GK3P	glycerol kinase 3 pseudogene
GLIPR2	GLI pathogenesis-related 2
GLUL	glutamate-ammonia ligase (glutamine synthetase)
GM2A	GM2 ganglioside activator
GMPR	guanosine monophosphate reductase
GNA15	guanine nucleotide binding protein (G protein), alpha 15 (Gq class)
GNPTG	N-acetylglucosamine-1-phosphate transferase, gamma subunit
GNS	glucosamine (N-acetyl)-6-sulfatase
GPNMB	glycoprotein (transmembrane) nmb
GPR171	G protein-coupled receptor 171
GPR56	G protein-coupled receptor 56
GPR65	G protein-coupled receptor 65
GPR84	G protein-coupled receptor 84

GPX1	glutathione peroxidase 1
GRIN3A	glutamate receptor, ionotropic, N-methyl-D-aspartate 3A
GRINA	glutamate receptor, ionotropic, N-methyl D-aspartate-associated protein 1 (glutamate binding)
GRN	granulin
GZMA	granzyme A (granzyme 1, cytotoxic T-lymphocyte-associated serine esterase 3)
GZMH	granzyme H (cathepsin G-like 2, protein h-CCPX)
HAPLN3	hyaluronan and proteoglycan link protein 3
HAVCR2	hepatitis A virus cellular receptor 2
HEXA	hexosaminidase A (alpha polypeptide)
HEXB	hexosaminidase B (beta polypeptide)
HK3	hexokinase 3 (white cell)
HNMT	histamine N-methyltransferase
HS3ST2	heparan sulfate (glucosamine) 3-O-sulfotransferase 2
HSD11B1	hydroxysteroid (11-beta) dehydrogenase 1
HSD17B14	hydroxysteroid (17-beta) dehydrogenase 14
HSD3B7	hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 7
HTRA4	HtrA serine peptidase 4
IDH1	isocitrate dehydrogenase 1 (NADP+), soluble
IFI30	interferon, gamma-inducible protein 30
IFNAR1	interferon (alpha, beta and omega) receptor 1
IFNG	interferon, gamma
IGF2R	insulin-like growth factor 2 receptor
IGSF22	immunoglobulin superfamily, member 22
IGSF6	immunoglobulin superfamily, member 6
IL15	interleukin 15
IL15RA	interleukin 15 receptor, alpha
IL17RA	interleukin 17 receptor A
IL18BP	interleukin 18 binding protein
IL31RA	interleukin 31 receptor A
INDO	indoleamine-pyrrole 2,3 dioxygenase
INDOL1	indoleamine-pyrrole 2,3 dioxygenase-like 1
IRAK3	interleukin-1 receptor-associated kinase 3
IRF1	interferon regulatory factor 1
IRG1	immunoresponsive 1 homolog (mouse)
ITFG1	integrin alpha FG-GAP repeat containing 1
ITGB2	integrin, beta 2 (complement component 3 receptor 3 and 4 subunit)
ITM2B	integral membrane protein 2B
JAKMIP1	janus kinase and microtubule interacting protein 1
KAL1	Kallmann syndrome 1 sequence
KCNE1	potassium voltage-gated channel, Isk-related family, member 1
KCNE3	potassium voltage-gated channel, Isk-related family, member 3
KCNJ10	potassium inwardly-rectifying channel, subfamily J, member 10
KCNJ2	potassium inwardly-rectifying channel, subfamily J, member 2
KCTD12	potassium channel tetramerisation domain containing 12
KLF3	Kruppel-like factor 3 (basic)
KLHDC8B	kelch domain containing 8B
LACTB	lactamase, beta

LAIR1	leukocyte-associated immunoglobulin-like receptor 1
LAMP1	lysosomal-associated membrane protein 1
LAMP2	lysosomal-associated membrane protein 2
LAP3	leucine aminopeptidase 3
LASS6	LAG1 homolog, ceramide synthase 6
LCP2	lymphocyte cytosolic protein 2 (SH2 domain containing leukocyte protein of 76kDa)
LGALS2	lectin, galactoside-binding, soluble, 2
LGALS3	lectin, galactoside-binding, soluble, 3
LGALS3BP	lectin, galactoside-binding, soluble, 3 binding protein
LGALS8	lectin, galactoside-binding, soluble, 8
LILRB1	leukocyte immunoglobulin-like receptor, subfamily B (with TM and ITIM domains), member 1
LILRB2	leukocyte immunoglobulin-like receptor, subfamily B (with TM and ITIM domains), member 2
LILRB3	leukocyte immunoglobulin-like receptor, subfamily B (with TM and ITIM domains), member 3
LILRB4	leukocyte immunoglobulin-like receptor, subfamily B (with TM and ITIM domains), member 4
LIMK2	LIM domain kinase 2
LIPA	lipase A, lysosomal acid, cholesterol esterase
LPCAT2	lysophosphatidylcholine acyltransferase 2
LRP1	low density lipoprotein-related protein 1 (alpha-2-macroglobulin receptor)
LST1	leukocyte specific transcript 1
LTA4H	leukotriene A4 hydrolase
LTBR	lymphotoxin beta receptor (TNFR superfamily, member 3)
LY96	lymphocyte antigen 96
M6PR	mannose-6-phosphate receptor (cation dependent)
MADD	MAP-kinase activating death domain
MAFB	v-maf musculoaponeurotic fibrosarcoma oncogene homolog B (avian)
MANBA	mannosidase, beta A, lysosomal
MAP3K3	mitogen-activated protein kinase kinase kinase 3
MARCO	macrophage receptor with collagenous structure
MCTP1	multiple C2 domains, transmembrane 1
ME1	malic enzyme 1, NADP(+)-dependent, cytosolic
MFSD1	major facilitator superfamily domain containing 1
MFSD7	major facilitator superfamily domain containing 7
MGAT4A	mannosyl (alpha-1,3-)-glycoprotein beta-1,4-N-acetylglucosaminyltransferase, isozyme A
MITF	microphthalmia-associated transcription factor
MPND	MPN domain containing
MPP1	membrane protein, palmitoylated 1, 55kDa
MRAS	muscle RAS oncogene homolog
MT1E	metallothionein 1E
MT1F	metallothionein 1F
MT1G	metallothionein 1G
MT1H	metallothionein 1H
MT1M	metallothionein 1M
MT1P2	metallothionein 1 pseudogene 2
MT1X	metallothionein 1X
MT2A	metallothionein 2A
MVP	major vault protein

MYO1F	myosin IF
NAGK	N-acetylglucosamine kinase
NAIP	NLR family, apoptosis inhibitory protein
NDRG1	N-myc downstream regulated gene 1
NFE2L2	nuclear factor (erythroid-derived 2)-like 2
NINJ1	ninjurin 1
NKG7	natural killer cell group 7 sequence
NLRC4	NLR family, CARD domain containing 4
NPC1	Niemann-Pick disease, type C1
NPC2	Niemann-Pick disease, type C2
NPL	N-acetylneuraminate pyruvate lyase (dihydrodipicolinate synthase)
NR1H3	nuclear receptor subfamily 1, group H, member 3
NRP2	neuropilin 2
NUB1	negative regulator of ubiquitin-like proteins 1
NUPR1	nuclear protein 1
OLFML2B	olfactomedin-like 2B
OSCAR	osteoclast associated, immunoglobulin-like receptor
OSTM1	osteopetrosis associated transmembrane protein 1
P2RX7	purinergic receptor P2X, ligand-gated ion channel, 7
P2RY13	purinergic receptor P2Y, G-protein coupled, 13
P2RY6	pyrimidinergic receptor P2Y, G-protein coupled, 6
P76	mannose-6-phosphate protein p76
PADI2	peptidyl arginine deiminase, type II
PECAM1	platelet/endothelial cell adhesion molecule
PECAM1	platelet/endothelial cell adhesion molecule
PILRA	paired immunoglobin-like type 2 receptor alpha
PLA1A	phospholipase A1 member A
PLA2G7	phospholipase A2, group VII (platelet-activating factor acetylhydrolase, plasma)
PLAUR	plasminogen activator, urokinase receptor
PLCL1	phospholipase C-like 1
PLD1	phospholipase D1, phosphatidylcholine-specific
PLD3	phospholipase D family, member 3
PLEKHB2	pleckstrin homology domain containing, family B (evectins) member 2
PLEKHG3	pleckstrin homology domain containing, family G (with RhoGef domain) member 3
PLEKHO2	pleckstrin homology domain containing, family O member 2
PLOD3	procollagen-lysine, 2-oxoglutarate 5-dioxygenase 3
PLSCR1	phospholipid scramblase 1
PLXDC2	plexin domain containing 2
PLXNC1	plexin C1
PPP2R2B	protein phosphatase 2 (formerly 2A), regulatory subunit B, beta isoform
PPT1	palmitoyl-protein thioesterase 1
PQLC3	PQ loop repeat containing 3
PRAM1	PML-RARA regulated adaptor molecule 1
PRCP	prolylcarboxypeptidase (angiotensinase C)
PRF1	perforin 1 (pore forming protein)
PSAP	prosaposin
PSEN1	presenilin 1
PSTPIP2	proline-serine-threonine phosphatase interacting protein 2
PTAFR	platelet-activating factor receptor

PTGER2	prostaglandin E receptor 2 (subtype EP2), 53kDa
PVRL2	poliovirus receptor-related 2 (herpesvirus entry mediator B)
PYCARD	PYD and CARD domain containing
RAB20	RAB20, member RAS oncogene family
RAB32	RAB32, member RAS oncogene family
RALB	v-ral simian leukemia viral oncogene homolog B (ras related; GTP binding protein)
RALGDS	ral guanine nucleotide dissociation stimulator
RAP2B	RAP2B, member of RAS oncogene family
RARRES1	retinoic acid receptor responder (tazarotene induced) 1
RARRES3	retinoic acid receptor responder (tazarotene induced) 3
RASSF4	Ras association (RalGDS/AF-6) domain family member 4
RBM47	RNA binding motif protein 47
RENBP	renin binding protein
RGL1	ral guanine nucleotide dissociation stimulator-like 1
RNF13	ring finger protein 13
RNF130	ring finger protein 130
RNF149	ring finger protein 149
RNF213	ring finger protein 213
RRAS	related RAS viral (r-ras) oncogene homolog
S100A11	S100 calcium binding protein A11
S100A11P	S100 calcium binding protein A11 pseudogene
S100A8	S100 calcium binding protein A8
S100A9	S100 calcium binding protein A9
SAMHD1	SAM domain and HD domain 1
SASH1	SAM and SH3 domain containing 1
SAT1	spermidine/spermine N1-acetyltransferase 1
SCARB2	scavenger receptor class B, member 2
SCO2	SCO cytochrome oxidase deficient homolog 2 (yeast)
SCPEP1	serine carboxypeptidase 1
SDC3	syndecan 3
SDCBP	syndecan binding protein (syntenin)
SDSL	serine dehydratase-like
SECTM1	secreted and transmembrane 1
SERPINA1	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1
SERPING1	serpin peptidase inhibitor, clade G (C1 inhibitor), member 1
SFXN3	sideroflexin 3
SGK3	serum/glucocorticoid regulated kinase family, member 3
SH3PXD2B	SH3 and PX domains 2B
SIGLEC1	sialic acid binding Ig-like lectin 1, sialoadhesin
SIGLEC11	sialic acid binding Ig-like lectin 11
SIGLEC16	sialic acid binding Ig-like lectin 16 (gene/pseudogene)
SIGLEC7	sialic acid binding Ig-like lectin 7
SIGLEC9	sialic acid binding Ig-like lectin 9
SIRPB1	signal-regulatory protein beta 1
SIRPB2	signal-regulatory protein beta 2
SLAMF7	SLAM family member 7
SLAMF8	SLAM family member 8
SLC11A2	solute carrier family 11 (proton-coupled divalent metal ion transporters), member 2
SLC16A3	solute carrier family 16, member 3 (monocarboxylic acid transporter 4)
SLC22A18	solute carrier family 22, member 18

SLC24A6	solute carrier family 24 (sodium/potassium/calcium exchanger), member 6
SLC27A1	solute carrier family 27 (fatty acid transporter), member 1
SLC27A2	solute carrier family 27 (fatty acid transporter), member 2
SLC29A3	solute carrier family 29 (nucleoside transporters), member 3
SLC2A9	solute carrier family 2 (facilitated glucose transporter), member 9
SLC30A1	solute carrier family 30 (zinc transporter), member 1
SLC31A1	solute carrier family 31 (copper transporters), member 1
SLC31A2	solute carrier family 31 (copper transporters), member 2
SLC38A6	solute carrier family 38, member 6
SLC39A8	solute carrier family 39 (zinc transporter), member 8
SLC46A1	solute carrier family 46 (folate transporter), member 1
SLC46A3	solute carrier family 46, member 3
SLC6A12	solute carrier family 6 (neurotransmitter transporter, betaine/GABA), member 12
SLC8A1	solute carrier family 8 (sodium/calcium exchanger), member 1
SLC8A1	solute carrier family 8 (sodium/calcium exchanger), member 1
SLCO2B1	solute carrier organic anion transporter family, member 2B1
SMPDL3A	sphingomyelin phosphodiesterase, acid-like 3A
SNX10	sorting nexin 10
SNX6	sorting nexin 6
SOAT1	sterol O-acyltransferase 1
SOD2	superoxide dismutase 2, mitochondrial
SPECC1	sperm antigen with calponin homology and coiled-coil domains 1
SPG20	spastic paraplegia 20 (Troyer syndrome)
SPNS1	spinster homolog 1 (Drosophila)
SPPL2A	signal peptide peptidase-like 2A
SPRY2	sprouty homolog 2 (Drosophila)
SQSTM1	sequestosome 1
STARD8	StAR-related lipid transfer (START) domain containing 8
STAT1	signal transducer and activator of transcription 1, 91kDa
STEAP3	STEAP family member 3
STOM	stomatin
STX4	syntaxin 4
TBC1D2	TBC1 domain family, member 2
TBC1D2B	TBC1 domain family, member 2B
TBXAS1	thromboxane A synthase 1 (platelet)
TCF7L1	transcription factor 7-like 1 (T-cell specific, HMG-box)
TCN2	transcobalamin II; macrocytic anemia
TEX2	testis expressed 2
TGM2	transglutaminase 2 (C polypeptide, protein-glutamine-gamma-glutamyltransferase)
TIFAB	TRAF-interacting protein with forkhead-associated domain, family member B
TIMP2	TIMP metallopeptidase inhibitor 2
TLR1	toll-like receptor 1
TLR2	toll-like receptor 2
TLR4	toll-like receptor 4
TLR5	toll-like receptor 5
TLR8	toll-like receptor 8

TMBIM1	transmembrane BAX inhibitor motif containing 1
TMEM127	transmembrane protein 127
TMEM140	transmembrane protein 140
TMEM144	transmembrane protein 144
TMEM176A	transmembrane protein 176A
TMEM176B	transmembrane protein 176B
TMEM50A	transmembrane protein 50A
TMEM51	transmembrane protein 51
TMEM86A	transmembrane protein 86A
TMSB10	thymosin beta 10
TNFAIP2	tumor necrosis factor, alpha-induced protein 2
TNFRSF1A	tumor necrosis factor receptor superfamily, member 1A
TNFSF10	tumor necrosis factor (ligand) superfamily, member 10
TNFSF12	tumor necrosis factor (ligand) superfamily, member 12
TNFSF13	tumor necrosis factor (ligand) superfamily, member 13
TNFSF13B	tumor necrosis factor (ligand) superfamily, member 13b
TNS1	tensin 1
TNS3	tensin 3
TSPAN14	tetraspanin 14
TSPAN4	tetraspanin 4
TSPO	translocator protein (18kDa)
TYMP	thymidine phosphorylase
TYROBP	TYRO protein tyrosine kinase binding protein
UBD	ubiquitin D
UBE2F	ubiquitin-conjugating enzyme E2F (putative)
VAC14	Vac14 homolog (S. cerevisiae)
VAMP3	vesicle-associated membrane protein 3 (cellubrevin)
VAMP5	vesicle-associated membrane protein 5 (myobrevin)
VAT1	vesicle amine transport protein 1 homolog (T. californica)
VPS37C	vacuolar protein sorting 37 homolog C (S. cerevisiae)
VSIG4	V-set and immunoglobulin domain containing 4
VWA5A	von Willebrand factor A domain containing 5A
WARS	tryptophanyl-tRNA synthetase
WDFY3	WD repeat and FYVE domain containing 3
ZDHHC7	zinc finger, DHHC-type containing 7
ZFYVE16	zinc finger, FYVE domain containing 16
ZMYND15	zinc finger, MYND-type containing 15
ZNF438	zinc finger protein 438
ZNF702P	zinc finger protein 702 pseudogene

Lying close to the macrophage cluster is a cluster enriched in genes associated with T cells. This is formed of 255 probesets representing 152 genes. It includes many T-cell surface markers including the pan-T cell antigens CD2, CD3 (the delta, epsilon and gamma components of the CD3 molecule), CD5 and CD7 as well as multiple components of the T cell receptor. Various elements of T cell receptor signalling are represented (FYN, ITK, LAT, LCP2, PRKCQ and ZAP70), as are cytotoxic molecules granzyme K and granzyme M.

Examination of the T cell expression profile across the dataset shows a pattern similar to that seen with the macrophage signature in that there is considerable variation in expression across the dataset but no obvious pattern to the expression (Figure 4.10). Again selected functional annotation is provided (Table 4.5).

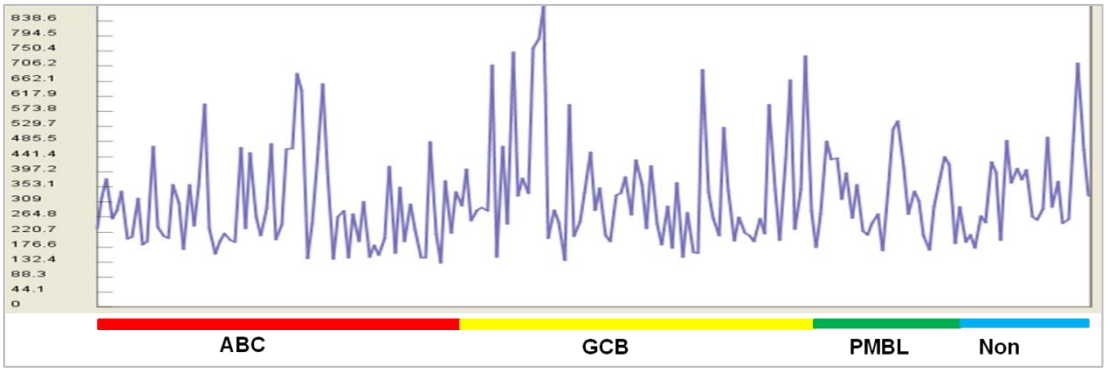


Figure 4.10 Mean intensity of the T cell cluster across the DLBCL dataset. There is considerable variation across the dataset but no obvious trend associated with tumour subtype or outcome. Individual tumours across x-axis with coloured bars indicating tumour subtypes.

Analysis tool	Source of annotation	Annotation	Significance (p-value)
GSEA	GSE10325	Genes upregulated in comparison of healthy CD4 T cells vs CD19 B cells	7.3×10^{-65}
GSEA	PID pathway	TCR signalling	4.1×10^{-20}
GSEA	GOTERM	Immune system process	9.9×10^{-20}
GSEA	KEGG pathway	TCR signalling	3.7×10^{-17}
DAVID	GOTERM_BP_FAT	T cell activation	2.4×10^{-18}
DAVID	SP_PIR_KEYWORDS	T cell	1.5×10^{-11}

Table 4.5 Functional annotation for the gene list from the 'T cell signature'. Selected annotation derived from DAVID and GSEA gene expression analysis tools. The T cell gene list was uploaded to DAVID and analysed through the functional annotation clustering tool and in GSEA using the molecular signatures database. Only selected annotation is shown here. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

There is a cluster lying adjacent to and between the macrophage cluster and T cell cluster that appears to represent the interferon response consisting of 83 probesets representing 58 genes (Figure 4.11, Table 4.6). These include the interferon regulatory factors IRF7 and IRF9 as well as multiple interferon inducible genes (IFI27, IFI35, IFI44, IFI44L, IFI6, IFIH1, IFIT2, IFIT3, IFIT5, IFITM1, IFITM2, IFITM3). Figure 4.12 demonstrates the relationships between the T cell, macrophage, interferon and ECM clusters.

Analysis tool	Source of annotation	Annotation	Significance (p-value)
GSEA	Hecker_IFNB1_targets	Genes transcriptionally modulated in the blood of patients with multiple sclerosis treated with IFNbeta1	3.2×10^{-73}
GSEA	Moserle_IFNA_response	Top 50 genes upregulated in ovarian cancer progenitor cells in response to IFNalpha	2.1×10^{-59}
GSEA	Bosce_IFN_induced_antiviral module	Genes representing interferon induced antiviral module in sputum during asthma exacerbations	1×10^{-58}
GSEA	Browne_interferon_responsive genes	Genes upregulated in primary fibroblast culture after treatment with IFNalpha for 6 hours	6.3×10^{-52}

Table 4.6 Functional annotation for the gene list from the 'IFN signature'. Selected annotation derived from DAVID and GSEA gene expression analysis tools. The IFN gene list was uploaded to DAVID and analysed through the functional annotation clustering tool and in GSEA using the molecular signatures database. Only selected annotation is shown here. P-values are for hypergeometric distribution test.

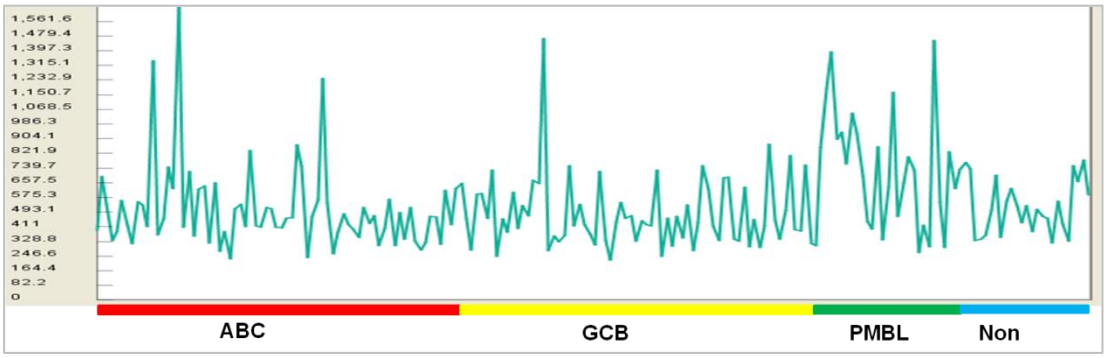


Figure 4.11 Mean expression intensity of the IFN cluster across the dataset. Individual tumours along the x-axis. Coloured bars demote tumour subtypes.

As well as these clearly immune clusters there are other clusters that lie in the 'stromal' area of the graph that appear to represent components of the extracellular

matrix and vasculature. The main ECM cluster forms a relatively large cluster close to but separate from the immune clusters. Expression levels of the genes in this cluster are higher in cases classified as PMBL compared to the other subtypes of DLBCL (Figure 4.12). This is in keeping with the recognised histological characteristics of PMBL, a tumour frequently characterised by a dense fibrotic stroma in which the tumour cells reside (Swerdlow et al 2008). The main ECM cluster contains 437 probesets representing 288 genes. This cluster contain genes for many structural proteins (collagens I, III,IV, V, VI, XI, XIII,XV and XVI, fibronectin 1, laminin alpha 4, beta 1 and 2 and gamma 1, caldesmon 1 and heparan sulphate proteoglycan 2). In addition to these, there are many modulators of the extracellular matrix including ADAM metallopeptidases (AMAM12, AMADTS12, ADAMTS5), matrix metallopeptidase 2 and several TIMPS (TIMP1, 2 and 3). Elements of the vascular signature are also included in this cluster including VEGFC, WT1, endothelial cell adhesion molecule and podoplanin. See table 4.7 for selected functional annotation. The relationships between the T cell, macrophage, interferon and ECM clusters are shown in Figure 4.13.

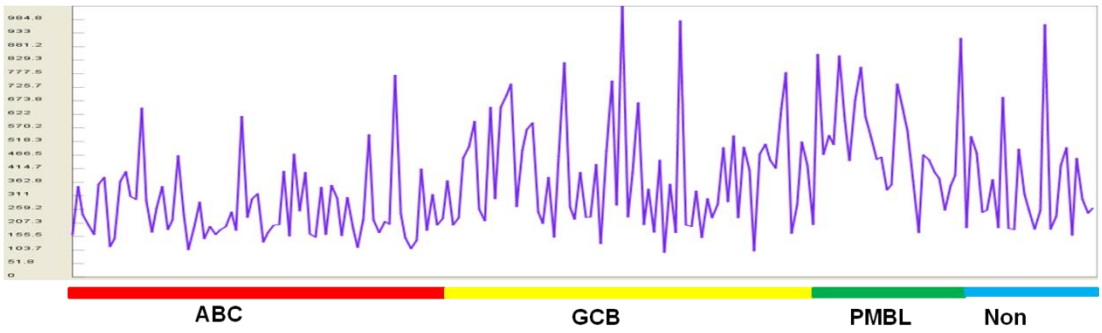


Figure 4.12 Mean expression intensity across the ECM cluster. Expression is higher in the tumours of PMBL subtype reflecting the histological appearances. There is no association with outcome. Individual tumours lie along the x-axis, with coloured bars representing tumour subtypes. Within each subtypes good outcome tumours lies to the left and poor outcome to the right.

Analysis tool	Source of annotation	Annotation	Significance (p-value)
GSEA	GO term	Extracellular matrix	1.8×10^{-40}
GSEA	Stanford tumour modules	ECM and collagens	1.1×10^{-122}
DAVID	SP_PIR_keywords	Extracellular matrix	2.1×10^{-46}
DAVID	GOTERM_BP_FAT	Cell adhesion	1.9×10^{-26}

Table 4.7 Functional annotation for the gene list from the 'ECM signature'. Selected annotation derived from DAVID and GSEA gene expression analysis tools. The ECM gene list was uploaded to DAVID and analysed through the functional annotation clustering tool and in GSEA using the molecular signatures database. Only selected annotation is shown here. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

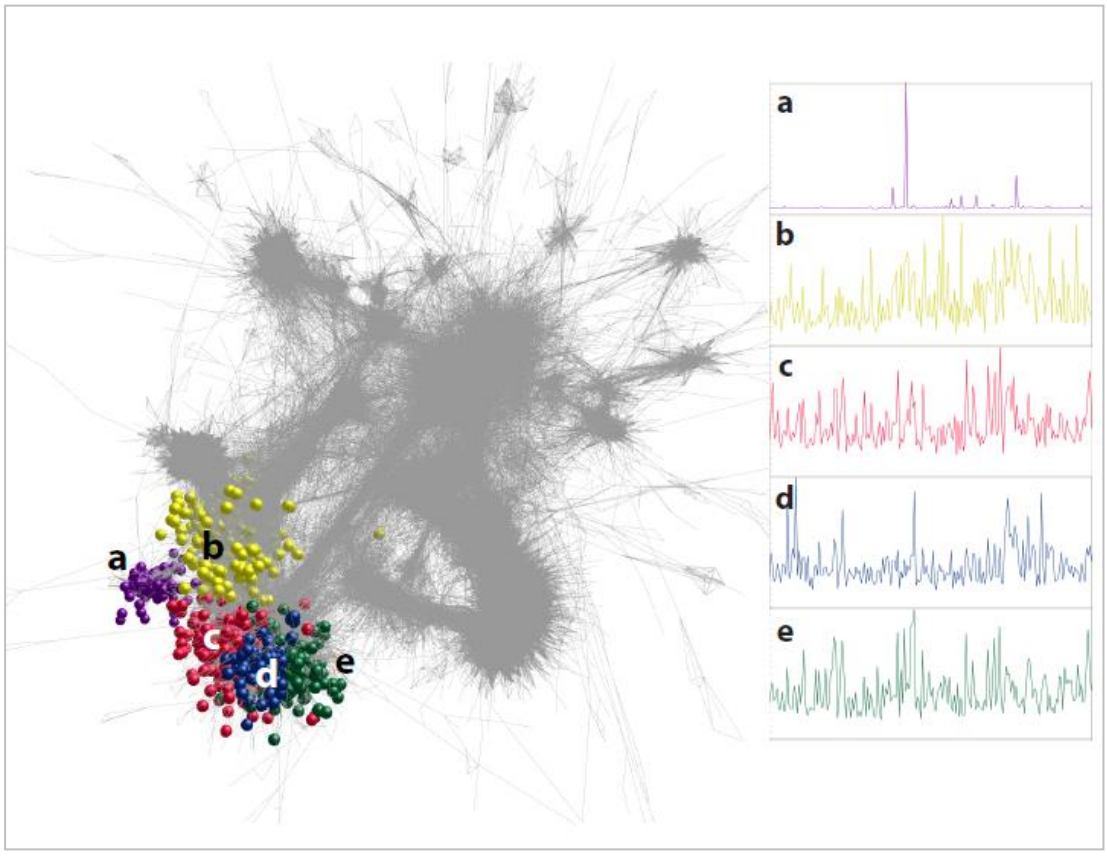


Figure 4.13 Further view of the graph network to show the relationships within the stromal cluster. The main graph is represented as edges only with no visible nodes. Selected nodes are made visible and enlarged. The right hand panel indicates mean gene expression in each cluster illustrated. A) The skeletal muscle cluster which lies separate from the main graph but connected to it. B) The ECM cluster lying close to but separate from the immune clusters. C) T-cell, D) Interferon response and E) Macrophage all lie very close together.

As well as elements of the endothelial signature that lie within the main ECM cluster there are is also a separate small cluster of endothelial genes. This lies closely associated with the ECM cluster and contains genes such as CD34, cadherin 5 (vascular endothelial cadherin), von Willibrand factor, endomucin and TEK, tyrosine kinase endothelial. A further cluster of ECM genes lies immediately adjacent to the main cluster and a cluster of adipocyte genes near but not abutting the main ECM cluster (Figure 4.14).

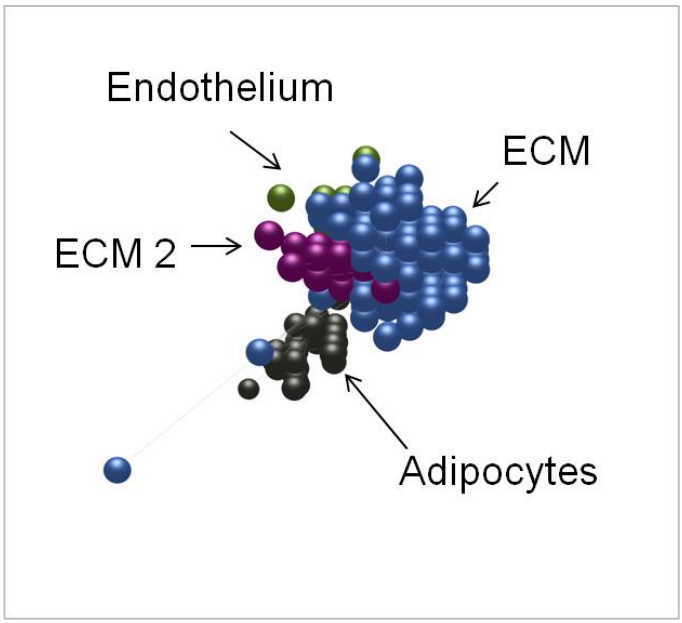


Figure 4.14 The clusters identified as ECM or endothelial genes cluster together within the graph network. The selected cluster are made visible while the rest of the graph is hidden to demonstrate cluster relationships.

4.4.2 Cell cycle related clusters

The other major area of the graph where it is possible to assign function to the clusters is in the area broadly represented by the cell cycle. There are several separate but closely related clusters in this area which group together. The main cell cluster contains 652 probesets representing 539 genes. This contains genes in which RNA binding activity and transcriptional regulators are prominent and include multiple RNA binding proteins (RBM14, RBM17, RBM22, RBM26, RBM33), splicing factors (SF1, SFRS1, SFRS2, SFRS4, SFRS7, SFRS8), nuclear riboproteins (HNRNPC, HNRNPF, HNRNPR, HNRNPU), and transcription factors (E2F1, E2F4,

SP1, YY1, ATF1, ATF2, HSF1, HSF2, NFYC) (see table 4.8 for selected functional annotation). A further cluster contains 184 probesets representing 143 genes. This appears to relate to the mitotic phase of the cell cycle (Table 4.9). The cluster includes the aurora kinases (AURKA and AURKB), centromere proteins (CENPA, CENPE, CENPF, CENPK), lymphoid specific helicase (HELLS), kinesin family members (KIF11,KIF14, KIF18A, KIF18B, KIF20A, KIF20B, KIF2C, KIFC1), DNA polymerases (POLA2, POLE3A) topoisomerase (TOP2A) and tubulins (TUBB2C, TUBB3) among many others. A complete list of genes in this cluster is provided in appendix 1. Several small related clusters that lie adjacent to these main clusters contain cell cycle related genes such as TATA box binding protein associated factor (TAF1B), RNA polymerase (POLR3H), splicing factor (SFRS7) and ribonuclease (RPP14). Analysis of the expression profile of the cell cycle genes shows variation across the dataset but no particular relationship with tumour type or outcome (Figure 4.15). This may in part reflect the treatment approach used in lymphoma where with current therapeutic regimens tumours with higher rates of proliferation including DLBCL are potentially curable in response to aggressive chemotherapy thus subtle differences in cell cycle in the tumour expression profile are negated by the intensive nature of the therapy received.

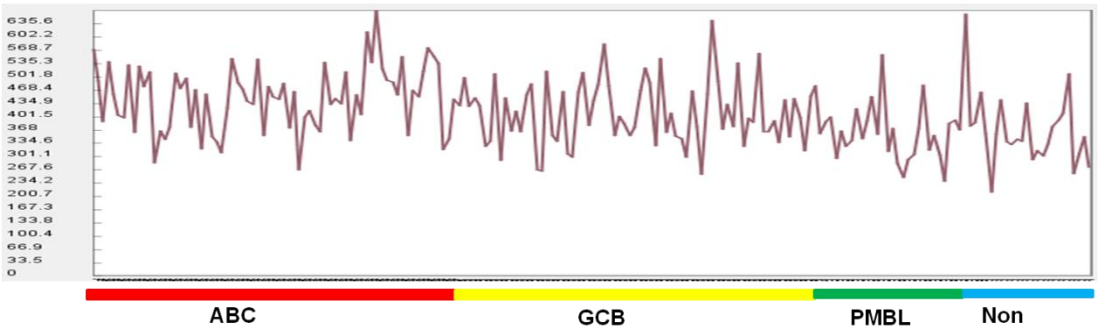


Figure 4.15 Cell cycle genes across DLBCL dataset. Mean expression of the genes forming the cell cycle clusters. Coloured bars represent tumour subtypes. Cell cycle activity shows similar expression across all tumours regardless of subtype or outcome.

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	SP_PIR_keywords	RNA binding	4.0×10^{-17}
DAVID	GOTERM_MF_FAT	Transcription factor binding	3.6×10^{-10}
DAVID	SP_PIR_keywords	Transcription	5.3×10^{-11}

Table 4.8 Functional annotation for the gene list from the largest cell cycle cluster. Selected annotation derived from DAVID and GSEA gene expression analysis tools. Only selected annotation is shown here. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Cell cycle	7.1×10^{-46}
DAVID	GOTERM_BP_FAT	M phase	1.8×10^{-46}
DAVID	SP_PIR_keywords	Mitosis	1.9×10^{-38}
GSEA	REACTOME	Cell cycle, mitotic	4.8×10^{-44}

Table 4.9 Functional annotation for the gene list from the cell cycle S/M phase cluster. Selected annotation derived from DAVID and GSEA gene expression analysis tools. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

4.4.3 Other non-stromal, non-cell cycle clusters

In addition, in this dataset, as in the others studied there were clusters of genes which could be assigned a function, but which were not of great interest in terms of insight into either stromal or malignant components of the dataset. These included clusters of ribosomal genes, genes involved in glycolysis, histone clusters, and haemoglobin clusters, among others.

The focus of this analysis was not on determining groups of genes that could subdivide types of tumour, but rather on looking for non-cancerous signatures, in particular those relating to the stroma, but analysis of the dataset did provide some

insights into the cancer related elements of each dataset. In this dataset, as in the others discussed here there are very few clusters that demonstrate clear difference in expression between subtypes of tumour or between clinical outcome, illustrating the difficulties in gene expression analysis directed to these questions i.e. that most of the data in a cancer gene expression study does not relate to factors specific to the tumour or to outcome. Where such clusters can be found however, they can be informative. In the DLBCL dataset, IRF4, one of the markers by immunohistochemistry of the ABC-subtype, does not lie in a cluster, but searching for the nearest neighbours of this draws in genes from the periphery of several small adjacent clusters which include FOXP1, PIM2 and CARD11, all described to be up-regulated in ABC-subtype of DLBCL, with amplifications or mutation affecting FOXP1 and CARD11 identified in 38% and 10% respectively of tumours studied (Figure 4.16) (Lenz et al 2008a, Davis et al 2010).

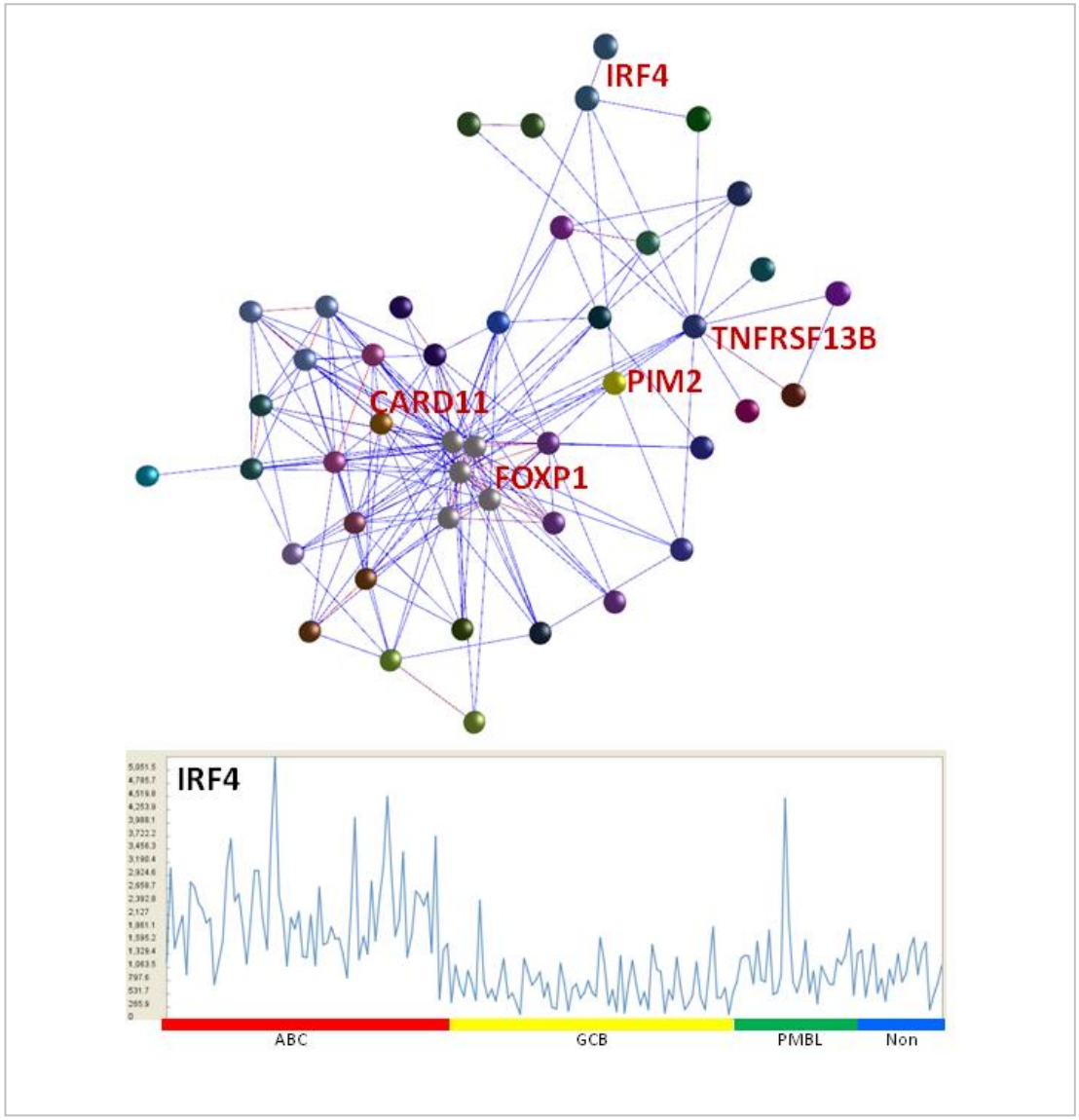


Figure 4.16 Tumour networks in DLBCL. IFR4 shows variation in expression between tumour subtypes being higher in ABC subtype than others. Searching for the nearest neighbours of this gene captures a small group of genes, many of which have been demonstrated to be mutated in tumours of ABC subtype. The other genes within this grouping are therefore also worthy of further study.

Having demonstrated that cell type and function signatures could be demonstrated by clustering analysis in the DLBCL dataset, this analysis was expanded to multiple further datasets to ensure the approach could be applies more universally and to provide signatures of TAMs in more than one tumour type to allow the final 'merge' analysis to be performed.

4.5 Gene signatures from Breast cancer

The breast cancer dataset consisted of 159 samples from patients treated in Sweden, of which 134 passed the QC step. These were stratified on the basis of molecular tumour type (Basal, HER2 positive, Luminal A or B, or normal-like), Ellis-Elston grade, survival outcome and recurrence outcome. The breast cancer network when laid out at $r \geq 0.65$ contains 19,246 nodes and 559,761 edges and has a polarised shape, with a large central core and clusters arising from opposite poles of this (Figure 4.17). The clusters at opposite ends of the graph are poorly characterised and designated housekeeping clusters, but within the main central area there are immune and cell cycle clusters. The immune clusters do not form a tight group, but rather separate into two main areas on opposite sides of the main cluster. There is considerable overlap of immune cell type signatures in these clusters with some B and T cell genes falling in a single cluster and T cell and macrophage genes in a single cluster. Cell cycle genes form multiple clusters related to expression pattern across tumour subsets with a large cluster formed of genes that go up with increasing aggressiveness of tumour and tumour grade, and others clusters with a different pattern of expression. This dataset seems particularly rich in tumours with 'spiked' profiles with many clusters formed of genes that are essentially expressed in only a single tumour. Despite this, multiple clusters of identifiable function are seen.

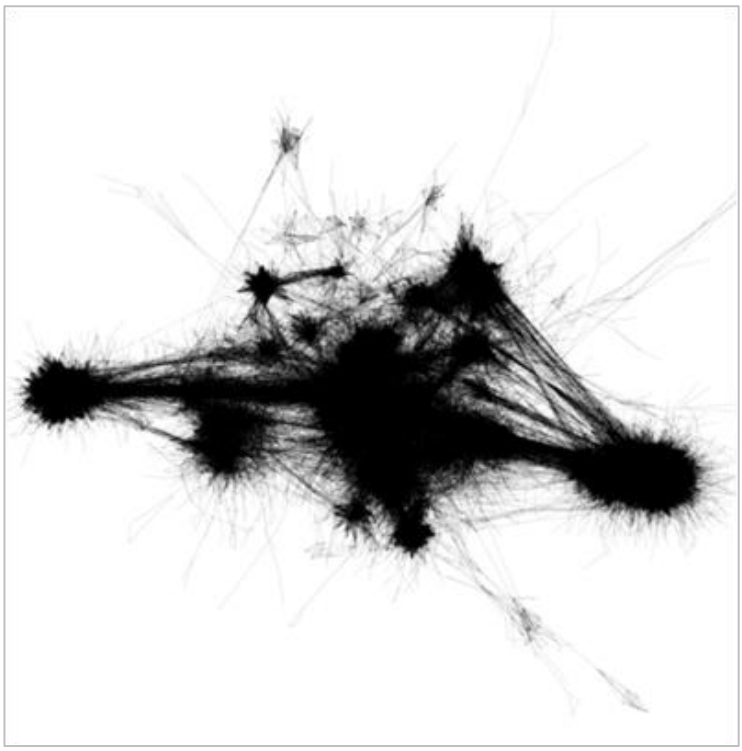


Figure 4.17 The breast cancer network laid out at r≥0.65 with nodes hidden and only edges visible. The stromal and cell cycle related clusters lie at sides of the main graph while less well characterised clusters make up the central part and extreme poles.

4.5.1 Stromal signatures

There are several distinct immune related signatures present in the dataset. These tend to fall within the central part of the graph. These break up into two main areas with several closely related clusters that appear to represent a lymphocyte signature and a macrophage signature falling closely together, and several other clusters which lie apart from these that represent T cells and macrophages (Figure 4.18). The interferon cluster breaks into three parts which joins these two main immune areas. Unlike in the DLBCL dataset, there is a distinct B cell/plasma cell signature which lies adjacent to the lymphocyte signature. These signatures all show considerable variation across the dataset, but no clear pattern of expression related to tumour type other than a suggestion of higher expression of all in the basal and EBB2 types, and reduced interferon response message in the normal-like tumours.

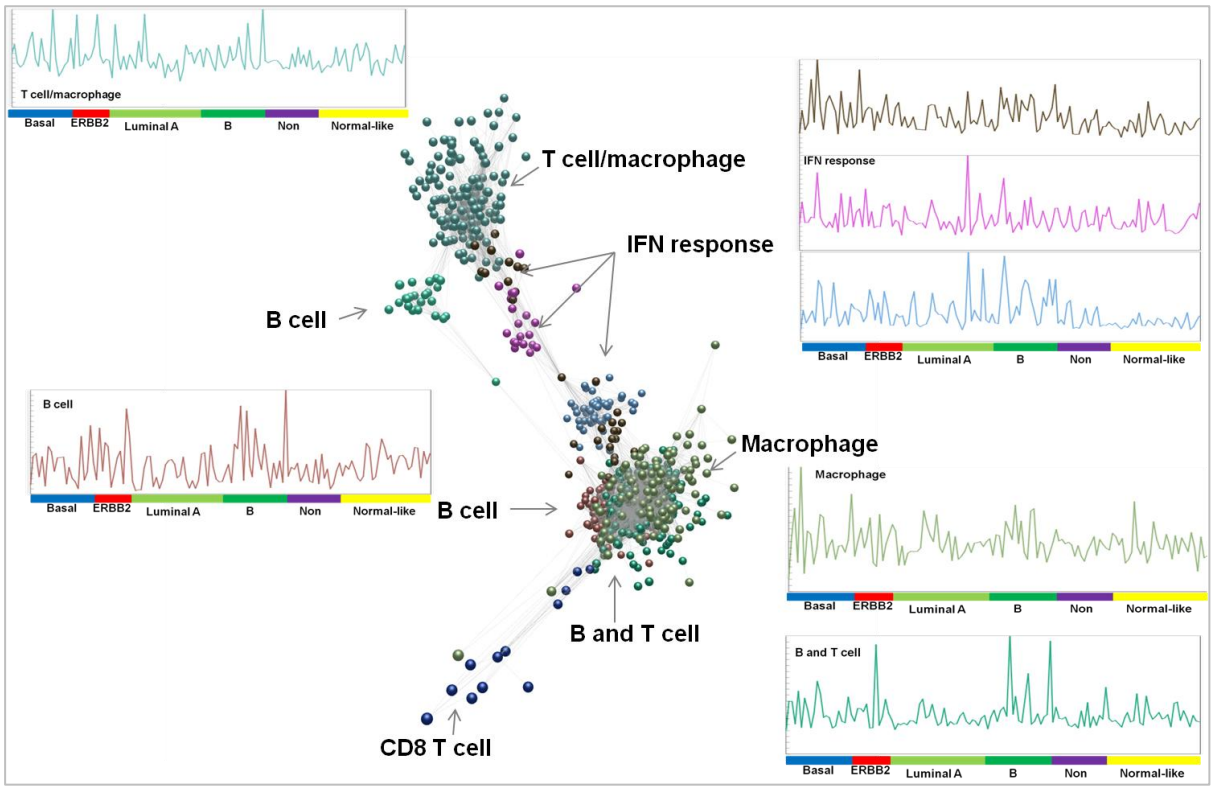


Figure 4.18. The immune clusters in the Breast dataset. The rest of the graph is hidden and only the selected immune clusters remain visible showing the division of the immune signatures into two broad areas. Side panels show the mean expression patterns in each of these clusters. Each cluster and the corresponding expression profile is in the same colour. Individual tumours represented on x-axis of these panels. Coloured bars indicate tumour subtype.

The macrophage only signature consists of 116 probesets representing 93 genes and contains markers of myeloid differentiation CD14, CD163, CD33 and CD4. The MCH class II genes fall within this cluster as do multiple Fc receptors (FCER1G, FCGR1A and FCGR1B). Allograft inflammatory factor, legumain, and mannose receptor C, type 1 also lie in this cluster. Analysis through DAVID identifies the cluster as being enriched in genes involved in defence response, antigen processing and presentation and phagocytosis (Table 4.10).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Defense response	7.0×10^{-20}
DAVID	GOTERM_BP_FAT	Antigen processing and presentation	2.2×10^{-19}
DAVID	GOTERM_BP_FAT	Phagocytosis	9.0×10^{-6}
GSEA	GSE22886	Genes downregulated in naive CD8 T cells vs monocytes	4.6×10^{-40}

Table 4.10 Functional annotation for the macrophage signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

The cluster with overlapping signatures of macrophages and T cells is slightly larger consisting of 147 probesets and 99 genes. This contains less obviously lineage specific genes and is identified largely from overlaps with T cell and macrophage clusters from other datasets. It does however contain CD5, CTLA4 and FYB as markers of T cell lineage as well as TLR8, IL-6 and IL-7 receptors and macrophage specific genes such as macrophage expressed 1 (MPEG1). Annotation from databases is less clear as to the nature of this cluster but does identify T cell activation and genes involved in the immune system (Table 4.11).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Regulation of T cell activation	9.9×10^{-7}
GSEA	REACTOME	Immune system	4.7×10^{-15}

Table 4.11 Functional annotation for the macrophage/T cell signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

The largest immune cluster (232 probesets representing 187 genes) contains elements common to both B and T lymphocytes. This shared lymphocyte signature appeared largely lost in the DLBCL signature where the only normal lymphocyte population identifiable was the T cell signature. This contains surface markers of both T and B cell lineage (CD2, CD3, CD7, CD52, CD19, CD22, CD79B) as well as elements of the T cell receptor and B cell receptor molecules (TRA, TRAC, IGMH). There are cytotoxic molecules granzymes A, H, K and M (GZMA, GZMH, GZMK, GZMM) and perforin 1 (PRF1). Multiple intracellular signalling molecule specific to each cell type or shared between them are present (BTK, FYN, FYB, ITK, ZAP70), as are numerous chemokines, cytokines and their receptors (CCL19, CCL5, CCR2, CCR5, CCR7, CXCL9, CXCR3, CXCR5, CXCR6, IL10RA, IL16, IL21R, IL23A, IL2RB, IL2RG, 1L32, IL7R). Selected functional annotation given in Table 4.12.

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Leukocyte activation	2.1×10^{-24}
DAVID	GOTERM_BP_FAT	Positive regulation of immune system process	2.4×10^{-21}
DAVID	KEGG PATHWAY	T cell receptor signalling pathway	4.7×10^{-13}

Table 4.12 Functional annotation for the lymphocyte signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values.

A small cluster (105 probesets of 53 genes) represents B cells or plasma cells, most likely plasma cells given the cluster contains CD38, a surface marker of plasma cells and IRF4, a transcription factor expressed in the late B cell/plasma cell stage of differentiation. This cluster is formed mostly of immunoglobulin genes. A further much small cluster of only 24 probesets formed also of B cell genes lies separately. See table 4.13 for selected functional annotation.

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	SP_PIR_KEYWORDDS	Immunoglobulin	1.5×10^{-19}
DAVID	GOTERM_BP_FAT	Antigen binding	3.3×10^{-11}
GSEA	GSE10325	Upregulated in B cells vs myeloid cells	1.1×10^{-35}

Table 4.13 Functional annotation for the B cell/plasma cell signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

The interferon response splits into three clusters that bridge the space between the two main immune grouping. In total these clusters represent 96 probesets and 68 genes. These show some variation with a subset of genes showing lower expression in normal-like breast cancer samples than others. Analysis through DAVID and GSEA support the categorisation as being associated with the interferon response but do not provide any insight into the nature of the differential expression between tumour types. This interferon cluster contains genes for multiple interferon induced proteins (IFI27, IFI35, IFI44, IFI44L, IFIH1, IFI6, IFIT1, IFIT2, IFIT3, IFIT5), interferon regulatory factors (IRF7, IRF9) and interferon stimulated exonuclease gene (ISG20).

The non-immune stromal signatures fall into two distinct and unattached areas, each of which contains areas representing extracellular matrix, endothelium and fat. (Figure 4.19) Those forming the tighter cluster of fat, endothelium and ECM tend to be higher in tumour of normal-like type.

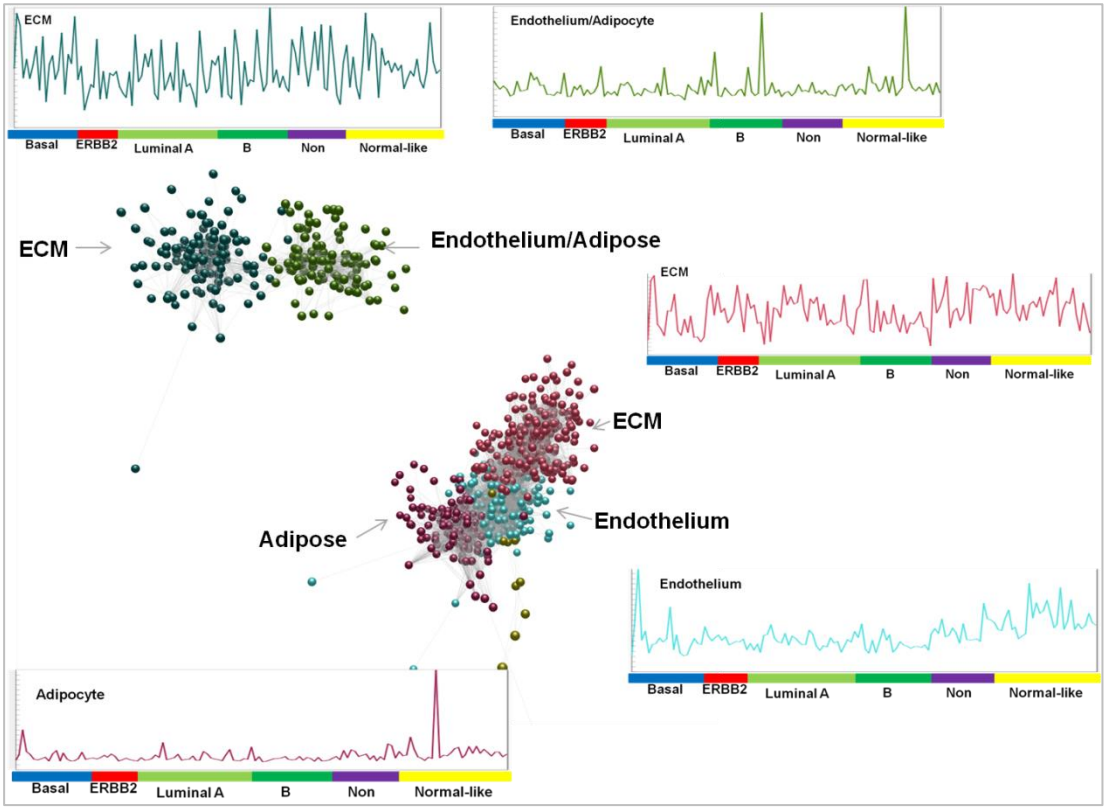


Figure 4.19. The stromal clusters in the Breast dataset. The rest of the graph is hidden and only the non-immune stromal signatures left visible. The stromal signatures divide into two broad, unconnected areas. The side panels around the network graph show the mean expression patterns in each of these clusters. Each cluster and the associated expression profile uses the same colour. Individual tumours represented on x-axis of these panels. Coloured bars indicate tumour subtype.

The ECM cluster in this area (red in figure 4.19) contains 273 probesets representing 195 genes and contains numerous collagens (COL1A1, COL1A2, COL3A1, COL5A1, COL5A2, COL6A1, COL6A2, COL6A3, COL8A2, COL10A1, COL11A1, COL16A1, COL18A1), caldesmon 1 (CALD1), laminins (LAMA4, LAMB1, LAMC1), vimentin (VIM) and versican (VCAN). Modifiers of the ECM are also present including ADAM metallopeptidases (ADAM12, ADAMTS2), matrix metallopeptidases (MMP11, MMP2), TIMP 3 and growth factors (TGFB3). There are elements of a vascular signature mixed into this cluster including angiopoetin-like 2 (ANGPTL2) and podoplanin arguing that expression of these molecules is regulated along with or closely influenced by factors also influencing

the extracellular matrix. Analysis of this cluster clearly identifies the enrichment for extracellular matrix genes (Table 4.14).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	SP_PIR_KEYWORDS	Extracellular matrix	2.9×10^{-42}
DAVID	GOTERM_BP_FAT	Cell adhesion	8.6×10^{-25}
DAVID	GOTERM_BP_FAT	Vasculature development	1.9×10^{-6}
GSEA	Stanford tumour modules	ECM and collagen	9.9×10^{-110}

Table 4.14 Functional annotation for the ECM signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

The main endothelial cluster in this area consists of 188 probesets representing 162 genes. It contains well-characterised endothelial associated genes such as PECAM1 encoding for CD31, CD248 (endosialin), cadherin 5 type 2 (vascular endothelium) (CDH5), caveolin 1, (CAV1), endomucin (EMCN), ERG, TEK endothelial tyrosine kinase and von Willebrand Factor (VWF). Analysis highlights this cluster to be enriched in genes expressed in the extracellular region and involved in angiogenesis (Table 4.15).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Blood vessel development	4.8×10^{-10}
DAVID	GOTERM_CC_FAT	Extracellular region	1.9×10^{-11}
DAVID	GOTERM_BP_FAT	Cell adhesion	8.3×10^{-8}

Table 4.15 Functional annotation for the endothelial signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

The adipocyte cluster consists of 123 probesets to 92 genes. It contains genes including lipoprotein lipase (LPL), hormone sensitive lipase (LIPE), CD36, adipocyte fatty acid binding protein (FABP4), growth hormone receptor, adiponectin (ADIPOQ) and elements of the PPAR receptor (PPARG). Annotation from GSEA in particular supports an adipocyte signature (Table 4.16).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
GSEA	NAKAYAMA_SOFT TISSUE TUMOURS	Adipocytic differentiation	2.0×10^{-62}
GSEA	STEGER_ADIPOGENESIS	Genes upregulated during adipogenesis of fibroblasts	1.7×10^{-22}
GSEA	KEGG PATHWAY	PPAR signalling pathway	1.45×10^{-14}
DAVID	GOTERM_BP_FAT	Response to insulin stimulus	1.5×10^{-6}

Table 4.16 Functional annotation for the adipocyte signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

The smaller ECM cluster lying at the opposite side of the graph consists of 132 probesets representing 87 genes. This contains smaller numbers of collagen genes (COL12A1, COL1A2, COL3A1), fibrillin 1 (FBN1) and several integrins (ITGA11, ITGBL1). Modifiers of the ECM are included in this cluster also (ADMATS12, ADMATS16, TIMP2). Analysis again confirms the enrichment of ECM genes in this cluster (Table 4.17).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	SP_PIR_KEYWORDS	Extracellular matrix	9.6×10^{-7}
DAVID	GOTERM_BP_FAT	Cell adhesion	1.9×10^{-4}
GSEA	Stanford tumour modules	ECM and collagens	8.4×10^{-9}

Table 4.17 Functional annotation for the ECM signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

There is a further cluster containing overlaps between endothelial and adipocyte signatures. This is of 115 probesets and 77 genes and contains genes involved in angiogenesis such as SOX17, endomucin (EMCN), endothelial cell-specific chemotaxis regulator (ECSCR) and ROBO4 and genes involved in lipid binding such as CD36, annexin A1 (ANXA1), and adipocyte fatty acid binding protein (FABP4) and genes involved in triglyceride biosynthesis (AGAT9, DGAT2, GPAM). This cluster does not demonstrate the highly significant enrichment of genes seen in other clusters but DAVID analysis does still draw some functional annotation (Table 4.18).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Triglyceride biosynthetic process	4.2×10^{-4}
DAVID	GOTERM_BP_FAT	Angiogenesis	1.5×10^{-2}
DAVID	GOTERM_MF_FAT	Lipid binding	1.9×10^{-2}

Table 4.18 Functional annotation for the adipocyte/endothelial signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values.

4.5.2 Cell cycle related signatures

The main cell cycle cluster contains 154 probesets representing 133 genes and appears to represent the mitotic phase of the cell cycle (see table 4.19). Expression of these genes is highest in the more aggressive tumour types basal-like and ERBB2 (Her2 positive). Of note the expression of these rises with increasing grade across tumour types where a spectrum of grades are represented in the dataset (luminal A and normal like) (Top panel figure 4.20). This correlates with the histological features of these tumours, the Ellis-Elston grade being calculated on mitotic count in a set area, nuclear pleomorphism and tubule formation. This cluster contains aurora kinase A (AURKA), cyclins A2 and B1, (CCNA2, CCNB1), cell division cycle associated genes 3,4 and 8 (CDCA3, CDCA4, CDCA8), various centromere proteins A, E, F, M (CENPA, CENPE, CENPF, CENPM), E2F transcription factors 1 and 8 (E2F1, E2F8), numerous kinesins (KIF11, KIF14, KIF15, KIF18B, KIF20A, KIF23, KIF2C, KIF4A, KIFC1), MKI67 and PCNA, (the proteins of which are both used immunohistochemically to quantify proliferation), topoisomerase 2 alpha (TOP2A) and DNA polymerase (POLA2).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	SP_PIR_keywords	Cell division	8.2×10^{-35}
DAVID	GOTERM_MF_FAT	M phase	1.6×10^{-43}
DAVID	GOTERM_BP_FAT	DNA replication	7.7×10^{-23}

Table 4.19 Functional annotation for the main cell cycle signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values.

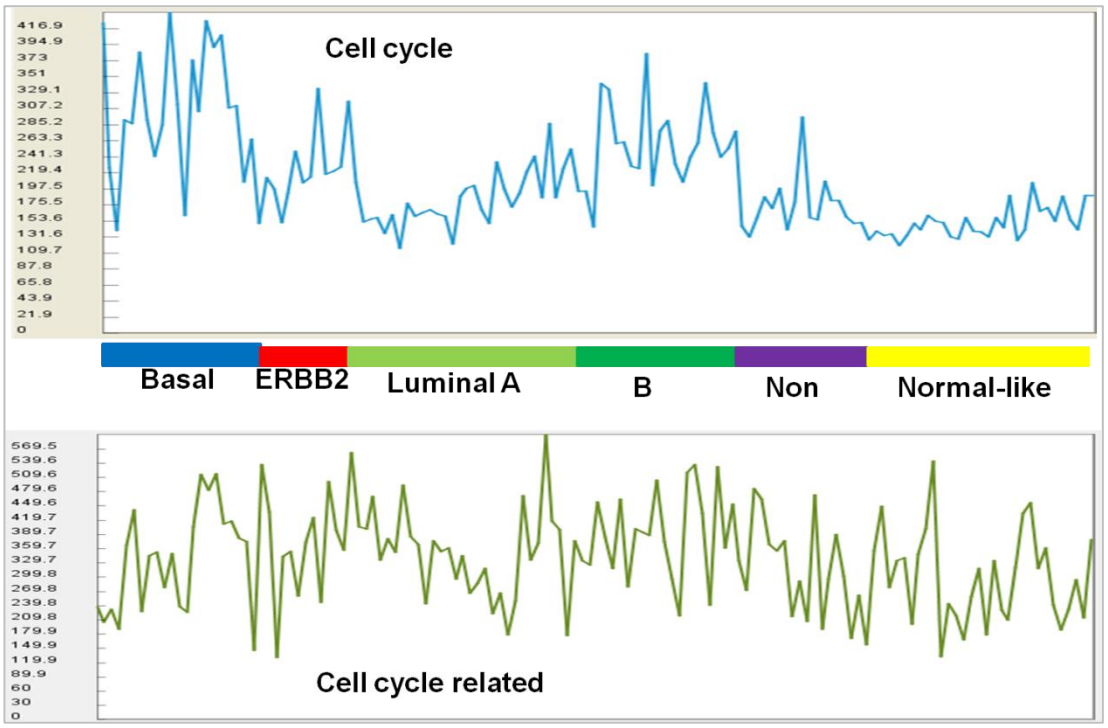


Figure 4.20. Top panel: The mean expression pattern of the main cell cycle cluster viewed across the dataset. The aggressive tumour subtypes basal-like and ERBB2 (Her-2 positive) lie to the left of the figure. Within each individual tumour type the cases are ordered from lowest grade at left to highest grade at right. The effect is most marked in the luminal A cases where all tumour grades from 1 (left) to 3 (right) are represented, although a similar pattern in seen in the normal-like cases as well. **Bottom panel:** The small cell cycle related cluster containing TP53. Individual tumours represented on x-axis of these panels. Coloured bars indicate tumour subtype.

There are several cell cycle associated clusters present. One of these, a much smaller cell cycle related cluster containing 39 probesets mapping to 35 genes has a pattern quite different from the main cell cycle cluster (lower panel figure 4.20). The genes within this cluster lie within cell cycle clusters in the merged dataset implying that in the other tumour types these genes show similar expression to cell cycle genes but on applying functional annotation using DAVID and GSEA little useful annotation is found. These genes show an expression pattern at variance with the mitotic phase cluster discussed above. Of interest, represented within this cluster is TP53 encoding the p53 protein.

4.5.3 Other signatures

As in other datasets, aside from cell cycle and stromal signatures there were other clusters, often small, where functional annotation could be assigned. These included clusters of ribosomal genes, genes involved in glycolysis, etc much as in the DLBCL dataset. These are not discussed further. Much of the graph in this as in other datasets cannot be interpreted as representing specific functions or cell types. In part this is due to the large number of genes about which little is known and which often cluster together providing no handle to approach the gene list from. In this dataset however, there is a relatively diverse group of tumours represented and there are clusters that provide information about these.

In the breast cancer dataset, searching for genes that differ between subsets, these tend to form small clusters in separate areas of the graph. One of these areas contains a sparse network of genes, separate from the main graph and containing a number of small clusters. These clusters represent genes whose profile is lower in expression in basal-like and ERBB2-positive tumours, tumours which normally lack expression of oestrogen receptor histologically and do not respond to anti-oestrogenic drugs. These clusters include ESR1 (the oestrogen receptor alpha), GATA3, FOXA1, and XBP1, and this cluster grouping appears to capture part of the oestrogen signalling transcriptional network, including modulators and downstream targets (Figure 4.21). GATA3 is involved in luminal differentiation in normal breast tissue and ESR1 and GATA3 have been demonstrated to reciprocally regulate each other (Asselin-Labat et al 2007, Eeckhoute et al 2007). GATA3 is an inducer of FOXA1, which in turn can induce XBP1, while ESR1 acts both upstream and alongside FOXA1 (reviewed in Nakshatri and Badve 2009). This small group of genes appears to capture multiple elements of the oestrogen receptor signalling pathway.

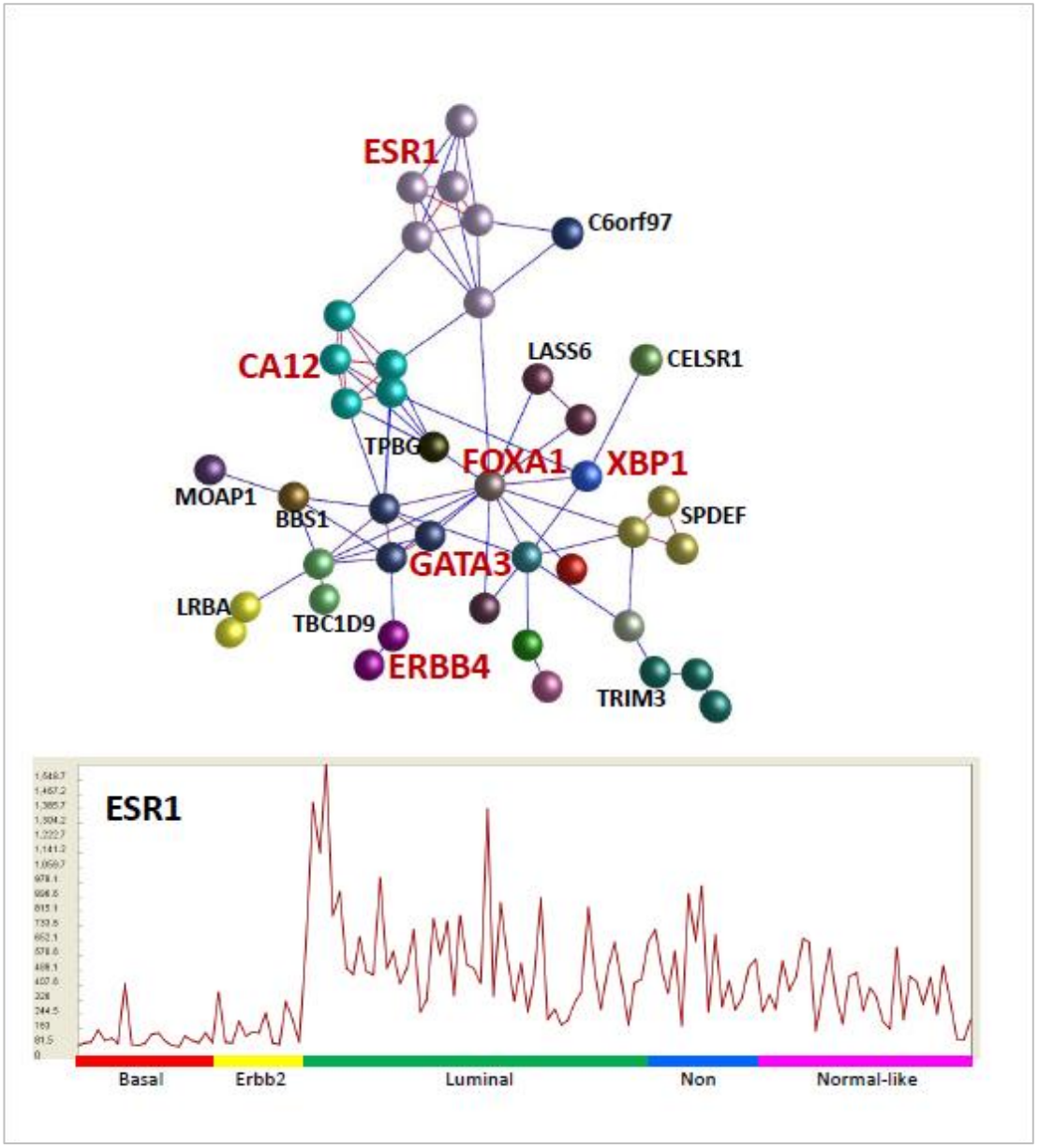


Figure 4.21. The oestrogen receptor (ESR1) lies in a small cluster outwith the main graph. Selecting the 'nearest neighbours' of this cluster pulls in other small clusters of interest. The expression pattern of the ERS1 cluster is shown in the bottom panel. Individual tumours represented on x-axis of this panel. Coloured bars indicate tumour subtype.

4.6 Gene signatures from ovarian cancer

The ovarian data set was of 285 cases from several hospitals in Australia and the Netherlands of which 265 passed the QC step and were used in the analysis. These were stratified based on whether they were malignant or of low malignant potential, histological type (endometrioid or serous), grade, stage and primary site (ovary, peritoneum or fallopian tube). The ovarian network laid out at r = 0. 65 has a rather 'hairy' appearance being formed of a large central area with numerous smaller clusters round the periphery (Figure 4.22). This is reflected in the relatively low number of edges as compared to nodes, with 19,415 nodes being joined by only 268,471 edges. Lying within the central area are clusters of immune genes, extracellular matrix genes and cell cycle genes. There are several clusters that represent immune related cells. These lie within the same area of the central part of the graph and lie near but not abutting the non-immune stromal elements. The cell cycle and cell cycle related clusters lie close together and separate from the stromal elements. The clusters of cell cycle genes show a clear difference in expression between the tumours of low malignant potential which show low expression and the malignant tumour which have high expression.

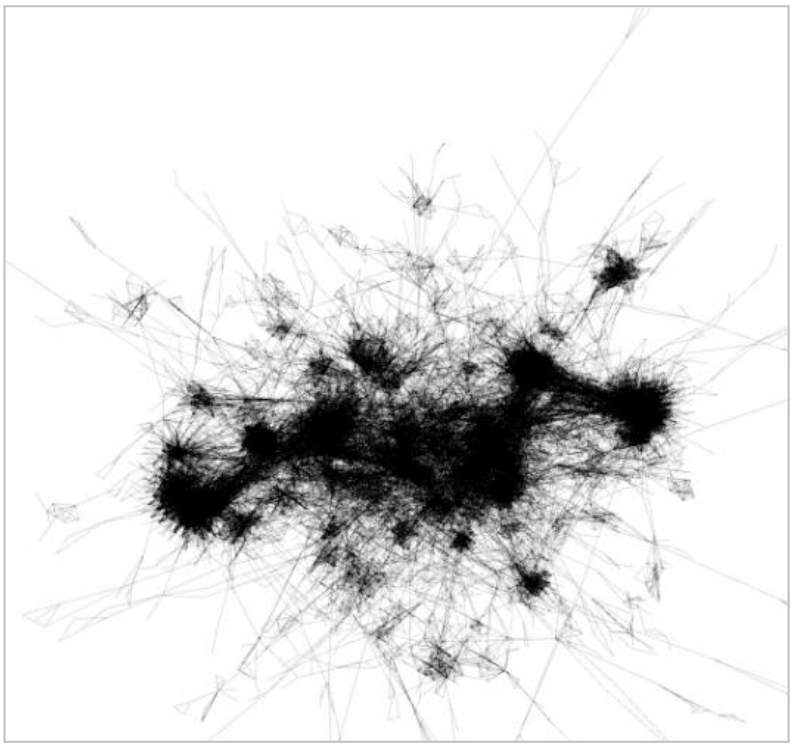


Figure 4.22. Ovarian carcinoma dataset laid out at r≥0.65 with nodes hidden and only edges visible.

4.6.1 Stromal signatures

The immune clusters fall together in one area of the graph (Figure 4.23). The largest cluster, which is enriched for macrophage and T cell genes, consists of 619 probesets representing 378 genes. There is lower expression in tumours of LMP and higher expression in tumours not arising in the ovary. This contains myeloid lineage markers such as CD14, CD163, CD68, CSF1R and T cell markers CD2, CD3 (delta, epsilon and gamma subunits), CD5, and CD52. There are multiple chemokines, cytokines, interleukins and their receptors (CCL3, CCL4, CCL5, CCR1, CCRL2, CXCL13, CXCL9, CXCR3, CXCR6, IL10RA, 1L12RB1, IL16, IL18RAP, IL21R, IL23R, IL2RB, IL2RG, IL7R and IFNG). Toll-like receptors are represented (TLR1, TLR2, TLR4, TLR7, TLR8, TLR10) as are some MHC class II molecules (HLA-DMA, HLA-DMB, HLA-DOA, HLA-DPA1, HLA-DPB1, HLA-DRA, HLA-DRB1, HLA-DRB4, HLA-DRB4, HLA-DRB6) and associated genes such as the class II, MHC, transactivator (CIITA) and CD74, the invariant chain of MHC II . There are

multiple elements of the T cell receptor signalling pathway (FYB, LAT2, TRA@, TRAC, TRAT1. TRAV20, TRGC2, TCRGV9, ZAP70). Cytotoxic granules are represented with multiple granzymes (GZMA, GZMB, GZMH,GZMK) and perforin 1 present. Analysis of the cluster confirms the enrichment for macrophage and T cell genes (Table 4.20).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Defence response	94.5×10^{-36}
DAVID	GOTERM_BP_FAT	T cell activation	5.3×10^{-23}
DAVID	GOTERM_BP_FAT	Antigen processing and presentation MHC II	2.1×10^{-15}
DAVID	GOTERM_BP_FAT	Myeloid leukocyte activation	5.1×10^{-8}
GSEA	Stanford tumour modules	Immune and inflammatory reponse	7.3×10^{-164}
GSEA	REACTOME	Immune system	2.8×10^{-16}

Table 4.20 Functional annotation for the macrophage/T cell signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

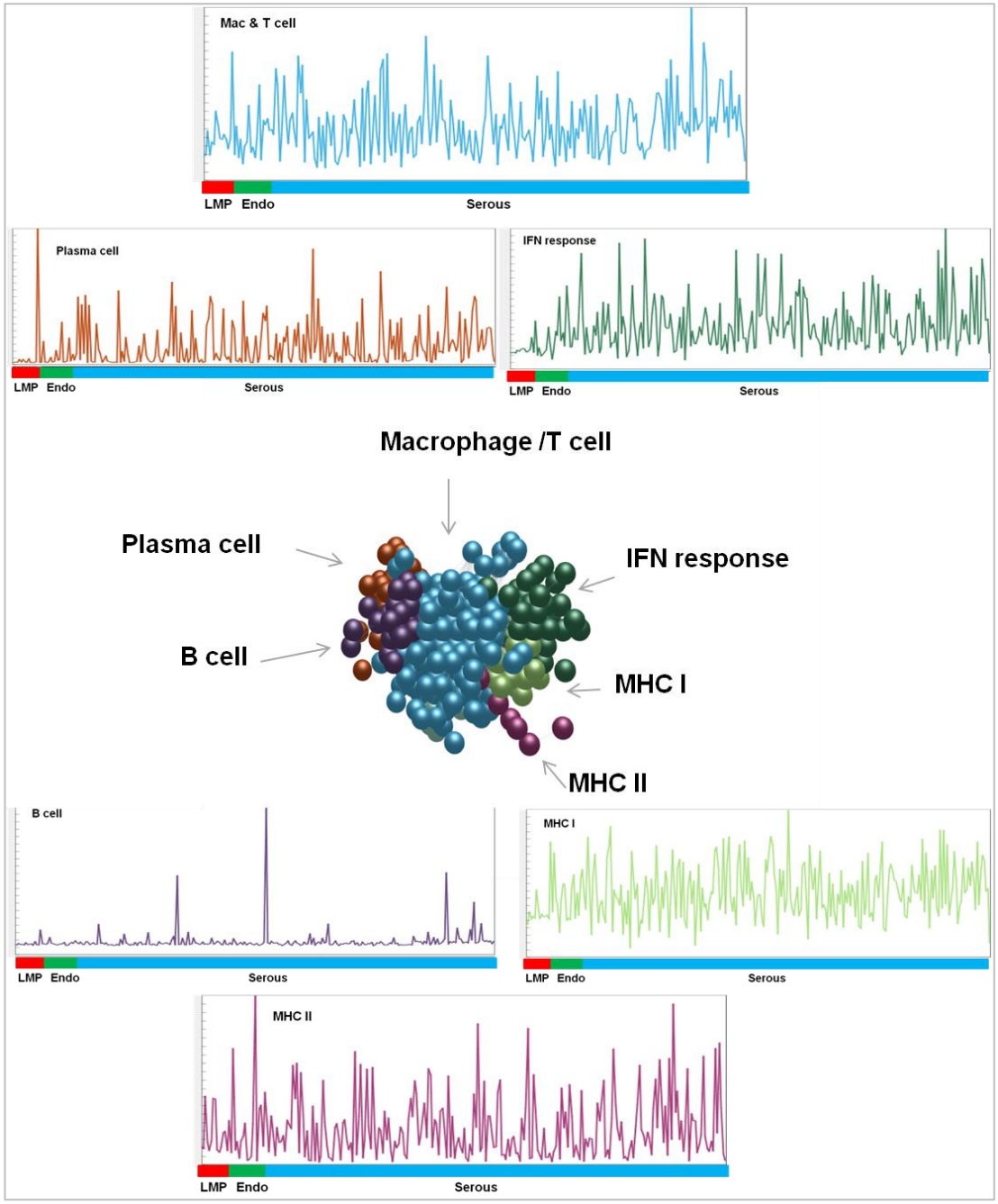


Figure 4.23. The immune cluster from the ovarian dataset with the rest of the graph hidden (central panel). Mean expression profiles for each cluster are shown with the colours of the clusters corresponding to the colours of the expression plots. Individual tumours are represented along the x-axis with the coloured bar indicating tumour types (red = LMP, green = endometrioid, blue = serous**).**

The interferon response (103 probesets representing 61 genes) and MHC class I signatures which fall next to the macrophage signature, show similar expression

patterns (Figure 4.23). The interferon response cluster contains numerous interferon induced proteins (IFI27, IFI35, IFI44, IFI44L, IFI6, IFIH1, IFIT1, IFIT2, IFIT3, IFIT5, MX1) and the interferon regulatory factors IRF7 and IRF9. The MHC class I cluster contains 57 probesets representing 29 genes. This contains the HLA molecules HLA-A, HLA-B, HLA-D, HLA-E, HLF-F, and HLA-G. Two components of the TAP transporter TAP1 and TAP2 are present as is the TAP binding protein (tapasin), two subunits of the proteasome (PSMB8, PSMB9) and beta-2-microglobulin (B2M). See tables 4.21 and 4.22 for selected functional annotation.

Analysis tool	Source of annotation	Annotation	Significance (p-value)
GSEA	HECKER_IFNB_TARGETS	Genes modulated in MS patients treated with subcut IFNB1	1.7×10^{-93}
GSEA	DER_IFN_ALPHA_RESPONSE_UP	Genes upregulated in HT1080 cells by treatment with IFNA	3.3×10^{-42}
DAVID	GOTERM_BP_FAT	Defence response	3.2×10^{-5}

Table 4.21 Functional annotation for the interferon response signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Antigen processing and presentation MHC I	8.6×10^{-17}
DAVID	SP_PIR_KEYWORDS	MHC I	3.6×10^{-13}
GSEA	KEGG	Antigen processing and presentation	3.6×10^{-21}

Table 4.22 Functional annotation for the MHC class I signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

The plasma cell signature is largely absent from the LMP tumours (Figure 4.23). It consists of 162 probesets representing only 81 genes. It contains some markers of B/plasma cell differentiation such as CD79A, CD38 and IRF4 but consists mostly of immunoglobulin genes for both heavy and light chains. See Table 4.23 for functional annotation.

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	SP_PIR_KEYWORDS	Immunoglobulin	6.5×10^{-17}
DAVID	GOTERM_MF_FAT	Antigen binding	2.8×10^{-14}
GSEA	GSE10325_B CELL_VS_MYELOID	Genes upregulated in B cells compared to myeloid cells	2.7×10^{-40}

Table 4.23 Functional annotation for the plasma cell signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

The B cell signature consisting only of 50 probesets to 27 genes is seen in only a small number of tumours (Figure 4.23). This contains the B cell marker CD79B, the BLK B lymphoid tyrosine kinase and several Fc-receptor like genes (FCRL1, FCRL2, FCRL3, FCRLA) as well as the immunoglobulin heavy chain, mu. See table 4.24 for functional annotation.

Analysis tool	Source of annotation	Annotation	Significance (p-value)
GSEA	GSE10325_B CELL_VS_MYELOID	Genes upregulated in B cells compared to myeloid cells	1.3×10^{-24}
DAVID	GOTERM_BP_FAT	Immune response	1.5×10^{-5}

Table 4.24 Functional annotation for the B cell signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

The relative absence of an immune signature in the tumours of low malignant potential likely reflects the behaviour of these tumours. Tending to remain confined to the ovary and not invading in a tissue destructive way, one could envisage that they would invoke little inflammatory /immune response.

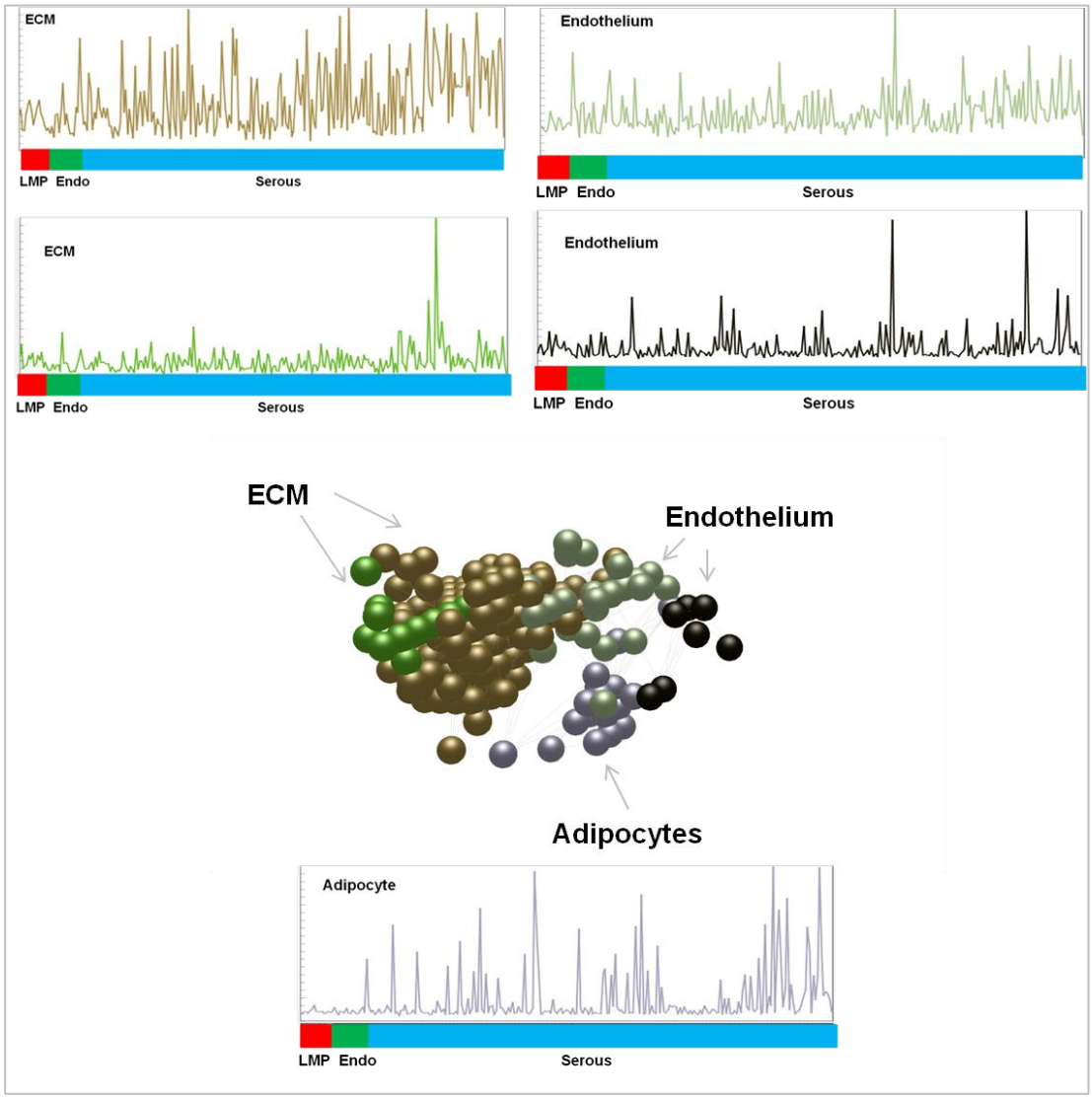


Figure 4.24. The non-immune stromal clusters shown with the rest of the network graph hidden. (Central panel) The mean expression pattern of the clusters are given in the surrounding panel with colours of the expression plots corresponding to the cluster colours. The tumour types are indicated along the bottom of each panel.

The non-immune stromal signatures consist of extracellular matrix which breaks into two closely related clusters, two endothelial clusters and an adipocyte signature (Figure 4.24). The main ECM signature consisting of 323 probesets to 180 genes shows lower expression in LMP tumours and higher but more variable expression in serous tumour with more consistently high expression in those serous tumours of primary peritoneal origin (right side of serous group in figure 4.24). This would be predicted given the increased likelihood of sampling of subperitoneal fibrous tissue/abdominal wall in these tumours. The cluster contains multiple collagens (COL1A1, COL1A2, COL3A1, COL5A1, COL5A2, COL6A2, COL6A3, COL8A1, COL10A1, COL11A1, COL16A1), and other structural proteins caldesom1, calponin2, fibrillin 1, laminin alpha 4, and tropomyosin. There are modifiers of the extracellular matrix (ADAM12, ADMATS12, AMADTSL1, MMP2, MMP11, TIMP2, TIMP3) as well as some growth factors and growth factor receptors (FGF1, TGFBR2, PDGFRB). Elements of the vascular signature are included in this cluster implying that expression of these genes is related to expression of the ECM genes. These include cerebral endothelial cell adhesion molecule (CERCAM), podoplanin, (PDPN) and vascular cell adhesion molecule 1 (VCAM1). See table 4.25 for selected functional annotation.

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	SP_PIR_KEYWORDS	Extracellular matrix	9.3×10^{-28}
DAVID	GOTERM_BP_FAT	Cell adhesion	3.7×10^{-12}
DAVID	GOTERM_BP_FAT	Vasculature development	7.5×10^{-8}
GSEA	NABA_CORE_MATRISOME	Ensemble of genes encoding core ECM.	1.0×10^{-42}

Table 4.25 Functional annotation for the extracellular matrix signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

The smaller ECM signature (bright green in figure 4.24) (22 probesets and 16 genes) shows low expression in most cases but slightly higher expression in those serous

tumours not of primary ovarian origin. This contains mostly collagen 6 components (COL6A1, COL6A2, COL6A6 AND COL29A1) and some modifiers of the ECM (AMADTS14, ADAMTS2, MMP13). See table 4.26 for selected functional annotation.

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	SP_PIR_KEYWORDS	Extracellular matrix	4.7×10^{-9}
GSEA	REACTOME	Extracellular matrix organisation	2.8×10^{-11}

Table 4.26 Functional annotation for the smaller ECM signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

The larger of the two small endothelial clusters (green in figure 4.24) (8 probesets and 7 genes) shows consistently lower expression in LMP tumours but is variable across the remaining dataset. This contains markers of endothelium such as CD248 (endosialin), CDH5, endothelial cell adhesion molecule (ESAM), and von Willebrand factor (VWF). See table 4.27 for selected functional annotation.

Analysis tool	Source of annotation	Annotation	Significance (p-value)
GSEA	GOTERM	Angiogenesis	1.4×10^{-6}
GSEA	GOTERM	Cell-cell adhesion	8.2×10^{-6}
DAVID	GOTERM_BP_FAT	Blood vessel development	1.2×10^{-4}

Table 4.27 Functional annotation for the endothelial signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

The smaller endothelial cluster (black in figure 4.24) (8 probesets and 4 genes) has a much spikier pattern, being expressed only in subsets of tumours and includes angiopoetin 2 (ANGPT2), endothelial cell-specific molecule (ESM1), and vascular endothelial growth factor receptor (FLT1). The cluster is too small to derive meaningful annotation from using GSEA and DAVID but from the known function of the genes is clearly involved in angiogenesis.

The adipocyte signature (22 probesets and 14 genes) has a spikey profile with many tumours essentially lacking this signature (Figure 4.24). Those of primary peritoneal origin are more consistently positive likely reflecting sampling of normal adipose tissue underlying the peritoneum. These genes present include CD36, adiponectin (ADIPOQ), adipocyte fatty acid binding protein (FABP4) and TIMP4. See table 4.28 for selected functional annotation.

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	KEGG PATHWAY	PPAR signalling	9.6×10^{-5}
GSEA	NAKAYAMA_SOFT TISSUE TUMOURS	Adipocytic differentiation	4.0×10^{-26}
GSEA	REACTOME	Transcriptional regulation of white adipocytic differentiation	1.0×10^{-11}

Table 4.28 Functional annotation for the adipocyte signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

4.6.2 Cell cycle signatures

The cell cycle signatures lie within the main part of the graph, but separate from the stromal signatures. As in other tumours there are clusters that represent the mitotic phase of the cell cycle and other related clusters (Figure 4.25). The cluster representing the mitotic phase is the largest cell cycle cluster formed of 162 probesets representing 108 genes (represented in black in figure 4.25). As would be expected, expression of this signature is low in the tumours of low malignant potential, characterised as they are histological by little visible mitotic activity. The frankly malignant tumours of any subtype show high expression of this signature. The cell cycle and related clusters separate into two main areas, the largest cluster forming the first area and a second area containing several smaller clusters. As illustrated in figure 4.25 these clusters also show a lower expression in the LMP tumours but to a much lesser degree.

The main cell cycle cluster contains aurora kinases A and B (AURKA, AURKB), cyclins A2, B1 and B2, (CCNA2, CCNB1, CCNB2), the cell cycle division genes CDC2, CDC25C, CDC6, CDCA2, CDCA3, CDCA5 and multiple genes for centromeric proteins (CENPQ, CENPE, CENPF, CENPI, CENPL, CENPM). There are multiple kinesin family members (KIF11, KIF14, KIF15, KIF18A, KIF18B, KIF20A, KIF23, KIF2C, KIF4A) and genes involved in DNA replication and repair including DNA polymerase (POLQ), DNA topoisomerase (TOP2A), exonuclease (EXO1) and an excision repair associated gene (ERCC6L). Analysis using DAVID and GSEA confirms the enrichment of this cluster for genes involved in the M phase of the cell cycle (Table 4.29).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	M phase	1.1×10^{-58}
DAVID	GOTERM_BP_FAT	Cell cycle	5.0×10^{-57}
GSEA	BENPORATH_CYCLING_GENES	Genes showing cell cycle stage specific expression	6.5×10^{-104}

Table 4.29 Functional annotation for the cell cycle signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

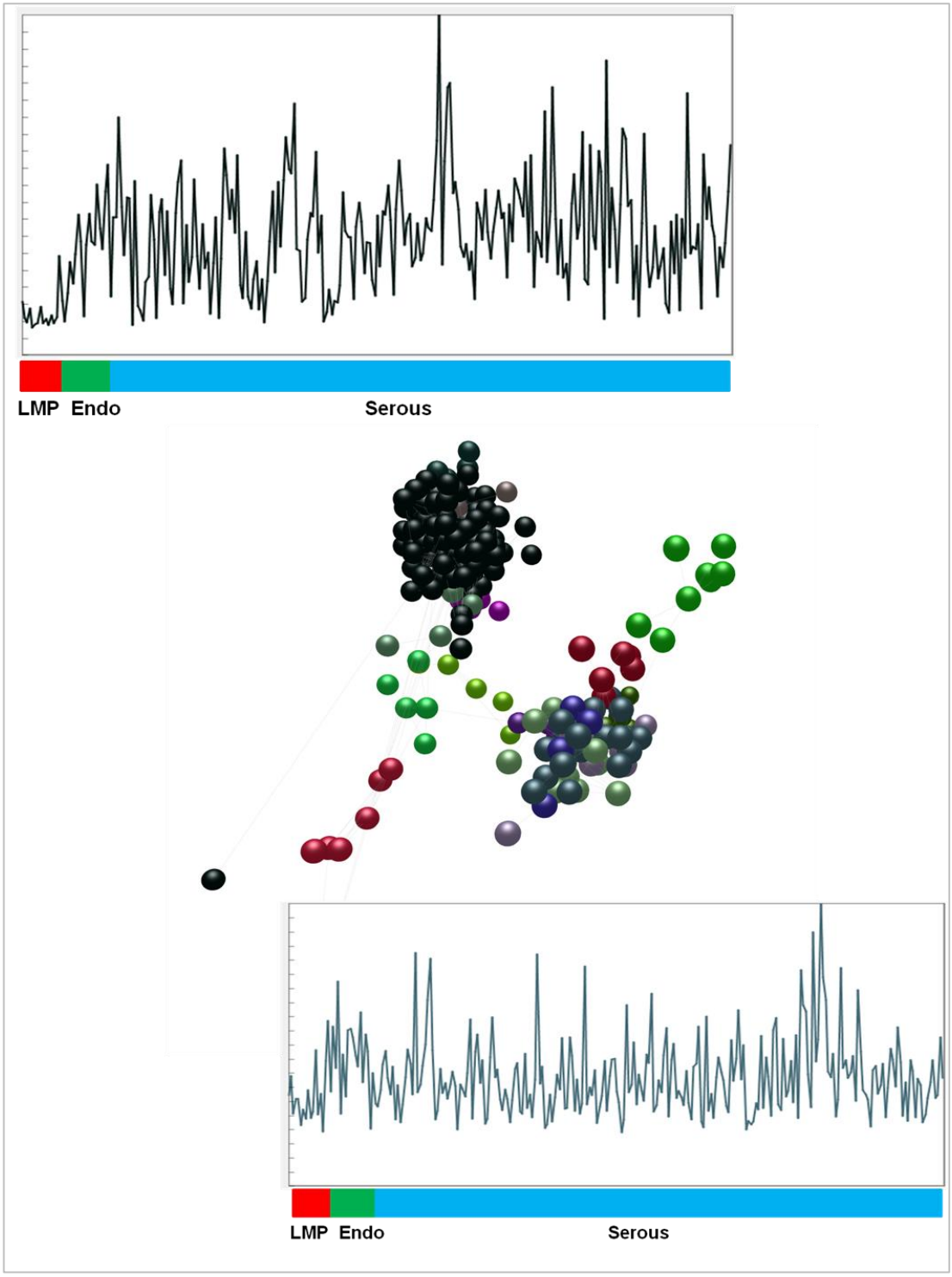


Figure 4.25. The cell cycle-related area of the graph consists of two main groups of clusters. Only selected expression profiles are illustrated for clarity. In the larger cluster area (top of central panel) there are clusters in which expression in markedly lower in LMP tumours (top panel). The second set of clusters also show lower expression in LMP tumours but the difference between subtypes is less marked (bottom panel).

The next largest cluster (blue in figure 4.25) containing 25 probesets and 19 genes lies in the second cell cycle area of the graph and contains several transcription factors (AHCTF1, SP1) and other genes that may be involved in transcription and RNA binding (AFF4, EWSR1, DHX, HNRNPU, ZFR). No highly significant annotation is associated with this cluster but that identified suggests a role in transcription (Table 4.30).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	SP_PIR_KEYWORDS	RNA binding	1.3×10^{-3}
DAVID	GOTERM_BP_FAT	Transcription	2.5×10^{-1}

Table 4.30 Functional annotation for the cell cycle/transcription signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

The rest of the cell cycle related clusters break into multiple small clusters none larger than 11 probesets and which contain such genes as CHEK1, DNA polymerase (POLE2), and cyclin E2.

4.6.3 Other signatures

As well as the cell cycle clusters there were many others clusters that separated tumours of low malignant potential from those of frank malignancy. As this was not the focus of the analysis these were not explored further. In addition, in this dataset as in all the others clusters representing histones, ribosomes, haemoglobin, glycolysis etc were seen.

4.7 Gene signatures from colorectal cancer

The colorectal dataset contained 155 cases of colorectal carcinoma derived from an Australian population, divided into microsatellite stable and unstable tumours. One hundred and fifty of these passed the QC stage and were used in subsequent analysis. This was the dataset with the least clinical annotation with no information available as to grade or stage. The colorectal dataset when laid out at $r \geq 0.65$ gives a graph of 22,687 nodes joined by 725,467 edges (Figure 4.26). One pole of the graph contains cell cycle genes, while immune genes lie within the main area of the graph. As would be predicted from histological appearances, the immune associated genes are present at higher level in microsatellite instable tumours, which are characterised histologically by a greater number of tumour infiltrating lymphocytes. In this dataset unlike many of the others the immune/inflammatory signatures contained a neutrophil signature reflecting the frequently high number of neutrophils seen in colorectal cancer in contrast to most other tumour types. An extracellular matrix cluster and endothelial cluster lie fairly centrally within the graph. Also present within this graph are "spiked" clusters where expression of the genes in the cluster is seen in only a few cases. These include a cluster of adipocyte related genes. In this dataset, as in other non-sex specific tumours, there are clusters of genes expressed only in the female or the male.

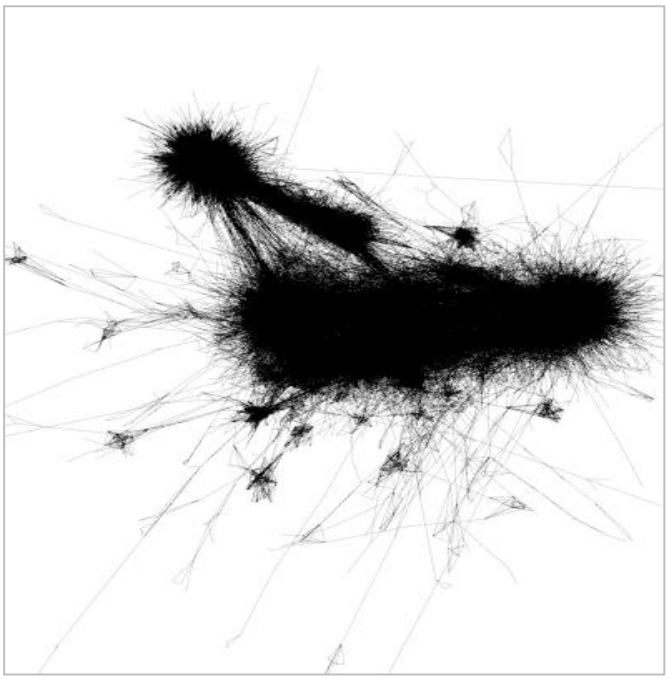


Figure 4.26. The colorectal cancer dataset when laid out at r≥0.65 with nodes hidden and only edges visible. The separate large cluster at 11 o'clock is of house-keeping genes. The stromal signatures lie within the centre of the graph and the cell cycle at one pole of the main graph.

4.7.1 Stromal signatures

There is an area of graph lying centrally in which both immune and non-immune stromal elements lie. The immune clusters lie close together and the non-immune stromal elements form a separate group of clusters. The immune signatures consist of one large cluster which contains macrophage and T cell signatures with small clusters surrounding this representing plasma cells, IFN response, MHC I and other clearly immune signatures that are none the less difficult to characterise further (Figure 4.27). As would be predicted based on the histology, these immune signatures tended to be higher is MSI tumours. In the absence of other annotation no comment can be made on the clinical features of much of the graph.

The large macrophage and T cell cluster consists of 699 probesets to 458 genes. This contains markers of monocytic differentiation such as CD14, CD163, and CSF1R as well as markers of T cell differentiation such as CD3, and elements of the T cell receptor complex. Both CD4 and CD8 are represented suggesting that the

cluster represents all subsets of T cells. The cluster contains multiple toll-like receptors (TLR1, TLR7, TLR8), cathepsins (CTSC, CTSO, CTSW), and granzymes (GZMA, GZMH, GZMK, GZMM). There are multiple MHC class II molecules (HLA-DMB, HLA-DOA, HLA-DPA1, HLA-DPB2, HLA-DQA1, HLA-DQB1, HLA-DRA, HLA-DRB1, HLA-DRB5) as well as CD74, the MHC class II invariant chain. Chemokines, cytokines, interleukins and their receptors are represented (CCL18, CCL4, CCL5, CCR1, CCR2, CCR5, CXCL13, CXCL9, CXCR6, IL16, IL10RA, IL18RAP, IL21R, IL23A, IL2RA, IL2RB, IL7R). Macrophage scavenger receptors are represented (MARCO, MRC1), as are apoliproteins (APOC1, APOE). Analysis using DAVID and GSEA confirms the enrichment for macrophage and T cell genes (Table 4.31).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Defense response	2.0×10^{-31}
DAVID	GOTERM_BP_FAT	Leukocyte activation	2.5×10^{-24}
DAVID	GOTERM_BP_FAT	Chemotaxis	6.2×10^{-11}
DAVID	GOTERM_BP_FAT	Phagocytosis	3.1×10^{-6}
GSEA	Stanford tumour modules	Immune and inflammatory response	1.2×10^{-141}
GSEA	REACTOME	Immune system	2.1×10^{-60}

Table 4.31 Functional annotation for the macrophage/T cell signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

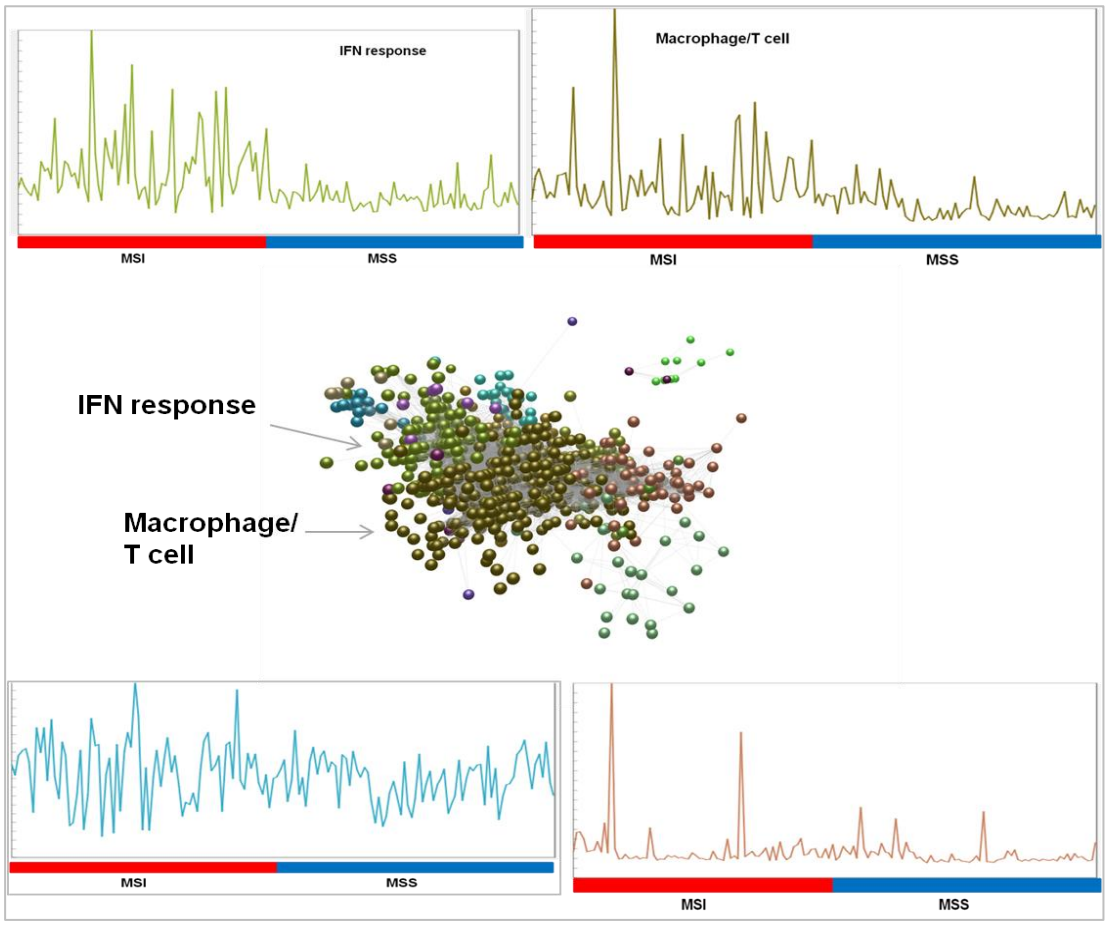


Figure 4.27. The immune clusters with the rest of the graph hidden. Top and bottom panel show mean expression profile of each cluster. The colours of the cluster correspond to the colours given in the expression plots. IFN response and macrophage/T cell in top panels, MHC class I in bottom left and other less well characterised immune cluster, bottom right.

The IFN response breaks into several clusters the largest of which (seen in figure 4 25 in green) is of 176 probesets and 121 genes (Figure 4.27). This cluster contains many interferon regulatory factors and interferon induced genes (IFI16, IFI35, IFI44, IFI44L, IFI6, IFIT1, IFIT2, IFIT3, IFIT5, IRF1, IRF7). Multiple components of the proteasome are present as well as elements of the antigen presentation machinery (CD74, MICB, TAPBP, TAP1, TAP2, PSMB9, PSMB10, PSME1, PSME2, HLA-E, HLA-DMA, HLA-DPA1, HLA-DRB1, HLA-DRB4). Selected functional annotation given in table 4.32.

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Antigen processing and presentation	2.6×10^{-10}
GSEA	BROWNE_INTERFERON REPONSIVE GENES	Genes upregulated in fibroblasts after culture with IFN alpha	1.0×10^{-60}
GSEA	HECKER_IFNB_TARGETS	Genes modulated in the blood of patients with MS treated with IFNB	2.6×10^{-55}

Table 4.32 Functional annotation for the interferon response signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

There is a plasma cell cluster containing 131 probesets and 67 genes. This contains mainly immunoglobulin genes which make up 17 of the 67 genes as well as markers of B cell/plasma cell differentiation CD79A, IRF4 and TNFRSF17. See table 4.33 for selected functional annotation.

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	SP_PIR_KEYWORDS	Immunoglobulin	4.5×10^{-17}
DAVID	GOTERM_BP_FAT	Lymphocyte activation	3.4×10^{-3}
GSEA	GSE10325	Genes downregulated in comparison of healthy T cell with healthy B cells	7.9×10^{-29}

Table 4.33 Functional annotation for the plasma cell signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

A small cluster of 85 probesets and 61 genes appears to represent a neutrophil signature. This signature is absent in the other datasets representing the paucity of

neutrophils in the inflammatory infiltrate associated with most established tumours. This contains the granulocyte CSF receptor (CSF3R) as well as the low affinity granulocyte-macrophage receptor (CSF2RB). There are the Fc receptors CD32 and CD16 (FCGR2A, FCGR2C, FCGR3B) as well as the chemotactic receptors C5AR1, FPR1, FPR2, and FPR3. See table 4.34 for selected functional annotation.

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Defense response	1.5×10^{-10}
DAVID	GOTERM_BP_FAT	Chemotaxis	2.1×10^{-5}
GSEA	GSE22886	Genes downregulated in comparison of naive B cells with unstimulated neutrophils	3.9×10^{-35}

Table 4.34 Functional annotation for the neutrophil signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

A small B cell cluster of 62 probesets to 41 genes is also present. This contains the B cell markers CD19, CD79A and MS4A1 (CD20) as well as B cell signalling molecules such as B lymphoid tyrosine kinase (BLK), and B cell scaffold protein with ankyrin repeats 1 (BANK1). Selected functional annotation given in table 4.35.

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	SP_PIR_KEYWORDS	B cell activation	2.0×10^{-4}
DAVID	GOTERM_BP_FAT	Defense response	5.5×10^{-4}
GSEA	GSE10325	Genes downregulated in comparison of healthy B cells with healthy myeloid cells	1.2×10^{-19}

Table 4.35 Functional annotation for the B cell signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

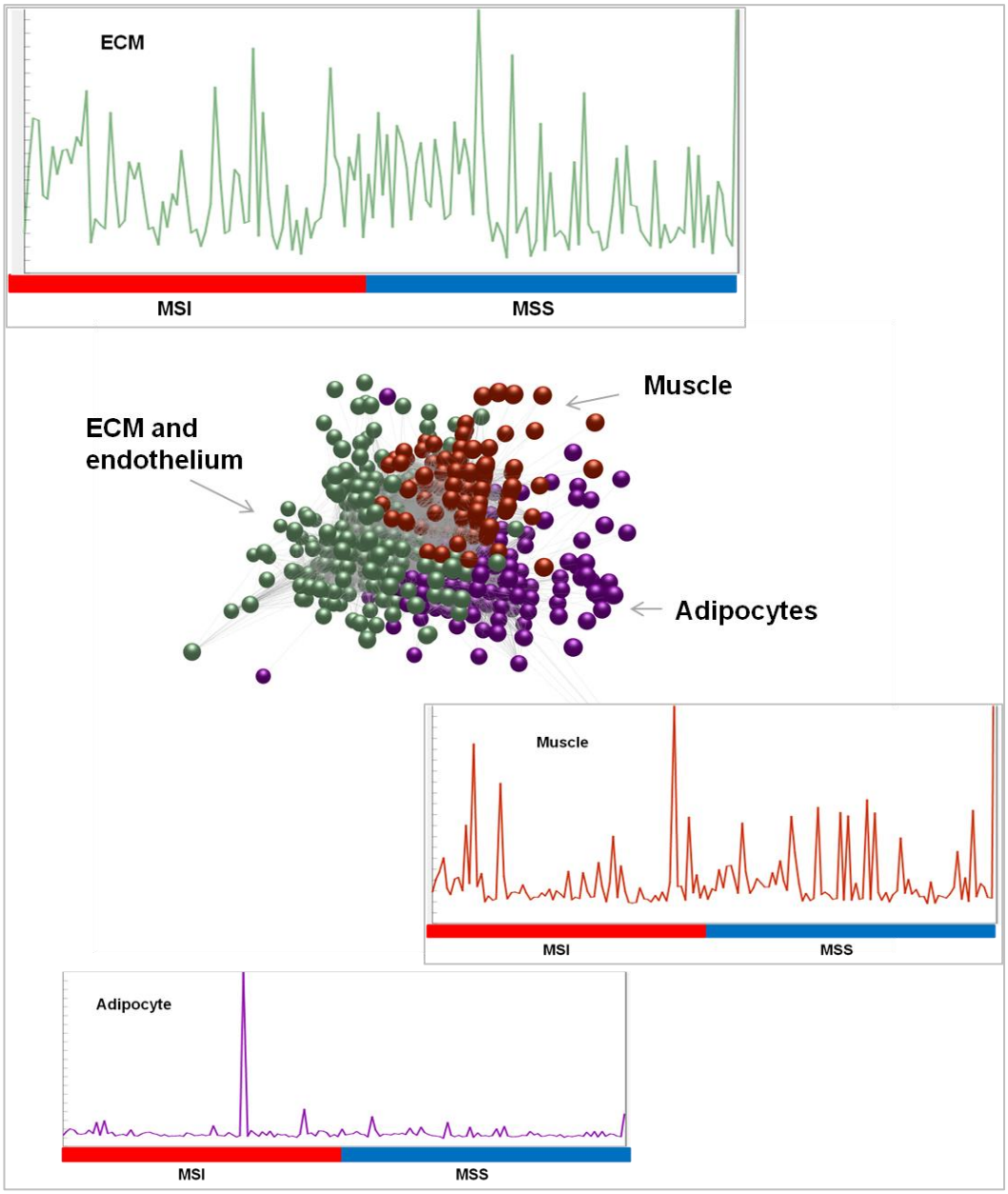


Figure 4.28. The non-immune stromal signature clusters with the rest of the graph hidden. Mean expression plots for each cluster in the surrounding panels. Colours of clusters correspond to colours of the expression plots. The x-axis of the plots represent individual tumours. Tumour subtypes are denoted by the coloured bars.

The non-immune clusters are formed of a large cluster that encompasses most of the ECM and endothelial signature, plus two smaller rather 'spiked' clusters in which there has clearly been sampling of muscle and fat (Figure 4.28). The large ECM cluster is of 481 probesets to 308 genes. This contains elements of the extracellular

matrix, part of the smooth muscle signature and elements of the endothelial signature and is seen in all cases. There is higher expression in subsets of MSI and MSS tumours but without further annotation the significance of this is lost. The cluster contains numerous structural proteins including collagens, fibulins, laminins and integrins, (COL10A1, COL11A1, COL12A1, COL15A1, COL18A1, COL1A1, COL1A2, COL3A1, COL4A1, COL4A2, COL5A1, COL5A2, COL6A1, COL6A2, COL6A3, FBLN1, FBLN2, FBLN5, LAMA4, LAMC1, ITGA11, ITGA5). There are modifiers of the extracellular matrix including ADAM metallopeptidases, matrix metallopeptidases and TIMP metallopeptidase inhibitors (ADAM12, ADAMTS12, ADAMST2, ADAMTS5, MMP2, MMP16, TIMP2, TIMP3). Growth factors and receptors are represented including connective tissue growth factor and platelet derived growth factor receptor (CTGF, PDGRFA, PDGFRB). There is a vascular signature contained within this cluster also implying tight association between the vasculature and the extracellular matrix. This includes the endothelial markers CD34 and CD31 (PECAM1), as well as angiopoetin 2 and similar molecules (ANGPT2, ANGPTL2, AMOTLO1), vascular endothelial cadherin (CDH5), endomucin, endothelial cell-specific molecule, and podoplanin (EMCN, ESM1, PDPN). Enrichment analysis confirms the cluster to be enriched in ECM and endothelial genes (Table 4.36).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_CC_FAT	Extracellular matrix	1.7×10^{-43}
DAVID	GOTERM_BP_FAT	Cell adhesion	1.6×10^{-29}
DAVID	GOTERM_BP_FAT	Vasculature development	2.9×10^{-17}

Table 4.36 Functional annotation for the ECM signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

The adipocyte cluster is of 388 probesets representing 262 genes. This has a spiked profile with expression being seen only in a subset of cases and presumed to represent sampling of the tumour where it has breached muscularis propria and is invading into fat outwith the bowel wall (Figure 4.28). The cluster contains

adipocyte expressed genes CD36, leptin, hormone sensitive lipase, lipoprotein lipase, and insulin-like growth factor 1 (CD36, LEP, LIPE, LPL, IGF1). There are numerous genes involved in lipid transport and metabolism (GM2A, GULP1, AKR1C1, CIDEA, CLU, GPIHBP1, DGAT2, GPAM). Enrichment analysis confirms the enrichment for adipocytic genes (Table 4.37).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
GSEA	NAKAYAMA_SOFT TISSUE TUMOURS	Associated with adipocytic differentiation	1.2×10^{-64}
GSEA	BOWUEST_STEMCELL CULTURED VS FRESH	Genes upregulated in cultured stem cells from adipose tissue vs freshly isolated cells	2.5×10^{-31}
DAVID	GOTERM_BP_FAT	Lipid localisation	4.3×10^{-3}

Table 4.37 Functional annotation for the adipocyte signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

The muscle cluster is 194 probesets to 137 genes. This also has a rather spiky profile but is present in a higher number of cases than the adipocyte genes, again most likely reflecting the method of sampling the tumour, in this case presumably including muscularis propria (Figure 4.28). This contains myosin light and heavy chains, actins and desmin (MYH11, MYL9ACTC1, ACTG2, DES) as well as many other muscle associated genes such as calponin 1, leiomodulin, myocardin, nicotinic cholinergic receptor, tropomyosin, vinculin (CNN1, LMOD1, MYOCD, CHRNA3, TPM2, VCL). Analysis demonstrates enrichment for genes associated with contractile fibres and the sarcomere (Table 4.38).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
GSEA	REACTOME	Muscle contraction	1.9×10^{-14}
DAVID	GOTERM_CC_FAT	Contractile fibre	5.2×10^{-12}
DAVID	GOTERM_CC_FAT	Sarcomere	6.9×10^{-7}

Table 4.38 Functional annotation for the muscle signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

4.7.2 Cell cycle signatures

The cell cycle clusters were dominated by one large cluster of 822 probesets containing 648 genes. This showed a distinctive expression pattern (top panel figure 4.29) but in the absence of clinical annotation this is difficult to interpret, although it is tempting to speculate that the higher peaks represent the higher grade tumours. There are several smaller associated clusters with less distinctive patterns of gene expression (one cluster for illustration (cluster 26), bottom panel figure 4.29). The remaining clusters are all small, containing less than 20 probesets per cluster and are not further discussed.

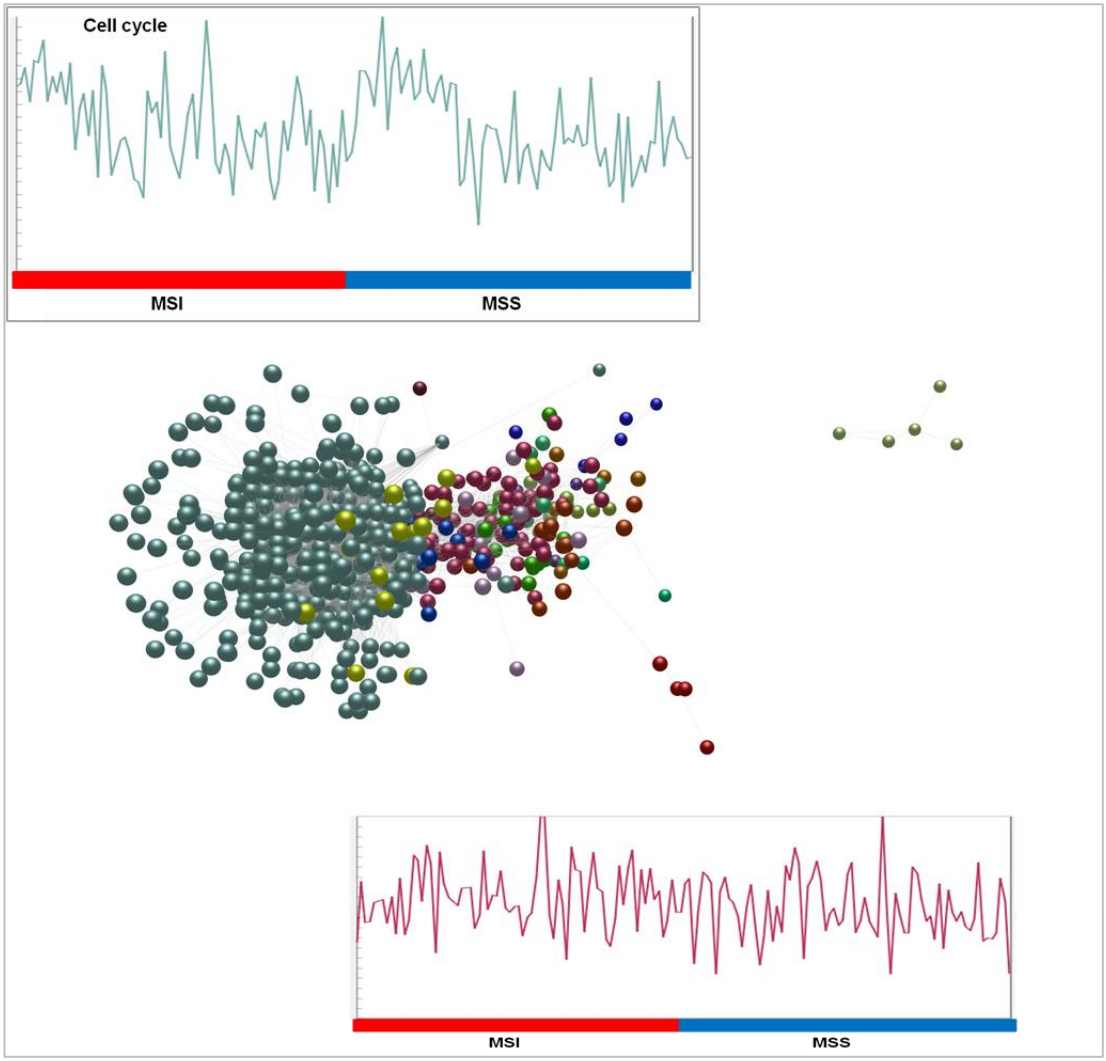


Figure 4.29. The cell cycle related clusters with the rest of the graph hidden. Expression plots for two selected clusters shown. Colours in clusters correspond to colours of expression plots. Individual tumours are represented along the x-axis in the expression plots with the coloured bars representing tumour subtype.

The largest cluster (pale blue in figure 4.29) appears to represent transcriptional and translational activity with multiple RNA binding proteins represented in the gene list including RNA binding motif proteins, eukaryotic translation initiation factors, heterogenous nuclear riboproteins, splicing factors and poly(A) polymerases (RBM14, RBM15, RBM17, RBM33, RBM34, RBM47, RBM9, EIF3J, EIF4E, EIF4EBP2, EIF4G1, EIF5, HNRNPC, HNRNPM, HNRNPU, HNRNPUL2, HNRNPDL, SFRS1, SFRS2, SFRS5, PAPOLA, PAPOLG) as well as genes involved in protein transport (ARF1, ARF6, RAB14, RAB18, RAB1A, AP1G1,

AP1S3, AP3B1, AP3D1, EXO5, EXO6, GOPC, GGA2, IPO11, IPO8, NUP133, TNPO2). See table 4.39 for selected functional annotation.

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Protein transport	7.6×10^{-13}
DAVID	SP_PIR_KEYWORDS	RNA binding	3.0×10^{-13}
DAVID	GOTERM_CC_FAT	Golgi apparatus	1.9×10^{-8}

Table 4.39 Functional annotation for the cell cycle transcription/translation signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

Cluster 26 (red in figure 4.29) contains 76 probesets to 67 genes. This, along with several of the smaller clusters appears to represent the mitotic phase of the cell cycle containing several cyclins, kinesin family members, helicase, and centrosomal and centromere proteins (CCNA2, CCNB1, CCNB2, KIF18A, KIF23, KIF2C, HELLS, CENPE, CEP55). Enrichment analysis confirms this enrichment for mitotic phase genes (Table 4.40).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	M phase	3.3×10^{-25}
DAVID	SP_PIR_KEYWORDS	Cell division	1.4×10^{-26}

Table 4.40 Functional annotation for the cell cycle mitotic phase signature (cluster 26) using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values..

4.8 Gene signatures from glioma

The glioma data set contained 271 tumours of which 253 passed the QC step, and which were stratified on the basis of histological type (astrocytoma, oligodendroglioma, glioblastoma), WHO grade and sex. The glioma dataset when laid out at r = 0.7 forms a network of 23,015 nodes joined by 623,591 edges in which the immune genes lie together (Figure 4.30). These tend to be higher in glioblastoma and lower in oligodendrogliomas. There are also multiple clusters formed of brain-specific genes, not seen in any other datasets and identified as brains specific from comparison with tissue atlases. As would be expected from the higher mitotic rate which forms part of the classification of glioblastoma, the cluster of cell cycle genes are expressed at the highest levels in this tumour, and as in other tumours the cell cycle and cell cycle related cluster fall in one area of the graph.

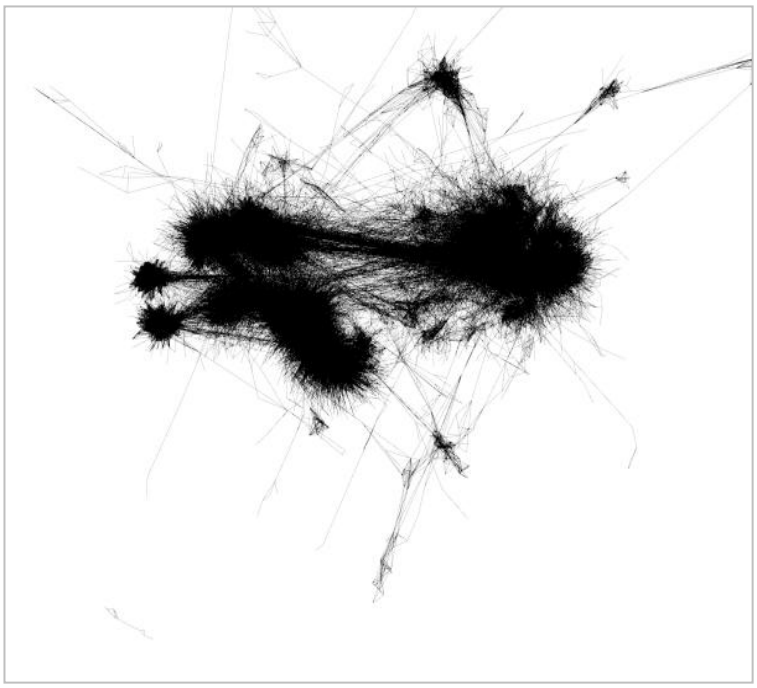


Figure 4.30. Glioma graph laid out at r≥0.7 with nodes hidden and only edges illustrated.

4.8.1 Stromal signatures

The immune clusters group together in one area of the graph. The largest cluster is enriched in macrophage genes and lies close to but separate from a cluster enriched in IFN associated genes (Fgiure 4.31). There is a small B cell signature with a rather

spiked profile in astrocytomas and glioblastomas but absent from oligodendrogliomas. Several other small clusters with spiked profiles are enriched in genes of broadly immune function but which cannot be further characterised.

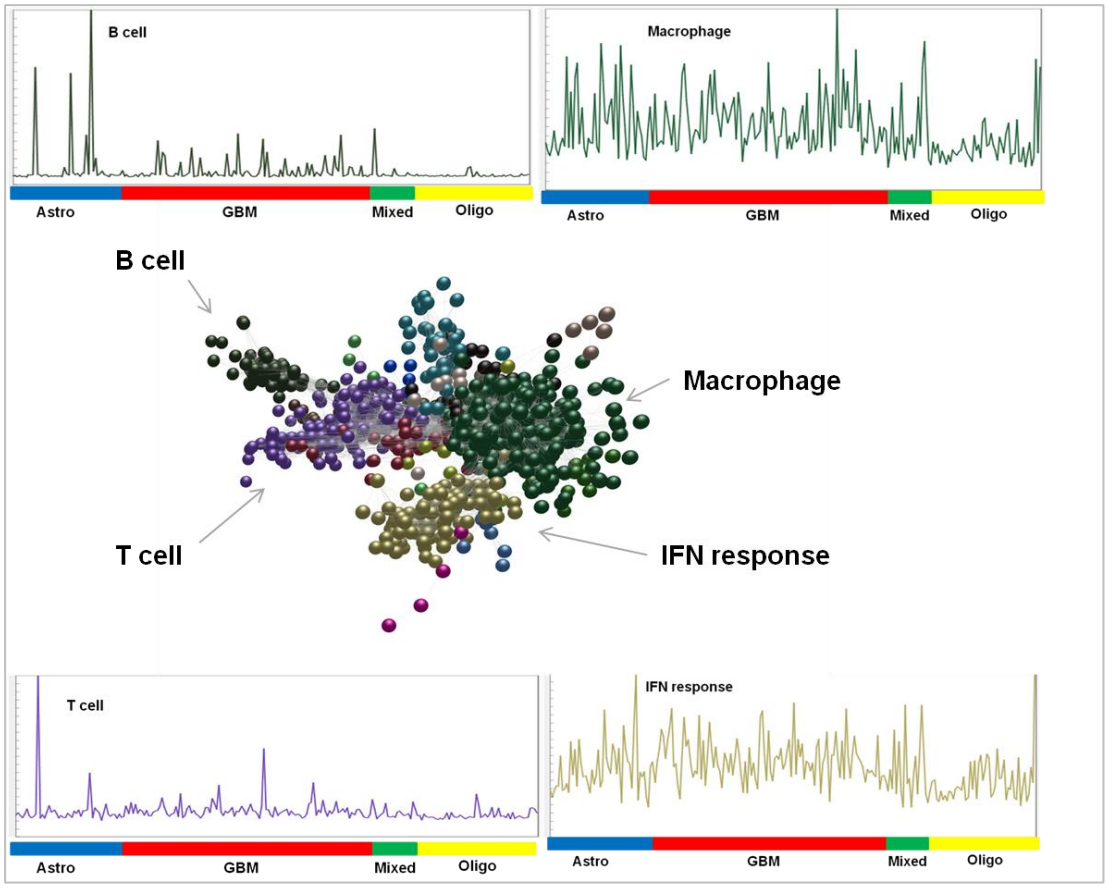


Figure 4.31. The immune clusters with the rest of the graph hidden. Mean expression plots for the clusters are given in the surrounding panels. Colours of clusters correspond to the colours of expression plots. Individual tumours are represented along the x-axis of the expression plots with coloured bars representing tumour types.

The largest cluster contains macrophage and T cell genes and consists of 513 probesets to 346 genes. It contains macrophage markers such as CD14, CD68, CD163 and CSF1R. There are multiple components of the class II antigen processing and presentation pathway with CD74, the class II invariant chain being present as well as multiple MCH class II genes and the class II transactivator (HLA-DMA, HLA-DMB, HLA-DOA, HLA-DPA,1 HLA-DPB1, HLA-DRA, HLA-DRB1, HLA-DRB4, HLA-DRB5, HLA-DRB6, CIITA). There are multiple toll-like

receptors (TLR1, TLR2, TLR5, TLR7, TLR8) as well as some macrophage scavenger receptor and apolipoprotein receptor (MSR1, APOB48R). Multiple Fc receptors are included in the cluster (FCER1G, FCGR1A, FCGR1C, FCGR2A, FCGR2B, FCGR2C, FCGRB3). Cytokines, and their receptors are represented (CCR1, IL10RA, IL12RB1, IL12RA1, IL17RA, IL18, IL21R, IL4R). There are complement receptors (C3A1, C5A1) as well as three chains of the C1q molecule (C1QA, C1QB, C1QC). There are several genes associated with the lysosome present (CTSB, CTSC, CTSD, CTSS, CTSZ, NAGA, IFI30, LGMN, LAPTM5, MAN2B1, MANBA, PLA2G15, SRGN, SLC15A3, STXBP2, ZNRF2). The enrichment for genes associated with immune response is confirmed on enrichment analysis (Table 4.41).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Defense response	6.4×10^{-27}
DAVID	GOTERM_BP_FAT	Leukocyte mediated immunity	1.1×10^{-12}
DAVID	GOTERM_CC_FAT	Lysosome	2.9×10^{-9}
DAVID	GOTERM_BP_FAT	Antigen processing and presentation	8.5×10^{-16}

Table 4.41 Functional annotation for the macrophage signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values.

A rather spiked cluster (purple in fig 4.31) is enriched in genes for T cells as well as containing some vascular genes. This contains 190 probesets representing 136 genes and appears to represent a cytotoxic T cell signature containing CD8, the cytotoxic granules granzymes A, B, H and K. ZAP70 is present but the other elements of the TCR signalling pathway are not seen. There is a vascular signature including CD34, PECAM1 (CD31), vascular endothelial cadherin, endomucin, endothelial cell-specific chemotaxis regulator, selectin P and thrombospondin 1 (CD34, PECAM1, CDH5, EMCN, ECSCR, SELP, THBS1). This close association between elements of the vascular signature and the T cell signature is not seen in other datasets and may represent an association unique to primary CNS tumours, but is more likely to

represent the skewing effect seen due to this expression pattern being seen in only a few tumours and therefore showing a distinctive gene expression profile. See table 4.42 for selected functional annotation.

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Blood vessel development	5.8×10^{-8}
DAVID	SP_PIR_KEYWORDS	T cell	3.2×10^{-8}
DAVID	GOTERM_BP_FAT	Immune response	7.3×10^{-8}

Table 4.42 Functional annotation for the T cell/vasculature signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values.

The interferon response cluster as in other datasets is closely associated with the macrophage cluster. This contains 123 probesets to 75 genes and is enriched in interferon induced proteins and regulatory factors (IFI27, IFI35, IFI44, IFI44L, IFI6, IFIH1, IFIT1, IFIT2, IFIT3, IRF1, IRF7, IRF9). The class I MHC molecules are represented as well as components of the proteasome and TAP transporters (B2M, HLA-A, HLA-B, HLA-C, HLA-E, HLA-F, HLA-G, PSMB8, PSMB9, PSME1, PSME2, TAP1, TAPBP). Enrichment analysis confirms the cluster to be enriched for IFN induced genes (Table 4.43).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Immune response	1.1×10^{-14}
DAVID	GOTERM_BP_FAT	Antigen processing and presentation	4.2×10^{-12}
GSEA	HECKER_IFNB1_TARGETS	Genes modulated in the blood of patients with MS treated with IFN beta	1.2×10^{-76}
GSEA	BROWNE_INTERFERON RESPOSIVE GENES	Genes upregulated in primary fibroblast culture after treatment with IFN alpha	2.9×10^{-68}

Table 4.43 Functional annotation for the interferon response signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

The small B cell cluster consists of 80 probesets to 39 genes. In this cluster, most of the genes with identified function encode immunoglobulin genes, with other elements of the B cell signature seen in other tumours largely absent. The informative genes that are present however include POU2AF1, which encodes the B cell transcription factor BOB-1 and TNFRSF17 encoding CD269, a marker of mature B cells.

There are several clusters representing the extracellular matrix. With the exception of the one 'spiked' cluster, these tend to show higher levels of expression in glioblastoma than in the other tumours (Figure 4.32).

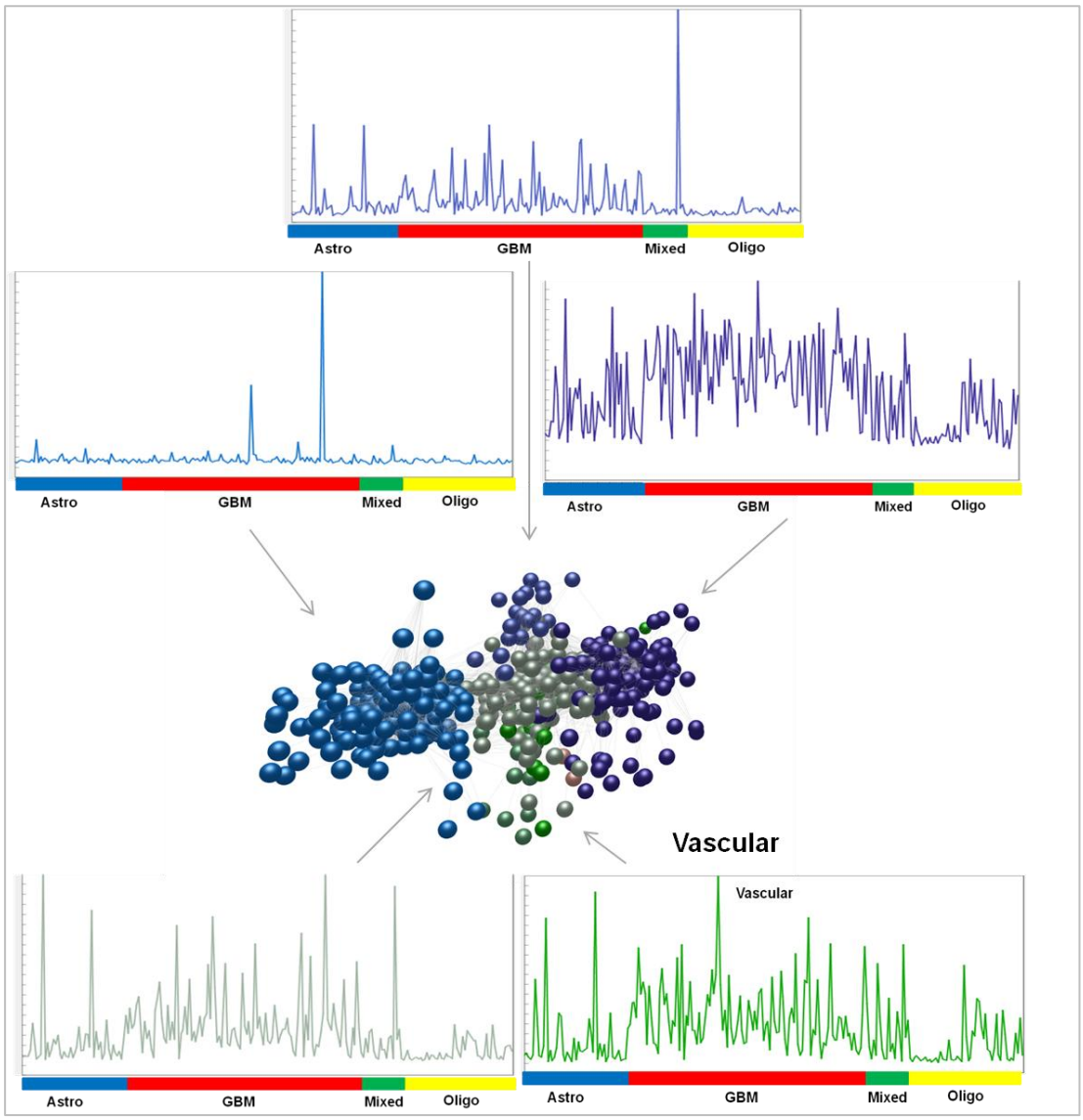


Figure 4.32. The non-immune stromal clusters with the rest of the graph hidden. Mean expression plots for selected clusters given in the surrounding panels. Colours of clusters correspond to colours of expression plots. Individual tumours represented along x-axis of expression plots with tumour types denoted by coloured bars.

The largest ECM cluster (blue in figure 4.32) is driven by expression in only a very small number of tumours. This contains 221 probesets representing 136 genes. The genes making up this cluster contain fewer collagens than seen in the ECM signature in most tumours reflecting the unique structure of the brain but contains extracellular proteins such as bone morphogenic protein 5, type 1 cadherin, cortexin 3, democollin 3, desmoglein 2, elastin, fibrillin 1 and 2, matrilin 3, and proteoglycan 4 (BMP5,

CHD1, DSC3, DSG2, ELN, FBN1, FBN2, MATN3, PRG4). See table 4.44 for selected functional annotation.

Analysis tool	Source of annotation	Annotation	Significance (p-value)
GSEA	GOTERM	Extracellular matrix	4.0×10^{-7}
GSEA	NABA_CORE_MATRISOME	Core extracellular matrix genes	2.3×10^{-11}
GSEA	NABA_ECM_GLYCOPROTEINS	Genes encoding ECM structural glycoproteins	5.1×10^{-9}

Table 4.44 Functional annotation for the ECM signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA (Blue cluster in figure). Selected annotation given only. Annotation derived from GSEA gives hypergeometric distribution p-values.

The next largest ECM cluster shows expression in the majority of tumours (purple in figure 4.32). Expression is lowest in the oligodendrogliomas, a pattern seen with several of the ECM clusters. This cluster is of 168 probesets to 124 genes. This again shows a less obvious ECM signature than that seen in other tumours and indeed it is only recognisable as such by comparison with ECM signatures derived from other analyses. This cluster contains among others caldesmon 1, chondroitin sulphate synthase, fibronectin 1, myosin light and heavy chains, myoferlin, transgelin and vimentin (CALD1, CHSY1, FN1, MYH9, MYL12A, MYL12B, MYOF, TAGLN2, VIM).

The next largest cluster is of 75 genes represented by 112 probesets. This is also shows higher expression in GBM than other tumours. There is more obvious enrichment for extracellular proteins in this cluster with multiple collagens represented as well as caldesmon1, fibrillin1, several laminins, and tropomyosins (COL1A1, COL1A2, COL3A1, COL5A1, COL5A2, COL18A1, CALD1, FBN1, LAMA3, LAMC1, TPM1, TPM2). There are modifiers of the extracellular matrix

including ADAM metallopeptidases (ADAM12, ADAMTS12, ADAMTS2) and MMP14. Enrichment of this cluster for ECM genes is confirmed on analysis with DAVID and GSEA (Table 4.45).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	SP_PIR_KEYWORDS	Extracellular matrix	9.5×10^{-19}
DAVID	GOTERM_BP_FAT	Cell adhesion	9.7×10^{-9}
GSEA	KEGG	ECM receptor interaction	1.3×10^{-20}

Table 4.45 Functional annotation for the smaller ECM signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

Lying adjacent to this is a small cluster of 28 probesets and 16 genes that appears to represent a vascular signature. This includes collagen 4, notch 3, angiopoetin 2 and the endothelial specific molecule (COL4A1, COL4A2, NOTCH3, ANGPT2, ESM1). This signature is higher in GBM than the other tumours, as would be predicted based on histology, these tumours being characterised by abundant neo-vascularisation. Despite the small number of genes present, analysis for enrichment demonstrates enrichment for genes associated with angiogenesis (Table 4.46).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Blood vessel morphogenesis	4.9×10^{-4}
GSEA	ROY_WOUND_BLOOD VESSEL	Genes upregulated in blood vessel cells from wound sites	4.4×10^{-7}

Table 4.46 Functional annotation for the vascular signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

4.8.2 Cell cycle signatures

The cell cycle signatures fall in one area of the graph although individual clusters are not particularly tightly related (Figure 4.33). Most of the cell cycle and cell cycle related signatures show higher expression in GBM reflecting the aggressive nature of that tumour.

The largest cluster (green on the left side of figure 4.32) appears to represent the mitotic phase of the cell cycle. This consists of 379 probesets to 260 genes. The cluster contains both A and B aurora kinases, multiple units of DNA polymerases and topoisomerases (AURKA, AURKB, POLA2, POLD3, POLE, POLE2, POLQ, TPO2A, TOPB1). There are multiple cyclins and kinesins (CCNA2, CCNB1, CCNB2, CCNF, KIF11, KIF14, KIF15, KIF18A, KIF18B, KIF20A, KIF22, KIF23, KIF2C, KIF4A, KIFC1). There are also many cell cycle division associated genes and centromere proteins (CDC6, CDC7, CDCA2, CDCA3, CDCA4, CDCA5, CDCA7, CDCA8, CENPA, CENPE, CENPF, CENPH, CENPI, CENPK, CENPL, CENPM, CENPN, CENPO). Enrichment analysis confirms the enrichment for genes associated with the mitotic phase of the cell cycle (Table 4.47).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Cell cycle	1.6×10^{-84}
DAVID	GOTERM_BP_FAT	M phase	5.4×10^{-78}
DAVID	KEGG_PATHWAY	DNA replication	3.2×10^{-19}
GSEA	BENPORATH_CYCLING_GENES	Genes showing cell cycle stage specific expression	7.4×10^{-169}

Table 4.47 Functional annotation for the cell cycle mitotic phase signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

The next largest clusters all contain cell cycle related genes with the next 3 largest illustrated in figure 4.33. One of these (illustrated in purple, bottom left of figure 4.33) containing 45 probesets to 36 genes appears to represent transcription/translation being rich in RNA binding proteins, RNA polymerase, heterologous nuclear riboproteins and splicing factors (RBM12, RBM8A, PLOR2D, HNRNPA2B1, HNRNPA3P1, HNRNPAB, HNRNPR, HNRNPU, HNRPLL, SFRS1, SFRS2, SFRS7, SFRS9).

There is a small cluster associated with these others with a rather spread out pattern within the graph (green cluster right hand side figure 4.33). This consists of 24 probesets and 20 genes. These show no clear functional annotation when considered as a cluster but frequently cluster with cell cycle related genes in other datasets suggesting a common function. This small cluster contains multiple units of the lysosomal H+ transporting ATPase.

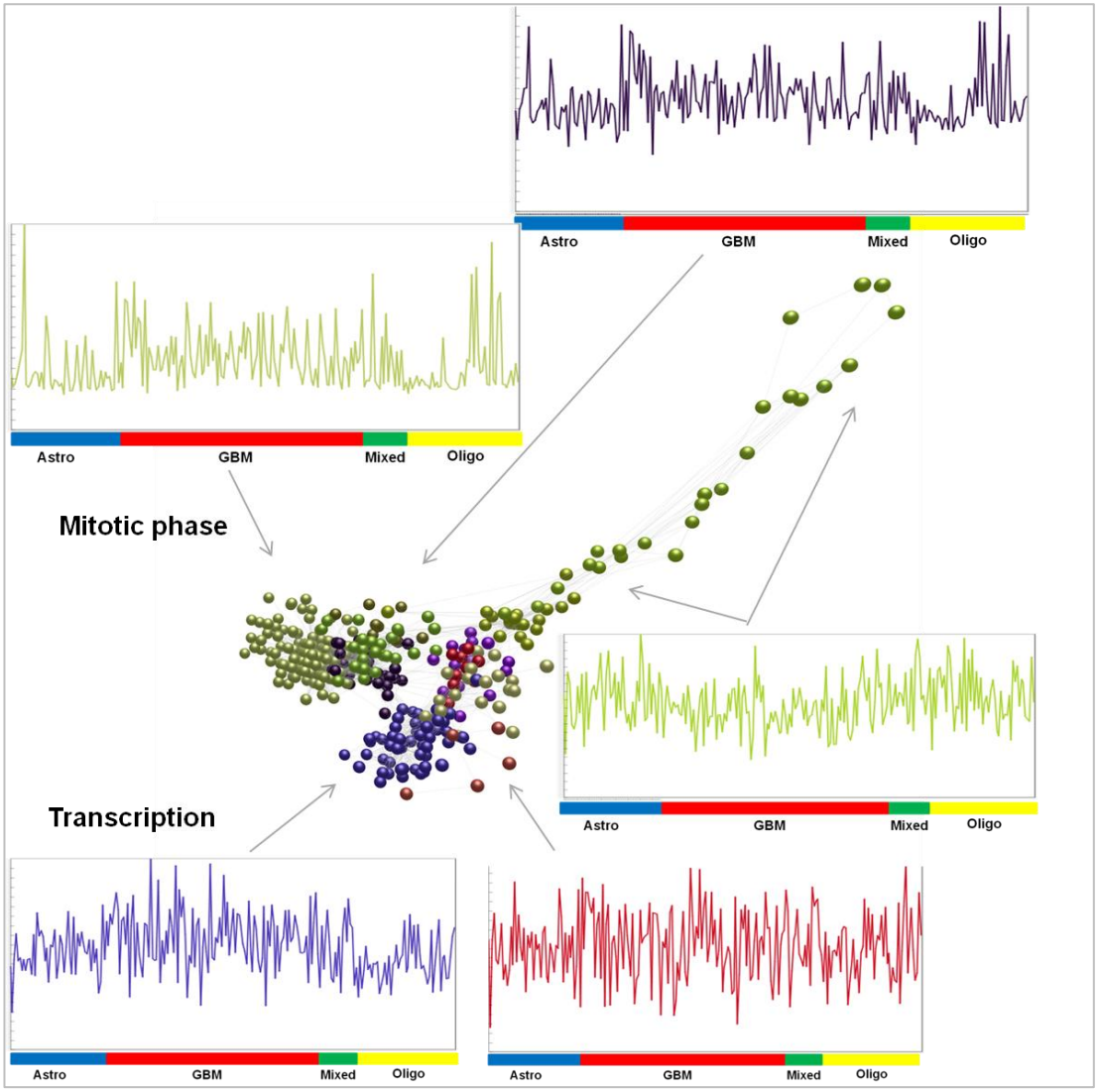


Figure 4.33. The cell cycle-related clusters with the rest of the graph hidden. Surrounding panels show mean expression profile in each cluster. Colours of clusters correspond to colours in expression plots. Individual tumours are represented along the x-axis of expression plots. Coloured bars denote tumour types.

4.8.3 Other signatures

As in other datasets there are clusters representing ribosomes, glycolysis, oxidative phosphorylation, histones, haemoglobin and similar. There are also clusters that appear from comparison with tissue libraries to be only expressed in the brain, as well as very many clusters that separate the tumour types from each other. These are not further explored.

4.9 Gene signatures from testicular tumours

The testicular dataset was formed of 107 samples from patients from the USA. Eighty six of these passed the QC step and were used in subsequent analysis. The dataset was of primary germ cell tumours and was stratified first on whether the tumours were pure or mixed histological types and then within each category on the constituent parts of the tumour using the WHO classification (seminoma, teratoma, embyronal carcinoma, yolk sac tumor, choriocarcinoma).

The testicular network has a very large number of edges in relation to nodes (945,082 and 18,934 respectively) (Figure 4.34). This reflects the complex mix of tumours in the dataset and specifically the inclusion of teratomas. These tend to form tight structure that shows very high internal connectivity but does not show high connectivity to the rest of the graph and on analysis often represent signatures of mature (non-testicular) tissue types in keeping with the histological appearances of teratoma. Because the signatures seen particularly in the teratomas are often unique to a single tumour, the graph structure is in part driven by 'spiked' signatures that form very tight clusters within the graph. The graph is laid out to include a smaller number of nodes than other graphs, as the high number of edges and complexity of the graph means this graph lies at the limit of computing power to manipulate the graph.

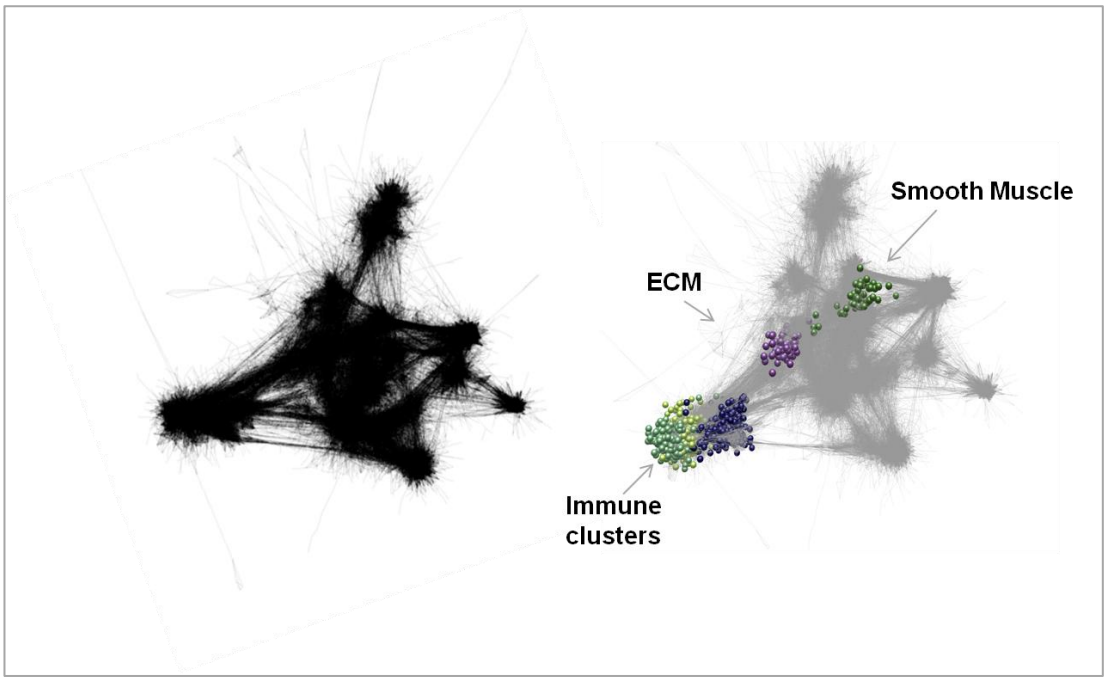


Figure 4.34. The testicular tumour graph laid out at r≥0.75. Left hand figure; only edges are shown to demonstrate the structure of the graph. Right hand figure; selected clusters are superimposed on the underlying structure. The immune clusters lie together at one edge of the graph. Other stromal signatures are dispersed dependent on whether they are common to all tumours (ECM signature here) or present only in individual teratomas (smooth muscle cluster).

4.8.4 Stromal signatures

The immune clusters, with the exception of the mast cell signature, lie together at one pole of the graph and cluster tightly together (Figure 4.34). All clusters show a relatively 'spiky' architecture reflecting the small number of tumours in the dataset. The macrophage signature tends to have a uniformly high expression in teratomas, and immune signatures all tend to be lower in yolk sac tumours but there is otherwise no close relationship with tumour type (Figure 4.35). The exception to this, and the cluster that lies entirely separate within the main graph, is the small mast cell cluster, which is more frequently seen in teratomas than other tumour types and bears no relation to the other immune signatures. The non-immune stromal signatures exist in two distinct patterns; there are those that are expressed in all tumour types and represent broadly the tumour stroma and vasculature and there are those that are discrete clusters of genes expressed in teratomas. The generic ECM signature is

expressed at higher levels in teratoma-containing tumours than others, in keeping with the histological appearances, most types of testicular tumour having an inconspicuous extracellular matrix histologically.

The largest immune cluster appears to represent a lymphocyte signature and is formed of 629 probesets to 428 genes (Figure 4.35). This contains markers of T cells including CD2, CD3 (delta, epsilon and gamma components), CD5, CD7, CD8 (A and B), CD52, and markers of B cells including CD19, CD22, CD79A and CD79B and CD20 (MS4A1). There are genes involved in lymphocyte activation, antigen recognition and immune function including the chemokines and receptors CCL21, CCR6, CCR7, CXCR4, CXCR5, MHC molecules including HLA-DOA, HLA-DOB, HLA-DQA1, HLA-DQB1, HLA-E, HLAF, CIITA, and interleukins IL15, IL16, IL23A along with interleukin receptors IL21R, IL2RB, IL2RG, IL4R, IL7R. Analysis via DAVID and GSEA confirms significant enrichment for lymphoid genes (Table 4.48).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Lymphocyte activation	7.3×10^{-35}
DAVID	GOTERM_BP_FAT	T cell differentiation	4.8×10^{-16}
DAVID	KEGG_PATHWAY	B cell receptor signalling	7.1×10^{-16}
GSEA	GSE29618	Genes upregulated in B cell compared to monocytes	3.9×10^{-65}

Table 4.48 Functional annotation for the lymphocyte signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

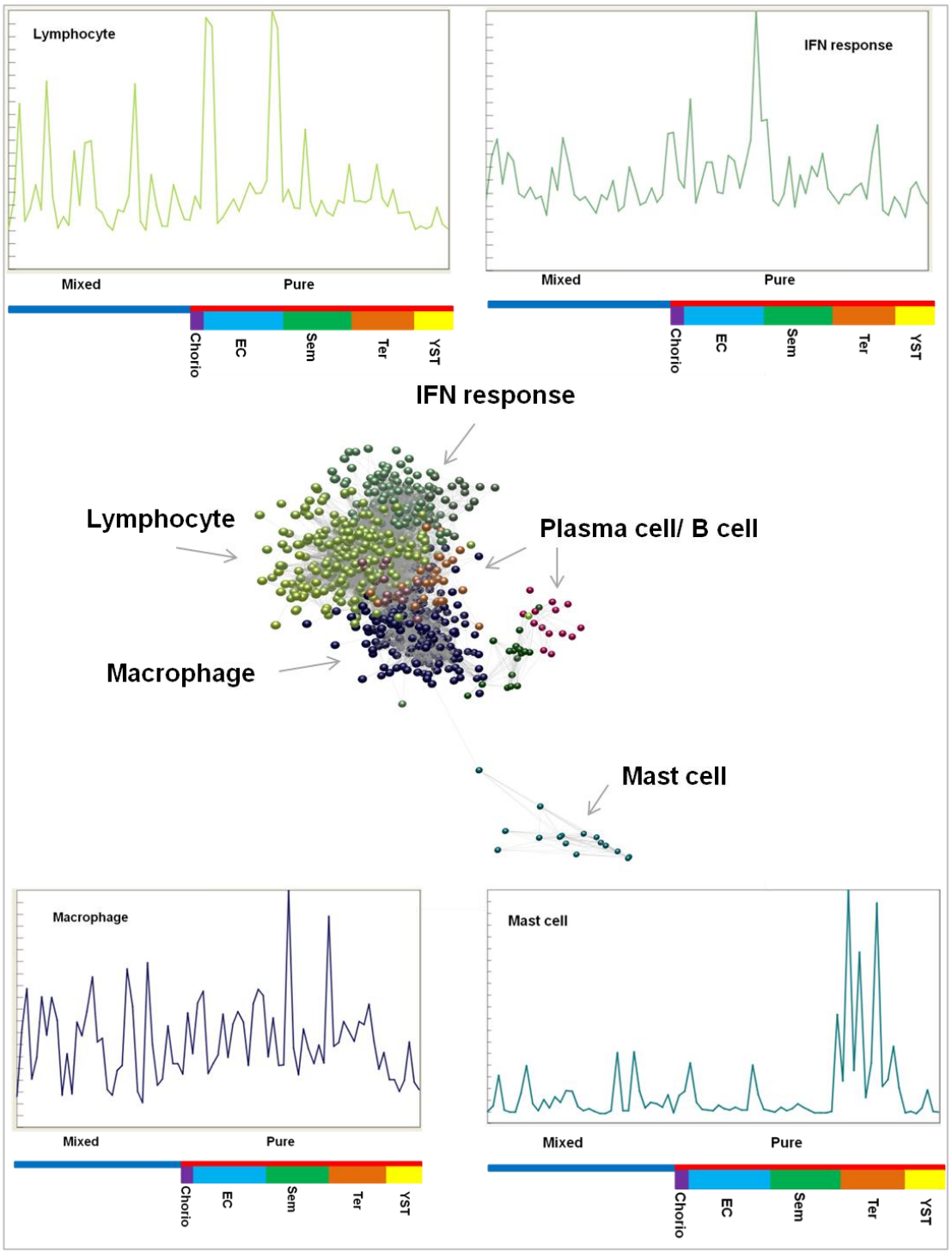


Figure 4.35. The immune clusters with the rest of the graph hidden. Surrounding panels show mean expression profile for each cluster. Colours of clusters correspond to colours of expression plots. Individual tumours are represented along the x-axis with coloured bars representing tumour groupings. Mixed tumours lie to the left and pure histological types to the right. (Key; Chorio = Choriocarcinoma, EC = Embryonal carcinoma, Sem = Seminoma, Ter = Teratoma, YST = Yolk Sac Tumour).

Adjacent to this is a large macrophage cluster of 376 probesets to 256 genes (Figure 4.35). This contains known macrophage markers CD14, CD68, CD163, CSF1R as well as multiple Fc receptors (FER1G, FCGR1A, FCGR1B, FCGR2A, FCGR2C), and toll-like receptors 2, 4 and 8. There are also multiple subcomponents of the C1q complement molecule, and the chemotaxis receptors C3AR1 and C5AR1. There are the cytokines CCL3, CCL4, CXCL16 and receptors CCR1, CCR5, CXCR3. Other macrophage associated genes such as allograft inflammatory factors 1 (AIF1), legumain (LGMN), and mannose receptor (MRC1) are also present. Analysis confirms enrichment for macrophage genes (Table 4.49).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Defense response	6.0×10^{-30}
DAVID	SP_PIR_KEYWORDS	Lysosome	4.6×10^{-17}
DAVID	KEGG_PATHWAY	Antigen processing and presentation	3.3×10^{-16}
DAVID	GOTERM_BP_FAT	Chemotaxis	3.3×10^{-8}
GSEA	GSE10325	Genes downregulated in B cell compared to monocytes	1.1×10^{-66}

Table 4.49 Functional annotation for the macrophage signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

The interferon response breaks into three clusters, the largest 155 probesets to 135 genes, the second 30 probesets to 20 genes and the third 19 probesets to 17 genes. As in other datasets, this cluster is closely related to both the T cell and macrophage signatures. The largest cluster contains genes such as interferon induced proteins 35, 44, interferon-induced protein with tetratricopeptide repeats 2 and 3, and interferon regulatory factor 7. Analysis confirms enrichment for interferon induced genes (Table 4.50).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
GSEA	HECKER_IFNB_TARGETS	Genes modulated in the blood of patients with MS on treatment with IFNB	4.7×10^{-41}
GSEA	BROWNE_INTERFERON RESPONSIVE GENES	Genes upregulated in primary fibroblasts after treatment with IFNA	2.4×10^{-35}
GSEA	REACTOME	Immune system	4.6×10^{-29}

Table 4.50 Functional annotation for the largest cluster of interferon response signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from GSEA gives hypergeometric distribution p-values.

There is are 3 plasma cell clusters, one of 59 probesets and 32 genes and further smaller clusters of 17 probesets and 10 genes and 15 probesets and 7 genes (Figure 4.34). These consist largely of immunoglobulin genes with 24 genes representing immunoglobulin loci and many others hypothetical proteins.

A further difficult to characterise immune cluster contains 31 probesets and 25 genes including the macrophage scavenger receptor MARCO, RANKL (TNFSF11), DC-SIGN (CD209), CCL23 and CCR1. It is difficult to identify a clear functional enrichment in this cluster although the signature is clearly broadly an 'immune' one.

A mast cell cluster consists of 16 probesets to 11 genes and contains an Fc receptor for IgE, KIT ligand, tryptase alpha and beta components, mast cell carboxypeptidase, and CXCL14 (MS4A2, KITLG, TPSAB1, TPSB2, CPA3, CXCL14). This has a markedly different pattern of expression than the other immune clusters and appears to represent part of a teratoma-associated signature rather than forming part of the immune response to the tumour (Figure 4.35). Enrichment analysis, despite the small number of genes, confirms this is enriched in mast cell genes (Table 4.51).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
GSEA	NAKAJIMA_MAST CELL	Top 50 mast cell specific genes	2.0×10^{-13}

Table 4.51 Functional annotation for the mast cell signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. DAVID analysis provides no informative functional annotation. Annotation derived from GSEA gives hypergeometric distribution p-values.

Present in this dataset but not seen in others is a small cluster representing a gamma-delta T cell signature. This consists of 14 probesets and 11 genes and contains the TCR gamma and delta chains, the cytotoxic granule components perforin 1 and granzymes A and H (TARP, TRD@, TRGC2, TRGV9, PRF1, GZMA, GZMH). This cluster also contains interferon gamma (IFNG) and CCL5.

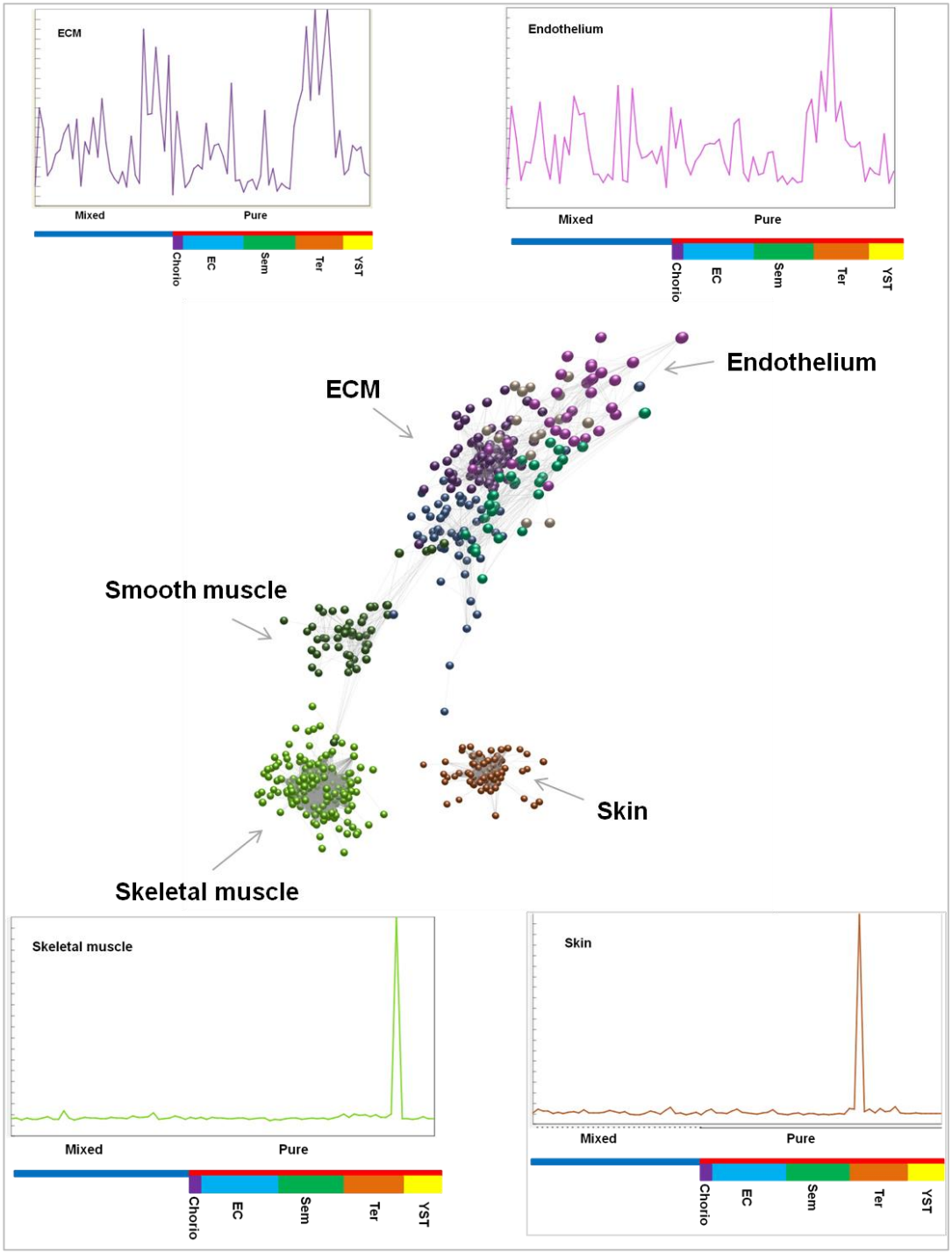


Figure 4.36. The non-immune stromal clusters with the rest of the graph hidden. The surrounding panels show mean expression profiles for each cluster. Cluster colours correspond to the colours of the expression plots. Individual tumours are represented on the x-axis of the expression plots. The coloured bars indicate tumour groupings. Mixed tumours lie to the left and pure histological types to the right. (Key; Chorio = Choriocarcinoma, EC = Embryonal carcinoma, Sem = Seminoma, Ter = Teratoma, YST = Yolk Sac Tumour).

The largest ECM clusters not associated with a specific teratoma, is of 77 probesets to 54 genes (Figure 4.36). This contains collagens I, III, V, VI and XII as well as biglycan, bone morphogenic protein, laminin alpha 4, and microfibrillar-associated protein 4 (COL1A1, COL1A2, COL3A1, COL5A1, COL5A2, COL6A1, COL6A2, COL6A3, COL12A1, BGN, BMP1, LAM4A, MFAP4). See table 4.53 for selected functional annotation.

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	SP_PIR_KEYWORDS	Extracellular matrix	4.4×10^{-18}
GSEA	STANFORD_TUMOUR _MODULES	ECM and collagens	1.7×10^{-46}
GSEA	NABA_CORE_MATRISOME	Genes encoding core extracellular matrix	2.6×10^{-33}

Table 4.53 Functional annotation for the ECM signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

There are two further small ECM clusters one of 29 probesets and 23 genes and one of 16 probesets to 14 genes (Figure 4.36). These contain genes such as fibulin 1, layilin, matrix remodelling associated 7, and TIMP2 (FBLN1, LAYN, MXRA7, TIMP2). These 3 ECM clusters show closely associated patterns of expression tending to be higher in tumours containing teratomatous elements.

There is a small endothelial cluster of 31 probesets to 25 genes. This has a similar profile to the generic ECM signature also tending to be higher in tumours with a teratomatous element (Figure 4.36). The cluster contains endosialin (CD248), endothelial cell specific chemotaxis regulator (ECSCR), endomucin (EMCN), CD31 (PECAM1) and TIE 1.

4.8.5 Cell cycle related signatures

In comparison with other datasets, the cell cycle cluster is smaller and rather more diffuse with much less obvious patterns between tumour types than in other datasets (Figure 4.37). In part this is likely to reflect the makeup of the dataset with the unique profiles in teratoma providing much highly correlated data forming complex structure within the graph and requiring a higher correlation threshold to be set to generate a visualisable graph than in other tumour types. Less highly correlated data, as much of this might be being driven by a variety of influences, has likely been lost from this dataset due to the limitation of the software.

The largest cell cycle signature (96 probesets, 87 genes) contains genes such as several general transcription factors (GTF2I, GTF3C2), mitogen activated protein kinases (MAPK1, MAP3K2, MAPK1IP1L), DNA polymerase interacting protein (POLDIP3) and retinoblastoma binding protein (RBBP9). Enrichment analysis identifies the cluster as being enriched in nuclear genes, genes involved in RNA binding and the cell cycle, but these enrichments are not highly significant (Table 4.53).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_CC_FAT	Nuclear lumen	5.0×10^{-7}
DAVID	GOTERM_MF_FAT	RNA binding	2.5×10^{-4}
DAVID	GOTERM_BP_FAT	M phase	5.4×10^{-3}

Table 4.53 Functional annotation for the largest cell cycle cluster using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

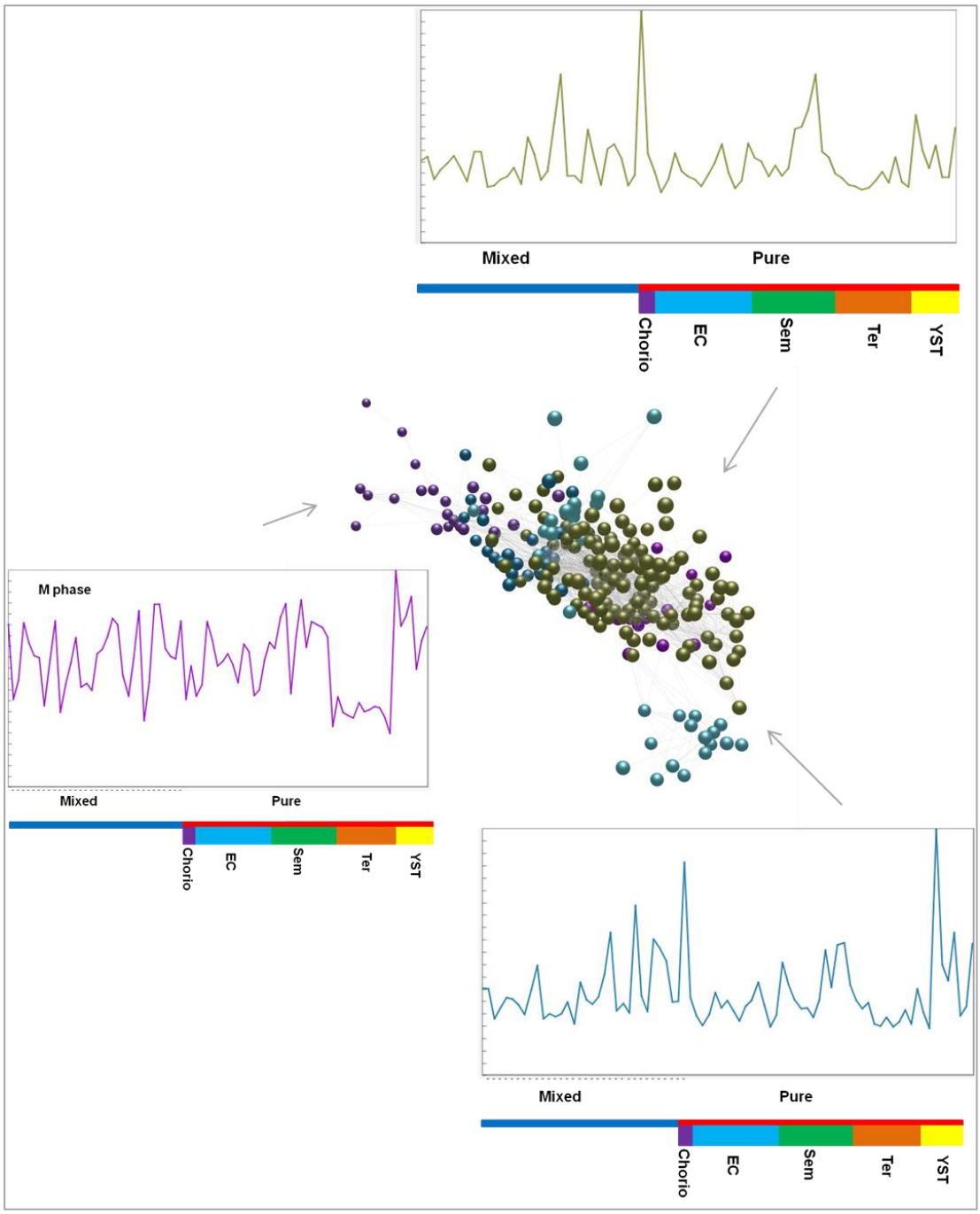


Figure 4.37. The cell cycle-related clusters with the rest of the graph hidden. Surrounding panel show mean expression profile for each cluster. The cluster colours correspond to the colours in the expression plots. Individual tumours are represented along the x-axis. Coloured bar denotes tumour groupings. Mixed tumours lie to the left and pure histological types to the right. (Key; Chorio = Choriocarcinoma, EC = Embryonal carcinoma, Sem = Seminoma, Ter = Teratoma, YST = Yolk Sac Tumour).

The next cell cycle signature (45 probesets, 42 genes) appears from enrichment analysis to be translation related containing genes associated with protein transport, mRNA processing and the endoplasmic reticulum (RAB5A, SEC23B, COPA, EXOC5, EXOC6, NUP133, STX16, XRN2, DHX35, CPSF2, PAPOLA, LASS6, PSEN1, SPTLC1, TMX1). Again these enrichments are not highly significant (Table 4.54).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Protein transport	4.0×10^{-5}
DAVID	GOTERM_BP_FAT	mRNA processing	3.8×10^{-2}
DAVID	SP_PIR_KEYWORDS	Endoplasmic reticulum	6.7×10^{-2}

Table 4.54 Functional annotation for the cell cycle translation related signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

Another small cluster of 31 genes appears to relate to DNA damage response pathways containing tumour protein p53 plus BRAC1 interacting protein, RAD21 homolog, and structural maintenance of chromosomes 3 (TP53, BRIP1, RAD21, SMC3).

There are only two small clusters that appear to represent the mitotic phase of the cell cycle (one of 17 and the other of 16 genes). These show a distinctive pattern of expression being much lower in teratomas compared to all other types of tumours as would be predicted from the histological appearances and clinical behaviour of the tumours (Figure 4.37).

4.8.6 Other signatures

The inherent complexity and variability of the graph demonstrates many signatures not seen in other datasets. Many of these, which appear to represent differentiated tissues are contributed by teratomas and include a skin signature, a neural tissue signature, a pancreas signature, a cartilage signature and a skeletal muscle signature. A liver signature is present in yolk sac tumours and there are three distinct but similar signatures of placenta, one of which represents chorionic somatotrophin, which are seen in choriocarcinoma-containing tumours. Many of the other signatures seen repeatedly in the datasets such as the housekeeping clusters, ribosomes, haemoglobin etc are also present.

The largest cluster in the graph (1349 probesets to 1051 genes), a position normally represented by a house-keeping cluster, represents testis specific genes, possibly reflecting the fact that the normal underlying gene expression profiles of testicular tissue has little in common with somatic tissues. Much of the cluster appears to related to spermatogenesis and contains such genes as the spermatogenesis associated genes 1, 3, 4, 6, 8, 9, 12, 16, 18, 20 and 22, sperm associated antigens 3, 6, 16 and 17, gametogenin, primary ciliary dyskinesia protein 1, outer dense fibre of sperm tails 1, 2 and 3-like, testis specific serine kinases 2, 4 and 6, and zona pellucid binding proteins (SPATA1, SPATA3, SPATA4, SPATA6, SPATA8, SPATA9, SPATA12, SPATA16, SPATA18, SPATA20, SPATA22, SPACA3, SPACA6,SPACA16, SPACA17, GGN, hCG_17324, ODF1, ODF2, ODF3L1, TSSK2, TSSK4, TSSK6, ZNBP, ZNBP2). Enrichment analysis confirms the enrichment for testis specific genes and genes involved in spermatogenesis (Table 4.55).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Sexual reproduction	1.9×10^{-58}
DAVID	GOTERM_BP_FAT	Spermatogenesis	1.4×10^{-56}
DAVID	GOTERM_CC_FAT	Cilium	1.6×10^{-11}
DAVID	GOTERM_BP_FAT	Meiosis	7.6×10^{-7}
GSEA	SU_TESTIS	Genes upregulated specifically in human testis	7.0×10^{-43}

Table 4.55 Functional annotation for the testis specific signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

The skeletal muscle cluster is of 519 probesets to 370 genes and contains genes including skeletal muscle actin, muscle creatine kinase, desmin, myoglobin, myocyte enhancer factor 2C, myosin binding protein C fast and slow types, multiple myosin light and heavy chains, myomesin family members, myotilin, myozenin 1 and 2, myopalladin, troponins and tropomyosins (ACTA1, CKM, DES, MB, MEF2C, MYBPC1, MYBPC2, MYH1, MYH2, MYH3, MYH4, MYH7, MYH8, MYL1, MYL2, MYL3, MYL5, MYL10, MYOD1, MYOD2, MYOD3, MYOT, MYOZ1, MYOZ2, MYPN, TNNC1, TNNC2, TNNI1, TNNI2, TNNT1, TNNT3, TPM1, TPM2, TPM3). The cluster shows enrichment for skeletal muscle genes (Table 4.56).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Muscle contraction	2.9×10^{-45}
DAVID	GOTERM_CC_FAT	Sarcomere	2.5×10^{-60}
DAVID	SP_PIR_KEYWORDS	Skeletal muscle	1.1×10^{-26}

Table 4.56 Functional annotation for the skeletal muscle signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

The neural tissue cluster (363 probesets, 281 genes) contains genes such as multiple glutamate receptors, and GABA receptors, synapsin 1 and 2, synaptic vesicle glycoproteins, neurotrophic tyrosine kinases, neurofilament, oligodendrocyte transcription factor 1 and glial fibrillary acidic protein (GRIA1, GRIA2, GRIA3, GRIA4, GRIK1, GRIN2A, GRM1, GRM3, GRM5, GRM7, GABBR2, GABRA1, GABRA2, GABRB1, GABRG1, SYN1, SYN2, SV2A, SV2C, NTRK2, NTRK3, NEFL, OLIG1, GFAP). The enrichment for neural genes is confirmed on analysis. (Table 4.57)

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Synaptic transmission	3.7×10^{-27}
DAVID	GOTERM_BP_FAT	Neuron differentiation	1.1×10^{-12}

Table 4.57 Functional annotation for the neural signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values.

The skin cluster (265 probesets, 206 genes) contains multiple keratins, keratinocyte differentiation associated protein, dermokine, democollin 1 and 3, desmoglein 1 and 3, epiregulin, late cornified envelope 2B and 3D and involucrin (KRT1, KRT2, KRT5, KRT6A, KRT6B, KRT10, KRT14, KRT15, KRT16, KRT17, KRT31, KRT77, KRT80, KRTDAP, DMKN, DSC1, DSC3, DSG1, DSG3, EREG, LCE2B, LCE3C, IVL). The enrichment for skin associated genes is confirmed on enrichment analysis (Table 4.58).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Epidermis development	7.3×10^{-48}
DAVID	GOTERM_CC_FAT	Desmosome	5.6×10^{-11}
GSEA	JAEGER_METASTASIS	Genes downregulated in metastases from malignant melanoma compared to primary tumours	1.8×10^{-120}

Table 4.58 Functional annotation for the skin signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

A smooth muscle signature (54 probesets, 31 genes) contains smooth muscle actin, smooth muscle calponin, leiomodulin, myosin light and heavy chains, myosin light chain kinase and tropomyosin (ACTA2, ACTG2, CNN1, LMOD1, MYH11, MYH6, MYL4, MHL9, MYLK, TPM1). See table 4.59 for selected functional annotation.

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	SP_PIR_KEYWORDS	Muscle protein	2.2×10^{-12}
GSEA	REACTOME_SMOOTH MUSCLE CONTRACTION	Genes involved in smooth muscle contraction	1.3×10^{-17}

Table 4.59 Functional annotation for the smooth muscle signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

A cartilage signature (52 probesets, 34 genes) is seen in several tumours containing a teratomatous component consistent with the frequent occurrence of areas of cartilaginous differentiation in teratomas (Figure 4.38). This signature includes collagens II, IX, XI and XXIV, chondroitin sulphate proteoglycan, aggrecan, and matrilin 1 cartilage matrix protein (VOL2A1, COL9A1, COL11A2, COL24A1, CSPG4, ACAN, MATN1). Enrichment for cartilaginous genes is confirmed on analysis via DAVID (Table 4.60).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Chondrocyte differentiation	2.3×10^{-6}
DAVID	GOTERM_BP_FAT	Cartilage development	3.7×10^{-6}

Table 4.60 Functional annotation for the cartilage signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values.

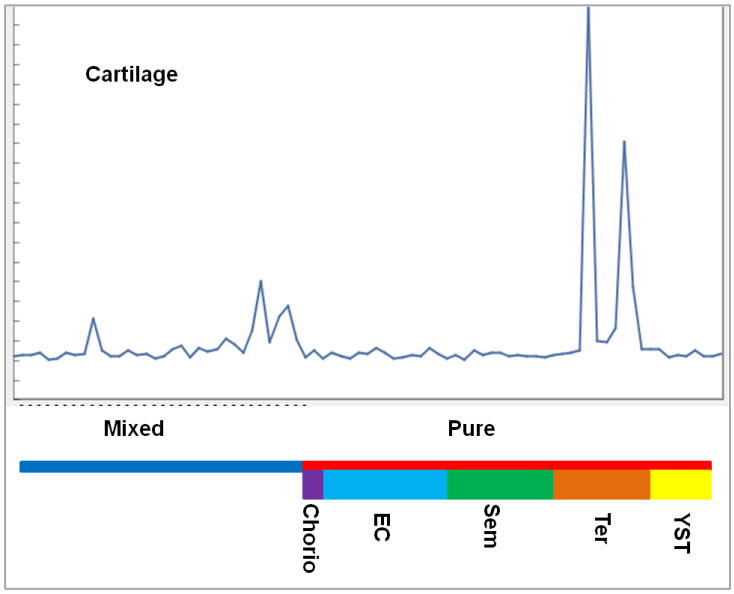


Figure 4.38. The genes in the cartilage cluster are expressed in only a few tumours all with a teratomatous component. Mean expression profile of genes forming cartilage cluster. Individual tumours are represented on the x-axis with the coloured bars denoting tumour groupings. Mixed tumours lie to the left and pure histological types to the right. (Key; Chorio = Choriocarcinoma, EC = Embryonal carcinoma, Sem = Seminoma, Ter = Teratoma, YST = Yolk Sac Tumour).

There is a 'liver' signature (140 probesets, 103 genes) which is expressed in tumours with a yolk sac component (Figure 4.39). This includes many of the synthetic products of hepatocytes including alpha-fetoprotein, albumin, multiple apolipoproteins, coagulation factors II and VII, fibrinogen and plasminogen (AFP, ALB, APOA1, APOA2, APOA4, APOB, APOC3, APOM, F2, F7, FGA, FGG, PLG). This enrichment for liver specific genes is confirmed on enrichment analysis (Table 4.61).

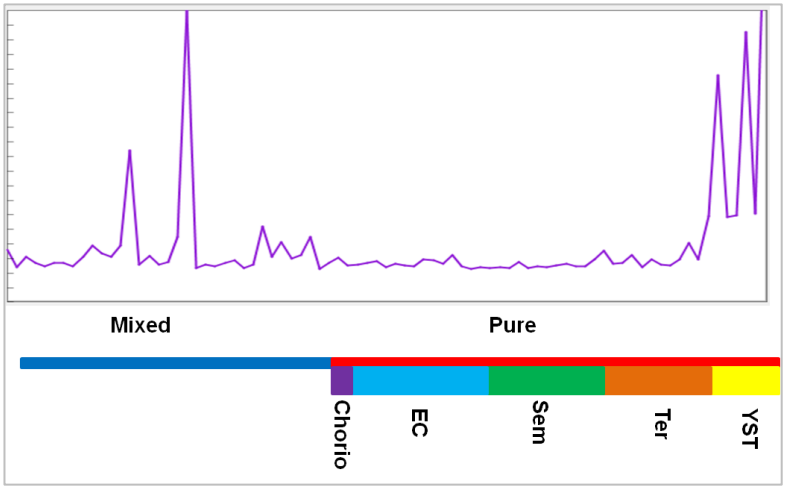


Figure 4.39. The 'liver' signature is expressed only in tumour with a yolk sac component. Mean expression profile of all genes in the signature. Individual tumours are represented on the x-axis with the coloured bar denoting tumour groupings. Mixed tumours lie to the left and pure histological types to the right. (Key; Chorio = Choriocarcinoma, EC = Embryonal carcinoma, Sem = Seminoma, Ter = Teratoma, YST = Yolk Sac Tumour)

Analysis tool	Source of annotation	Annotation	Significance (p-value)
GSEA	HSIAO_LIVER SPECIFIC GENES	Liver selective genes	2.9×10^{-50}
DAVID	GOTERM_BP_FAT	Haemostasis	5.3×10^{-7}
GSEA	REACTOME	Genes involved in lipoprotein metabolism	1.2×10^{-15}

Table 4.61 Functional annotation for the 'liver' signature using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

There are three placental clusters (155 probesets and 103 genes, 86 probesets and 63 genes and 55 probesets to 31 genes) (Figure 4.40). These are show expression only in a small number of mixed tumours that contained elements of choriocarcinoma, cases of pure choriocarcinoma and one case labelled pure embryonal carcinoma that presumably contains a component of choriocarcinoma not identified in the tissue selected for histological examination and classification and hence mis-classified. The first cluster contains genes such as ABP1 (amiloride binding protein 1, which is expressed at the feto-maternal interface in humans), and placental growth factor (PGF) and is not obviously enriched in placental specific genes simply by reading the gene list. Comparison with a tissue atlas and DAVID and GSEA analysis confirms however that there is significant enrichment for placentally expressed genes (Table 4.62).

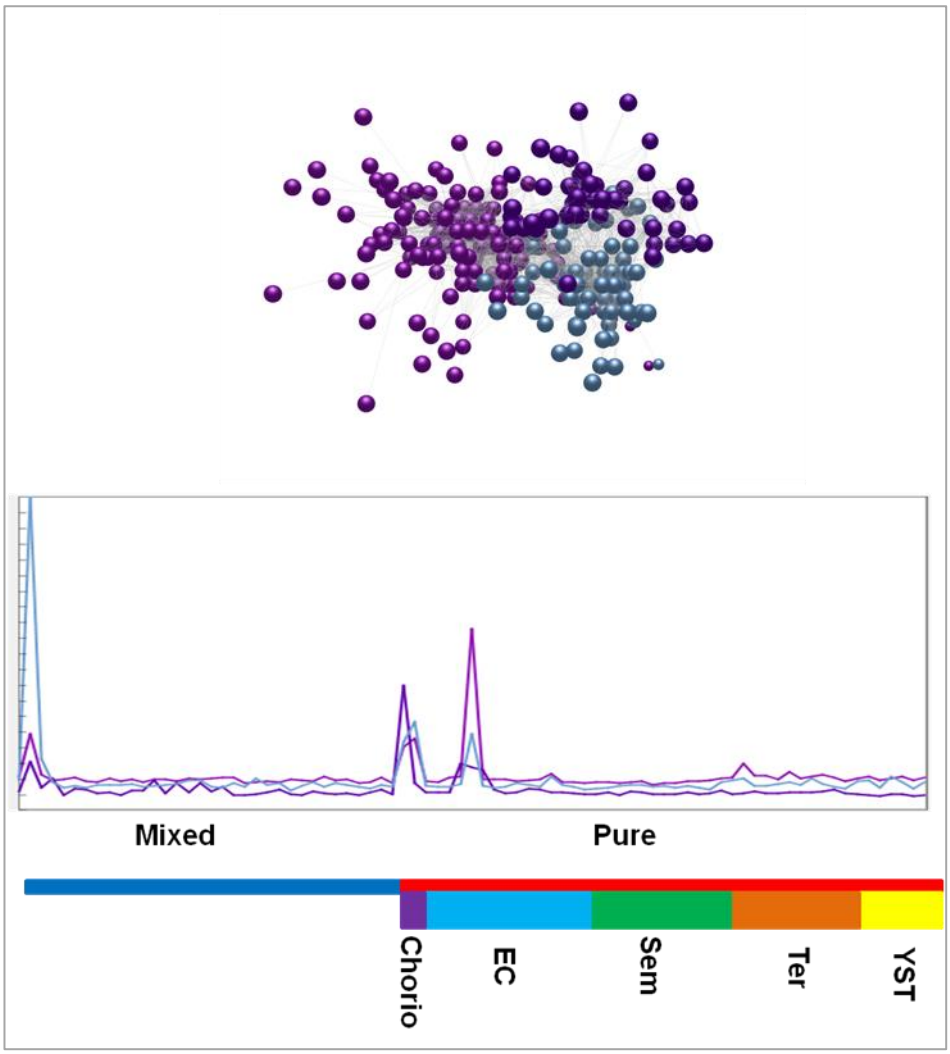


Figure 4.40. Three closely related placental signatures seen in tumours with a choriocarcinamatous component. These form a group three closely related clusters. Top panel: Network graph with all other clusters hidden. Bottom panel: Mean expression profile of all genes in each signature. Colours of clusters correspond to colours of expression plots. Individual tumours are represented on the x-axis with the coloured bar denoting tumour groupings. Mixed tumours lie to the left and pure histological types to the right. (Key; Chorio = Choriocarcinoma, EC = Embryonal carcinoma, Sem = Seminoma, Ter = Teratoma, YST = Yolk Sac Tumour).

The second placental cluster much more obviously contains placental specific genes, including pregnancy specific beta-1 glycoproteins 1,2,3,4,5,6,7 and 9, prolactin receptor, chorionic gonadotrophin beta polypeptide, placental insulin-like 4 luteinizing hormone beta polypeptide and placenta specific 4 genes (PSG1, PSG2, PSG3, PSG4, PSG5, PSG7, PSG9, PRLR, CGB, INSL4, LHB, PLAC4). The third

placental cluster consists of chorionic somatomammotrophin hormones and the closely related growth hormones (CSH1, CSH2, CSHL1, GH1, GH2).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
	Placental cluster 1		
GSEA	STANFORD_TUMOUR_MODULES	Placental genes	6.8×10^{-15}
GSEA	SU_PLACENTA	Genes upregulated specifically in human placenta	5.5×10^{-11}
	Placental cluster 2		
DAVID	GOTERM_BP_FAT	Female pregnancy	5.6×10^{-9}
GSEA	STANFORD_TUMOUR_MODULES	Placental genes	3.3×10^{-14}
	Placental cluster 3		
GSEA	REACTOME	Growth hormone receptor signalling	4.5×10^{-7}

Table 4.62 Functional annotation for the placental signatures using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

4.9 Features common to all datasets

In all six primary datasets studied, the analysis method generated biologically believable clusters in which functional enrichment of genes could be demonstrated. Having validated the approach in terms of its ability to identify genes expressed in specific cell types and functional processes within a given dataset, further work was performed to assess how these signatures were preserved across the different tumour types studied here with the assumption being that the elements which should be in common between tumours of different histogenesis should either be elements essential to tumour growth regardless of cell of origin, i.e. genes fundamental to replicating cells, or representative of the less variable elements of the tumour i.e. the non-malignant cells within the tumour. For instance is the immune signature in DLBCL the same as the signature in gliomas, or is the profile of the immune infiltrate in different environments highly heterogeneous, such that there is little overlap between these signatures? Or do all tumours have conserved signatures of cells or functions that exist regardless of the tissue from which they arise i.e. is there a core common cancer transcriptome?

Analysis of each tumour type's transcriptome demonstrated the clusters generated to fall into two broad groups; those which contain signatures which are specific to origin of the tumour and those with signatures that are conserved across all or most datasets. The common or conserved signatures tended to fall into the broad categories of stromal cell signatures, proliferation related signatures and 'house-keeping' signatures. There were clusters in every dataset which represented signatures of various immune cells and the interferon response. The extracellular matrix and other stromal signatures including the vasculature were represented in every dataset and many contained other stromal signatures dependent on the make-up of the individual tissue sampled. Cell-cycle and cell cycle related clusters were present in every dataset as were less biologically interesting clusters such as clusters representing ribosomes, histones, haemoglobin etc. In all datasets there were preserved clusters that were designated as 'house-keeping' clusters in the absence of informative annotation of the genes involved in those clusters. This suggested that

there were common ‘core’ signatures present that defined for example, the tumour associated macrophage, regardless of the histogenesis of the tumour. Comparison of clusters between datasets to identify conserved signatures proved difficult; genes present in all datasets were readily identifiable but for those only in a proportion of datasets there was the possibility they lay immediately adjacent in the network graph but did not fall within the cluster, thus they may appear falsely ‘absent’ from the cluster when they were in effect ‘almost present’. Individual genes could be analysed by this method where the exact position (or absence of the gene from the network graph) could be established but it was not suitable for analysis of the whole dataset. To overcome these difficulties a different approach was developed which is presented in the next section.

4.10 Gene signatures from merged datasets

To determine to what extent the signatures seen in all tumour types were conserved, a method was developed that allowed merging of the data from the six datasets discussed above. Normally data from different labs, run under different conditions, even as in this case data on the same platform, cannot be directly compared as technical differences affecting measured fluorescence and the effect of normalisation will make direct comparison of expression values impossible. To overcome this problem Pearson correlations generated for analysis of individual datasets were used to construct a network based on mean probeset to probeset correlations. This preserves the co-expression associations, which are what drive the construction of the transcriptome network, while overcoming the difficulties involved in attempting to merge values. The inference was made that signatures which persisted when treated in this way represented conserved signatures as weaker correlations or tumour specific correlations would tend to fall below the selected correlation threshold.

Layout of a graph derived from the mean Pearson values ($r = \geq 0.6$) across the six different tumour types resulted in a network of 9,882 nodes connected by 134,563 edges (Figure 4.41), a much smaller graph than those derived from individual tumours at this threshold. Clustering of the graph using the MCL algorithm resulted in 639 clusters ranging in size from 1008 to 4 nodes (minimum cluster size set to 4).

When analysed in this way a number of clusters were again shown to be highly enriched with genes associated with expression in individual cell types and/or associated with specific cellular functions and were annotated as such . A full list of clusters and the genes that they contain is given in appendix 1.

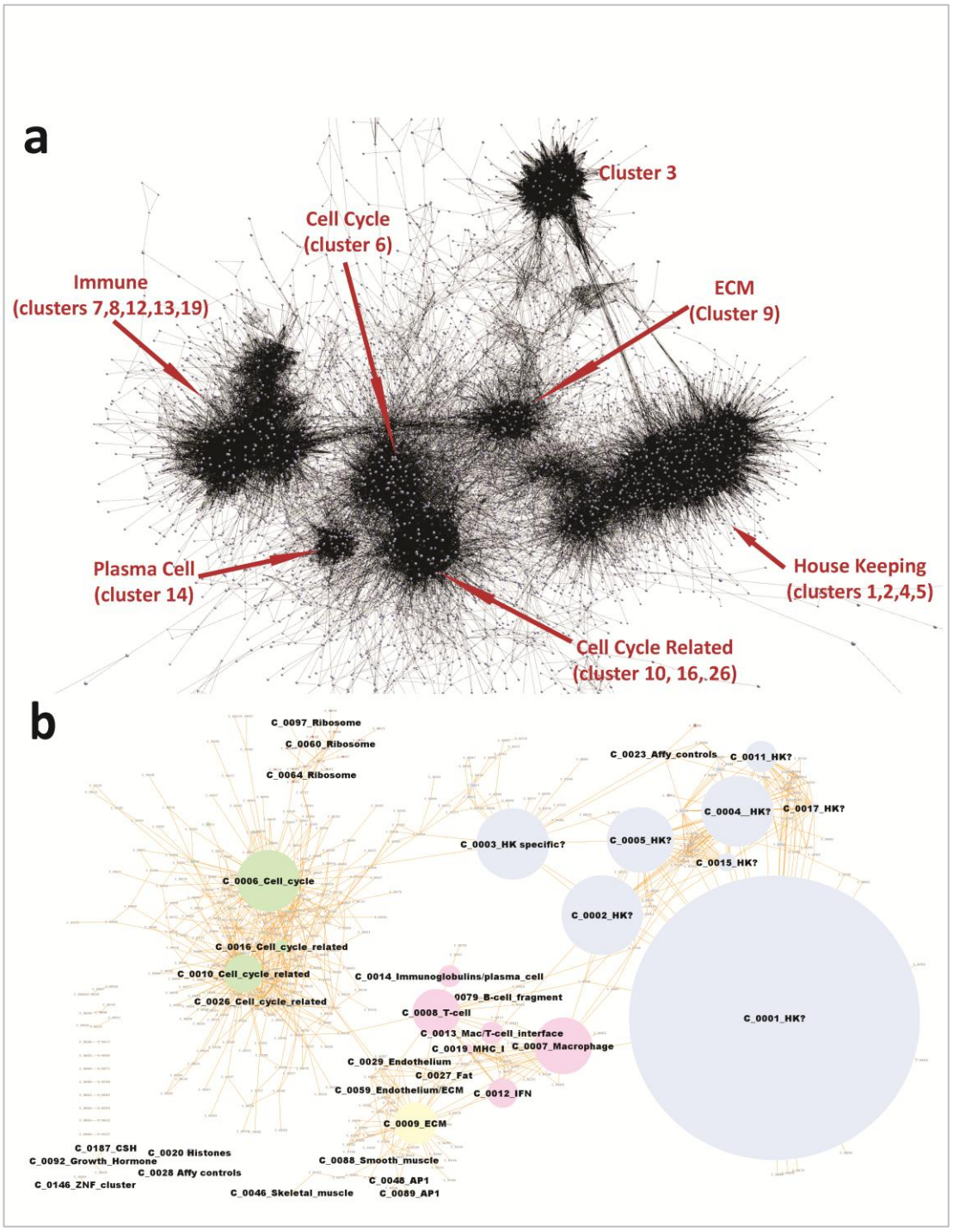


Figure 4.41 Network graph of conserved transcription signatures in cancer. **a)** 3D graph layout with labelling of main features in the network's topology. A graph of 9,882 nodes and 184,563 edges was created at Pearson threshold r≥0.6. Clustering of the graph using the MCL algorithm resulted in 639 clusters ranging in size from 1,008 nodes to 4 nodes. A number of large clusters were shown to be highly enriched in genes associated with expression in individual cell types and/or associated with specific cellular functions. See appendix 1 for details. **b)** Collapsed cluster diagram where single nodes represent a gene cluster and are sized according to the number of transcripts within the cluster, edges represent connections between members of each cluster. Nodes representing the main clusters have been coloured according to functional groupings. Blue - clusters represent those associated with housekeeping functions; green - clusters of genes which are directly involved in cell cycle progression or whose expression is in way some linked to it; pink - genes associated with the immune component of the tumour and yellow - other stromal elements. Smaller clusters enriched with genes of known function are also shown.

As predicted, the signatures identified by this approach were those that were observed in all or most datasets, with loss of those that were tumour-type or tumour-tissue type-specific. The conserved signatures consisted predominantly of the three categories discussed briefly above; housekeeping clusters, pathways of common cellular activity (i.e. cell cycle, ribosome etc), and stromal signatures. This approach allows analysis of the 'core' of the cancer transcriptome; those signatures that are conserved regardless of the histogenesis of the tumour.

The overall structure of the graph could be broken down into three main groupings which draw together clusters enriched in genes representing broadly similar functions. The largest grouping of genes within the graph contained four of the five largest clusters being comprised of between 1,101 and 254 nodes/transcripts as well as many other smaller clusters. Many of the genes in this area of the graph are poorly characterised and GO enrichment and pathway analysis was of little use in better defining the functional significance of the clusters. However, the majority of the genes in these clusters were relatively uniformly expressed across all samples in all datasets and were therefore designated "house-keeping" (HK) clusters. (Cluster 3, containing genes that clustered together tightly in all datasets and which appears to be a very well-conserved signature fell apart from the other HK clusters, but was also designated an HK cluster in the absence of meaningful annotation).

A second portion of the graph was highly enriched with genes encoding cell-cycle or cell-cycle related proteins. Cluster 6 (229 transcripts) of this grouping being highly enriched in genes associated with mitosis and are commonly found to be highly expressed in proliferating cells. This cluster contains multiple cyclins (CCNA2, CCNB1, CCNB2, CCNE2, CCNF), kinesins (KIF2C, KIF4A, KIF11, KIF14, KIF15, KIF18A, KIF18B, KIF20A, KIF20B, KIF22, KIF23, KIFC1) and minichromosome maintenance complex components (MCM2, MCM3, MCM4, MCM5, MCM6, MCM7, MCM10). There is also expression of members of the E2F transcription factor family (E2F1, E2F8), DNA polymerases (DNA2, POLA2, POLE, POLE2, POLQ) and topoisomerase (TOP2A). The Aurora kinases (AURKA, AURKB), BUB1 and both CHEK1 and 2 checkpoint proteins are also present along with many others. (See appendix 1 for full list). Selected functional annotation confirming the enrichment for cell cycle genes is given in table 4.63.

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Cell cycle	2.3×10^{-74}
DAVID	GOTERM_BP_FAT	M-phase	1.2×10^{-66}
DAVID	GOTERM_BP_FAT	DNA replication	3.4×10^{-44}
GSEA	BENPORATH_CYCLING GENES	Genes showing cell cycle stage specific expression	4.1×10^{-162}
GSEA	STANFORD TUMOUR MODULES	Cell cycle	6.9×10^{-195}

Table 4.63 Functional annotation for the cell cycle signature from the merged datasets using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

The cell cycle cluster is also associated with clusters of genes whose expression would appear to be tied to it (clusters 10, 16 & 26) and involved in mRNA/protein processing. Associated with the cell cycle cluster are further smaller clusters enriched in mitochondrial, ribosomal and glycolysis-related genes which were designated cell-cycle related gene clusters due to their close proximity to cluster 6. It is perhaps unsurprising that signatures relating to replication are shared across multiple tumours, given the proliferative nature of tumour cells. The presence of such a well-preserved set of cell cycle genes argues for the fundamental importance of the cell cycle in malignant cells, such that although elements of regulation may be missing in individual tumours, taken as an average across multiple tumour types the signature is well preserved. Due to the method of analysis used, which will tend to preserve the signatures contributed by normal cells, it is reasonable to assume that part of this signature is contributed by replicating non-malignant cells within the tumour, but when this signature is examined in individual tumour type, there is a relationship with more aggressive tumours, arguing that this signature is derived mostly from the tumour cells.

Finally, a third area of the graph was enriched with genes whose expression is associated with different elements of the tumour stroma. Within this component the clusters associated with tumour-associated immune activity are in close proximity with each other, those representing other stromal components being somewhat more distant.

The macrophage signature (cluster 7, 220 transcripts) contains many genes considered to be specific to the myeloid lineage including CD68, CD14, CD163 and CSF1R. There is enrichment for lysosomal genes (CD68, LAPTM5, SCL15A3, ACP2, CTSB, CTSC, CTSD, CTSL1, CTSS), multiple genes involved in chemotaxis are also present (CMKLR1, PTAFR, FCER1G, C5AR1, C3AR1, CCR1, CCRL2, ITGAM, ITGB2, FPR1), as are multiple toll-like receptors (TLR1, TLR2, TLR7, TLR8). The signature also contains scavenger receptors CD163, MARCO, MSR1, as has previously been described by many groups for tumour-associated macrophages

(reviewed in Allavena et al, 2010). Finally, within the macrophage cluster are multiple components of the MHC class II antigen processing machinery (HLA-DMA, HLA-DMB, HLA-DOA, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB1, HLA-DRA, HLA-DRB1, HLA-DRB4, HLA-DRB5, HLA-DRB6, CD74, LGMN). Interestingly among the genes also expressed is CD86, the co-stimulatory molecule suggesting that these cells may be able to efficiently present antigen to T cells. The presence of genes involved in phagocytosis, MHC class II antigen processing and T cell co-stimulation all suggest a macrophage able to act as an antigen-presenting cell. The presence of scavenger receptors and genes involved in lipid metabolism hint at a role in apoptotic cell clearance by TAMS.

A full list of genes from the macrophage signature and selected enrichment analysis is provided below (Tables 4.64 and 4.65). Other signatures are provided in appendix 1

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Defense response	8.5×10^{-26}
DAVID	GOTERM_BP_FAT	Antigen processing and presentation via MHC II	6.1×10^{-19}
DAVID	SP_PIR_KEYWORDS	Lysosome	2.5×10^{-12}
GSEA	GSE 10325	Genes downregulated in comparison of healthy B cells with healthy myeloid cells	2.8×10^{-67}
GSEA	STANFORD TUMOUR MODULES	Immune and inflammatory response	1.5×10^{-108}

Table 4.64 Functional annotation for the macrophage signature from the merged datasets using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

Table 4.65 Macrophage signature from merged dataset laid out at Pearson correlation ≥ 0.6 with MCL inflation values of 2.2.

Gene Symbol	Description
ACP2	acid phosphatase 2, lysosomal
ADAMDEC1	ADAM-like, decysin 1
ADAP2	ArfGAP with dual PH domains 2
ADAP2	ArfGAP with dual PH domains 2
ADORA3	adenosine A3 receptor
AIF1	allograft inflammatory factor 1
AOAH	acyloxyacyl hydrolase (neutrophil)
APOL1	apolipoprotein L, 1
APOL2	apolipoprotein L, 2
APOL3	apolipoprotein L, 3
ARRB2	arrestin, beta 2
ATP8B4	ATPase, class I, type 8B, member 4
BCL2A1	BCL2-related protein A1
C1orf38	chromosome 1 open reading frame 38
C1orf54	chromosome 1 open reading frame 54
C1QA	complement component 1, q subcomponent, A chain
C1QB	complement component 1, q subcomponent, B chain
C2	complement component 2
C3AR1	complement component 3a receptor 1
C5AR1	complement component 5a receptor 1
CAPG	capping protein (actin filament), gelsolin-like
CCR1	chemokine (C-C motif) receptor 1
CCRL2	chemokine (C-C motif) receptor-like 2
CD14	CD14 molecule
CD163	CD163 molecule
CD300A	CD300a molecule
CD4	CD4 molecule
CD68	CD68 molecule
CD74	CD74 molecule, major histocompatibility complex, class II invariant chain
CD84	CD84 molecule
CD86	CD86 molecule
CD97	CD97 molecule
CECR1	cat eye syndrome chromosome region, candidate 1
CHIT1	chitinase 1 (chitotriosidase)
CIITA	class II, major histocompatibility complex, transactivator
CLEC10A	C-type lectin domain family 10, member A
CLEC2B	C-type lectin domain family 2, member B
CLEC7A	C-type lectin domain family 7, member A
CMKLR1	chemokine-like receptor 1
CSF1R	colony stimulating factor 1 receptor

CSF2RB	colony stimulating factor 2 receptor, beta, low-affinity (granulocyte-macrophage)
CTSB	cathepsin B
CTSC	cathepsin C
CTSD	cathepsin D
CTSL1	cathepsin L1
CTSS	cathepsin S
CYBB	cytochrome b-245, beta polypeptide
CYTH4	cytohesin 4
DPYD	dihydropyrimidine dehydrogenase
DRAM	damage-regulated autophagy modulator
DSE	dermatan sulfate epimerase
EMR2	egf-like module containing, mucin-like, hormone receptor-like 2
FCER1G	Fc fragment of IgE, high affinity I, receptor for; gamma polypeptide
FCGR1A	Fc fragment of IgG, high affinity Ia, receptor (CD64)
FCGR1B	Fc fragment of IgG, high affinity Ib, receptor (CD64)
FCGR2A	Fc fragment of IgG, low affinity IIa, receptor (CD32)
FCGR3B	Fc fragment of IgG, low affinity IIIb, receptor (CD16b)
FCGR3B	Fc fragment of IgG, low affinity IIIb, receptor (CD16b)
FGL2	fibrinogen-like 2
FGR	Gardner-Rasheed feline sarcoma viral (v-fgr) oncogene homolog
FPR1	formyl peptide receptor 1
FPR3	formyl peptide receptor 3
GBP2	guanylate binding protein 2, interferon-inducible
GCH1	GTP cyclohydrolase 1
GPNMB	glycoprotein (transmembrane) nmb
GPR65	G protein-coupled receptor 65
GPX1	glutathione peroxidase 1
HCK	hemopoietic cell kinase
HK3	hexokinase 3 (white cell)
HLA-DMA	major histocompatibility complex, class II, DM alpha
HLA-DMB	major histocompatibility complex, class II, DM beta
HLA-DOA	major histocompatibility complex, class II, DO alpha
HLA-DPA1	major histocompatibility complex, class II, DP alpha 1
HLA-DPB1	major histocompatibility complex, class II, DP beta 1
HLA-DQA1	major histocompatibility complex, class II, DQ alpha 1
HLA-DQB1	major histocompatibility complex, class II, DQ beta 1
HLA-DRA	major histocompatibility complex, class II, DR alpha
HLA-DRA	major histocompatibility complex, class II, DR alpha
HLA-DRB1	major histocompatibility complex, class II, DR beta 1
HLA-DRB4	major histocompatibility complex, class II, DR beta 4
HLA-DRB5	major histocompatibility complex, class II, DR beta 5
HLA-DRB6	major histocompatibility complex, class II, DR beta 6 (pseudogene)
HS3ST2	heparan sulfate (glucosamine) 3-O-sulfotransferase 2
IFI30	interferon, gamma-inducible protein 30

IGSF6	immunoglobulin superfamily, member 6
IL15RA	interleukin 15 receptor, alpha
IL18BP	interleukin 18 binding protein
ITGAM	integrin, alpha M (complement component 3 receptor 3 subunit)
ITGAX	integrin, alpha X (complement component 3 receptor 4 subunit)
ITGB2	integrin, beta 2 (complement component 3 receptor 3 and 4 subunit)
LACTB	lactamase, beta
LAIR1	leukocyte-associated immunoglobulin-like receptor 1
LAIR1	leukocyte-associated immunoglobulin-like receptor 1
LAPTM5	lysosomal multispanning membrane protein 5
LCP2	lymphocyte cytosolic protein 2 (SH2 domain containing leukocyte protein of 76kDa)
LGALS9	lectin, galactoside-binding, soluble, 9
LGMN	legumain
LHFPL2	lipoma HMGIC fusion partner-like 2
LILRB1	leukocyte immunoglobulin-like receptor, subfamily B (with TM and ITIM domains), member 1
LILRB2	leukocyte immunoglobulin-like receptor, subfamily B (with TM and ITIM domains), member 2
LILRB4	leukocyte immunoglobulin-like receptor, subfamily B (with TM and ITIM domains), member 4
LIPA	lipase A, lysosomal acid, cholesterol esterase
LST1	leukocyte specific transcript 1
LY86	lymphocyte antigen 86
LY96	lymphocyte antigen 96
MAFB	v-maf musculoaponeurotic fibrosarcoma oncogene homolog B (avian)
MAN2B1	mannosidase, alpha, class 2B, member 1
MARCO	macrophage receptor with collagenous structure
MFSD1	major facilitator superfamily domain containing 1
MGAT1	mannosyl (alpha-1,3-)-glycoprotein beta-1,2-N-acetylglucosaminyltransferase
MNDA	myeloid cell nuclear differentiation antigen
MPP1	membrane protein, palmitoylated 1, 55kDa
MS4A4A	membrane-spanning 4-domains, subfamily A, member 4
MS4A6A	membrane-spanning 4-domains, subfamily A, member 6A
MS4A7	membrane-spanning 4-domains, subfamily A, member 7
MSR1	macrophage scavenger receptor 1
MYO1F	myosin IF
NAGK	N-acetylglucosamine kinase
NCF2	neutrophil cytosolic factor 2
NCKAP1L	NCK-associated protein 1-like
NPL	N-acetylneuraminate pyruvate lyase (dihydrodipicolinate synthase)
NR1H3	nuclear receptor subfamily 1, group H, member 3
P2RY13	purinergic receptor P2Y, G-protein coupled, 13
PILRA	paired immunoglobin-like type 2 receptor alpha
PLA2G7	phospholipase A2, group VII (platelet-activating factor acetylhydrolase, plasma)
PLCB2	phospholipase C, beta 2

PLEK	pleckstrin
PLEKHO2	pleckstrin homology domain containing, family O member 2
PTAFR	platelet-activating factor receptor
RHOG	ras homolog gene family, member G (rho G)
RNASE6	ribonuclease, RNase A family, k6
RNF130	ring finger protein 130
SCO2	SCO cytochrome oxidase deficient homolog 2 (yeast)
SCPEP1	serine carboxypeptidase 1
SECTM1	secreted and transmembrane 1
SIGLEC1	sialic acid binding Ig-like lectin 1, sialoadhesin
SIGLEC7	sialic acid binding Ig-like lectin 7
SLAMF8	SLAM family member 8
SLC15A3	solute carrier family 15, member 3
SLC31A2	solute carrier family 31 (copper transporters), member 2
SLC7A7	solute carrier family 7 (cationic amino acid transporter, y+ system), member 7
SLCO2B1	solute carrier organic anion transporter family, member 2B1
SNX10	sorting nexin 10
SPI1	spleen focus forming virus (SFFV) proviral integration oncogene spi1
SRGN	serglycin
TBXAS1	thromboxane A synthase 1 (platelet)
TCIRG1	T-cell, immune regulator 1, ATPase, H+ transporting, lysosomal V0 subunit A3
TFEC	transcription factor EC
TLR1	toll-like receptor 1
TLR2	toll-like receptor 2
TLR7	toll-like receptor 7
TLR8	toll-like receptor 8
TM6SF1	transmembrane 6 superfamily member 1
TMEM140	transmembrane protein 140
TNFAIP2	tumor necrosis factor, alpha-induced protein 2
TNFRSF14	tumor necrosis factor receptor superfamily, member 14 (herpesvirus entry mediator)
TNFRSF1B	tumor necrosis factor receptor superfamily, member 1B
TNFSF13B	tumor necrosis factor (ligand) superfamily, member 13b
TREM2	triggering receptor expressed on myeloid cells 2
TRPV2	transient receptor potential cation channel, subfamily V, member 2
TYMP	thymidine phosphorylase
TYROBP	TYRO protein tyrosine kinase binding protein
VAMP5	vesicle-associated membrane protein 5 (myobrevin)
VSIG4	V-set and immunoglobulin domain containing 4

The T cell signature (cluster 8, 181 transcripts) contains pan-T cell markers (CD3, CD2, CD7) and elements of the T cell receptor signalling cascade (ZAP70, LCK, VAV, ITK). There are many chemokines, cytokines and their receptors in the cluster (CXCL9, CCL19, CCL5, LTB, CXCR3, CXCR6, CCR7, CCR2, CCR5, IL2RB, IL2RG, IL17R, IL10RA). Interferon gamma, the prototypical 'classical' macrophage activator is also expressed in this cluster. The T-cell signature is suggestive of an active state with expression of cytotoxic molecules granzymes and perforin (GZMA, GZMH, GZMK, GZMM, PRF1) as well as markers of activation (CD69). Selected enrichment analysis given in table 4.66

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	T cell activation	1.7×10^{-22}
DAVID	KEGG PATHWAY	T cell receptor signalling	1.1×10^{-12}
GSEA	GSE 22886	Genes upregulated in naive CD8 cells vs monocytes	7.7×10^{-60}

Table 4.66 Functional annotation for the T cell signature from the merged datasets using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

Lying adjacent to the T cell and macrophage clusters there is the signature of an interferon response (cluster 12, 115 transcripts) containing elements of the proteasome and multiple interferon regulatory factors (IRF1, IRF2, IRF7 & IRF9) and interferon inducible proteins (IFI6, IFI27, IFI35, IFI44). See table 4.67 for selected functional annotation.

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Response to virus	2.1×10^{-21}
GSEA	HECKER_IFNB_TARGETS	Genes transcriptionally modulated in the blood of MS patients treated with subcut IFNB	2.8×10^{-83}
GSEA	BROWNE_INTERFERON RESPONSIVE GENES	Genes upregulated in primary fibroblast culture after treatment with IFN alpha	6.6×10^{-72}

Table 4.67 Functional annotation for the interferon response signature from the merged datasets using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

The preservation across all tumour types of a cluster of genes associated with an interferon response and the presence of *IFNG* in the T cell signature suggests that activation of this pathway forms a consistent part of the response to a tumour. This is also in keeping with data derived from murine models in which it was shown that TAMs express many interferon inducible genes (Biswas et al., 2006). Taken together, these data do not entirely support the view that TAMs have a so-called M2 (or alternative activation) (Mantovani et al 2009) phenotype characterised by dominant actions of interleukin 4; there is clear evidence of M1-type activation involving type I and type II interferons. Nor does the analysis support the view that recognition of tumour-associated antigens is compromised by a lack of antigen-presenting cells. Though it should be noted, that as with any study based on gene expression signatures, the signatures derived represent an average cell within the tissue and cannot address the differences that may be seen in different areas of any tumour.

The largest non-immune-related element of the stromal signature is a cluster of genes that are associated with extracellular matrix (cluster 9, 163 transcripts). As such the cluster appears to represent the fibroblast/myofibroblast component of the tumours and contains many structural proteins including collagens (COL1A1, COL1A2, COL3A1, COL5A1, COL5A2, COL6A1, COL6A2, COL6A3, COL11A1, COL12A1) as well as cadherins (CDH11), laminins (LAMA4, LAMB1), fibrillin (FBN1) and integrins (ITGA5, ITGAV, ITGB1). The signature also contains modifiers of the extracellular matrix such as MMP2, LOXL1, ADAMTS12, ADAMTS2 and receptors for growth factors (PDGFRB). See table 4.68 for selected functional annotation.

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	SP_PIR_KEYWORDS	Extracellular matrix	3.2×10^{-41}
DAVID	GOTERM_BP_FAT	Cell adhesion	2.2×10^{-20}
GSEA	NABA_CORE_MATRISOME	Genes encoding core ECM proteins	3.8×10^{-63}

Table 4.68 Functional annotation for the extracellular matrix signature from the merged datasets using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

Histologically, the presence of a desmoplastic tumour stroma is a well recognised phenomenon occurring in many tumour types. However, like many other elements of the microenvironment the precise role played by this reactive stroma has been difficult to assess: is the role of the stroma to contain the tumour or is it yet another factor recruited to promote the survival of the malignant cells? Recent data from studies of DLBCL suggest that in this tumour at least the answer may be that different elements of the stroma contribute to both a good and a poor prognosis (Lenz et al., 2008a) whereas work in small cell carcinoma has established the role of interactions between the ECM and tumour cells in resisting chemotherapy-induced death (Hodkinson et al., 2006). More recently work by Kim et al. (2009) in a lung

carcinoma model highlighted the role that ECM components, in this case versican, can play in activating other elements of the microenvironment suggesting that as for other elements of the tumour microenvironment, cross talk between elements is likely to be of great importance.

The vascular signature fragments into four closely aligned clusters, three of which (clusters 29, 38, and 49, 50 transcripts in total) appear to represent endothelium and the fourth (cluster 59) associated with the basement membrane/extracellular matrix component (Table 4.69). These clusters contain classical and well characterised markers of vascular differentiation such as CD34, PECAM1 (CD31), VWF, KDR and CDH5. In addition it contains many genes that have been identified as endothelial specific genes by alternative bioinformatic analysis approaches (ECSCR, EMCN, ROBO4, TEK, EPAS1, GPR116), components of the Notch signalling pathway (NOTCH3, NOTCH4) and other endothelial genes which have been demonstrated in normal and tumour associated endothelium such as PLVAP (plasmalemmal vesicle associated protein) (Ghilardi et al 2008, Herbert et al 2008, Huminieki and Bicknell 2000, Wallgard et al 2008, Strickland et al 2005). It is worth noting that these alternative bioinformatic approaches to determining endothelial signatures used either purified cell populations for gene expression data or characterised profiles based on similarity to known markers in contrast to the approach presented here which uses unpurified tumour material and an unsupervised clustering approach. This analysis also identifies genes not previously identified by a bioinformatic approach, such as PLVAP (plasmalemmal vesicle associated protein), which has been well characterised as an endothelial gene expressed in certain normal tissues and in some tumours (Strickland et al 2005). Expression of this protein is associated with diaphragm formation in fenestrated endothelium and has been reported in tumour endothelium in a wide spectrum of tumours. In our data, Notch 4, which is implicated in endothelial sprouting, lies in one of the endothelial clusters, while Notch 3, the causative gene for CADASIL (cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy) and which is expressed in vascular smooth muscle lies in the related cluster enriched in ECM and basement membrane proteins. A recent study has investigated the crosstalk between

endothelial and mural cells via Notch 3 signalling and shown a reduction in angiogenesis in an *in vitro* co-culture system when Notch 3 is knocked out in mural cells (Liu et al 2009).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Blood vessel development	1.6×10^{-11}
DAVID	SP_PIR_KEYWORDS	Angiogenesis	1.1×10^{-10}

Table 4.69 Functional annotation for the endothelial signatures from the merged datasets using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

Finally there is a small cluster that contains many adipocyte specific genes (cluster 27, 22 transcripts) including ADH1B, ADIPOQ, FABP4 and LPL.

Other small clusters or groupings of small clusters of note are comprised of the Affymetrix control probes (cluster 23 contains all hybridization controls and cluster 28 all 'spike-in' control probes); histone complexes (cluster 20 contains 24 histone genes); AP1 transcription factors/early response genes (clusters 48, 89 and 336) containing the genes JUN, JUNB, FOS, EGR1, EGR3, IER2, NR4A1, NR4A2, ATF3, CTGF and DUSP1; and growth-related hormones GH1, CSH1, CSH2, CSHL2 (clusters 92 and 187).

It is of interest to consider how the different signatures relate to each other. The fact that macrophage, T cell, ECM and endothelial-specific genes cluster independently of itself indicates that there is not a tight causal relationship between them. However it is clear that the functions of these cells may be closely related, with the T cell signature and macrophage signature in particular, demonstrating a similar expression pattern and close spatial relationship in the network graph. The exact nature of the relationship between these elements of the stroma may differ between individual

tumours with the vascular signature lying in the same cluster as the T cell signature in some tumours, arguing a very close pattern of co-expression of these genes. While the expression pattern is correlated is not possible to be sure of the exact nature of the relationship; do factors secreted by lymphocytes drive angiogenesis?, or does the presence of the vasculature allow lymphocytes to reach and infiltrate the tissue?, or are both of these signatures driven by a separate, closely associated signatures such as that contributed by the tumour associated macrophage. This analysis, providing a single 'snapshot' view of the individual tumours, does not allow this question to be answered. The absence of a conserved pattern of tight association between signatures however argues that these relationships are not absolute and are influenced by many factors which are likely to be of variable strength in different tumour types.

Despite the frequent association of macrophage number with microvessel density in solid tumours (Sickert et al 2007, Lissbrant et al 2000, Eerola et al 1999, Orre and Rogers 1999), these signatures are clearly separate in the tumours studied here suggesting that there are other factors involved in the recruitment of neovasculature to the tumour in addition to the role played by TAMs.

4.11 Conservation of 'core' signatures in independent dataset of skin cancers

In order to confirm that the 'core' signatures generated from the 6 datasets are generally applicable to cancer datasets rather than specific to the tumours studied, an independent tumour dataset, not used in derivation of the core signatures was analysed and the signatures generated from our merged analysis mapped to this. This final dataset of 82 samples was derived from various types of skin cancer including basal cell carcinoma, squamous cell carcinoma, primary melanoma, including melanoma in-situ, and melanoma metastatic to subcutaneous tissue, lymph node, brain and adrenal gland (Riker et al 2008).

This graph was laid out at Pearson correlation threshold ≥0.8 giving a graph of 21,293 nodes and 601,710 edges (Figure 4.42). As with each of the previous datasets, this dataset had its own peculiarities relating to the individual tumour types, with many of the clusters representing genes with marked differences in expression between the three different tumour types and clusters with a tissue specific expression, in this case skin, adrenal gland and other sites of metastasis. The data tended to show a spiked pattern, reflecting the relatively small number of tumours, a large number of which were metastatic melanoma. In addition however, there was preservation of many of the signatures seen in the 'core' dataset, with clusters present in this new dataset which were highly significantly enriched for genes of the 'core' signatures. As in other datasets the spatial relationship between clusters was preserved with immune signatures falling in one part of the graph, non-immune stromal clusters lying together and cell cycle and related genes clustering together. On analysis of some of the individual expression patterns seen in the skin cancer data set, clusters of genes representing macrophages and lymphocytes showed increased expression in melanoma compared to BCC and SCC in keeping with the inflammatory infiltrate frequently seen in these tumours.

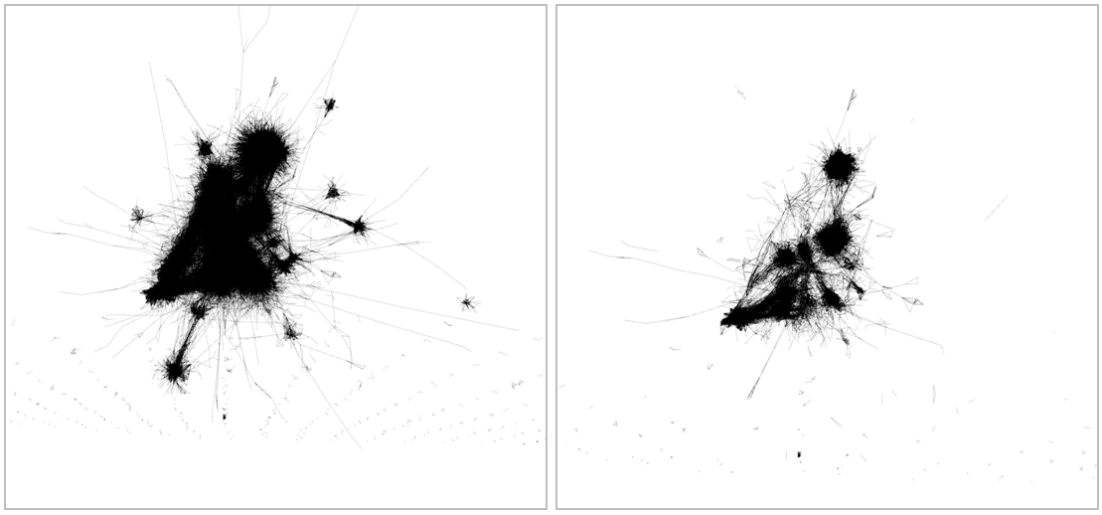


Figure 4.42. Clusters present in the merged signature were conserved in the skin dataset. Left panel: The skin dataset laid out at p≥0.8 with nodes hidden and only edges visible. Right: The same network graph but with only those probesets that formed part of a cluster in the merged signature shown.

4.11.1 Stromal signatures

As in other datasets, there was a cluster representing the tumour associated macrophage (Figure 4.43). This was of only 110 probesets to 62 genes and contained markers of macrophages including CD14 and CD163. There are toll-like receptors 2 and 4, macrophage scavenger receptor 1 and single Fc receptor (CD64). Multiple genes involved in chemotaxis are present (C3AR1, CCL13, CCR1, CMKLR1, FPR3, ITGAM) as well as genes frequently seen in the macrophage signature in other datasets including AIF1, HCK, LAIR1, and TNFSF13. Enrichment analysis via DAVID and GSEA confirms this to represent a macrophage signature (Table 4.70).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Defense reponse	1.0×10^{-12}
DAVID	GOTERM_BP_FAT	Chemotaxis	2.6×10^{-4}
GSEA	GSE 10325	Genes downregulated in healthy B cells vs myeloid cells	2.6×10^{-41}
GSEA	GSE 29618	Genes upregulated in monocytes vs myeloid DCs	9.2×10^{-35}

Table 4.70 Functional annotation for the macrophage signature from the skin dataset using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

A further large cluster of 368 probesets representing 245 genes contains lymphocyte genes (Figure 4.43). This contains surface markers of both B and T cells including CD22, CD79a and CD79b, MS4A1 (CD20), CD3, CD5, CD7, and TCR alpha. There are chemokines, cytokines and receptors including CCL21, CCR6, CCR7, IL2RG, IL21R and IL7R. Elements of the T cell receptor signalling pathway are present including CD28, Ikaros family zinc finger (IKZF1), ZAP70, lymphocyte specific protein tyrosine kinase (LCK) and Wiskott-Aldrich syndrome gene (WAS). Enrichment analysis confirms the signature to be enriched in T and B cell associated genes (Table 4.71).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Lymphocyte activation	1.1×10^{-23}
DAVID	GOTERM_BP_FAT	T cell activation	2.5×10^{-15}
DAVID	GOTERM_BP_FAT	Regulation of B cell activation	6.5×10^{-5}
GSEA	GSE 10325	Genes upregulated in healthy B cells vs myeloid cells	3.4×10^{-55}
GSEA	GSE 29618	Genes upregulated in B cells vs monocytes	5.2×10^{-48}

Table 4.71 Functional annotation for the lymphocyte signature from the skin dataset using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

There is a cluster representing plasma cells of 86 probesets to 24 genes. This contains multiple immunoglobulin heavy and light chain genes, PIM2, expressed in B cells and TNFRSF17, which is expressed only in mature B cells. Enrichment analysis does not provide much useful annotation for this cluster, with enrichment for 'antigen binding' and immunoglobulins (Table 4.72).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_MF_FAT	Antigen binding	6.4×10^{-12}
DAVID	SP_PIR _KEYWORDS	Immunoglobulin	1.1×10^{-12}
GSEA	GSE 29618	Genes upregulated in B cells vs monocytes	2.7×10^{-26}

Table 4.72 Functional annotation for the plasma cell signature from the skin dataset using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

As in all datsetsets analysed, there is an interferon response here represented as 28 probesets and 21 genes (Figure 4.42). This contains all 3 components of the 2,5-oligoadenylate synthetase (OAS1, OAS2, OAS3), and many interferon induced proteins (IFI44, IFI44L, IFIH1, IFIT1, IFIT2, IFIT3, IFI27, IFI6, IFIT2). Enrichment analysis confirms this signature to be interferon related (Table 4.73).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Immune reponse	1.9×10^{-6}
DAVID	GOTERM_MF_FAT	RNA binding	3.6×10^{-4}
DAVID	GOTERM_BP_FAT	Response to virus	4.1×10^{-6}
GSEA	HECKER_IFNB_TARGETS	Genes modulated in the blood of patients with MS treated with subcut IFNB	2.6×10^{-37}

Table 4.73 Functional annotation for the interferon response signature from the skin dataset using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

A small cluster of mast cell genes of only 11 probesets to 7 genes is also present. This contains mast cell carboxypeptidase, and tryptase alpha and beta as well as TWIST2 and OSR2 which do not appear to be particularly associated with Mast cells from the literature. Enrichment analysis confirms only enrichment for protease activity reflecting the small number of genes in the cluster (Table 4.74).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	SP_PIR_KEYWORDS	Protease	3.8×10^{-3}

Table 4.74 Functional annotation for the mast cell signature from the skin dataset using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

The non-immune components of the stromal signature include extracellular matrix, endothelial and skeletal muscle signatures.

The largest ECM signature is of 57 probesets to 39 genes. This contains no collagen molecules in contrast to ECM signatures seen in every other dataset. There are

structural protein genes fibulin 1 and 2, fibrillin 1, elastin and laminin, alpha 2. (FBLN1, FBLN2, FNB1, ELN, LAMA2). There are growth factor receptors including insulin like growth factor 1 (IGF1) and platelet-derived growth factor receptor like (PDGFRL). Enrichment analysis confirms this to represent an ECM signature (Table 4.75).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	SP_PIR_KEYWORDS	Extracellular matrix	2.4×10^{-14}
GSEA	NABA_CORE_MATRISOME	Core extracellular matrix genes	1.5×10^{-14}

Table 4.75 Functional annotation for the ECM signature from the skin dataset using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

The endothelial signature contains 24 genes including CD31 (PECAM1), von willibrand factor (VWF), vascular endothelial cadherin (CDH5), ERG and endothelial tyrosine kinase (TEK). Enrichment analysis confirms enrichment for genes associated with angiogenesis (Table 4.76).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	GOTERM_BP_FAT	Cell adhesion	5.7×10^{-9}
DAVID	GOTERM_BP_FAT	Blood vessel development	7.8×10^{-7}

Table 4.76 Functional annotation for the endothelial signature from the skin dataset using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values.

The skeletal muscle signature is of 68 probesets to 61 genes and includes multiple myosins and their binding proteins (MYH1, MYH2, MYH7, MYL1, MYL2, MYL3, MYLPF, MYBPC1, MYBPC2), actins (ACTA1, ACTC1), muscle creatine kinase (CKM), troponins and tropomyosins (TNNC1, TNNC2, TNNI1, TNNI3, TPM1,

TPM2, TPM3). Enrichment for skeletal muscle genes is confirmed on analysis (Table 4.77).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	SP_PIR_KEYWORDS	Muscle protein	1.6×10^{-41}
DAVID	SP_PIR_KEYWORDS	Skeletal muscle	3.1×10^{-22}
GSEA	REACTOME	Genes involved in striated muscle contraction	5.8×10^{-38}

Table 4.77 Functional annotation for the skeletal muscle signature from the skin dataset using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

4.11.2 Cell cycle signatures

The cell cycle signature is difficult to discern in the this dataset, tending to fragment into small clusters, presumably reflecting the marked differences in proliferation between tumour types compounded by the tendency of normal proliferation in the epidermis to pull these genes into skin specific clusters. The largest of the cell cycle clusters containing only 19 probesets and 16 genes appears to represent the mitotic phase of the cell cycle. It contains cyclin B2, centrosomal protein 152kDa, centromere protein F, and the minichromosome maintenance complex components 6 and 10 (CCNB2, CEP152, CENPF, MCM6, MCM10). Despite the small number of genes, enrichment analysis demonstrates enrichment for genes associated with mitosis (Table 4.78).

Analysis tool	Source of annotation	Annotation	Significance (p-value)
DAVID	SP_PIR_KEYWORDS	Cell division	4.2×10^{-8}
DAVID	SP_PIR_KEYWORDS	Mitosis	2.9×10^{-7}
GSEA	REACTOME	Genes involved in cell cycle, mitotic	7.0×10^{-14}

Table 4.78 Functional annotation for the cell cycle signature from the skin dataset using the DAVID functional annotation clustering tool and MSigDB tool in GSEA. Selected annotation given only. Annotation derived from DAVID gives modified Fisher Exact p-values and GSEA hypergeometric distribution p-values.

4.11.3 Conservation of signatures

To establish whether the signatures seen in this independent dataset showed conservation of the signatures seen in the 'merged' data, clusters derived in this analysis were tested for enrichment of the genes in the conserved signatures using Fisher's test adjusted for multiplicity of testing. The main macrophage cluster in the skin cancer dataset (cluster 16) was significantly enriched for genes in the merged macrophage cluster (cluster 7) with an adjusted p-value of 9.0^{E-120} by Fisher's test with correction for multiplicity of testing. Similarly other of the signatures from the merged graph were conserved in the skin dataset (Table 4.79 and figures 4.42 and 4.43).

Cluster name	Gene number in merged cluster(s)	Genes in skin cluster	Significance of enrichment in skin clusters (Adjusted Fisher's test)
Macrophage	*163*	*62*	9.0^{E-120}
T cell	*145*	*245*	4.9^{E-90}
IFN response	*73*	*21*	6.9^{E-53}
MHC class I	*16*	*27*	6.8^{E-65}
Plasma cell	*36*	*24*	3.9^{E-175}
Mast cell	*4*	*7*	1.4^{E-27}
Extracellular matrix	*100*	*39*	6.3^{E-6}
Endothelium	*17, 13, 11*	*24*	1.7^{E-11}, 2.9^{E-16}, 7.5^{E-13}
Skeletal muscle	*15*	*61*	8.9^{E-32}
Cell cycle	*182*	*16*	8.3^{E-21}
Ribosomes	*12*	*16*	1.6^{E-36}
Histones	*26*	*10*	9.1^{E-32}
Glycolysis	*5*	*5*	2.9^{E-30}

Table 4.79. Signatures from the merged dataset were conserved in an independent dataset. Clusters from the skin dataset were tested for enrichment of genes in clusters in the merged dataset using Fisher's test with adjustment for multiplicity of testing.

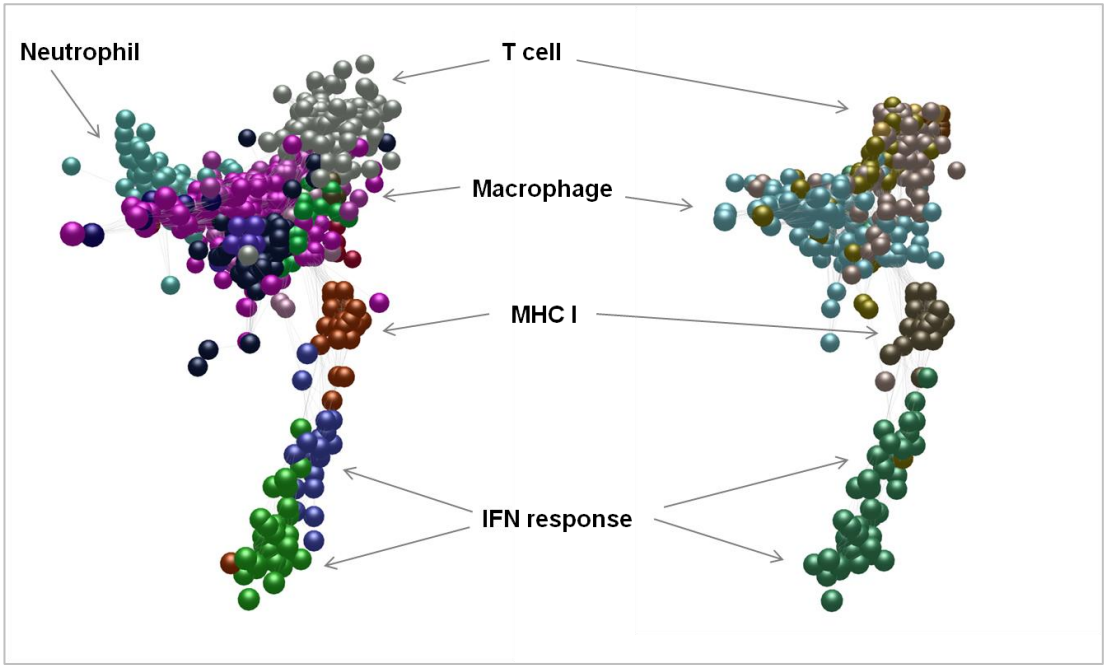


Figure 4.43. Conservation of the 'core' immune signatures in the skin cancer dataset. The left-hand graph shows only the immune clusters in the skin cancer dataset with the rest of the graph hidden. In the right hand graph, the same area of the graph is illustrated but with the cluster annotation from the merged graph applied, and nodes which represent genes not seen or not forming clusters in the merged graph absent. The colours represent the individual cluster colours from on the left, the MCL clustering of the skin dataset and on the right, the imported cluster annotation from the merged clusters. The clusters get smaller when only the conserved signatures are seen, but there is good preservation of most immune clusters. Some, such as the neutrophil cluster which is not seen in the merged graph, are lost when applying the merged annotation.

Analysis of tumour type-specific clusters (i.e. unconserved clusters) also provided insight into tumour biology in this dataset; for example if we chose to look at the nearest neighbours of SILV, a gene whose product (gp100) is detected by the antibody HMB45 which is used clinically in the diagnosis of melanoma, being expressed only in junctional melanocytes and melanoma, we pull out the genes TYR (tyrosinase) the key enzyme in melanin biosynthesis, MLPH (melanophilin), which plays a role in melanosome transport, GPR 143 expressed on the melanosome membrane, and MLANA (melan-a), as well as MITF, a melanocytic transcription factor and transcriptional regulator of many of these genes (Figure 4.44)(reviewed in Steingrımsson et al 2004). SNCA (alpha-synuclein) also lies within this grouping, and while there is no literature regarding the expression of this in melanoma, it does

play a role in dopamine metabolism in the CNS, and therefore may also function in melanin metabolism. In addition, CDK2, a gene downstream of and regulated by MITF and known to be required for melanoma growth and proliferation (Du et al 2004) lies within these 'near neighbours', as does BIRC7 , another target of MITF (Dynek et al 2008), and anti-apoptotic protein contributing to melanoma survival and resistance to therapy (Vucic et al 2000, Lazar et al 2010).

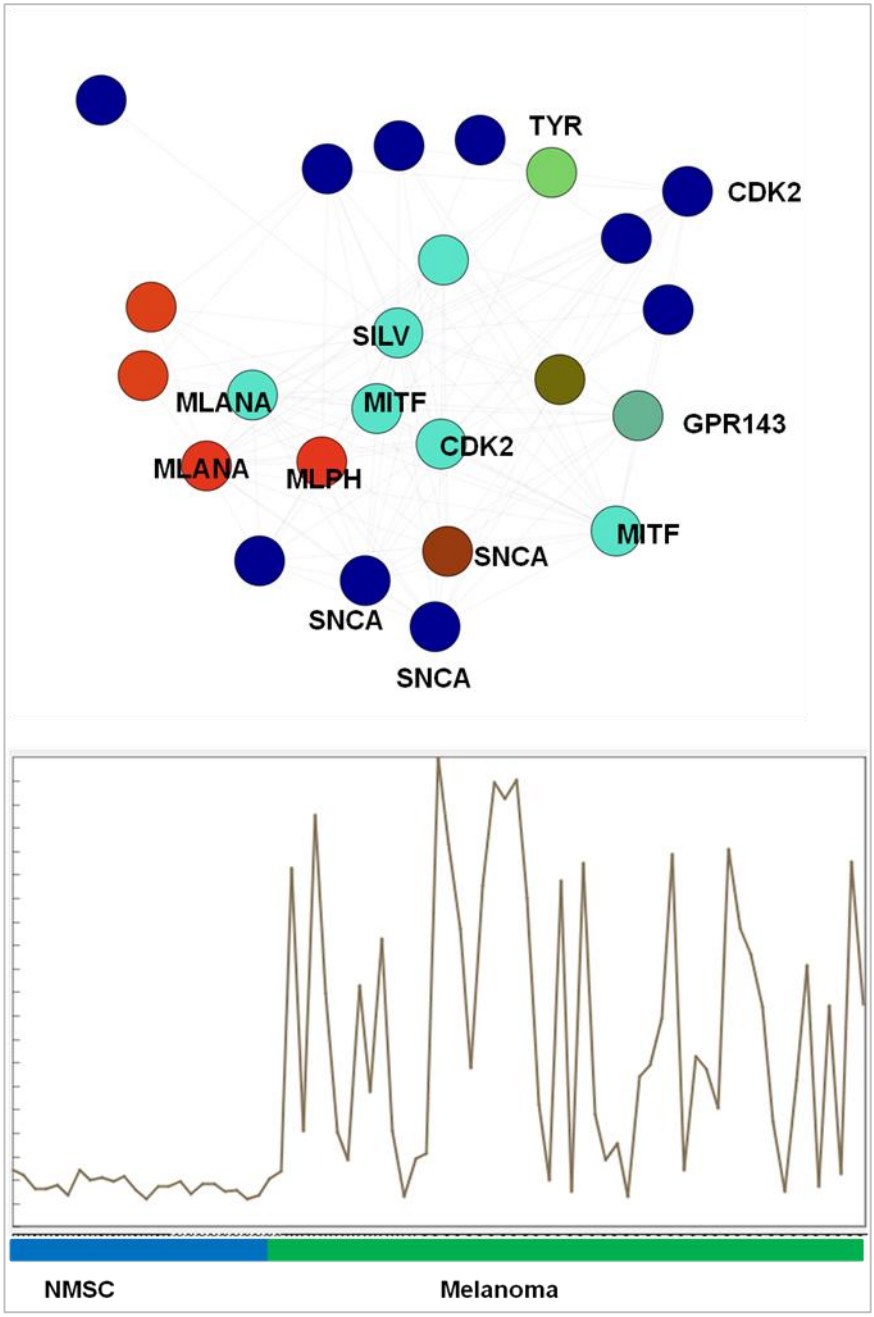


Figure 4.44. Nearest neighbours of SILV. Selecting the nodes closest to SILV, the gene encoding the protein recognised by the antibody HMB45 generates a group of genes defining elements of melanocyte biology. The mean expression of these genes is seen in the expression plot. Individual tumours are represented on the x-axis. The coloured bar indicates tumour groupings. NMSC= non-melanocytic skin cancer.

Additional analysis by Freeman of further datasets confirmed conservation of the 'core' signatures in other unrelated tumour types including Hodgkin lymphoma and gastric carcinoma (data not shown).

Conclusions

This chapter presents a novel approach to isolating functional signatures, including cell-specific signatures, from large human datasets. This was made possible by recent advances in graphical presentation and analysis of data that allow much larger datasets to be analysed than would previously have been possible as well as the public provision of these large datasets. Signatures are defined on the basis of patterns of co-expression of genes within the dataset. The analysis presented here applys no preconceived notions of what data is important in creation of the network graph allowing only the correlation threshold to decide what is included in the graph. The approach differs fundamentally from much of the gene expression clustering analysis presented in the literature which is often focussed on defining differences between groups.

The data presented above demonstrates that it is possible to determine process and cell specific signatures from complex datasets, and thus define features of the cancer transcriptome in situ in genuine human tumours. Clearly the data derived from this analysis cannot capture all the elements of an individual signature. Features that are common to many cell types and processes such as intracellular signalling pathways will be represented in many cell types/processes and, unless there is a dominant driver in the dataset such that all other influences are insignificant, then the signature seen will be a sum of all the influences on the signature and will therefore not readily correlate with any one cell type or process. Discrete processes that are identified such as stromal signatures or cell cycle pathways will also represent a sum of all the influences on them and thus the overall signature of, for example, the tumour associated macrophage. As in any analysis method that looks at homogenized tissue/data it is impossible to localise the signal and know where in the complex

three-dimension structure of the tumour the signature resides. Some of the limitations in spatial descriptions of the signatures will be addressed in a limited way in the next chapter where elements of the TAM signature are characterised in human DLBCL tissue samples. What the analysis presented here does tend to isolate are those items of a signature that are unique to a cell type or process such that the only or dominant signature comes from that cell type or process. Thus, while the analysis cannot capture all the elements of an individual cell type signature or process, it will capture the elements that make the signature unique from other cell types or processes.

As with all approaches that analyse cells or signatures by levels of mRNA, any genes whose expression is regulated at a different point, be that rapid mRNA degradation, high protein turnover or sequestration of protein in intracellular compartment, will not be accurately represented in this data.

Having identified signatures, among others, that relate to TAMs, the following chapter seeks to validate elements of this signature in an independent cohort of human diffuse large B cell lymphoma cases with emphasis on genes that formed part of the tumour associated macrophage signature.

Chapter 5

Validation of TAM gene signature

Having identified a signature of the tumour associated macrophage in DLBCL using a bioinformatic approach, further work was undertaken to validate the signature by analysis of expression of selected elements of that signature. As the studies in the previous chapter were performed using publically available datasets, it was not possible to validate any of the signatures by independent analysis of genes expressed in the macrophage signature on a gene by gene basis. Instead, immunohistochemical studies were undertaken on an unrelated set of diffuse large B cell lymphomas as well as benign reactive lymphoid tissue to assess protein expression of elements of the signature. This moved the work from an *in silico* analysis of human tumour tissue back into human tumour tissue sections. Initially nine genes from the signature were selected for further study.

5.1 Creation of tissue microarrays for benign lymph nodes and DLBCL.

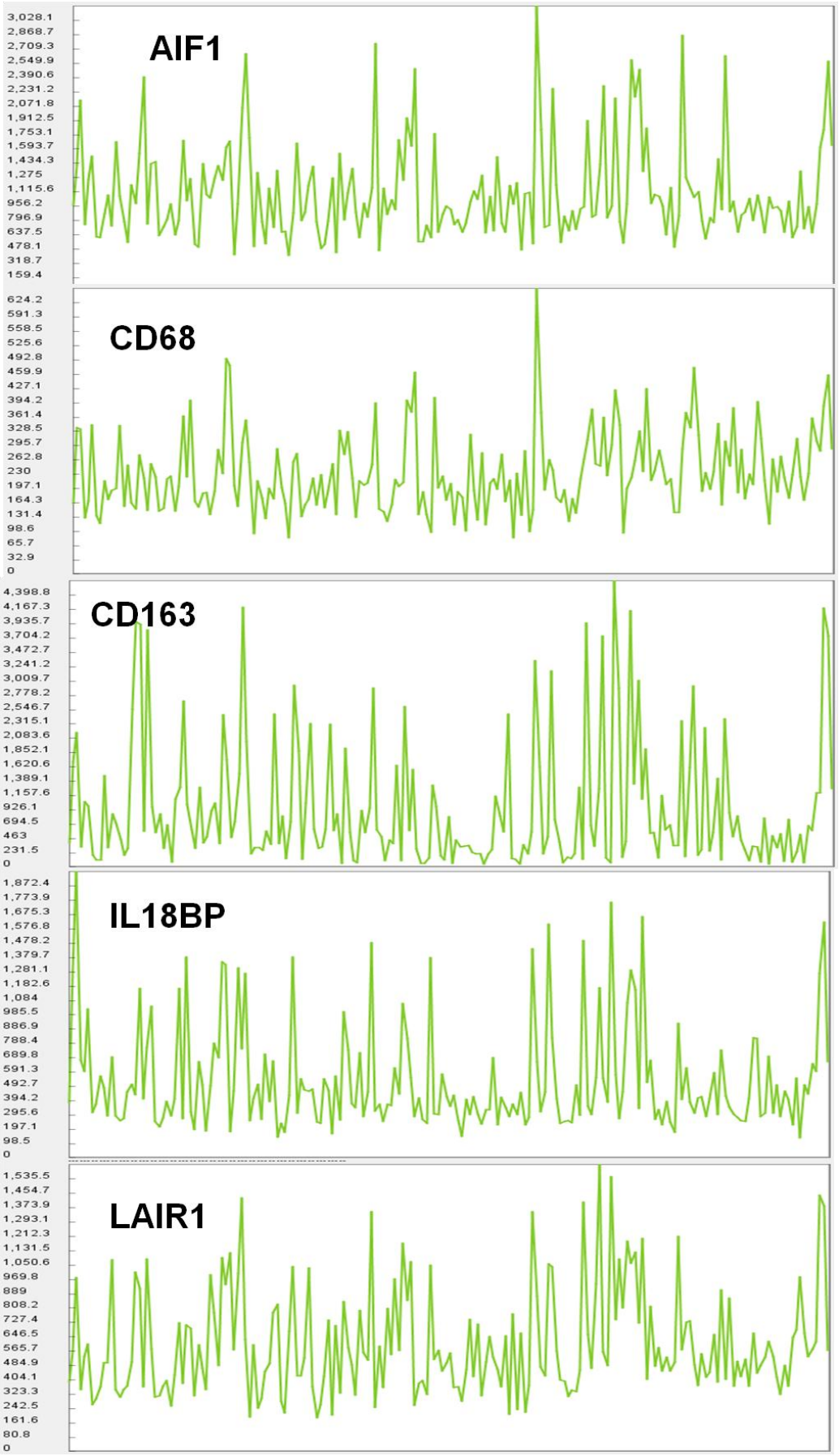
To allow a large number of cases to be studied tissue microarrays were created using benign and malignant lymph nodes. Benign lymph nodes were identified by searching the pathology archive over a 10 year period using the SNOMED codes 'lymph node' and 'hyperplasia'. All cases so identified were reviewed and only those cases not associated with a malignancy selected, i.e. cases of isolated lymphadenopathy where there was no evidence of lymphoma. Histological review of these cases confirmed the reactive nature of the lymph nodes in all cases selected for inclusion. Malignant lymph nodes containing DLBCL were identified through the records of the Scotland and Newcastle Lymphoma Group database. Again all cases were reviewed to confirm the diagnosis. Selected representative areas of nodes were marked for inclusion in the tissue microarray.

Creation of tissue microarrays allows large numbers of specimens to be examined on a limited number of slides which provides several advantages; archival human tissue can be used without compromising the requirement to leave tissue in the block for future diagnostic use; removal of individual variation in staining between slides as well as savings in reagents required to examine the tissue and savings in time to analyse the cases without the need to move from slide to slide. Large numbers of cases can be stained simultaneously without the requirement for access to large automated stainers capable of dealing with hundreds of slides simultaneously, an approach which removes some of the variability inherent in staining slides in batches, but may still show slide-to-slide variation. The use of TMAs in the study of lymphoma is well established and has been validated in several studies (Hedvat et al 2002, Rassidakis et al 2002, Tzankov et al 2003). In the cases reported below, some whole sections were also stained to ensure uniformity of expression across the tissue. The use of 2mm cores also provided sufficient tissue to study so that localisation to anatomical compartments within the node was possible in the reactive lymph nodes. Sixty-one cases were included in the benign/reactive TMA and seventy-four in the DLBCL TMA.

5.2 Immunohistochemistry for selected markers from TAM profile

Initially nine genes from the signature were selected for validation at the protein level (Figure 5.1). All showed marked variation in gene expression across the tumour dataset. These were selected from genes present in both the DLBCL TAM signature and the 'core' TAM signature and cases that were only in the DLBCL TAM signature and not the 'core' signature. These included well characterised markers of macrophages CD68 and CD163 as well as less-well characterised genes PLAUR, IL-18BP, LGALS3BP, TYMP, AIF1, LAIR1 and SIGLEC1 (CD169). Aside from CD163 and CD68, which were selected as being known macrophage markers, the selection was based on there being some knowledge of the function of the genes suggesting that their role in tumours may be of interest, and availability of reagents. Antibodies which claimed to work on FFPE tissue were available for all targets selected. After extensive preliminary work assessing different methods of

antigen retrieval and staining protocols PLAUR, IL-18BP and SIGLEC1 wererejected as specific staining in control tissue was not achieved. The remaining six targets were assessed by immunohistochemical expression in benign and malignant lymph node tissue (See tables 2.2 and 2.3 for details of clones, suppliers and antigen retrieval methods). In the reactive lymph nodes, three major compartments containing macrophages could be recognised; tingible body macrophages in reactive germinal centres, sinus histiocytes and 'stromal' macrophages which lay in the nodal parenchyma but not in reactive germinal centres or sinuses. To allow comparison of expression patterns of TAM markers within individual cases, all staining was performed on sequential 3µm sections of TMA.


AIF1
3,028.1
2,868.7
2,709.3
2,549.9
2,390.6
2,231.2
2,071.8
1,912.5
1,753.1
1,593.7
1,434.3
1,275
1,115.6
956.2
796.9
637.5
478.1
318.7
159.4
CD68
624.2
591.3
558.5
525.6
492.8
459.9
427.1
394.2
361.4
328.5
295.7
262.8
230
197.1
164.3
131.4
98.6
65.7
32.9
0
CD163
4,398.8
4,167.3
3,935.7
3,704.2
3,472.7
3,241.2
3,009.7
2,778.2
2,546.7
2,315.1
2,083.6
1,852.1
1,620.6
1,389.1
1,157.6
926.1
694.5
463
231.5
0
IL18BP
1,872.4
1,773.9
1,675.3
1,576.8
1,478.2
1,379.7
1,281.1
1,182.6
1,084
985.5
886.9
788.4
689.8
591.3
492.7
394.2
295.6
197.1
98.5
0
LAIR1
1,535.5
1,454.7
1,373.9
1,293.1
1,212.3
1,131.5
1,050.6
969.8
889
808.2
727.4
646.5
565.7
484.9
404.1
323.3
242.5
161.6
80.8
0

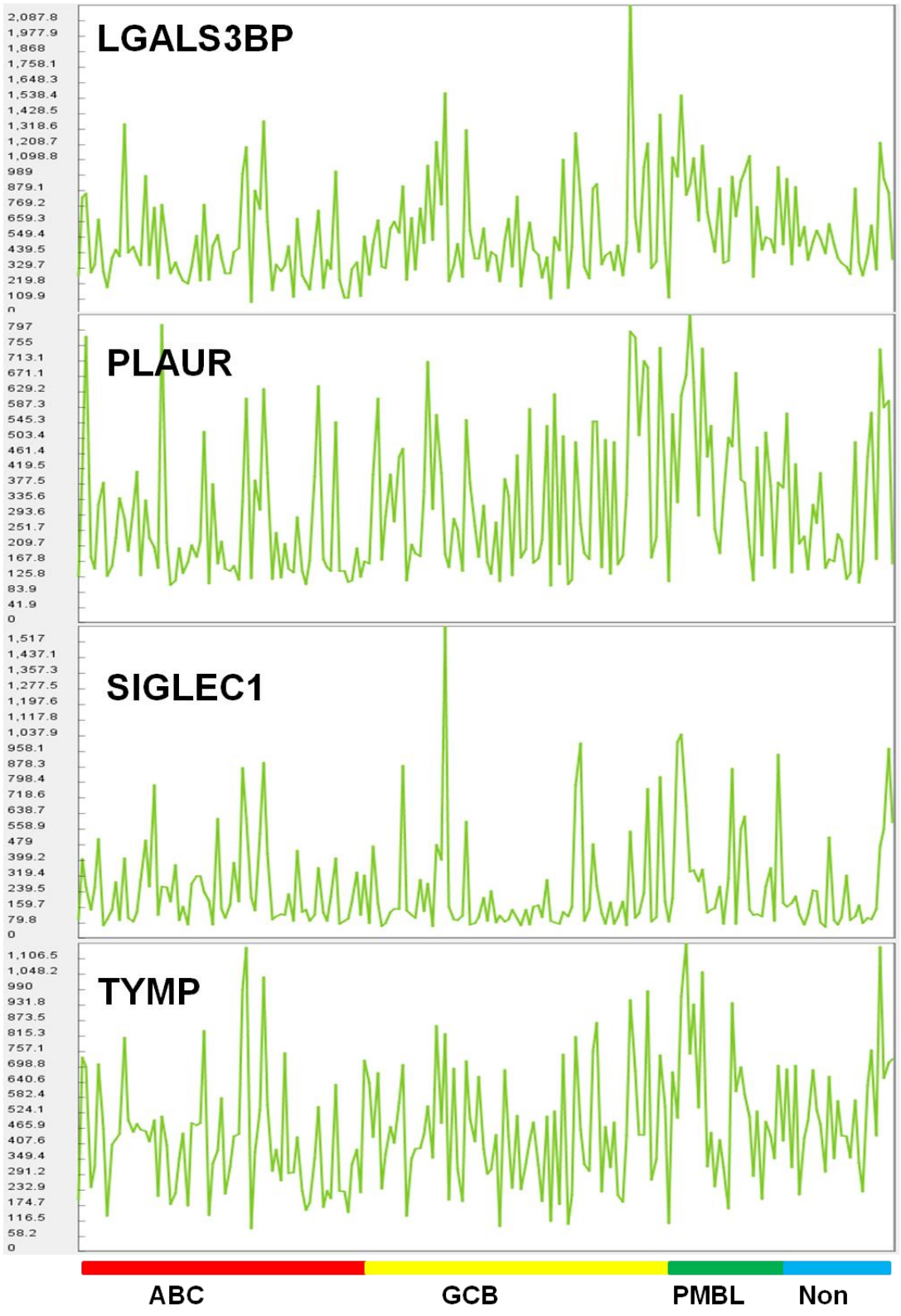


Figure 5.1 Expression pattern of the nine selected genes in the DLBCL cohort from chapter 4. Individual tumours are represented on the x-axis, with the coloured bars denoting tumour subtype. Mean fluorescence values for each gene are given on the y-axis.

5.2.1 CD68 expression in lymphoid tissue.

CD68 showed granular cytoplasmic staining and was seen in all compartments of the reactive lymph node (Figure 5.2). Staining tended to be strong with little variation from case to case.

Expression was seen in all tumour cases, showing a similar granular cytoplasmic staining to that seen in the reactive nodes (Figure 5.2).

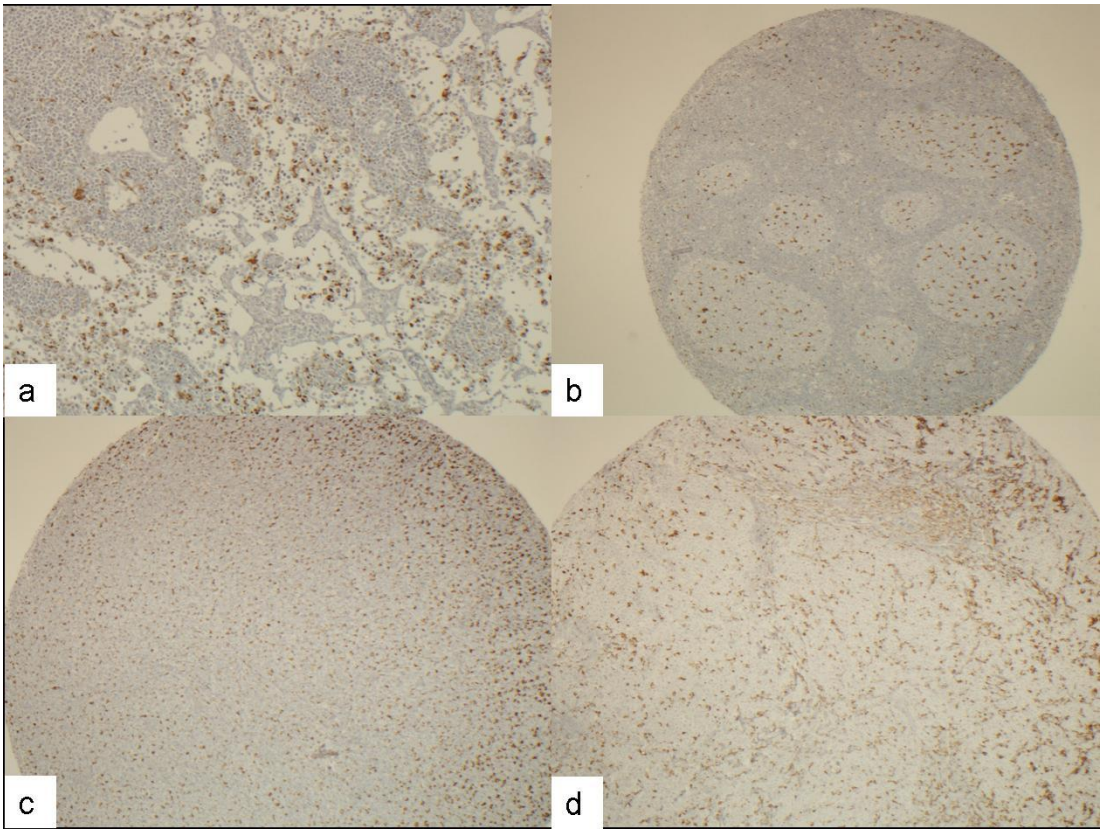


Figure 5.2. Expression of CD68 was seen in all compartments of the reactive lymph node. a) in sinus histiocytes x200 magnification, b)expression in tingible body macrophages in the germinal centre and stromal macrophages x100 magnification. All tumour cases contained CD68 positive cells. c) and d)representative cases of DLBCL x100 magnification.

5.2.2 CD163 expression in lymphoid tissue

CD163, a macrophage scavenger receptor, showed membranous staining and was expressed by sinus histiocytes with a uniformly strong expression in reactive lymph nodes (Figure 5.3). It was negative or occasionally weakly positive in tingible body macrophages and almost entirely absent in 'stromal' macrophages.

Expression was seen in macrophages in most tumours, although there was considerable variation between tumours with some containing large numbers of positive cells and others very occasional weakly staining cells. Some tumours were entirely negative (Figure 5.4).

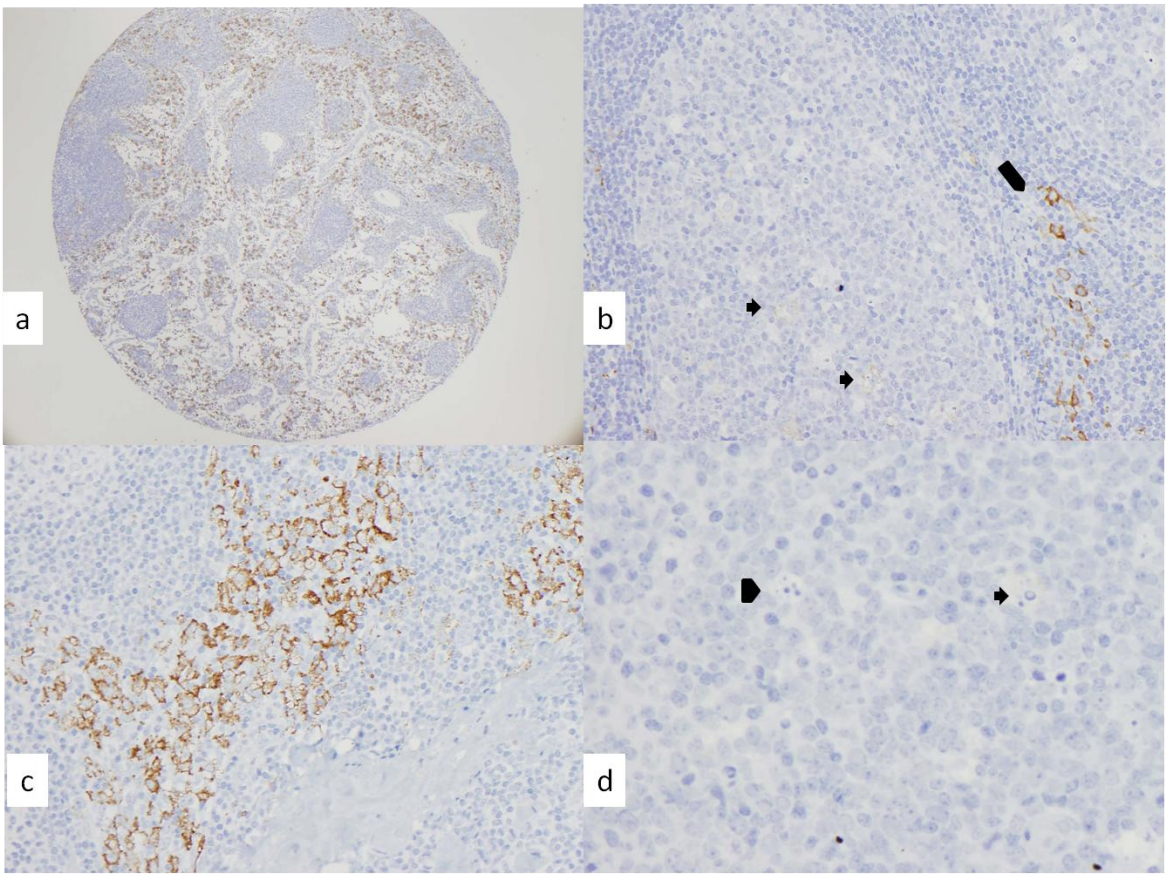


Figure 5.3 Expression of CD163 in reactive lymph nodes. Strong staining was seen in sinus histiocytes a) x40 magnification and c) x200 magnification. There was weak or absent staining in tingible body macrophages. b) x100 magnification . Arrows indicate tingible body macrophages and arrow head sinus histiocytes. d) x200 magnification. Arrow indicates weak membranous expression in tingible body macrophage and arrow head absent expression in tingible body macrophage in same germinal centre.

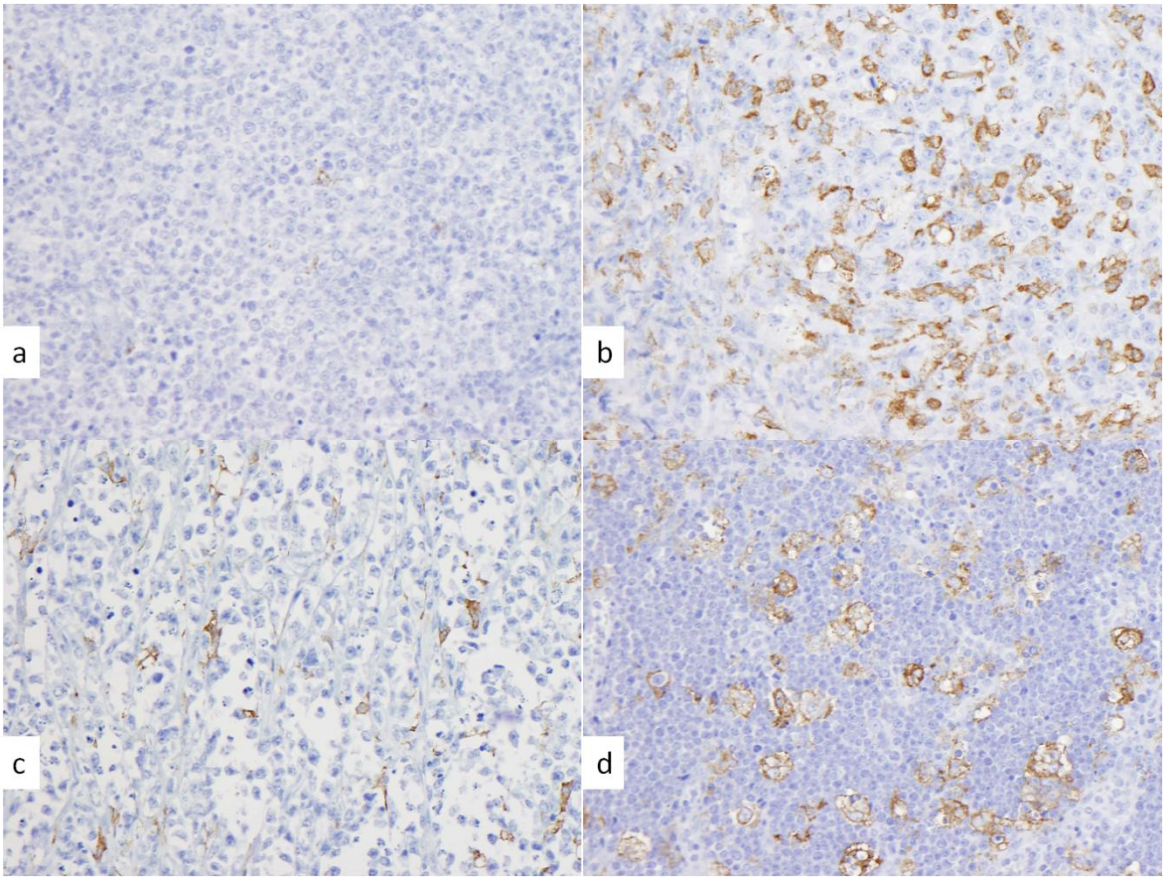


Figure 5.4 Expression of CD163 in lymphoma. Cases of DLBCL were stained with anti-CD163. There were positive macrophages in almost all cases, but with marked variation from case to case as to the number of positive cells. Macrophages of varying morphology were positive including those containing abundant apoptotic bodies as in d). All images at x200 magnification.

5.2.3 AIF1 expression in lymphoid tissue

Allograft inflammatory factor – 1 (AIF1), a calcium binding protein whose expression is upregulated in response to interferon gamma, showed membranous and cytoplasmic staining in reactive lymph nodes with macrophages in all compartments being positive (Utans et al 1995) (Figure 5.5). Staining tended to be more intense in sinus histiocytes than other macrophage types, but all cell types were positive.

In DLBCL, AIF1 was expressed in macrophages in all cases (Figure 5.5). There was a mixture of patterns of staining, the majority of cases showing membranous and

cytoplasmic staining, some cases showing membranous staining only, and a small number of cases showing membranous, cytoplasmic and nuclear staining.

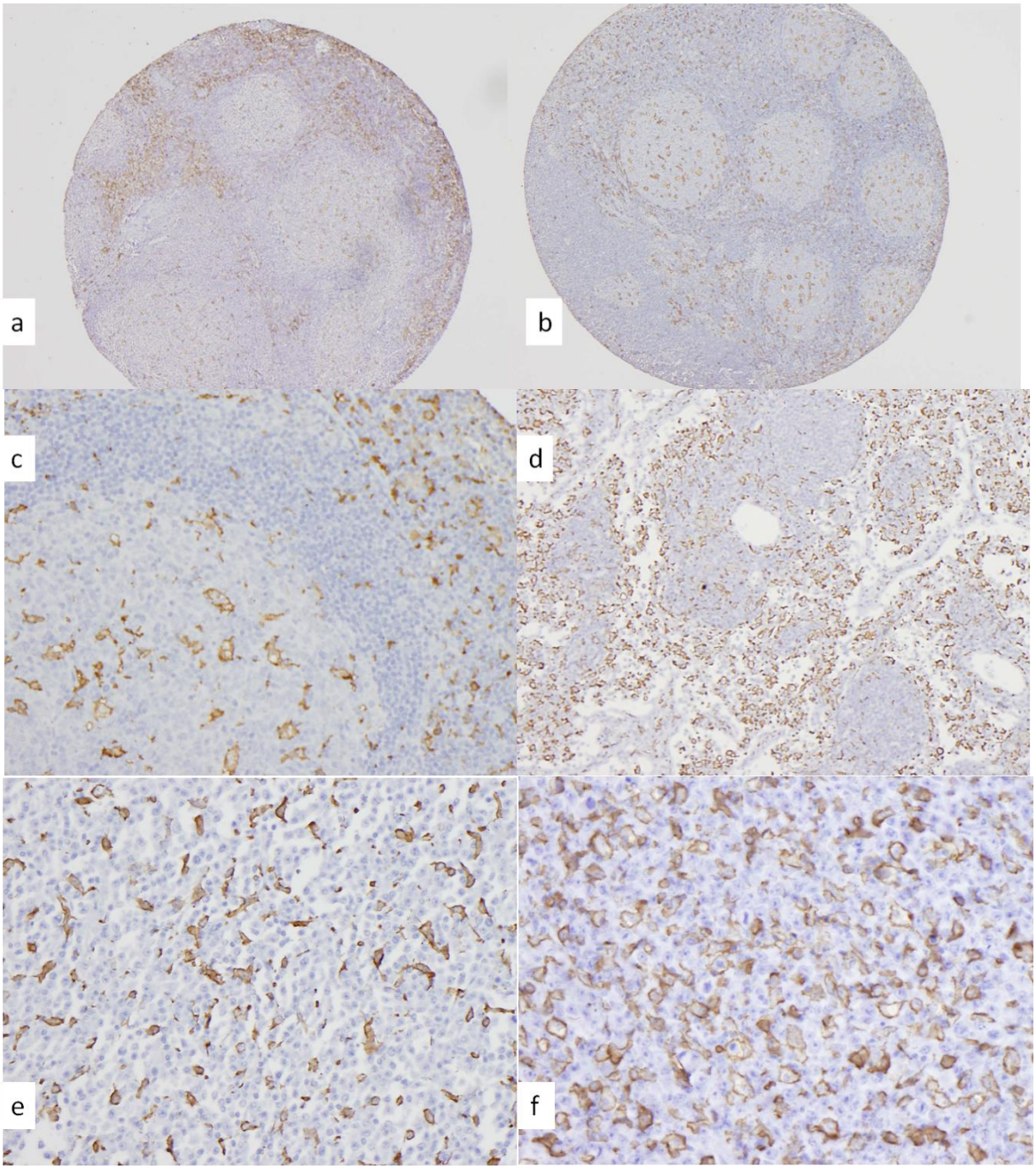


Figure 5.5 AIF1 expression is seen in all compartments of reactive lymph nodes. a) x40 magnification to show staining in sinus histiocytes and tingible body macrophages in germinal centres. b) x40 magnification to show 'stromal' macrophages and tingible body macrophages. c) x200 magnification. Staining in germinal centre and 'stromal' macrophages with less positive cells in mantle zone. d) x200 magnification. Sinus histiocytes show strong expression. e) and f) x200 magnification in DLBCL.

5.2.4 LAIR1 expression in lymphoid tissue

Leukocyte associated immunoglobulin-like receptor–1 (LAIR1) (CD305) expression was seen in all compartments of the reactive lymph node (Figure 5.6). Staining showed a membranous pattern with some occasional weak granular cytoplasmic staining.

Staining was seen in TAMs in all cases of DLBCL analysed. Staining showed a similar pattern to that observed in the reactive lymph nodes with membranous and occasional cytoplasmic staining (Figure 5.6).

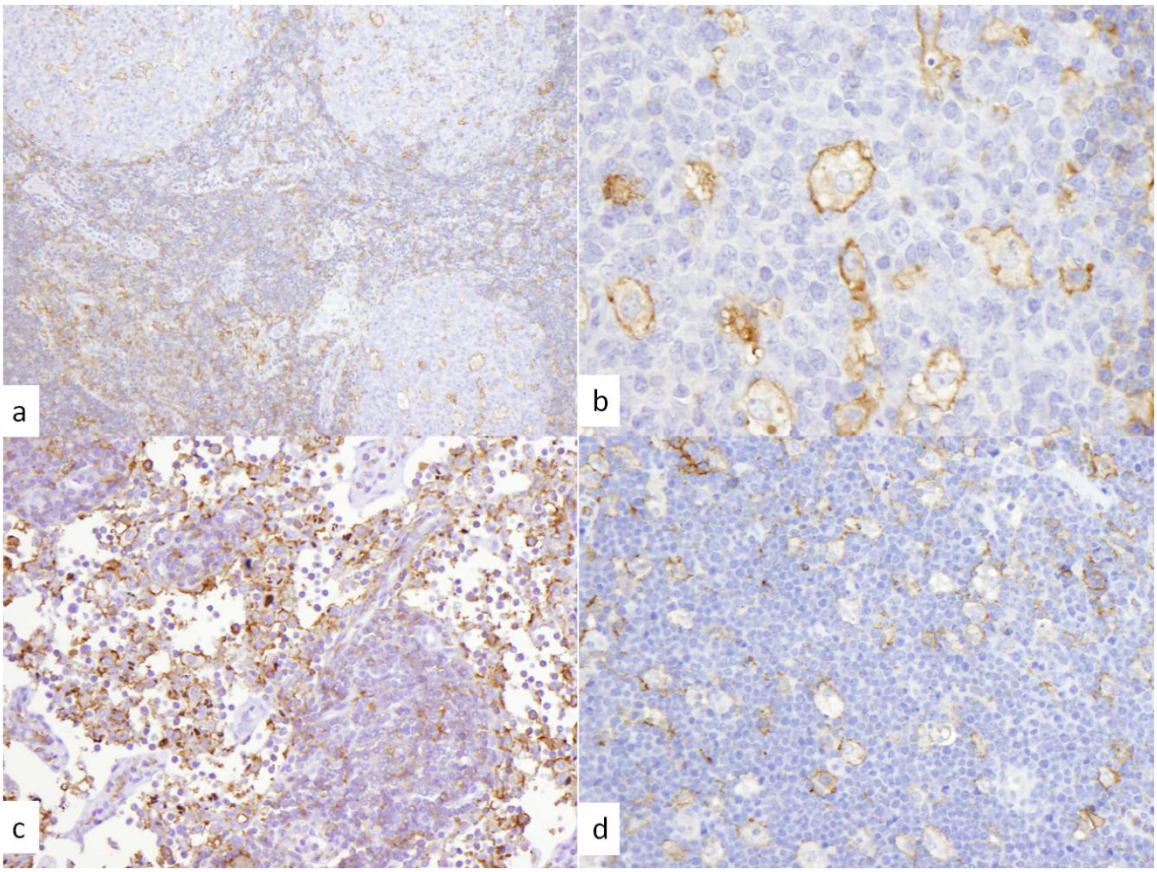


Figure 5.6 a) x100 magnification. LAIR 1 expression was seen in tingible body macrophages, sinus histiocytes and stromal macrophages. b) x400 magnification. LAIR1 expression in tingible body macrophages shows membranous staining and some weak cytoplasmic staining. C) x200 magnification. Staining in sinus histiocytes. d) x200 magnification. Representative case of DLBCL, in this case showing macrophages full of apoptotic debris.

5.2.5 LGALS3BP expression in lymphoid tissue

Lectin galactoside-binding, soluble, 3 binding protein (LGALS3BP, Mac2 binding protein, 90K) was expressed in a large proportion (70%) of reactive lymph nodes studied, but not all. The pattern of expression was relatively restricted with expression in all positive cases in sinus histiocytes, expression in stromal macrophages in only 4% of cases and complete absence in tingible body macrophages. Where expression was seen it was mostly weak and showed a granular cytoplasmic pattern (Figure 5.7).

In lymphoma, the majority of cases were negative. Only 30% of cases showed expression in TAMs and this expression tended to be very weak and focal.

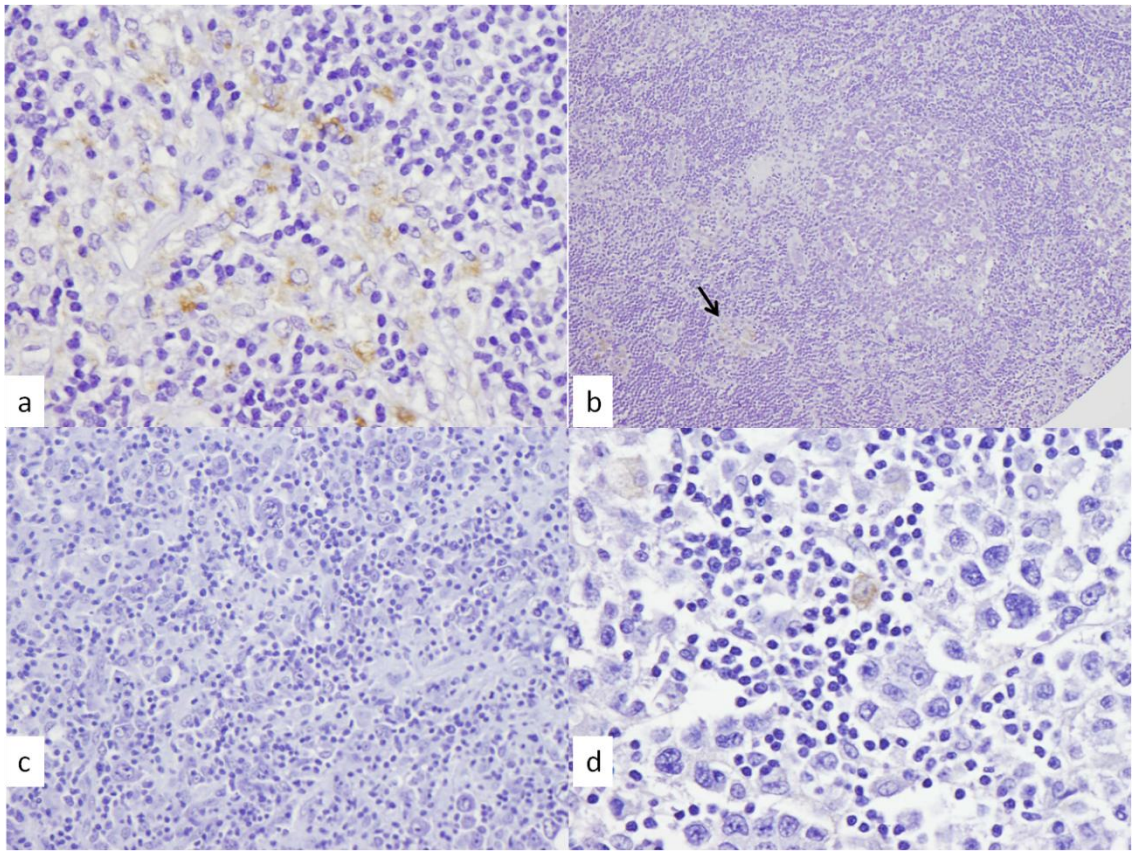


Figure 5.7 LGALS3BP expression was seen in the majority of cases of benign lymph nodes but almost always only in sinus histiocytes and with weak expression. a) x400 magnification. Granular cytoplasmic expression in sinus histiocytes in reactive lymph node. b) x100 magnification. Weak expression in sinus histiocytes in reactive lymph node (arrow). Adjacent reactive germinal centre is negative. c). X200 magnification. Absence of staining in DLBCL. d). X 400 magnification. Occasional TAMs show weak granular cytoplasmic staining in DLBCL.

The expression pattern of LGALS3BP was of particular interest. The known functions of the molecule will be discussed further in the next chapter, but much of the literature reflects its role as a secreted protein, measurable in serum or forming a linking molecule in the ECM which connects cells and the structural matrix. Thus the position in the macrophage signature was intriguing. However, work on the role of the molecule in HIV has demonstrated expression at the mRNA level in both T cells and macrophages, but protein expression only in macrophages (Lodermeyer 2013). It is entirely conceivable therefore that it may form part of the macrophage signature and its presence in subsets of macrophages in the non-malignant lymph node would tend to confirm this to be a macrophage expressed protein. The possibility that what the antibody used for immunohistochemistry is binding to is LGALS3BP secreted protein bound to galectin-3 on the surface of macrophages cannot be excluded however. It is notable that not all TAMs, and in fact only small proportion, showed positive staining. There are several possible explanations for this; the protein may be expressed by macrophages but secreted and hence not localised to the macrophage; there may be a disconnect between mRNA and protein in macrophages in this situation as has been described in T cells in culture; expression may be mostly by non-TAM macrophages which are poorly represented in the TMA from lymphomas but are likely to represent part of the mean macrophage profile in the gene expression studies; protein expression may be below the level of detection of the assay or the inclusion of LGALS3BP in the signature may be wrong. The absence of 'stromal' staining on immunohistochemistry would not support a secreted, ECM-binding role for the molecule. The presence of positive staining in

some macrophages would argue that this does not lie in the macrophage signature by error. The level of protein expression in absolute terms or relative to the mRNA level may account for the expression patterns seen but this was not explored further.

5.2.6 TYMP expression in lymphoid tissue

Thymidine phosphorylase (TYMP, also known as platelet derived endothelial cell growth factor) was seen in all macrophage compartments in the reactive node as well as in high endothelial venules (Figure 5.8). Staining was nuclear and cytoplasmic in pattern.

There was expression in TAMS in all tumours which showed considerable variation from tumour to tumour. Nuclear staining appeared to be a constant finding in TAMs while cytoplasmic staining was not seen in all cells.

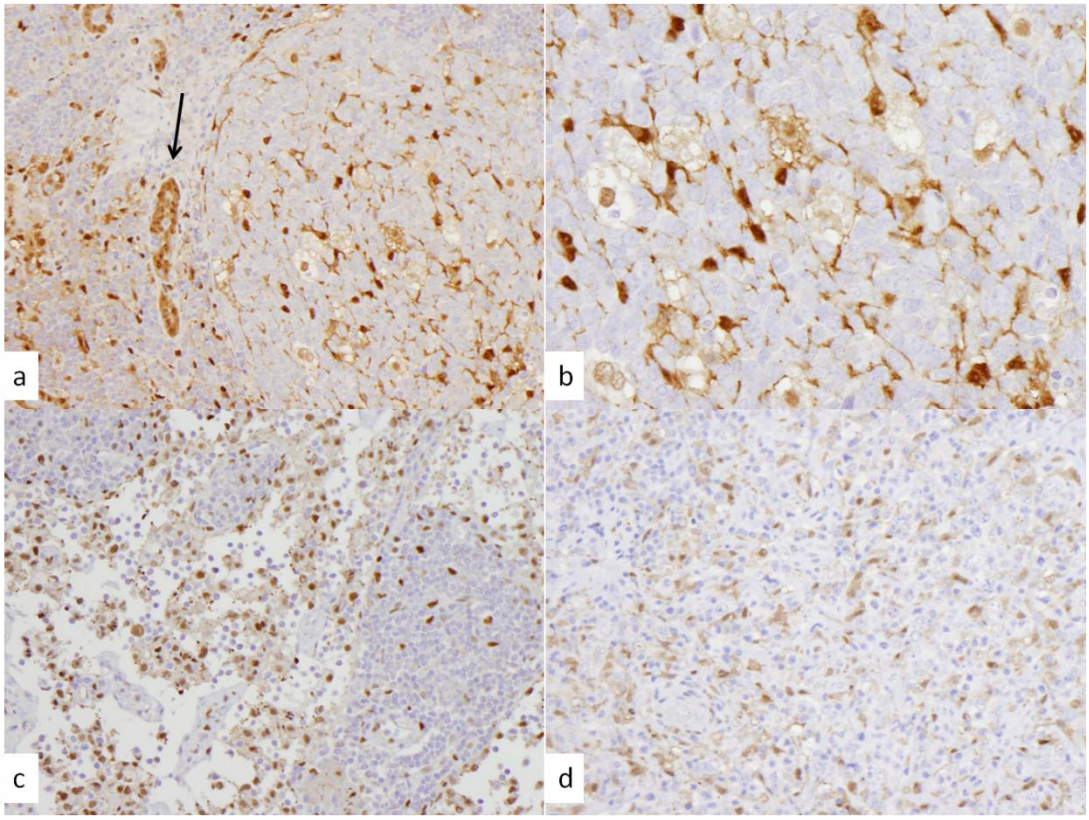


Figure 5.8 a) x 100 magnification. TYMP expression was seen in macrophages in the germinal centre (right hand side of image) as well stromal and sinus macrophages (left hand side of image). TYMP expression also seen in high endothelial venules (arrow). b)x 400 magnification. Tingible body macrophages in the germinal centre showing nuclear and occasional cytoplasmic staining. c) x200 magnification. Expression was also seen in sinus histiocytes. d) x200 magnification. Representative case of DLBCL showing TYMP expression in TAMs.

5.2.7 Expression patterns in individual cases

To examine distribution of staining of individual markers composite images were examined of serial sections in individual cases. These demonstrated differences in staining patterns between antigens, although all stained in macrophage-like cells morphologically. Not all macrophages expressed all antigens. Some examples are given below in figures 5.9 - 5.12.

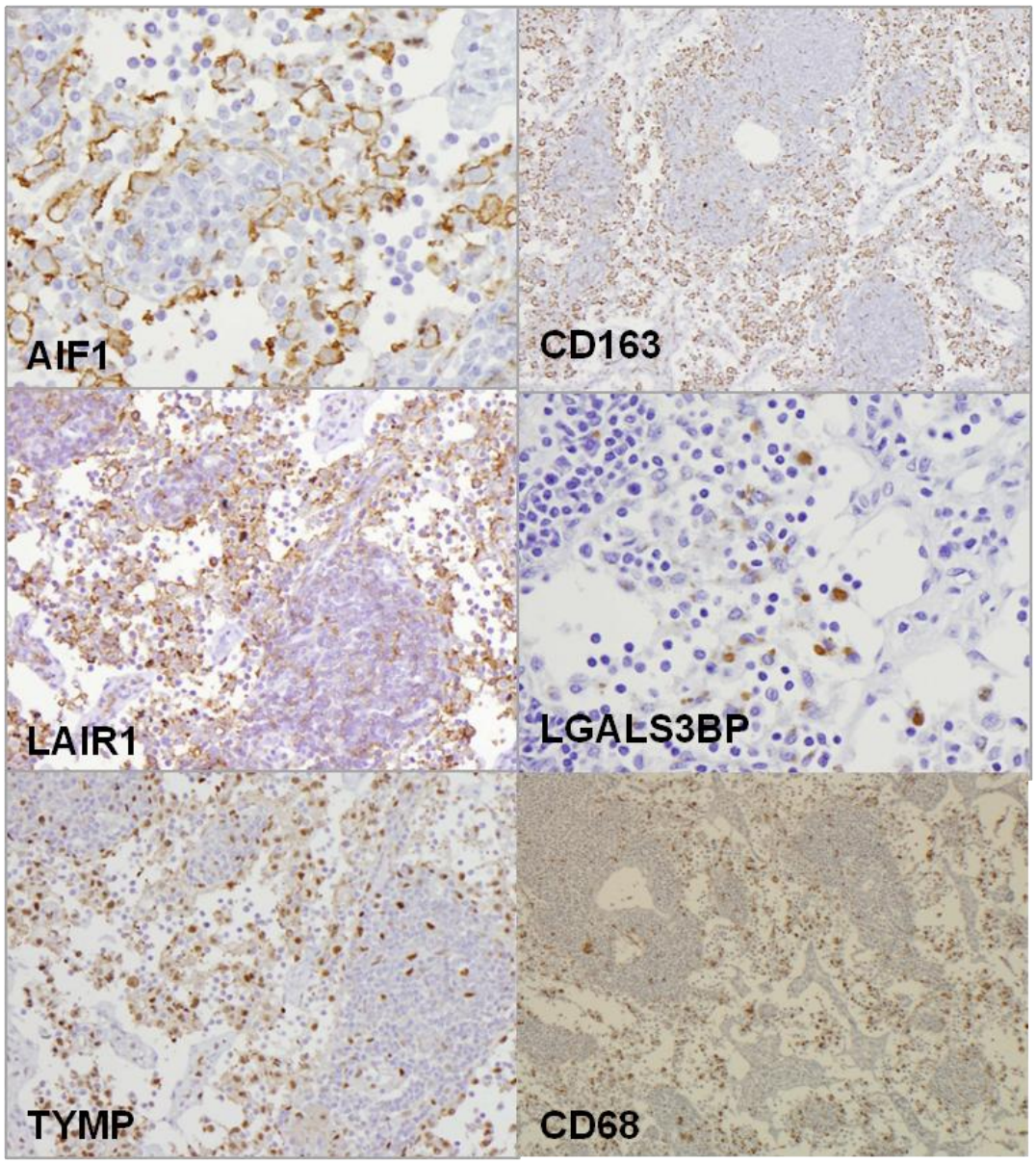


Figure 5.9. Representative reactive lymph node to show pattern of staining in sinus histiocytes.

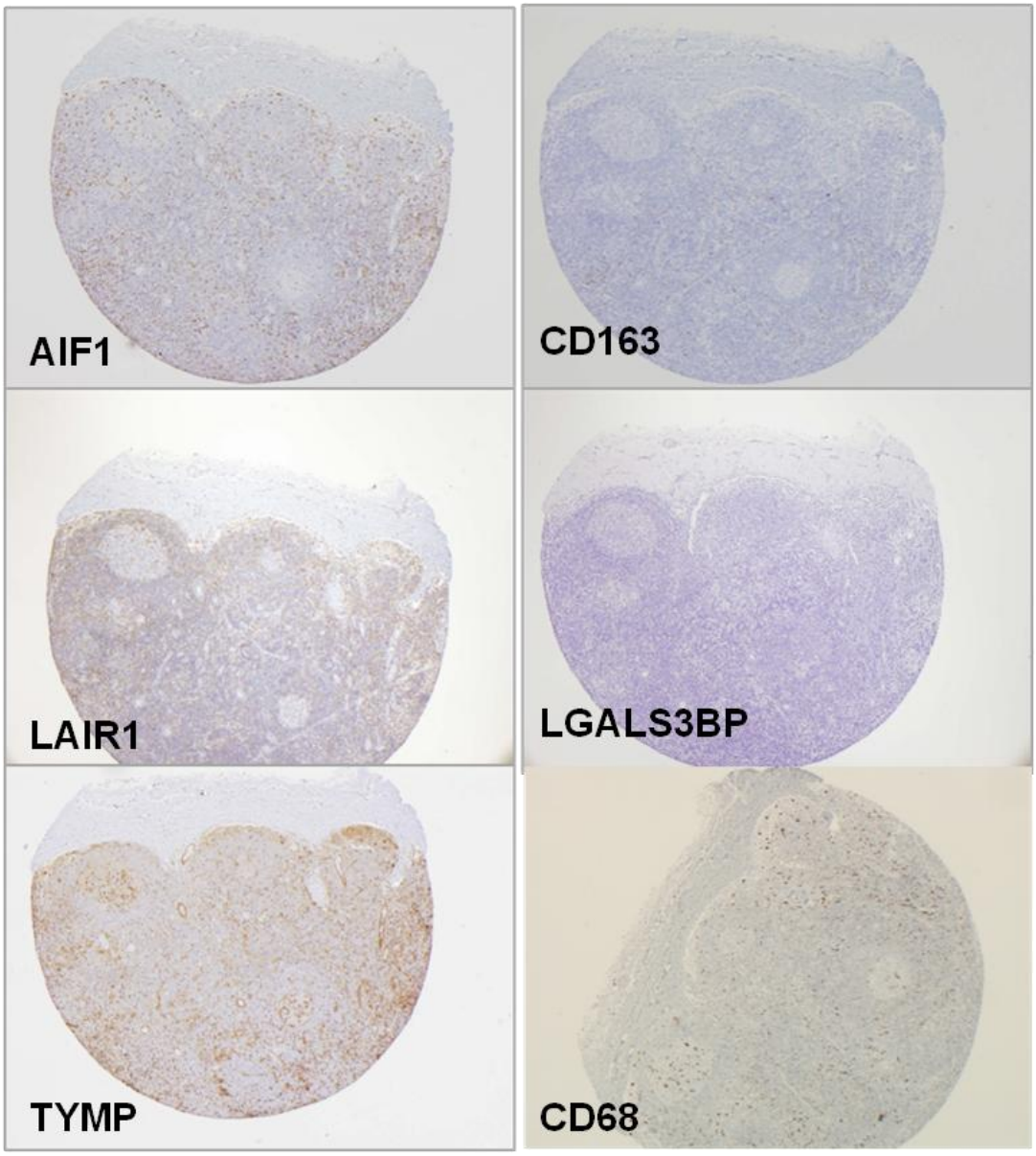


Figure 5.10. Representative reactive lymph node showing peripheral sinus, reactive germinal centres and stromal macrophages.

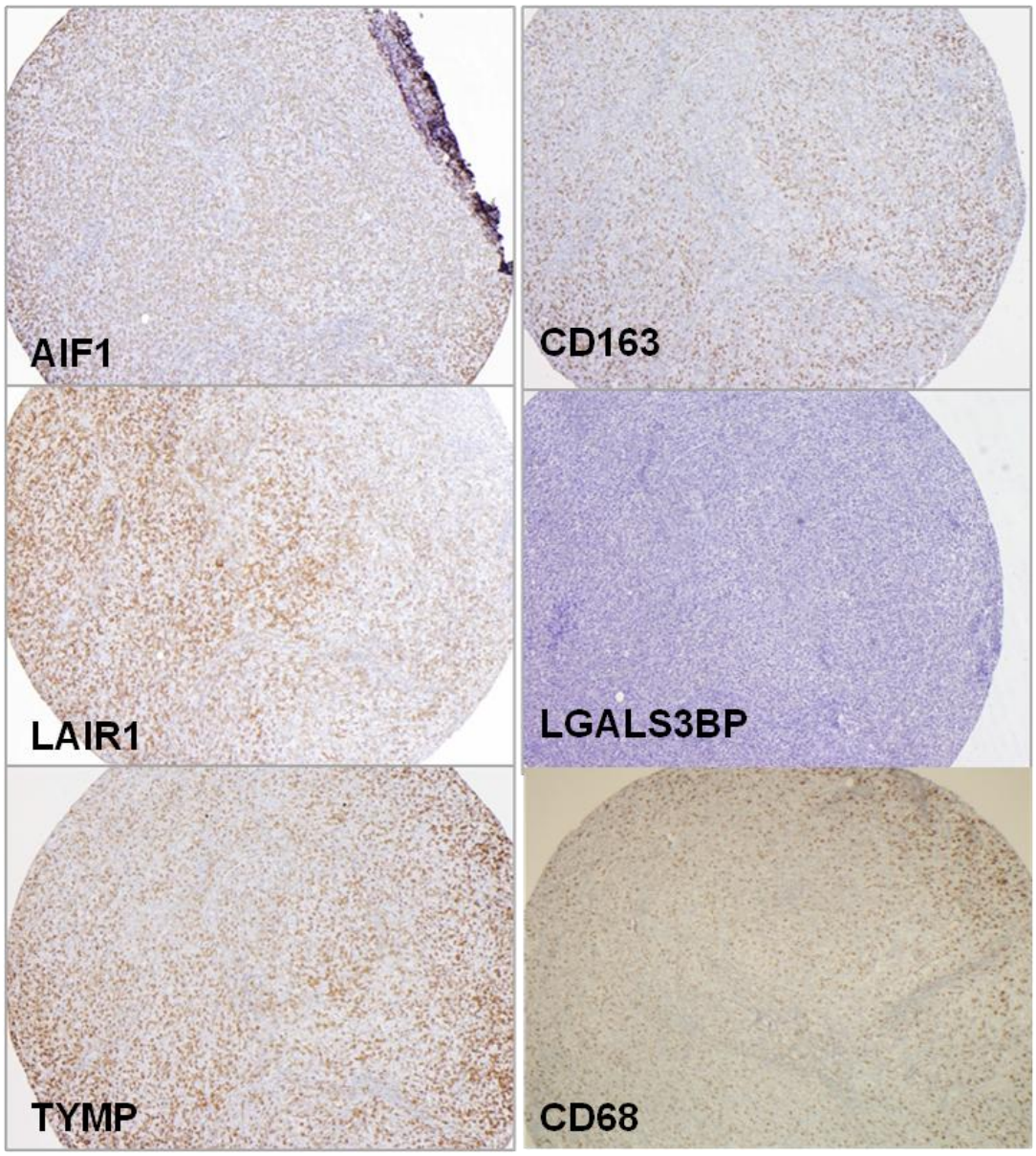


Figure 5.11. Representative example of DLBCL showing pattern of distribution of 6 makers studied.

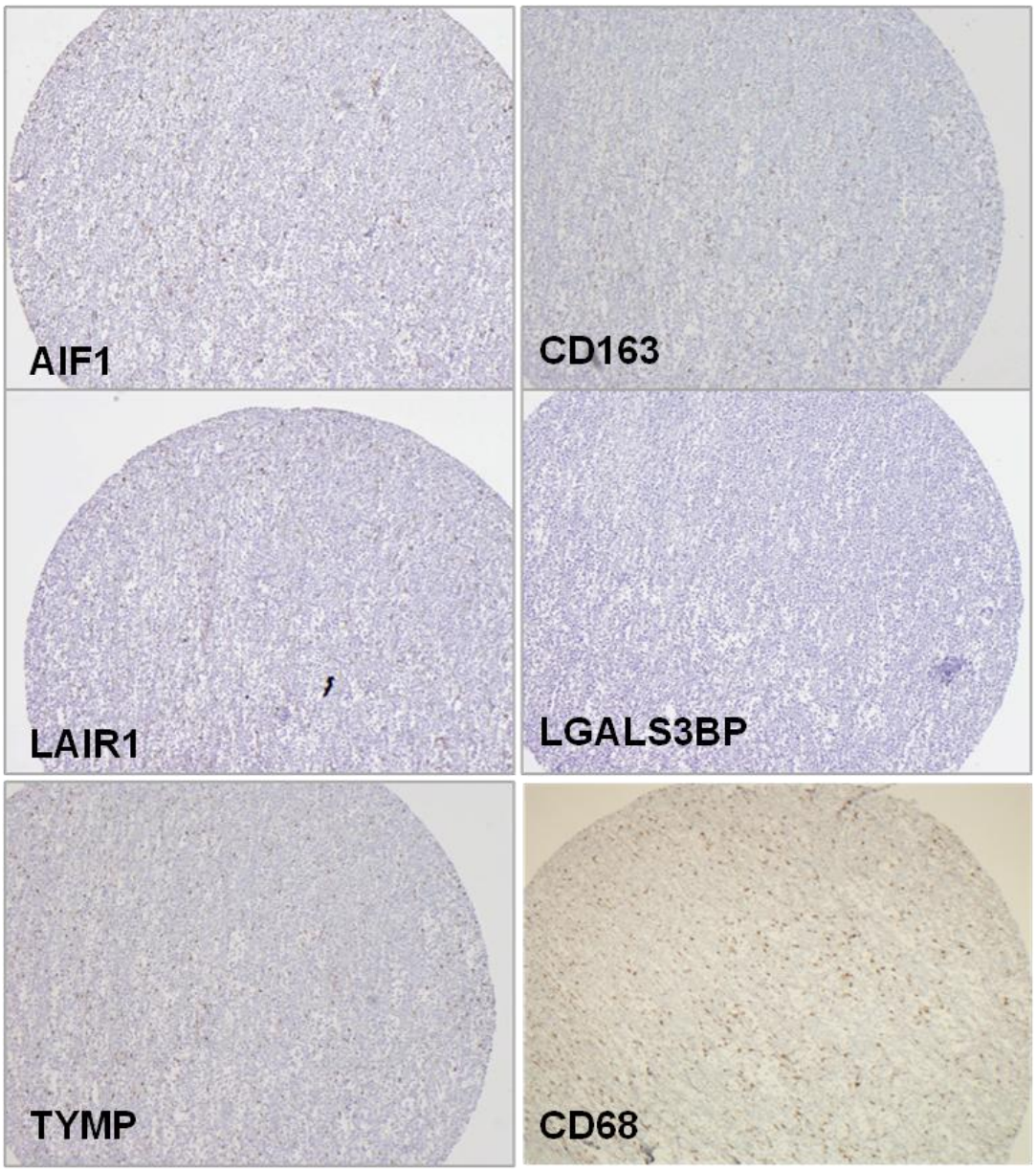


Figure 5.12. Further representative example of DLBCL staining in sequential sections.

5.2.8 Double labelling of macrophage markers

To determine whether the macrophages markers selected for study in this chapter were being expressed by the same cells, a limited panel of double staining was undertaken. As the archival human tissue available to study was all formalin fixed and paraffin embedded, fluorescence staining was not possible due to high auto-fluorescence of the tissue, so chromogenic labelling was undertaken using two different chromogens. This approach was limited as many of the antigens were expressed in the same intracellular compartments, but staining with CD163, a purely membranous antigen, and TYMP, which showed a cytoplasmic and nuclear expression pattern was possible.

In both the reactive lymph node and in TAMS in DLBCL, some cells with the morphology of macrophages showed double labelling with TYMP and CD163, but there were also cells that expressed one without the other. Given the differences in observed expression pattern described above, this was to be expected with CD163 appearing on sequential sections to have a much more limited pattern of expression than TYMP.

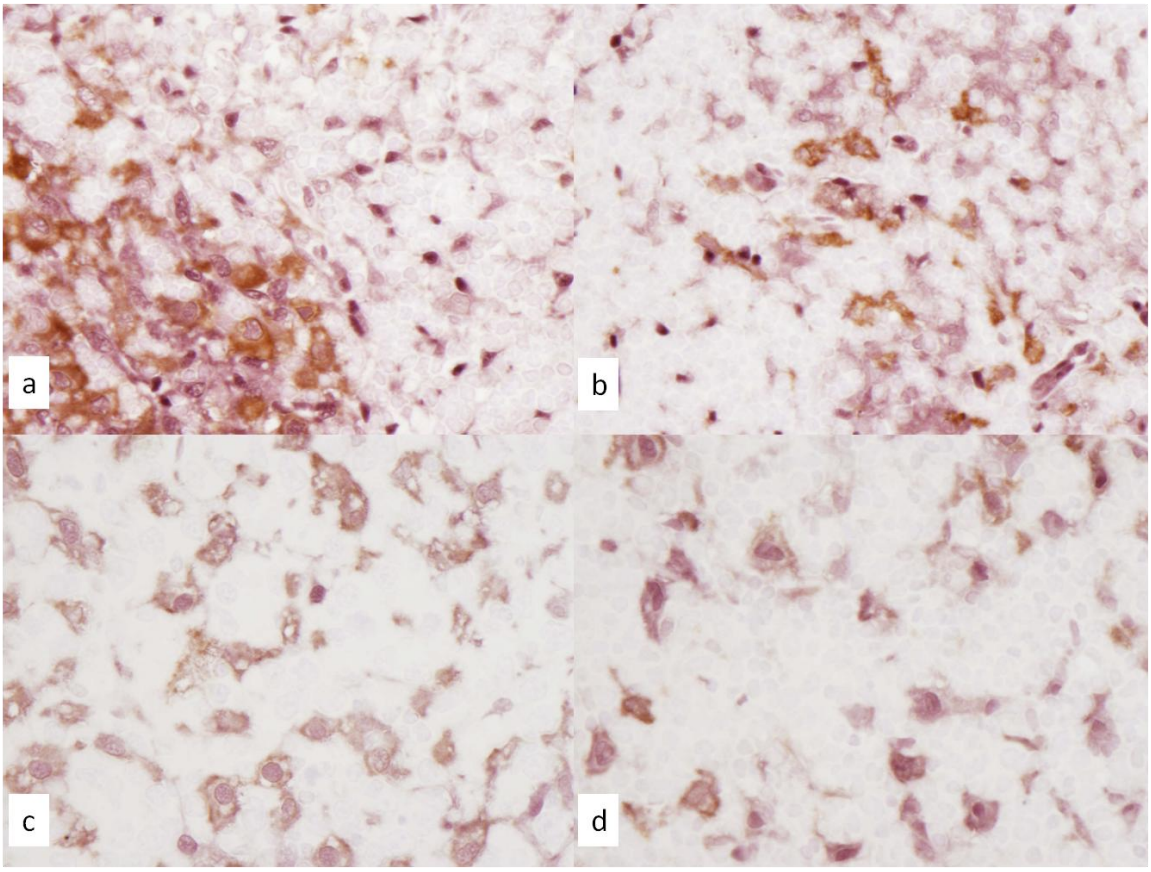


Figure 5.13 Double labelling with TYMP (purple) and CD163 (brown). a) and b) x 200 magnification. Reactive lymph nodes showing more widespread expression of TYMP than CD163 but a proportion of cells showing expression of both. c) and d) x 400 magnification. TAMs in DLBCL, with some dual expression of TYMP and CD163.

Conclusions

All markers of macrophages selected from the 'macrophage' signature generated in the previous chapter, and in which immunohistochemistry was successfully performed were shown to be expressed in macrophages in both benign and malignant conditions, although in the case of LGALS3BP, expression was seen in only in lymph nodes from some individuals. Not all markers were expressed by all macrophages with clear differences in expression patterns in benign lymph nodes between macrophages in different anatomical compartments of the node and therefore presumably with different functions. This same pattern was observed, though with less obvious spatial distribution, in the macrophages in DLBCL. In sequential sections, there appeared to be differing patterns of expression, although all

markers were expressed in cells with macrophage-like morphology. The double immunolabelling highlighted this also, with a proportion of cells being positive for one of other of the proteins and a proportion being double-positive.

This would imply that the signature derived from the data analysis is a 'global' signature of the macrophages within the tumour, with sufficient similarities between cells for them to cluster together, regardless of whether all the markers in the signature are expressed by individual cells. It is to be expected the microenvironment of the tumour is not a uniform space, but reflects minute differences in the cellular composition, blood supply, and extracellular matrix of the tumour as well as local differences in availability of oxygen and nutrients. Macrophages as cells of marked plasticity would be expected to reflect these environmental changes. It also implies that these elements of the signature are not simply genes that are expressed at all times in all macrophages, but rather are modulated in response to stimuli and therefore, of potential interest for further study. Given the complexity of the tumour microenvironment and the plasticity of the macrophage, there is unlikely to be a single tumour associated macrophage signature in DLBCL or any other malignancy, but rather a collection of signatures that add up to the TAM signature and reflect the balance of activation and suppressive influences on each individual cell.

Chapter 6

Relationship of expression of selected TAM markers with clinical outcome in DLBCL

Aims of the chapter.

Having established that selected markers from the macrophage RNA expression profile were indeed present in human macrophages at the protein level, both in macrophages in reactive lymphoid conditions and in macrophages in DLBCL, this chapter will focus on whether differences in the pattern and intensity of staining in TAMs and differences in absolute numbers of those macrophages correlates with outcome in a cohort of patients treated with a standard chemotherapeutic regimen. This is an independent cohort of patients from those discussed in chapter 3. The clinical characteristics and demographic details of the patient cohort will be presented, then each of a panel of markers whose known functions suggest they may play a role in tumour response and their relationship to clinical outcome will be discussed. For each protein studied, the pattern of antibody staining was assessed by eye to determine whether there appeared to be differences in intensity, expression pattern or number of positive cells between tumours and variables selected for scoring on this basis. Four molecules were studied in relation to clinical outcome; Thymidine phosphorylase (TYMP), leukocyte associated Ig-like receptor-1 (LAIR1), allograft inflammatory factor 1 (AIF1) and soluble galectin-3 binding protein (LGALS3BP).

6.1 Clinical features of DLBCL – CHOP treated cohort

A cohort of patients were identified from the Scotland and Newcastle Lymphoma database. The initial selection criteria to identify patients from the database were

that they had a diagnosis of Diffuse Large B-cell lymphoma, and that they were treated at one of the three hospitals in Lothian to which there was access to the stored formalin-fixed paraffin embedded tissue on which the diagnosis was based. This identified 276 cases spanning a period from 1991 to 2002. From this initial cohort only those patients treated with CHOP chemotherapy (cyclophosphamide, doxorubicin, vincristine and prednisolone), which was the standard treatment at this time, were included to create a cohort of patients in which any differences in outcome were not due to differences in chemotherapeutic regimen.

One hundred and sixty five patients were identified who had received CHOP chemotherapy. Diagnostic material from these patients was reviewed and cases excluded on any of the following grounds; formalin-fixed paraffin embedded material was missing from the pathology archives or there was insufficient diagnostic material present to allow sampling for inclusion in a TMA and preservation of material for future clinical use.

Seventy four cases were identified suitable for inclusion in a TMA. Clinical outcome data was available for 71 of these. The clinical features and survival data for these cases is given in Table 6.1. The patients ranged in age from 25 to 79 years. There were similar numbers of males and females (37 males, 34 females). The majority of patients (52 of 71) fell into ECOG performance status 1 or 2, indicating that they were ambulatory and self-caring , with only 12 patients falling in ECOG groups 3-5, a bias reflecting the fact that this is a cohort of patients selected on the basis that they are fit enough to undergo chemotherapy. Treatment success was defined as ongoing complete or partial remission with treatment failure defined as relapsed disease or death. Forty-one of seventy one patients relapsed or died in the period of follow up. For the patients in the treatment failure group, the period of follow up lasted from 1.7 to 86 months with a median of 24 months. In the treatment success group follow-up was from 10 to 141 months with a median of 47 months.

For patients who relapsed or died, the mean time to treatment failure was 11.8 months, with a median of 0 months. The mean time to death in this group was 26.5 months with median of 21 months. (See Table 6.2).

Characteristic	Classification	Total group (n=71)	Treatment failure (n=41)	Treatment success (n=30)
		Number (Percentage)	Number (Percentage)	Number (Percentage)
Sex	Male	37 (52)	22 (54)	15 (50)
	Female	34 (48)	19 (46)	15 (50)
Age (in years)	Mean [Range]	58.3 [25-79]	61.7 [34 - 79]	53.6 [25 – 77]
	Median [IQR]	61 [50.5-69.5]	63 [56 - 71]	57 [41 – 66]
Clinical stage (Ann Arbor stage)	1	19 (27)	6 (15)	13 (43)
	2	15 (21)	8 (20)	7 (23)
	3	22 (31)	17 (41)	5 (17)
	4	15 (21)	10 (24)	5 (17)
ECOG Grade (Eastern Cooperative Oncology Group Perfomance Status)	1	32 (44)	15 (37)	17 (57)
	2	20 (28)	14 (34)	6 (20)
	3	7 (10)	3 (7)	4 (13)
	4	4 (6)	3 (7)	1 (3)
	5	1 (1)	1 (2)	0 (0)
	Not known	7 (10)	5 (12)	2 (7)
Bulk disease	Yes	22 (31)	14(34)	8 (27)
	No	44 (62)	26 (63)	18 (60)
	Not known	5 (7)	1 (2)	4 (13)
Extranodal involvement	Yes	26 (37)	16 (39)	10 (33)
	No	42 (60)	22 (54)	20 (67)
	Not known	3 (4)	3 (7)	0 (0)
Marrow Involvement	Yes	7 (10)	5 (12)	2 (7)
	No	59 (83)	33 (80)	26 (87)
	Not known	5 (7)	3 (7)	2 (7)
International Prognostic Index (IPI)	1	20 (28)	5 (12)	15 (50)
	2	14 (20)	12 (29)	2 (7)
	3	8 (11)	6 (14)	2 (7)
	4	1 (1)	0 (0)	1 (3)
	Not known	28 (39)	18 (44)	10 (33)
Treatment failure	Failure	41 (59)		
	Success	30 (42)	-	-
Overall survival	Dead	40 (56)		
	Alive	31 (44)	-	-

Table 6.1. Demographic data for group as a whole, then subdivided by the basis of outcome.

	Mean (Range)	Median (Interquartile range)
Time to treatment failure in months	11.8 (0 - 72)	0 (0 - 14.3)
Time to death in months	26.5 (1.7 – 86)	21 (6.2 – 42.4)

Table 6.2. Outcome data for patients in whom treatment failed (defined as relapse of disease or death).

6.2 TYMP

6.2.1 Introduction to TYMP

Thymidine phosphorylase (TYMP) is an evolutionarily conserved molecule, originally isolated from bacteria, that plays a vital role in pyrimidine salvage, converting thymidine and phosphate to thymine and 2-D-deoxyribose- 1- phosphate (reviewed in Bronckaers et al 2009). It was independently identified by several groups in different settings and was also given the alternative names of platelet-derived endothelial cell growth factor and gliostatin (Friedkin and Roberts, 1954, Miyazono et al, 1987, Asai et al, 1992) The name platelet derived endothelial cell growth factor hints at its role in angiogenesis, although the name is a misnomer, based on initial experiments carried out before the molecule was recognised as thymidine phosphorylase in which endothelial cell proliferation was measured using incorporation of radioactive thymidine. Subsequent work has show TYMP to be a chemotactic factor for endothelial cells rather than a growth factor (Moghaddam and Bicknell 1992, Ishikawa et al 1989).

Angiogenesis is a relatively uncommon phenomenon in the adult being confined to certain physiological and pathological processes including the endometrial growth cycle, inflammation, wound healing and neoplasia. While TYMP has been shown to be an important angiogenic molecule in endometrial proliferation much of the work on TYMP has been in relation to tumorogenesis (Zhang et al 1997, Fujimoto et al 1998, Moghaaddam et al 1995, Arima et al 2000, O'Brien et al 1996, Yoshikawa et al 1999, Takebayashi et al 1996a, Takebayashi 1996b, O'Byrne et al 2000, Takebayashi et al 1999, Fujimoto et al 1999, Yamamoto et al 1998, Mainou-Fowler et al 2006). Tumours as rapidly growing tissues require to lay down new blood vessels to progress beyond a very small size. In the absence of effective angiogenesis, tumours will fail to progress or infarct when they outgrow their blood supply. TYMP appears to be regulated by a variety of factors including hypoxia, radiation damage and inflammation (Griffiths et al 1997, Blanquicett et al 2002, Sawada et al 1998, Sawada et al 1999). Interferon gamma induces TYMP expression

and indeed, there is an interferon response element in the TYMP promoter (Schwartz et al 1998).

In human tumours TYMP has been postulated to have a dual role in promoting angiogenesis and inhibiting apoptosis. Much of the work on TYMP in human tumours has been correlative, looking at the association between TYMP expression and prognosis, with the majority of studies in the majority of types of tumour showing an association between high levels of TYMP and poor outcome (reviewed in Bronckaers et al 2009). Some studies have also addressed the putative angiogenic role of TYMP in tumours by demonstrating an association between high levels of TYMP and microvessel density (Relf et al 1997, Kawahara et al 2010, Morita et al 2001). TYMP expression can be seen to correlate with expression of other angiogenic factors, notably VEGF, but data from HIF1 and HIF2 knockout mice, suggest that these may be under different regulation, although both are responsive to hypoxia (Toi et al 1996, Brown et al 1995, O'Byrne et al 2000, Sivridis et al 2002). In many of the earlier studies of TYMP in human tumours, expression of TYMP was analysed by RT-PCR or enzyme activity. More recent studies have shown that in many of these tumours, the expression of TYMP is not by the tumour itself, but by TAMs implying a role for the microenvironment of the tumour in driving angiogenesis (Kim et al 2009, Kawahara et al 2010). Data from a variety of sources have shown that the enzymatic activity of TYMP is essential for its role in angiogenesis, implicating a product of this enzymatic activity as the key signalling molecule, although the exact nature of the signal remains unclear, as the putative active molecule 2-deoxy-d-ribose does not appear to bind to a surface receptor (Moghaddam et al 1995, Miyadera et al 1995, Brown and Bicknell 1998, Hotchkiss et al 2003).

TYMP however, also seems to have functions unrelated to angiogenesis showing the ability to inhibit apoptosis in various malignant cell lines. There appears to be several facets to this, with different data suggesting the TYMP enzymatic activity is required or not required dependent on the mechanism driving the apoptosis (reviewed in Akiyama et al 2004). It seems though that in certain settings in vitro, TYMP can exerting a direct anti-apoptotic effect via its metabolite 2DDR on tumour

cells, in an autocrine fashion (Ikeda et al 2006). The significance, if any, of this role in vivo in unclear, but given that in most tumours, the majority of TYMP expression is in the TAMs, it could be hypothesised that a similar paracrine effect could be seen with TAMs suppressing apoptosis of tumour cells.

In treated tumours TYMP has an additional role to that which it may play in tumorogenesis; TYMP may be induced by treatment with radiotherapy and various chemotherapeutic compounds causing upregulation of expression (Kim et al 2009, Liekens et al 2007) . In addition there are a class of chemotherapeutic agents whose action relies on conversion of the pro-drug to the active form by the action of TYMP (reviewed in Leikens et al 2007). Thus in cancer TYMP may be a double-edged sword helping promote cancer growth, but providing an avenue for chemotherapeutic attack. This potential weakness is being targeted by current clinical approaches, with strategies designed to upregulate TYMP expression prior to commencement of chemotherapies reliant on TYMP activity. Several of the drugs used in the treatment of DLBCL rely on the activity of TYMP, but there is currently no data as to whether TYMP expression is a good or adverse prognostic factor in DLBCL.

TYMP was present in the TAM profile on gene expression analysis. It can be hypothesised that in TAMs TYMP plays a dual role in supporting tumour cell growth by both promoting angiogenesis and protecting tumour cells from apoptosis. In this hypothesis, expression of TYMP by TAMS should be a marker of poor prognosis.

6.2.2 Scoring of TYMP expression

Using the TMA of DLBCL cases, expression of TYMP by TAMs was assessed by a single observer blinded to the clinical outcome. The total number of TYMP-positive macrophages in a set area of 3HPF was counted. Of the seventy four cases present in the array, data was obtained for 55 cases. Cases were excluded if all replicates of the case were lost during the staining protocol, or if too little tumour was represented to allow counting of 3 HPF. Based on distribution of cases on a dot plot, the cases

were divided into high or low numbers of TAMs with a cut off value of 300 positive cells (Figure 6.1).

		Time to treatment failure in months Mean (± SE) *Median (± SE)*	Overall survival in months Mean(± SE) *Median (± SE)*
Number of positive cells	High (n=31)	63 (± 12) *34 (± 13)*	81 (± 11) *86 (± 25)*
	Low (n=24)	49 (± 11) *36 (± 6)*	61 (± 10) *56 (± 7)*

Table 6.3 TYMP expression by macrophages in DLBCL and outcome in a CHOP-treated cohort of patients. A TMA containing 74 cases was stained with anti-TYMP antibody (HPA001072) and the number of positive TAMs stained in 3 HPFs was counted and divided into high or low numbers of TYMP-positive TAMs with a cut-off value of 300 cells/3hpf.

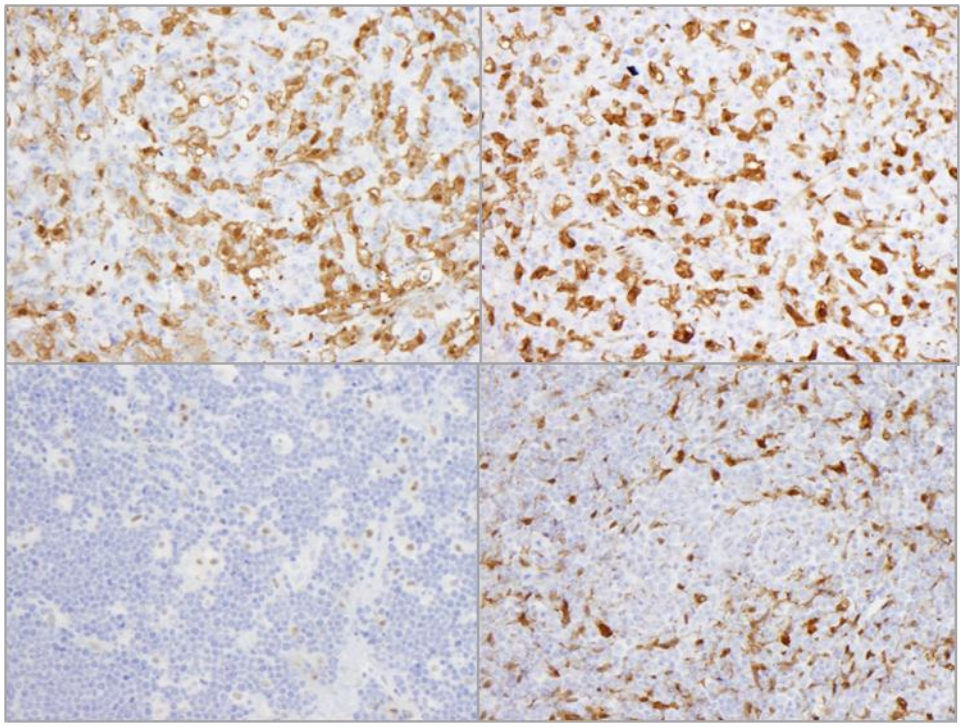


Figure 6.1 Sections of TMA were stained with anti-TYMP antibody (HPA001072) and number of positive cells counted in a set area (3HPF). There was variation in staining patterns between tumours with differential staining in nucleus and cytoplasm. No difference

in outcome was detected between tumours based on expression pattern, or staining intensity (data not shown). All images x100 magnification.

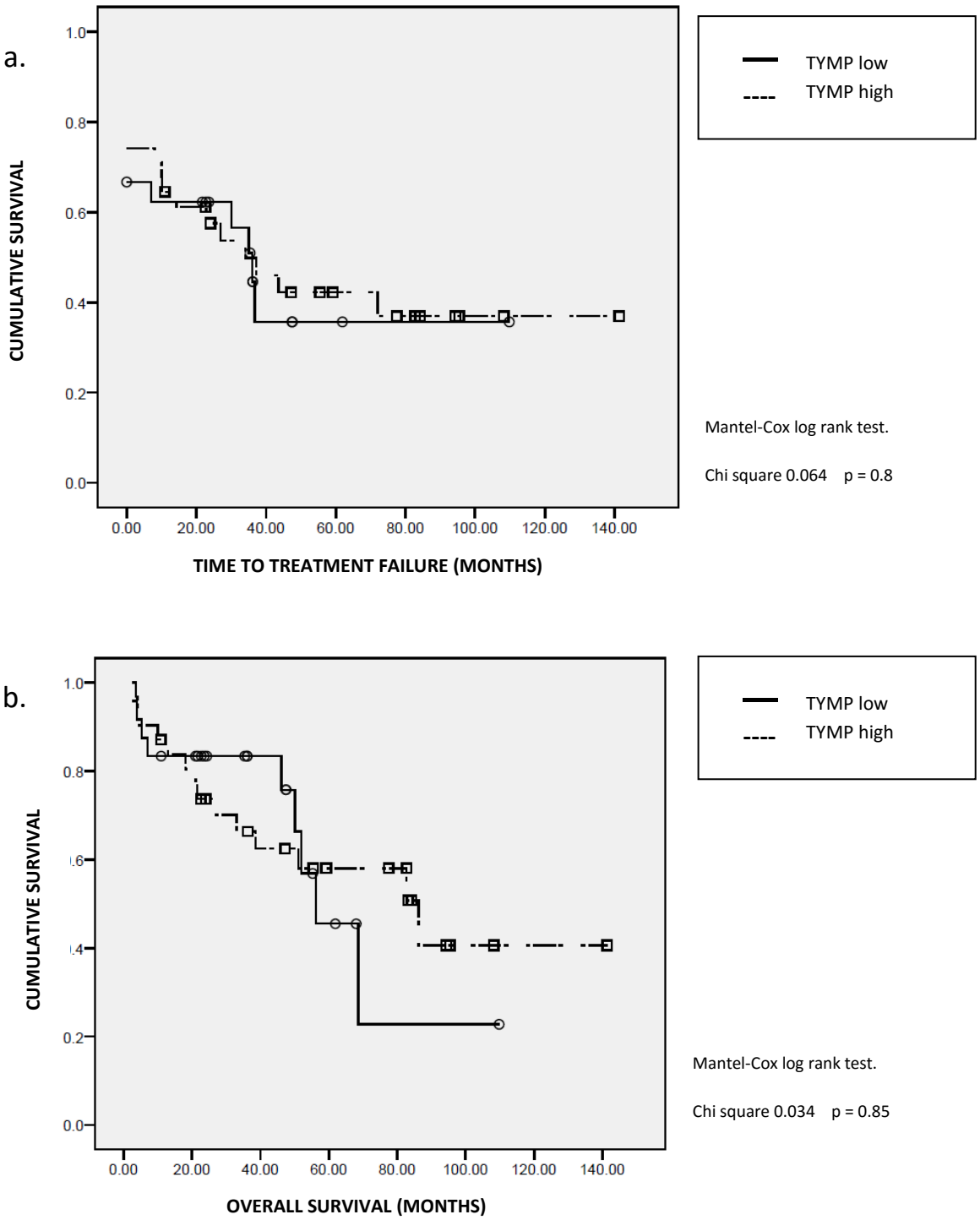


Figure 6.2 The number of TYMP expressing macrophages was counted in a set area (3HPF) in a cohort of DLBCL treated with CHOP chemotherapy. Cases were divided into those with high or low numbers of TYMP positive cells and Kaplan-Meier survival curves plotted for a) time to treatment failure and b) time to death.

6.2.3 Relationship to clinical outcome

Outcome, assessed both in terms of time to treatment failure and overall survival, was calculated using the data of the SNLG database dividing cases into those with high or low numbers of TYMP-expressing TAMs. There was no difference between the two groups on Kaplan-Meier survival curves for either disease-free survival or overall survival with a mean time to treatment failure of 63 months in the TYMP-high group and a mean time to treatment failure of 49 months in the TYMP-low group (Figure 6.2, Table 6.3). Mean overall survivals were 81 and 61 months respectively.

In this cohort of tumours, there was no observed difference in outcome between tumours with high or low numbers of TYMP positive TAMs, or differences related to expression pattern or staining intensity implying that in DLBCL the presence of increased numbers of TYMP-expressing TAMs does not predict aggressive behaviour. In considering the effect that TYMP expression has on tumour outcome, the role of chemotherapy also needs to be considered, the pyrimidine salvage pathway being important in the action of several chemotherapeutic drugs, including those used in CHOP chemotherapy. Given that all the patients in this cohort were treated with CHOP chemotherapy following the biopsy, the possibility cannot be entirely excluded that there were differences in behaviours of the tumours, which were masked by differences in response to chemotherapy. Given that all patients who are fit enough are treated with CHOP chemotherapy and that this has been the standard therapy for many years, this possibility is currently unassessable on the basis of available archival material.

6.3 AIF1

6.3.1 Introduction to AIF1

Allograft inflammatory factor (AIF) 1 is a 17kD cytoplasmic calcium binding protein which was originally identified in the context of organ transplantation as a factor expressed in both human and rat cardiac allografts in the context of rejection (Utans et al 1995, Utans et al 1996). It lies in the MHC class III region of the genome in humans along with other important molecules in the immune response, such as TNF Alpha, NFκB and complement components (Iris et al 1993, Deininger et al 2002). An evolutionarily conserved molecule, with extensive homology between species, increased expression is seen in sponges when they are allografted (Zhao et al 2013, Muller et al 2002). In humans it appears to be a marker of organ rejection, with differential expression separating patients into those with and without rejection of renal allografts (McDaniel et al 2013). The distribution of the molecule in the normal situation has been studied in the mouse, in which, outwith the testis, it appears in the normal state to be expressed only by cells of the macrophage/monocyte lineage, but is not expressed by all macrophages being absent in alveolar macrophages (Kohler 2007). It appears therefore to be a molecule which is relatively specific to the macrophage lineage, expressed only in certain situations and involved in mounting an immune response in allografts.

The exact mechanism of action of AIF1 is not yet well understood. Expression of AIF1 appears to lead to expression of pro-inflammatory cytokines, and may play a role in endothelial cell activation. Broadly, AIF1 appears to be a marker of a 'pro-inflammatory' state. Studies in different setting yield different outcomes, but expression of AIF1 appears to be upregulated by various pro-inflammatory molecules such as IFNγ, IL-1, TNFα and oxidised lipoproteins and upregulation of the molecule leads to increased expression of iNOS, IL-6, IL-12 and other such pro-inflammatory mediators (Watano et al 2001, Yang et al 2005, Tian et al 2006, Utans et al 1995, Yan and Chen 2010, Nagawaka et al 2004). Expression of AIF1 by macrophages decreased levels of macrophage apoptosis and increased the migratory capacity of the cells (Yang et al 2005). Conversely knockdown of AIF1 expression

decreased macrophage migration, inflammatory mediator expression and survival (Yang et al 2005). Such a tight relationship of expression with the function of the activated macrophage argues that AIF1 expression represents a fundamental component of macrophage activation rather than a molecule expressed simply as a non-essential bystander.

Outwith the transplant situation, there is some further data as to the significance of the molecule in the immune response. It appears to play a role in autoimmune disease with polymorphisms in the molecule associated with systemic sclerosis and a variant form of the molecule associated with Type 1 diabetes mellitus in Norwegian families (DelGaldo et al 2006, Alkassab et al 2007, Eike et al 2009). It has also been detected in the synovium of patients with rheumatoid arthritis, but not those with osteoarthritis (Kimura et al 2007). AIF1 expressing macrophages have been identified in other diseases as well, with AIF1 expression seen in macrophages in the atherosclerotic plaque, where activation appears to be secondary to the phagocytosis of oxidised LDL and where there is evidence for a role in promoting the proliferation of vascular smooth muscle cells (Tian et al 2005, Mishima et al 2008, Somerville et al 2012).

Very little is known of its role, if any, in cancer. It is expressed by a subset of microglial cells and macrophages in glioma, suggesting that in this setting either expression is regulated, or a specific subset of cells is recruited to the tumour site (Deininger et al 2000). Higher numbers of AIF1 positive cells are seen in higher grade tumours, but given the propensity for higher grade tumours of any histological subtype to contain larger numbers of macrophages it is unclear how specific this effect is (Deininger et al 2000). Outwith the setting of the tumour associated macrophage, it has been shown to be expressed by malignant epithelium in breast cancer and expression correlated with higher grade and stage (Liu et al 2008). In vitro expression in breast cancer cells has been shown to increase proliferation of those cells by an undetermined mechanism (Liu et al 2008). In contrast, recent data from gastric carcinoma indicates that lower levels of expression in epithelium correlates with poorer outcome (Ye et al 2014). In giant cell tumour of bone, an unusual but usually benign tumour possibly derived from cells of the

monocyte/macrophage lineage, higher levels of AIF1 correlate with more aggressive behaviour (Conti et al 2011).

AIF1 appears to represent a macrophage-expressed molecule with features of 'classical' activation, with a role in autoimmune disease and transplant rejection but little studied in other settings. Whilst there is little mechanistic data as to control of expression and downstream effectors, it can be hypothesised that increased expression of AIF1 expression macrophages in tumours would represent evidence of a 'classically ' activated, rejecting macrophage and as such would be associated with better prognosis.

6.3.2 Scoring of AIF1 expression

Using the TMA of DLBCL cases, expression of AIF1 by TAMs was assessed by a single observer blinded to the clinical outcome. Cases were excluded if all replicates in the TMA were lost on processing. Staining was assessed on a 4 point visual scale by eye (0 for no staining, through to 3 for strong staining). Cases were then further grouped for mean staining intensity ≤ 2 or >2 and relationship to clinical outcome sought. Cases were also assessed by staining pattern and were divided into those with and without nuclear expression of AIF1.

6.3.2 Relationship to clinical outcome

Intensity of AIF1 staining in TAMs was not associated with difference in outcome measured either as disease free survival or overall survival with cases with high intensity staining having a mean disease-free survival of 48 months compared to the 62 months in the low intensity staining group. Mean overall survivals were 68 and 84 months respectively (Table 6.4). There were no differences in the Kaplan-Meier survival curves between groups (Figures 6.3 and 6.4). While there was a shorter disease free survival and overall survival in those patients' whose TAMS showed higher expression of AIF1, a relationship at odds to that hypothesised, this did not

reach statistical significance. Overall, this data does not suggest that AIF1 plays a significant role in the support or suppression of DLBCL growth by TAMS. An interesting relationship was also observed with the pattern of expression of the protein. In those patients in whom the macrophages showed expression of AIF in the nucleus as well as the cytoplasm, there were no deaths due to disease. This represented a very small group of patients (11 patients in total), and again was not statistically significant. A review of the literature does not produce any data as to nuclear expression of AIF1. It is possible that the antibody used also binds an unidentified nuclear protein, but it could also be hypothesised that an alternative variant of AIF1 exists that can translocate to the nucleus. The nature of the molecule being detected in the nucleus in these tumours was not explored further.

		Time to treatment failure in months **Mean (± SE)** ***Median (± SE)***	**Overall survival in months** **Mean(± SE)** ***Median (± SE)***
Intensity of staining	High (n=32)	48 (± 13) *34 (± 13)*	68 (± 8) *82 (± 18)*
	Low (n=26)	62 (± 9) *36 (± 7)*	84 (± 13) *69 (± 0)*
Nuclear staining	Yes (n=11)	90(± 20) -	101 (± 18) -
	No (n=48)	44 (± 7) *30 (± 12)*	66 (± 16) *69 (± 16)*

Table 6.4. Intensity of staining for AIF1 in macrophages in DLBCL and outcome in a CHOP-treated cohort of patients. A TMA containing 74 cases was stained with anti-AIF1 antibody (Ab5076) and the scored by eye on a 4 point scoring system (0-4). Cases were divided into two categories with cases with mean staining intensity >2 allocated to the high group. Cases were also categorised on the basis or presence/absence of nuclear staining.

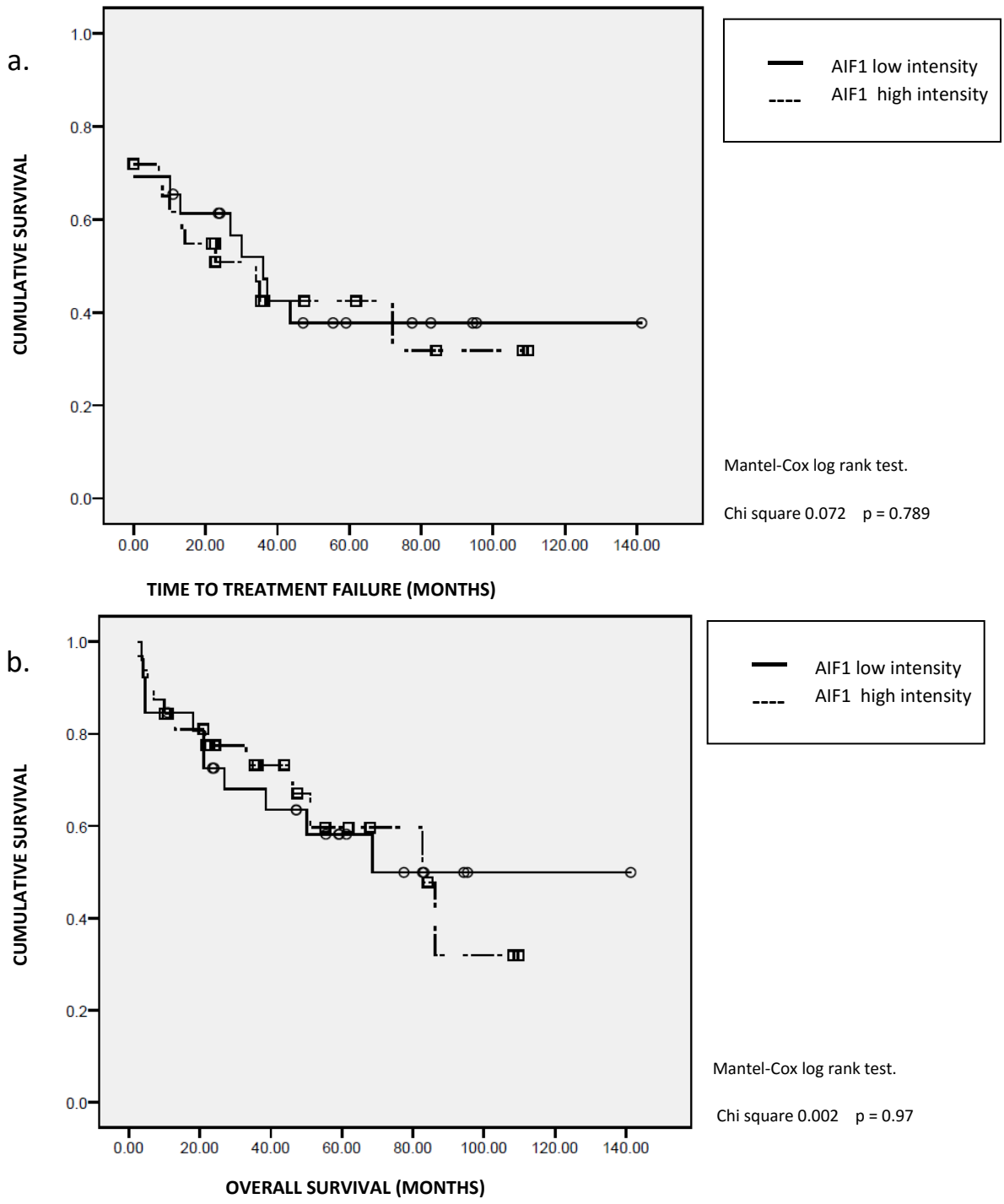


Figure 6.3 The intensity of staining of TAMs in a cohort of patients with DLBCL treated by CHOP chemotherapy was scored visually by a single observer and cases divided into those with high or low intensity staining regardless of the distribution of staining within the cell. Kaplan-Meier survival curves were plotted for a) time to treatment failure and b) time to death and demonstrated no differences between the groups.

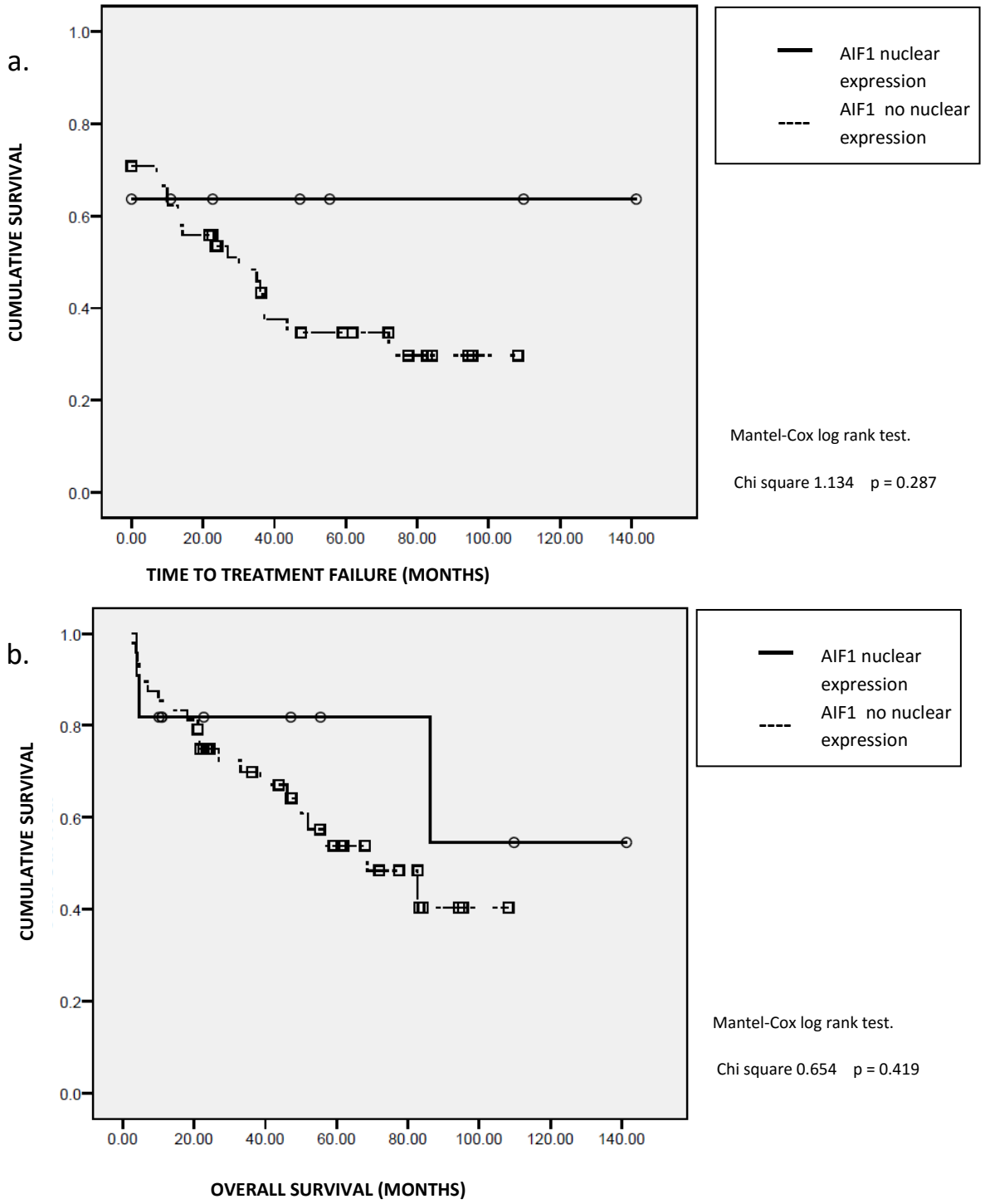


Figure 6.4 The localisation of staining for AIF1 in TAMs was assessed in a cohort of cases of DLBCL treated by CHOP chemotherapy. A small group of cases demonstrated nuclear staining for AIF1, and within this group in those patients who did not have primary treatment refractory disease, there were no cases of treatment failure. This difference did not meet statistical significance. Kaplan-Meier survival curves plotted for a) time to treatment failure and b) time to death.

6.4 LGALS3BP

6.4.1 Introduction to LGALS3BP

Galectin-3 binding protein (LGALS3BP), also known as Mac-2 binding protein or 90K, was originally identified in the serum of patients with cancer and postulated to be a cancer secreted molecule (Iacobelli et al 1986, Scambia et al 1988, Ullrich et al 1994). It has been identified in the serum of patients with a wide variety of malignancies and is usually a marker of poor prognosis (Fornarini et al 2000, Fusco et al 1998, Iacovazzi et al 2001, Marchetti et al 2002, Strizzi et al 2002). It was subsequently shown to also be detectable in the serum of patients with HIV progressing to AIDS in the absence of neoplastic complications of AIDS suggesting its role was wider than simply being a biomarker of malignancy (Iacobelli et al 1991, Natoli et al 1993). The protein is of 567 amino acids, is subject to extensive post-translational glycosylation and has a molecular weight of between 90 and 95 KDa dependent on glycosylation status (Koths et al 1993, Ullrich et al 1994). The protein appears to exist in serum as a large oligomeric complex (Sasaki et al 1998). It contains a scavenger receptor motif, but unlike many other molecules with this motif which are membrane based has the structural characteristics of a secreted protein. The protein has been identified in the malignant cells in many types of tumour, but is also expressed by cells of the immune system including macrophages (Becker et al 2008, Ulmer et al 2006).

The mechanisms of actions of the molecule are poorly understood and much of the data relates to the secreted form of the molecule. As well as galectin-3, this appears to bind to multiple components of the extracellular matrix and to integrins expressed on cells and is postulated to play a role in tumour metastasis by providing anchorage for circulating tumour cells, protection from apoptosis of tumour cells by the direct action of binding of tumour cells to the molecule and promotion of angiogenesis by VEGF-dependent and independent pathways (Inohara et al 1996, Piccolo et al 2012, Sasaki et al 1998, Stampolidis et al 2015). In HIV, expression of the molecule by cells of the monocytic lineage renders them less susceptible to infection by HIV-1 and promotes expression of IL-1 and IL-6 by the cells (Lodermeyer et al 2013).

Expression in this context is upregulated by IFN gamma. Expression of IL-6 has also been seen in bone marrow stromal cells on binding of LGALS3BP (Fukaya et al 2008).

The relatively limited, but suggestive data as to the function of this molecule, a poor prognostic indicator in many types of cancer in serum, but whose expression at a local level influences macrophage function made this an interesting molecule to study further. Is this an immune-related molecule whose normal function has been subverted by tumour cells to promote survival and metastasis? Of interest also was the very little functional data relating to this molecule in lymphoma where binding by Jurkat T lymphoma cells to LGALS3BP protected the cells from chemotherapy induced apoptosis suggesting a possible role in response to therapy (Fornarini et al 2000).

6.4.2 Scoring of LGALS3BP expression

Using the TMA of DLBCL cases, expression of LGALS3BP by TAMs was assessed by a single observer blinded to the clinical outcome. Cases were excluded if all replicates in the TMA were lost on processing. Fifty nine cases were available for assessment. LGALS3BP was expressed in macrophages in only a small number of tumours, and where present often showed only weak staining (Table 6.5 and Figure 6.5). Cases were assessed only for presence or absence of staining in TAMs.

6.4.3 Relationship to clinical outcome

There was no relationship between expression of LGALS3BP by TAMS and outcome with the mean time to treatment failure being 54 months in the cohort with positive staining in TAMs and 57 months in the cohort with negative staining in the TMAs. The overall mean survivals in these groups were 67 and 81 months

respectively. The Kaplan-Meier survival curves demonstrated no differences in survival (Figure 6.6).

		Time to treatment failure in months Mean (± SE) *Median (± SE)*	Overall survival in months Mean(± SE) *Median (± SE)*
LGALS3BP expression in macrophages	Positive (n=19)	54 (± 11) *36 (± 20)*	67 (± 10) *68 (± 14)*
	Negative (n=40)	57 (± 10) *30 (± 14)*	81 (± 11) *86 (± 14)*

Table 6.5 Presence or absence of staining in TAMs in DLBCL and outcome in a CHOP-treated cohort of patients. A TMA containing 74 cases was stained with anti-LGALS3BP antibody (HPA000554) and then scored by eye for presence or absence of staining in macrophages.

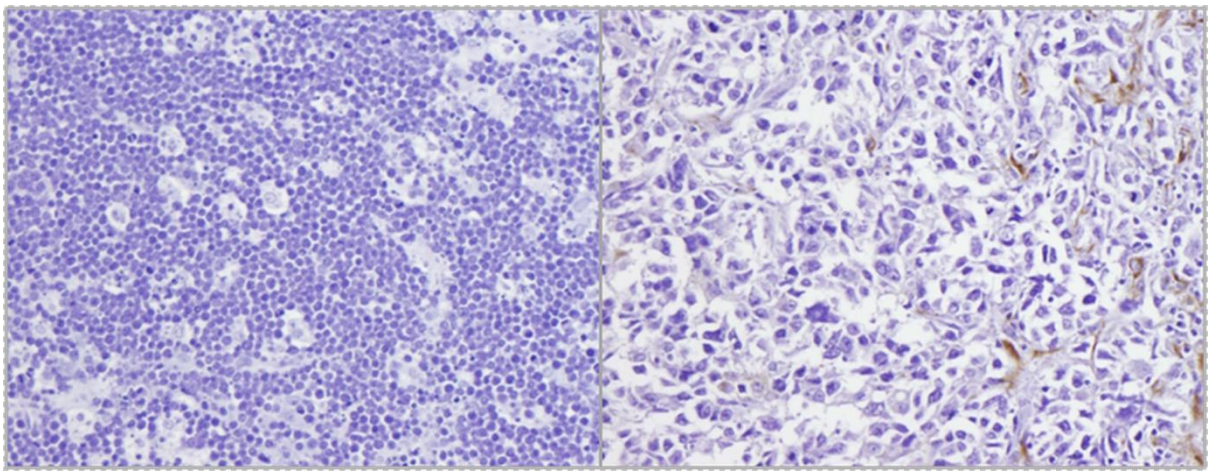


Figure 6.5 Sections of DLBCL were stained with an antibody against LGALS3BP (HPA000554). In many cases there was no expression in TAMs, while a proportion of cases showed staining which was mostly weak and focal (Shown here in macrophages with a rather spindle cell morphology). X200 magnification.

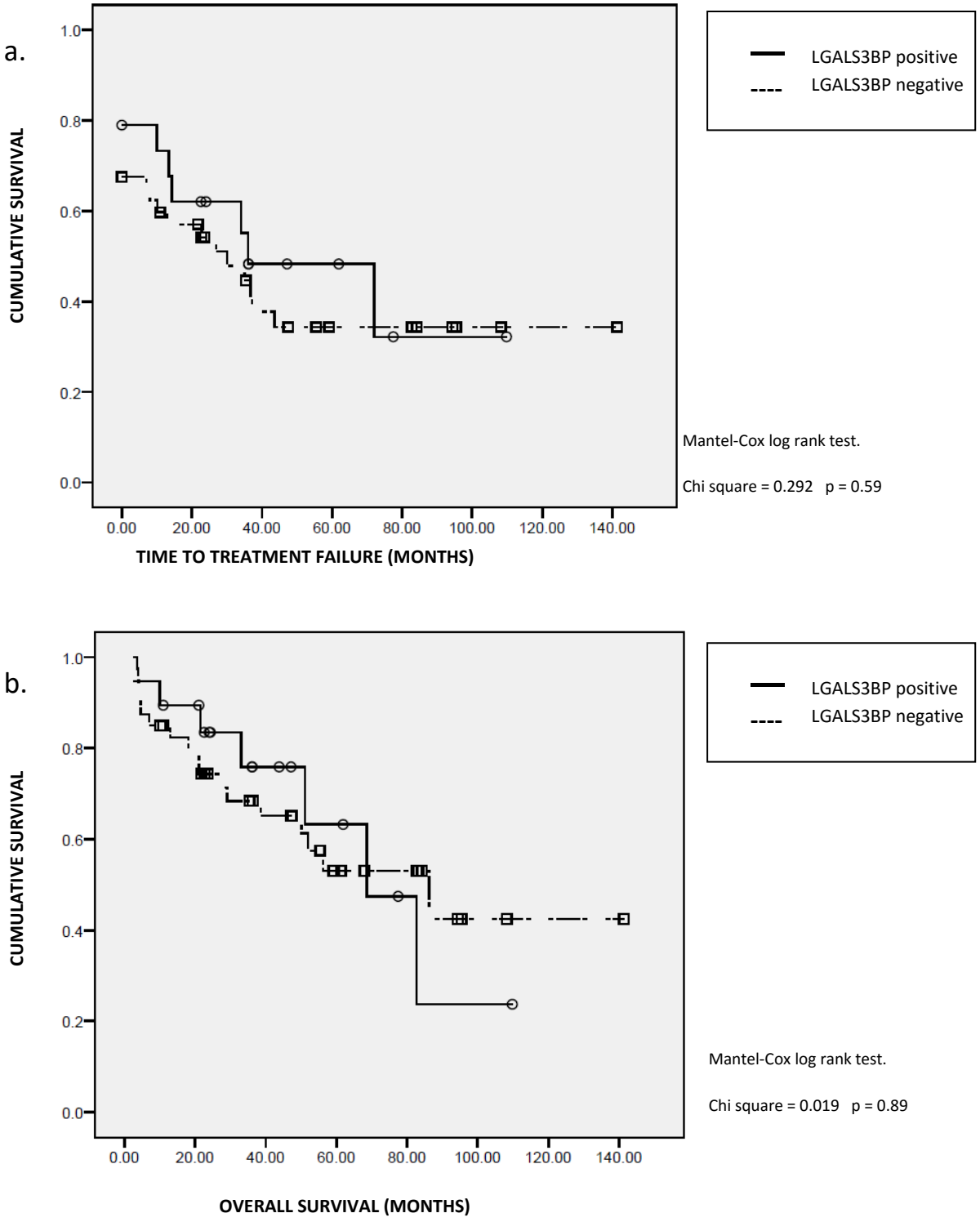


Figure 6.6 The presence of absence of staining for LGALS3BP in TAMs was assessed in a cohort of cases of DLBCL treated by CHOP chemotherapy. A small group of cases showed positive staining in TAM. Kaplan-Meier survival curves plotted for a) time to treatment failure and b) time to death in patients with or without LGALS3BP staining in TMAs demonstrated no difference between the two groups.

6.5. LAIR1

6.5.1 Introduction to LAIR1

LAIR1 is a 32kDa transmembrane protein of the immunogloblulin superfamily. It is a 278 amino acid-containing inhibitory receptor containing two intracytoplasmic ITIM motifs, and a single extracellular C2 immunoglobulin-like domain (Meyaard et al 1997). Homologues have been identified both in mouse and rat (Lebbink et al 2007, Lebbink et al 2005). In humans, but not in rodents there is a closely related molecule LAIR 2, which has a high degree of homology to LAIR1 but which lacks the transmembrane and cytoplasmic domains (Lebbink et al 2008). This gene lies close to LAIR1 but is transcribed in the opposite direction. The LAIR1 protein was first cloned by Meyaard et al in 1997 and its distribution on all peripheral blood mononuclear cells described (Meyaard et al 1997). The inhibitory nature of signalling through LAIR1 was recognised well in advance of the identification of its ligand, with early work focussing particularly on inhibition of tumour cytotoxity by NK cells (Meyaard et al 1997). Recently however, the specific ligand for this receptor has been identified as a motif almost entirely specific to and present on all collagen molecules, a hydroxyproline motif in the form of Gly-Pro-Hyp repeats (Lebbink et al 2006). This ligand, which binds to LAIR1 with high affinity, seems to be involved in the formation of the collagen triple-helix and is present on both extracellular and transmembrane collagen proteins (Lebbink et al 2006).

Collagen binding receptors are widely expressed in human biology including integrins, DDR1 and 2 and the platelet receptor, and the functions encompassed by these receptors is very broad, perhaps reflecting the fundamental importance of collagen in mammalian biology, but to date LAIR1 is the first described inhibitory collagen receptor (reviewed in Leitinger at al 2007). As such, it provides an intriguing insight into the ways in which the structural microenvironment can control immune cell activation in the context of both innate and adaptive immunity. With the exception of while cells of haemopoetic origin are in circulation, almost any conceivable immune interaction is likely to place in the immediate vicinity of collagen.

Initial work on LAIR1 was carried out prior to identification of the receptor, but can now be interpreted in the light of current knowledge. Crosslinking of the LAIR1 receptor by antibodies specific for the receptor has been shown to result in inhibition of NK- cell tumour cell lysis, inhibition of T cell activation by either CD3 cross-linking or antigen presentation, and downregulation of BCR-signalling driven calcium mobilisation and immunoglobulin production (Meyaard et al 1997, Merlo et al 2005). As well as signals mediated through ITAM-containing receptors, LAIR1 also inhibits signalling via cytokines with inhibition of GM-CSF driven differentiation of peripheral blood monocytes to dendritic cells and GM-CSF driven proliferation in primary leukaemia cells (Poggi et al 1998, Poggi et al 2000).

Given the ubiquity of the ligand, it would seem apparent that the downstream signalling from LAIR1 must be more complex than an on-off signal, and the available data supports this: studies in B-cells have demonstrated that the inhibitory signal of LAIR1 can be overcome in the presence of a sufficient activation signal (Merlo et al 2005). In addition, levels of expression of the receptor itself may also play a role. In B-cells, LAIR1 expression is down-regulated as B cells are activated, such that naive B-cells express high levels of LAIR1 compared to memory B-cells or plasma cells, suggesting that the activation threshold of these cells will be lowered (Meyaard et al 2008). Similarly in T cells, expression is high in naive cells but lower and more heterogeneous in subsets of memory cells suggesting active regulation of LAIR1 expression (Maasho et al 2005, Jansen et al 2007). There is limited in vitro data about the effect of T cell activation on LAIR1 expression, but in T cells, CD3 and CD28 activation have been shown to downregulate LAIR 1 expression (Jansen et al 2007), while CD3 and IL-2 stimulation have been shown to upregulate expression (Maasho et al 2005). While the data is limited, it is tempting to speculate that in T cells activation in the absence of an appropriate second signal, leads to upregulation of LAIR1, increasing the threshold for activation. The method by which LAIR1 is regulated in T cell activation has been studied. Regulation in T cells does not appear to be at the mRNA level, or by shedding of the protein as a soluble form, but rather by internalisation of the protein to intracellular stores (Jansen et al 2007).

The mechanisms by which LAIR1 regulates T cell activation have been partially elucidated; Src-homology-2(SH2) domain-containing protein tyrosine phosphatise (SHP-1) is constitutively associated with the intracellular ITIM motifs of LAIR1 in Jurkat cells and in primary human T cells suggestive of a role for LAIR1 in setting the activation threshold of the cells (Sathish et al 2001). On crosslinking of LAIR1 both ITIM-motifs are phosphorylated and this phosphorylation of both motifs is required for optimal LAIR1 activation (Verbrugge et al 2003). The inhibitory effects of LAIR1 signalling in T cells appear related to the surface density of the molecule (Saverino et al 2002). In these studies, the specific cytokine responses of T cells was also analysed, showing not a blanket inhibition of T cell responses but rather different effects dependent on the individual cytokine, with IL-2 and IFNγ production being inhibited in the presence of LAIR1 activation and TGFβ being enhanced (Saverino et al 2002).

In addition to the mechanism described above, there appears to be a role, albeit poorly understood for soluble forms of the receptor. Human LAIR1 has been reported to be found in the supernatant of stimulated PBMC suggesting that the molecule may be shed on activation (Ouyang et al 2004). It is worth noting in the report above where soluble LAIR 1 was measured, that measurement was made using an ELISA technique directed against an extracellular component of the LAIR1 which is identical to the extracellular component of LAIR2. LAIR2 (which lacks a transmembrane and intracytoplasmic domain) has been reported to have an antagonistic effect on LAIR1 signalling (Lebbink et al 2008). Currently there is little evidence to establish whether LAIR 1 is shed during immune activation and whether LAIR2 is secreted on immune activation. Of note though there is no mouse or rat homologue of LAIR2 suggesting that it arose more recently, and is not necessary for the role LAIR1 plays in other mammals. In vivo LAIR 2 has been shown to interfere with LAIR1 binding to collagen (Lebbink et al 2008). The same group also demonstrated that in PMBC stimulated with PMA and ionomycin, CD4 positive T cells, though not other cell types, secreted LAIR2. LAIR2 was also found in large quantities in the urine, though not the plasma of pregnant women, as well as being

found in synovial fluid of patients with rheumatoid though not osteoarthritis, all suggesting a possible role for LAIR2 in immune modulation (Lebbink et al 2008).

In the context of primary human monocytes and macrophages, there is almost no data as to the expression of LAIR1 and the effect of activation save for a single study in which CD14+ precursors were driven in the direction of dendritic cell differentiation, with a resultant downregulation of LAIR1 (Poggi et al 1998).

There is limited data from acute myeloid leukaemia (AML) cell lines, suggesting a role of LAIR 1 in myeloid cell differentiation. A study by Poggi et al (Poggi et al 2000) suggested that signalling through LAIR1 in AML cell lines induced apoptosis. In primary myeloid leukaemia cells, LAIR1 crosslinking stops the AML cells from responding to the proliferative signal of GM-CSF and in AML cell lines LAIR 1 cross-linking induces apoptosis (Zocchi et al 2001). These data however, being derived from cells which are inherently abnormal in their proliferation and differentiation, while suggestive of a role for LAIR1 in monocyte/macrophage function cannot be regarded as conclusive. A recent gene expression study of the effect of hypoxia on monocytes demonstrated that LAIR1 was one of the genes whose expression was downregulated on exposure to hypoxic conditions, implying that monocytes may be more readily activatable in hypoxic conditions (Bosco et al 2006). From a biological point of view, this would seem plausible, implying that the activation thresholds for monocytes/macrophages may be lower in hypoxic (and therefore) diseased tissue than when in circulation.

There is very little data as to the role, if any, that LAIR1 plays in human disease, save for a single case report of a case of chronic active EBV in a boy whose NK cells lacked LAIR1 (Aoukaty et al 2003). Presence or absence of LAIR1 expression on other cell types was not commented on in this case.

Expression of LAIR1 by TAMs provides an intriguing potential mechanism for immune cell modulation by tumours. Deposition of collagen by tumours (or the

induction of collagen expression by other non-tumour cells by the tumour) would provide a mechanism to downregulate immune cell activation. In this context, the desmoplastic response seen in many tumours may represent a mechanism of immune escape. Coupled with the hypoxic environment that exists in many if not all tumours, the tumour microenvironment may provide multiple elements to activate these inhibitory receptors. It can be hypothesised that in tumours where there is strong expression of LAIR1 on TAMs, these cells will be less readily activatable and hence less able to support tumour growth.

The expression pattern and intensity of staining for LAIR1 in human TAMS in DLBCL was examined to determine whether differences in LAIR1 expression by TAMs correlated to differences in clinical outcome.

6.5.2 Scoring of LAIR1 expression

Using the TMA of DLBCL cases, expression of LAIR1 was scored in macrophages, both in terms of intensity of expression and number of macrophages expressing the protein in a given area. A histoscore, combining staining intensity and number of stained cells was calculated where possible. Of the seventy four cases present in the array, data was obtained for 56 cases for staining intensity and in 34 cases it was possible to calculate a histoscore. Cases were excluded if all replicates of the case were lost during the staining protocol, and in the case of the histoscores if counting of cells in 5HPFs was impossible. This group included cases where there was too little tumour present to count a sufficient area of tumour, cases with tumour cell staining as well as macrophage staining and cases where the morphology of the macrophages did not allow an accurate count to be made.

Cases were scored for intensity on 5 point grading scale, by a single observer blind to outcome, where 0 was no staining and 4 was very intense staining (Figure 6.7). Multiple scores for a single tumour were converted to a mean score. Cases were divided into high or low intensity (low <3, high ≥ 3) and high or low histoscore

(<500, ≥500) based on examination of dots plots of individual cases to determine what subgroups existed within the whole dataset (Table 6.6).

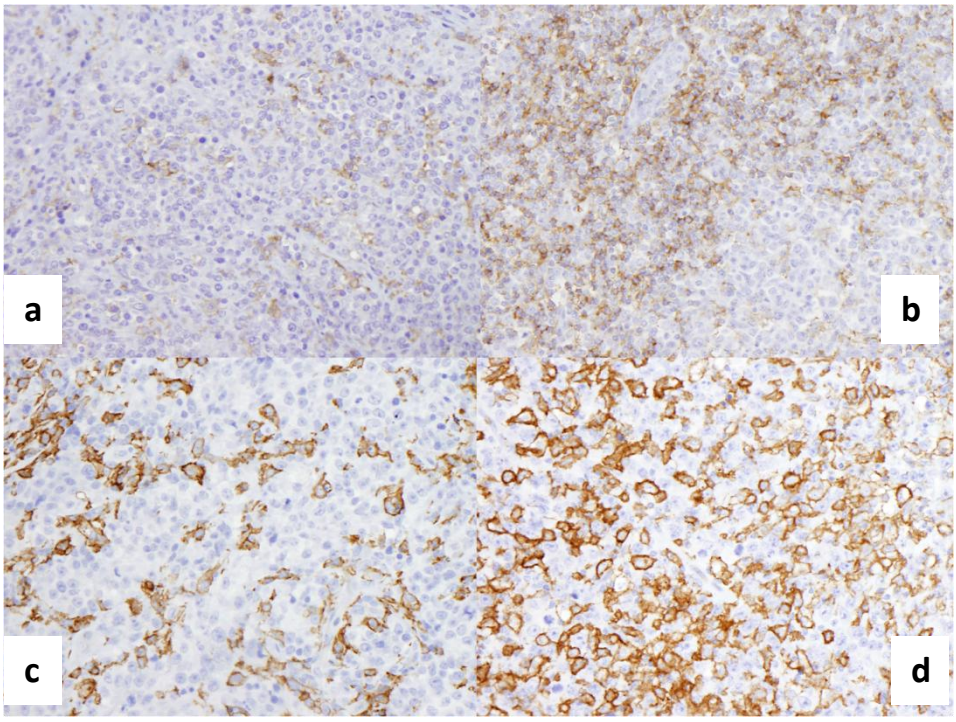


Figure 6.7 Sections of TMA were stained with anti-LAIR1 antibody (HPA001155) and intensity of staining was quantified by eye using a 5 point scoring system (0-4). Examples are given of intensity of staining scoring 1-4. a) Weak staining =1. b) Moderate staining = 2, c) Strong staining = 3. d) Very strong staining = 4. All images x100 magnification.

6.5.3 Relationship to clinical outcome

Outcome was assessed both in terms of time to treatment failure, a disease specific measure of success of therapy, and time to death, with the data from the SNLG database not recording whether death was disease specific or unrelated. Kaplan-Meyer survival curves were plotted and the Mantel-Cox log rank test applied to determine whether the groups differed in their outcome (Figure 6.8 and 6.9). For intensity of staining, there was better outcome in terms of time to treatment failure and time to death in those patients whose TAMs showed low intensity staining with LAIR1, though this difference reached significance only for overall survival. There were no differences in outcome between groups with high and low histoscores. This would support the hypothesis that in macrophages with a lower activation threshold, these cells are more readily activatable to support tumour growth. Alternatively, it could be argued that these macrophages are more activatable in the face of therapy and hence relate to better outcome because of cytotoxic effects on the tumour cells in the face of therapy. As all patients were uniformly treated, it is not possible to differentiate between these two possibilities.

Of interest, was the lack of expression of LAIR1 by the tumour cells. Data from PBMC indicates that circulating B cells express LAIR1 at variable levels dependent on the status of the cell, expression being higher in naive cells than memory cells (Meyaard et al 1997). LAIR1 staining was not observed in B cells in reactive lymph nodes (Figure 5.6) and was only rarely seen in B cells in lymphoma. This absence in reactive lymph nodes would argue against loss of expression being a tumour specific phenomenon. Two major possible explanations exist; that there are differences in LAIR 1 expression intensity and/or sensitivity to antigen retrieval techniques between macrophages and other cell types, or that expression of LAIR1 by lymphocytes is altered by the microenvironment of the lymph node. Of note is that previous studies of LAIR1 expression by lymphocytes have only analysed those in the peripheral blood where expression has been seen to be downregulated on B cell activation (Meyaard et al 2008).

		Time to treatment failure in months Mean (± SE) *Median (± SE)*	Overall survival in months Mean(± SE) *Median (± SE)*
Intensity of staining	High (n=39)	39 (± 8) *14 (± 9)*	58 (± 8) *50 (± 9)*
	Low (n=17)	76 (± 16) *72 (± 26)*	96 (± 14) *86 (± 4)*
Histoscore	High (n=22)	46 (± 9) *34 (± 7)*	62 (± 8) *83 (± 18)*
	Low (n=12)	45 (± 11) *36 (± 1)*	62 (± 9) *68 (± 0)*

Table 6.6 LAIR 1 expression by macrophages in DLBCL and outcome in a CHOP-treated cohort of patients. A TMA containing 76 cases was stained with anti-LAIR 1 antibody (Atlas antibodies) and scored for intensity of staining in TAMs on a five point scale. Cases were divided into high or low intensity staining (<3 or ≥3) and comparison made with outcome data. A Histoscore combining intensity of staining with number of positive cells was also calculated and divided into high or low histoscore (<500, ≥500) and comparison made with outcome data.

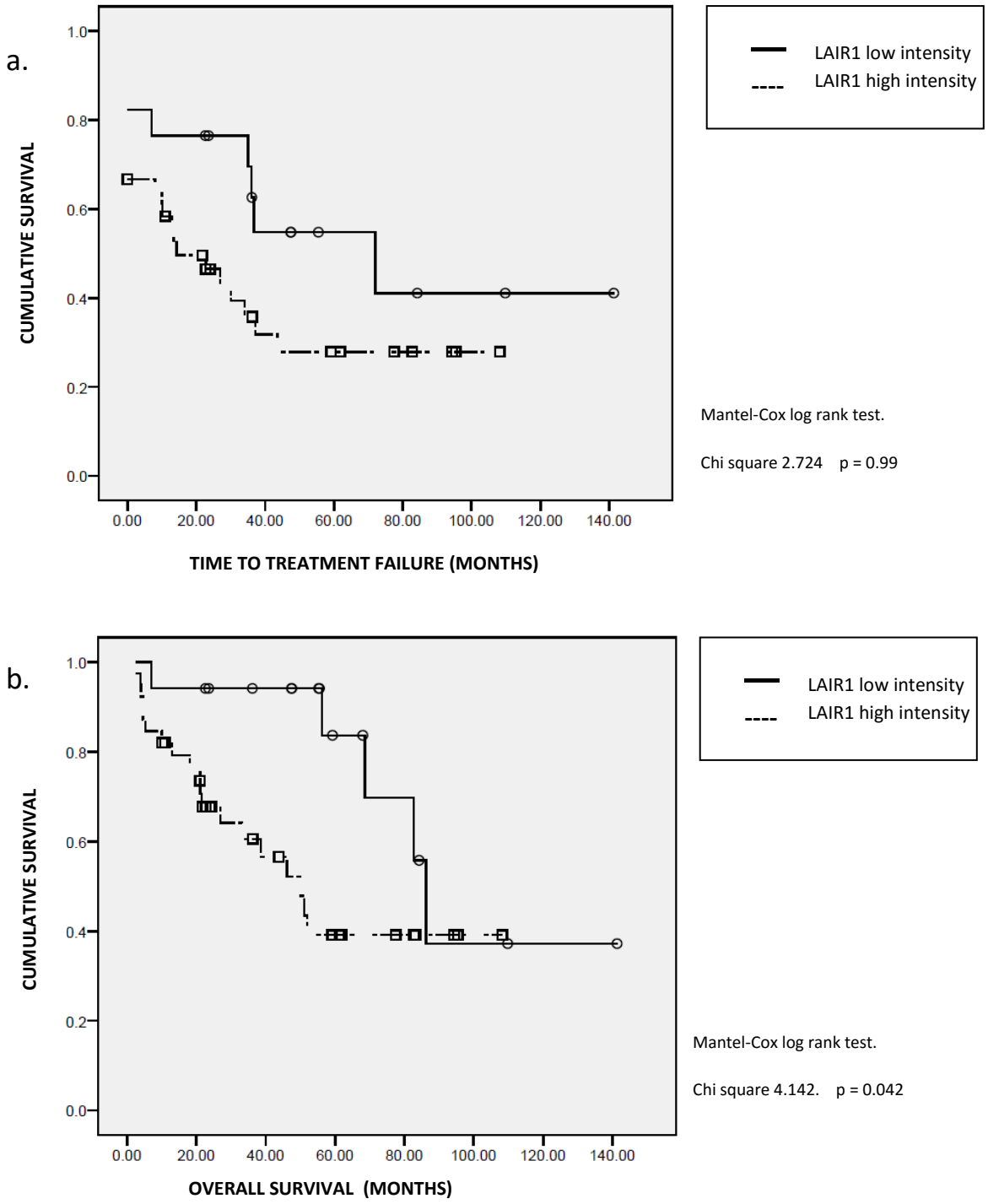


Figure 6.8 Diffuse large B cell lymphoma from a cohort of CHOP-treated patients were stained with an anti-LAIR1 antibody and scored for intensity of staining in TAMs. Kaplan-Meier survival curves were plotted to show (a) time to treatment failure and (b) time to death in patients with high or low intensity staining.

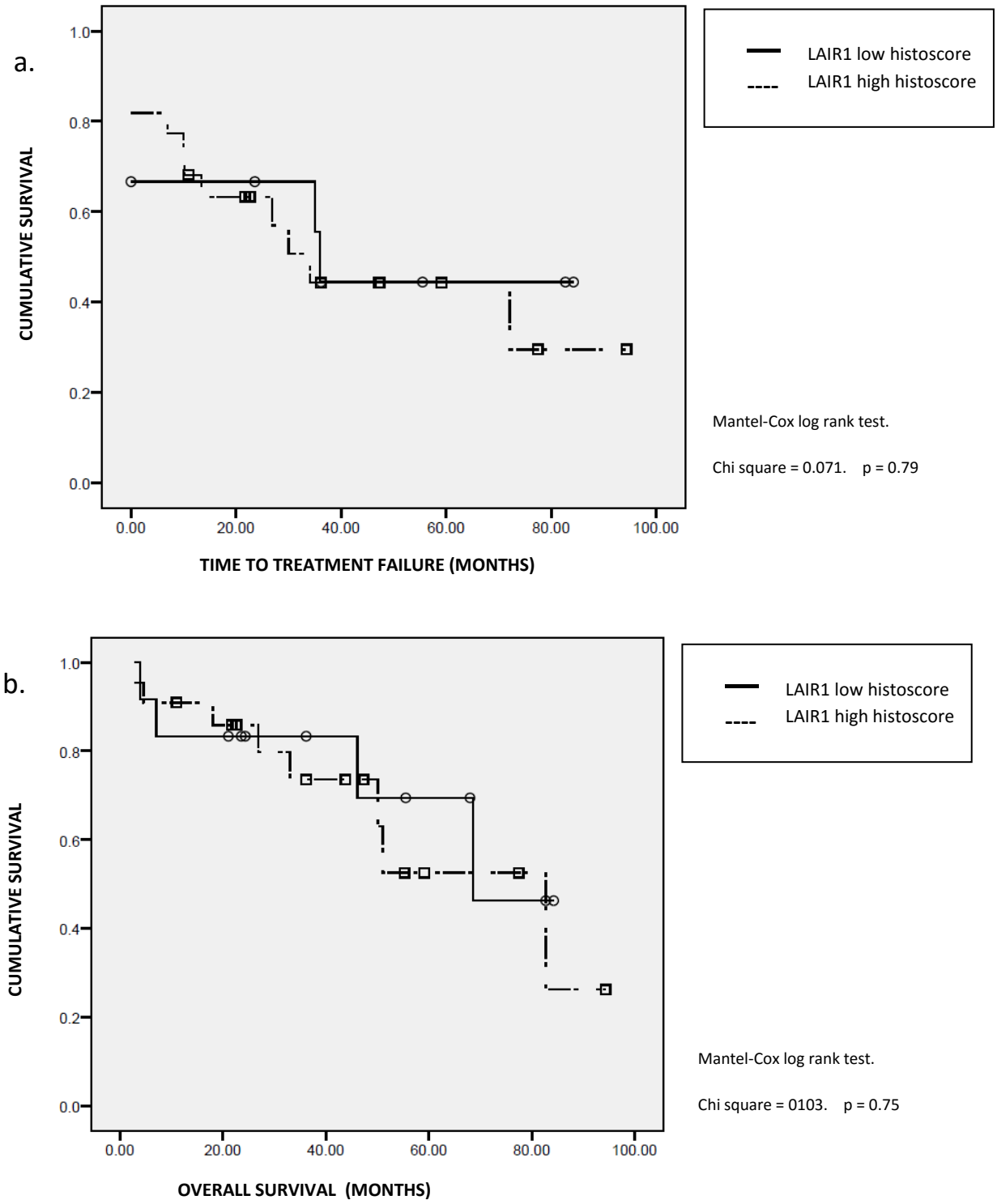


Figure 6.9 Diffuse large B cell lymphoma from a cohort of CHOP-treated patients were stained with an anti-LAIR1 antibody and histoscores calculated by multiplying number of positive cells by staining intensity. Histoscores greater than 500 are grouped into the high histoscore category. Kaplan-Meier survival curves were plotted to show (a) time to treatment failure and (b) time to death in patients with high or low histoscores.

6.6 Conclusions

Four genes from the *in silico* gene expression signature of TAMs in diffuse large B cell lymphoma whose known function suggested they may have a role to play in the macrophage response to the tumour were examined in an independent cohort of cases of DLBCL treated with a CHOP regimen and relationship to outcome studied. For AIF, TYMP and LGALS3BP there was no relationship of expression of these molecules at the protein level with outcome. Only for LAIR1 was a significant relationship seen with expression pattern and outcome and while the trend was similar for disease free survival and overall survival, only in the overall survival data was the result significant. The failure to reach significance in the more specific measure of outcome, i.e. disease free survival, raises the possibility that this represent a spurious association rather than a true effect on outcome, but it is notable that in the Kaplan-Meir curves for disease free survival, the two curves do not come together despite the absence of significant difference between them. It is possible that examination of a large cohort of cases of DLBCL may answer the question as to whether the expression of LAIR1 is truly associated with outcome.

Chapter 7

In vitro studies of LAIR1

Following the observation that differences in staining intensity of TAMs in DLBCL were associated with better or worse outcomes in a standardly treated cohort of patients, a series of in vitro experiments were carried out to investigate the effects of LAIR1 expression by TAMs on the ability of these to support or suppress tumour growth. A co-culture system was developed using THP-1 cells and BL2 cells and the expression of LAIR-1 characterised in this system. THP-1 cells were selected as a source of human macrophage-like cells to provide a pure population of cells of the monocyte/macrophage lineage. The use of BL2 cells was well established within the group in a co-culture model of tumour growth and provided a well-characterised model of tumour growth. The expression of LAIR1 in this model system was studied, as well as the effect of polarisation of macrophages by the addition of cytokines, the effect collagen had on the expression of LAIR-1 expression and tumour cell growth, as well as the influence of apoptotic tumour cells on LAIR-1 expression. Limitations of time allowed for completion of this thesis meant much of the data presented here represents preliminary experiments requiring to be repeated.

7.1 Characterisation of LAIR1 expression in BL2 and THP-1 cell lines.

7.1.1 LAIR 1 is expressed on the surface of THP-1 cells but not BL2 cells

Surface expression of LAIR1 was studied first on undifferentiated THP-1 cells and BL2 cells using 2 different antibodies directed to the extracellular part of the molecule (clones DX26 and NKAT255) (Figure 7.1). There was clear expression of LAIR1 on the surface of THP-1 cells by flow cytometry, but no expression on BL2 cells. This was in keeping with the findings from the reactive lymph node and

DLBCL immunohistochemical studies, which showed expression to be confined to the macrophage component of the tissue and not expressed in benign or malignant B cells. This is in contrast to earlier work on the expression levels of the molecule which showed clear expression on naive B cells with lower levels being present on memory B cells or plasma cells (Merlo et al 2005.) In most of the studies which have looked at LAIR1 expression in human leukocyte populations, peripheral blood has been used as a source of cells, and it is possible that the expression profile of leukocytes in circulation differs from that in solid tissues. In the original description of LAIR1 by Meyaard et al in 1997, they also noted an absence of expression of LAIR1 in many malignant cell lines of haemopoetic origin, despite expression in primary human cells (Meyaard et al, 1997). Whilst it is possible that loss of LAIR-1 expression may be common to malignant cell lines as a consequence of growth in a collagen-free environment, the absence of expression in the solid tissues studied in Chapter 5 suggests that this not a purely in vitro phenomenon.

7.1.2 Differentiation of THP-1 cells alters the expression level of LAIR -1 on THP-1 cells

To create macrophage-like cells, THP-1 cells were differentiated by addition of PMA for three days, and then allowed to mature either in the absence of cytokines or in the face of IFNγ and LPS, IL-10 or IL-4 for 5 days to stimulate differentiation towards the polarised states of M1 or M2 macrophages . Cell morphology was assessed by growing cells on sterile glass coverslips and THP-1 cells were assessed at day 0, day 3 and day 6 (Figure 7.2). Cells became larger, with more voluminous cytoplasm and more spread-out morphology typical of macrophages during the course of the differentiation process. Immunocytochemistry on permiabilised intact cells showed a decrease in the intensity of the staining for LAIR1 as the differentiation process progressed as well as a slightly granular pattern to the staining in the more differentiated cells hinting at the possibility that there may be sequestration of the antigen in an intracellular compartment in the more differentiated cells.

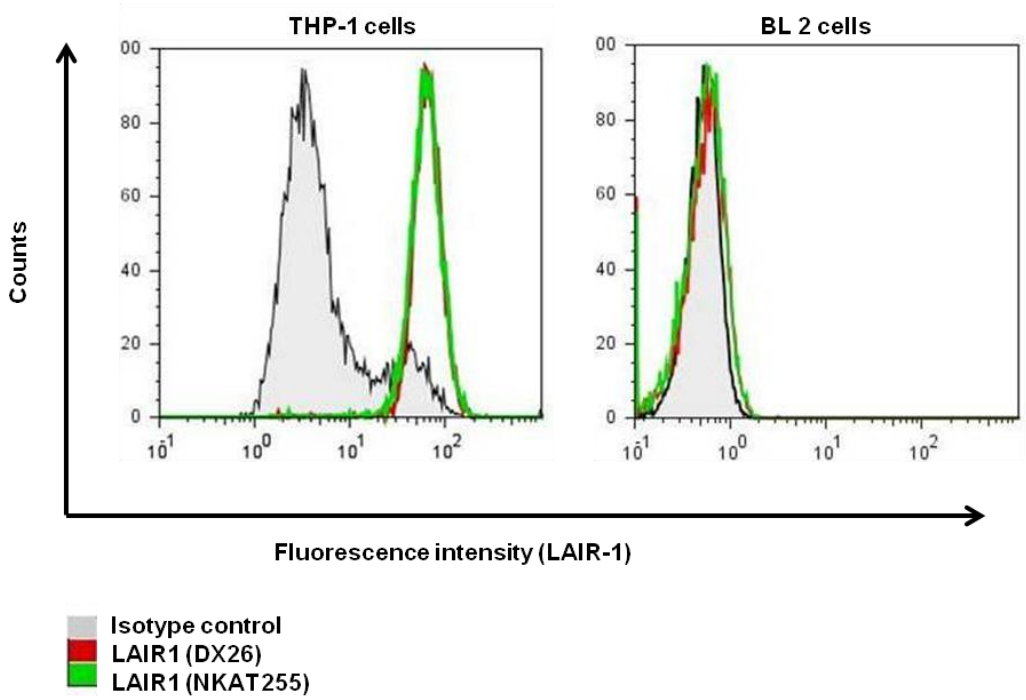


Figure 7.1 Surface expression of LAIR1 on undifferentiated THP-1 cells and BL2 cells.

THP-1 cells and BL2 cells were stained with antibodies to the extracellular component of LAIR-1, and then labelled with a secondary $F(ab')_2$ fragment conjugated to Alexa-Fluor 488 and fluorescence intensity analysed by FACS. (Isotype control show in grey. DX26 clone in red and NKAT255 clone in green).

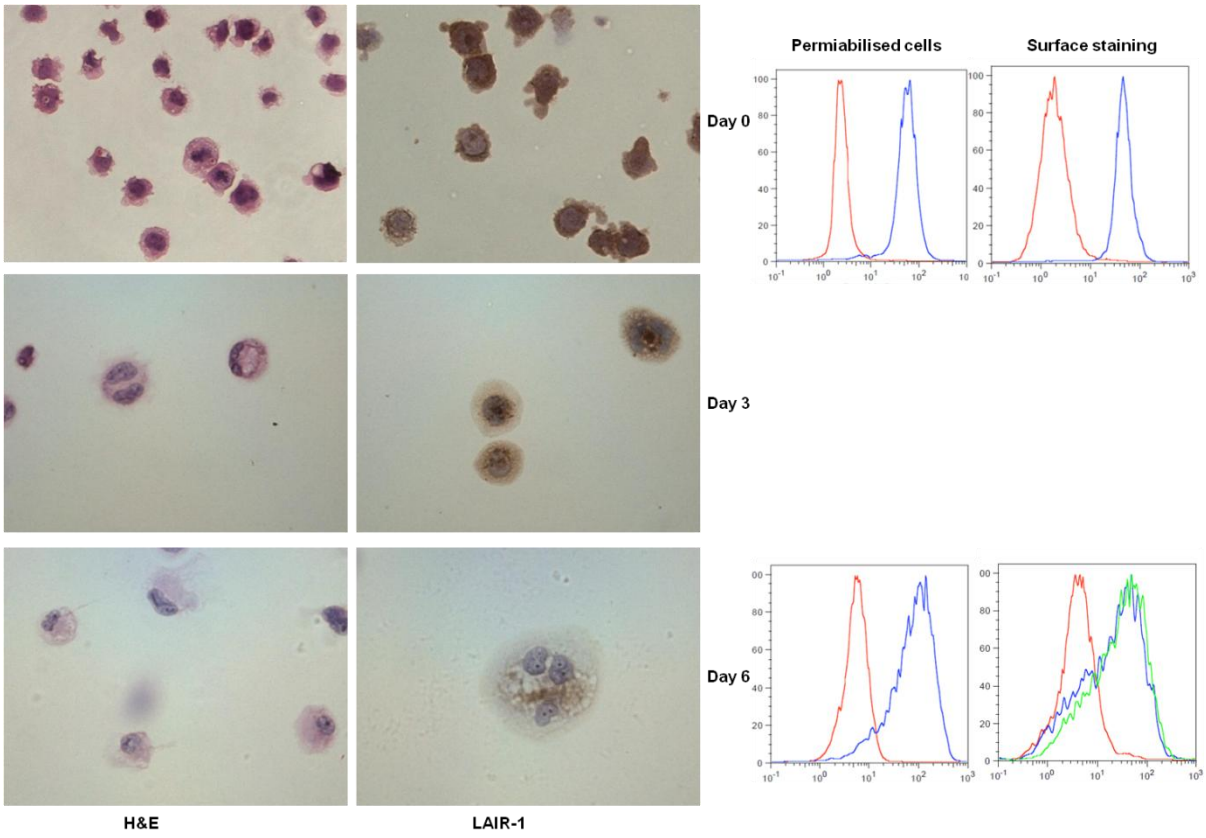


Figure 7.2 LAIR-1 is expressed on differentiated THP-1 cells

THP-1 cells were differentiated by the addition of PMA and plated down onto chambered slides. These were incubated for 3 or 6 days and then stained directly on the slides with haematoxylin and eosin (H&E) or anti-LAIR1 antibody (HPA011155) directed against an intracellular component of the molecule. The day 0 slides represent cytospins. The cell became larger, with increased cytoplasm over the course of the differentiation and the expression of LAIR 1 appeared less intense. LAIR-1 expression was assessed at day 0 and day 6 by flow cytometry on permiabilised cells stained with an antibody recognising an intracellular epitope (HPA011155) and intact cells with antibodies to a surface epitope (blue represents clone DX26 and green clone NKTA255). Red represents the isotype control in all cases.

Expression of LAIR1 in differentiating cells was further assessed by flow cytometry using antibodies against both the extracellular and intracellular components of the molecule (Figure 7.2). In undifferentiated cells all three antibodies showed a similar pattern with the histogram showing strong expression of LAIR1 by THP-1 cells with very little overlap of the histogram with the isotype controls. In contrast, in the differentiated cells there was a much wider histogram with considerable overlap with the isotype control indicating that individual cells showed variable levels of expression of LAIR1. This effect was seen with both the intracellular and extracellular component of the molecule, although the effect was more marked with the extracellular component regardless of the clone used. In all the experiments using differentiated THP-1 cells, there is also a relatively wide peak in the histogram for the isotype control despite the use of a secondary $F(ab')_2$ fragment rather than a secondary whole immunoglobulin molecule to minimise non-specific binding of antibodies to surface receptors on macrophages.

7.1.3 THP-1 expression of LAIR1 shows only minor differences regardless of the polarisation status of the cell.

Expression of LAIR-1 expression was assessed by flow cytometry on differentiated and polarised THP-1 cells. Cells were prepared as previously described by growth in tissue culture plates in the presence of PMA for three days, washed and fresh medium added with either no cytokines or IFNγ/LPS, IL-4 or IL-10. Prior to staining with antibodies, cells were lifted from the culture plates by first adding cold detachment buffer and gently shaking the cells at 4°C, and then gently lifting with a cell scraper. By day 8 in the differentiation process, the cells were firmly adherent and detachment of the cells resulted in considerable cell death both by assessment by trypan blue exclusion and by gating on the flow cytometer for viable and non-viable cells by forward and side scatter. Assessment of LAIR1 expression was analysed only for those cells falling into the viable gate on flow cytometry. As seen in figure 7.3 THP-1 cells, regardless of their polarisation state showed a broad histogram

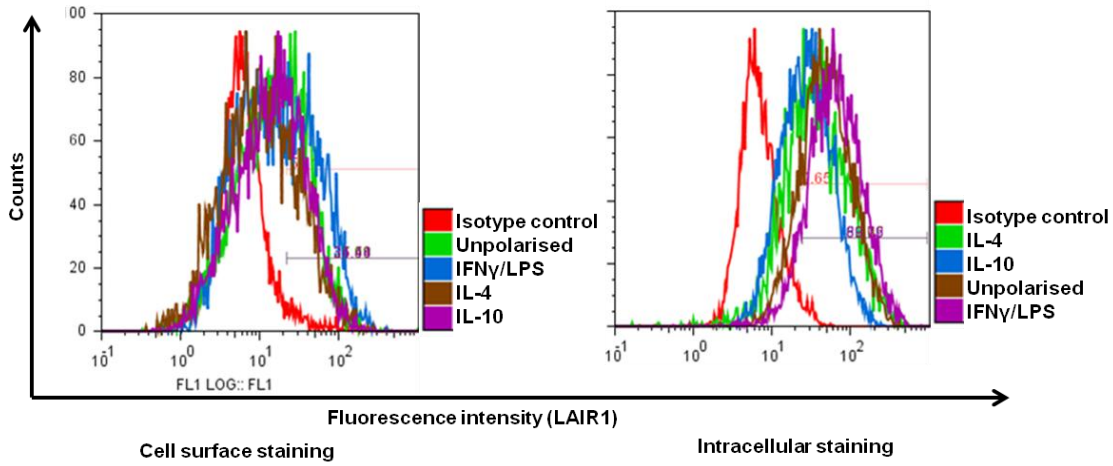
which overlapped with the isotype control for expression of LAIR1 for both extracellular and intracellular staining. There did not appear to be reproducible differences in expression between any polarisation state, although all these experiments were confounded by the large rates of cell death seen in preparation of the cells for flow cytometry.

Expression of LAIR1 was also assessed by Western blotting (Figure 7.3). Cells were differentiated and polarised as above and then protein extracted from whole cell lysates. Equal quantities of protein from each experimental condition were loaded onto precast 4-12% Bis-Tris gels and western blotting performed. All three antibodies against LAIR1 were used. All three antibodies detected a protein band of about 40kDa of similar intensity in all polarisation states. The predicted molecular weight of LAIR-1 being 32kDa based on the protein structure, this would be consistent with the weight of the protein plus post-translational modifications, presumably glycosylation (Meyaard et al, 1997). Interestingly, the two antibodies directed at the extracellular component of the LAIR1 molecule DX26 and NKTA255, showed a single band on western blotting while the antibody directed against an intracellular component of the molecule also showed several bands of lower molecular weight, being smaller than 14kDa. This suggested that there might perhaps be a cleavage product of the molecule present, containing the intracellular component of the molecule, raising the possibility of surface shedding of the molecule. Possible surface shedding has been reported in a variety of inflammatory conditions in humans including rheumatoid arthritis and pregnancy although in these studies it is unclear if the molecule being measured was cleaved LAIR1 or secreted LAIR2 (Lebbink et al, 2008). In T cells, where expression of LAIR1 has been better studied, there does not appear to be cleavage of LAIR1 from the surface of cells, but rather internalisation of the molecule (Jansen et al, 2007). While internalisation of the molecule might also happen in THP-1 cells, the presence of a shorter form of the molecule detected by an antibody to the intracellular component of the molecule is suggestive of cleavage occurring. It is of interest that this potential cleavage product band was of greater intensity in cells treated with IFNγ/LPS than other polarisation states, hinting at a possible positive feed-forward loop in ‘classically activated’ cells reducing inhibitory signalling and making cellular activation more easy. This result

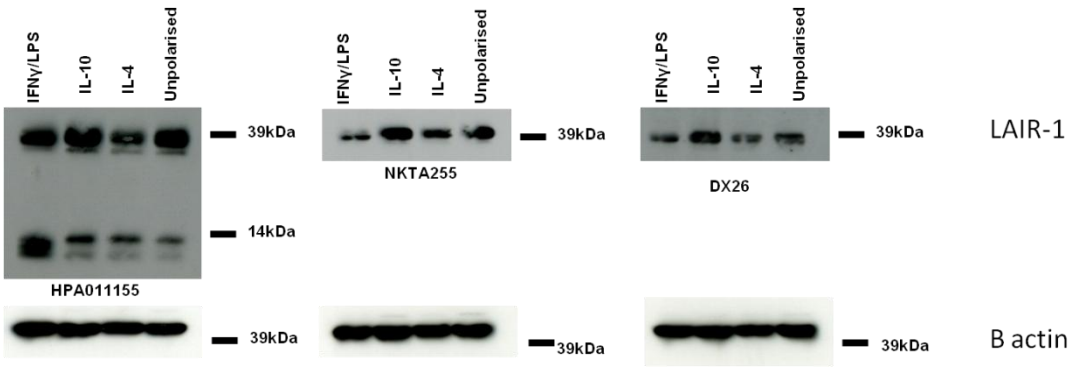
is however in contrast with the cell surface expression data from flow cytometry which did not detect a difference between the levels of expression of LAIR 1 on the surface regardless of polarisation state.

Expression of LAIR1 was also assessed at the level of mRNA in differentiated and polarised THP-1 cells (Figure 7.3). Very similar levels of expression of LAIR1 mRNA was seen in undifferentiated THP-1 cells, differentiated but unpolarised THP-1 cells, and polarised THP-1 cells. This suggests that if there is regulation of LAIR1 expression it is not at the level of transcription.

a


Counts
FL1 LOG:: FL1
Isotype control
Unpolarised
IFNγ/LPS
IL-4
IL-10
Isotype control
IL-4
IL-10
Unpolarised
IFNγ/LPS
Fluorescence intensity (LAIR1)
Cell surface staining
Intracellular staining

b


IFNγ/LPS
IL-10
IL-4
Unpolarised
39kDa
14kDa
HPA011155
NKTA255
DX26
LAIR-1
B actin

c

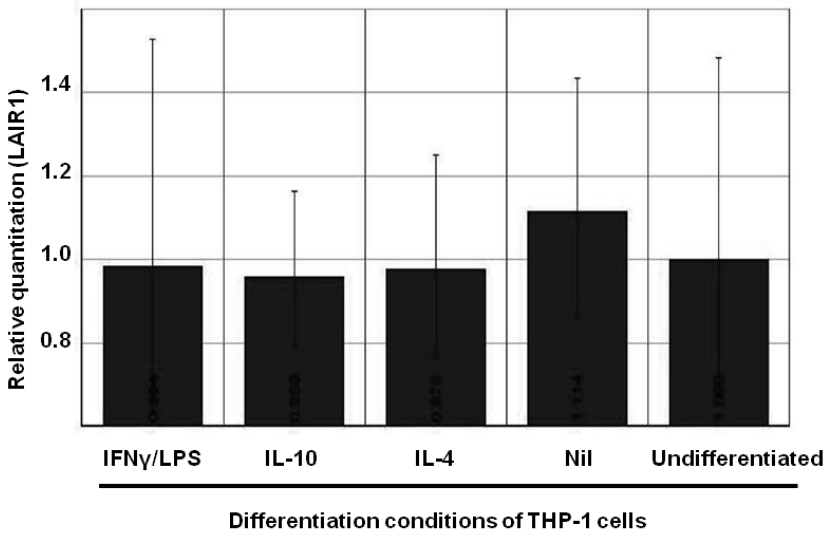

Relative quantitation (LAIR1)
1.4
1.2
1.0
0.8
IFNγ/LPS
IL-10
IL-4
Nil
Undifferentiated
Differentiation conditions of THP-1 cells

Figure 7.3 Polarisation status of THP-1 cells does not influence LAIR-1 expression.

THP-1 cells were differentiated by the addition of PMA for 3 days, then cytokines as indicated for 5 days. a). Expression of LAIR-1 was assessed by flow cytometry both by surface staining using clone DX26 directed to the extracellular component of the molecule and using clone HPA011155 directed to the intracellular component of the molecule in permiabilised cells. b). Western blotting was performed on whole cell lysates using three antibodies (HPA011155, NKTA255, DX265) against LAIR1. Equal quantities of protein were used in each case and an antibody against β-actin used as a loading control. c). RNA was extracted from the THP-1 cells and quantitative RT-PCR performed. 18S was used as the endogenous control and results were normalised to undifferentiated THP-1 cells.

7.1.4 Addition of collagen to THP-1 cells does not alter expression of LAIR-1

To assess the effect of collagen on expression of LAIR1 by THP-1 cells, cells were cultured in the presence or absence of collagen. Preliminary experiments in which collagen was added at a late time point failed as the diluent for the collagen proved toxic to the cells, and similarly differentiating the cells in the absence of collagen and lifting and replating on collagen resulted in loss of large numbers of THP-1 cells. Eventually, a method was arrived at in which all stages of the experiment were performed with or without collagen. Tissue culture plates were coated with collagen or diluent and dried in tissue culture cabinets under UV light to cross-link and sterilise the collagen. The plates were then washed to remove the diluent and THP-1 cells added along with PMA, and cultured for 3 days, before the addition of fresh media with cytokines and culture for 5 days. Cells were lifted as described previously and surface and intracellular expression of LAIR1 assessed by flow cytometry (Figure 7.4). There did not appear to be any difference in expression of LAIR 1 in the presence or absence of collagen regardless of the polarisation state of the THP-1 cells implying that any action of the LAIR1 ligand on macrophages is not through alteration of surface expression of the receptor.

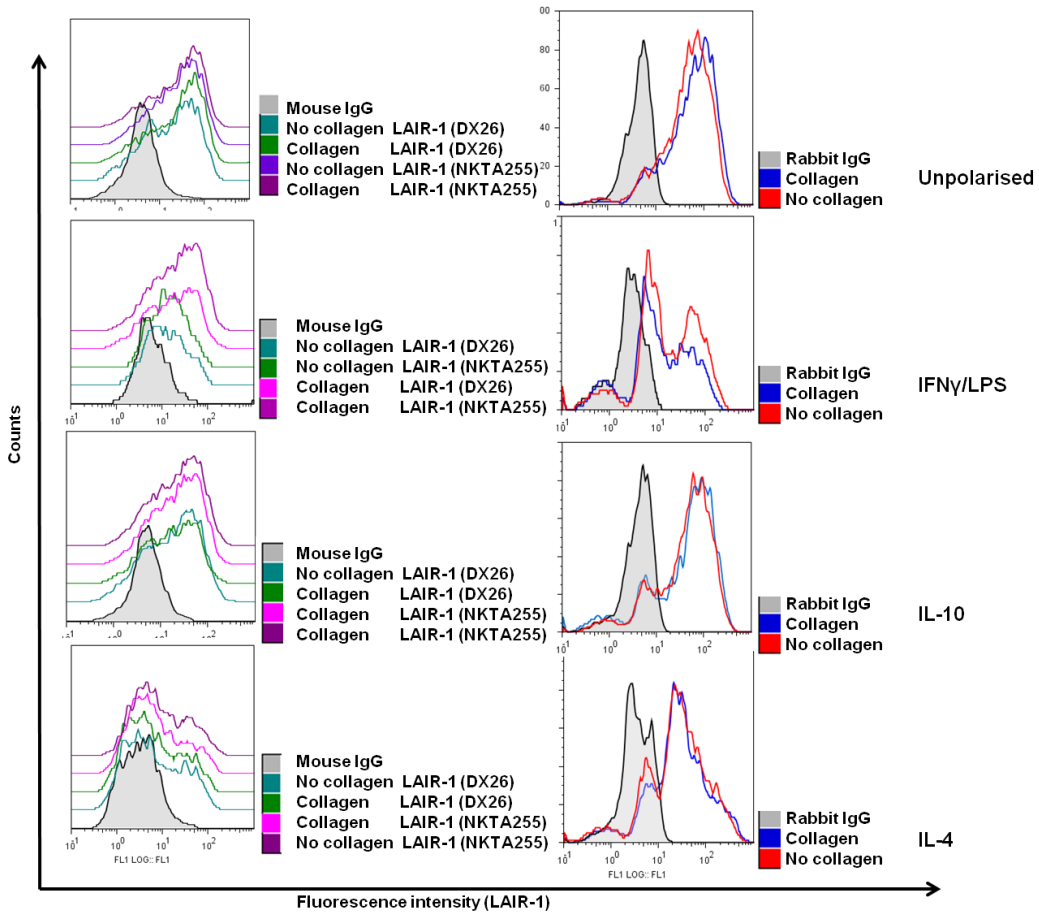


Figure 7.4 Addition of collagen to culture conditions does not alter LAIR-1 expression.

THP-1 cells were differentiated and polarised by the addition of PMA for 3 days then cytokines for 5 days either on untreated tissue culture dishes or dishes that had first been coated with human collagen IV. Cells were lifted and stained with antibodies against the extracellular component [left panel] (DX26 and NKTA255) or intracellular component of the LAIR1 molecule [right panel] and expression analysed by flow cytometry. Only viable cells based on forward and side scatter were analysed.

Work by Poggi et al using AML cell lines suggested that crosslinking of LAIR1 was fatal to THP-1 cells (Poggi et al, 2000). Although on morphological assessment of cells grown in tissue culture plates, this did not appear to be the case, more detailed morphological assessment of THP- 1 cells grown on or off collagen was undertaken. Chambered slides were coated with collagen or diluent and treated as the tissue culture plates above. THP-1 cells were differentiated by the addition of PMA then allowed to mature, before fixing and staining with H&E. In both conditions, cells reached near confluent growth and showed similar morphology suggesting that the presence of collagen was not toxic to THP-1 cells (Figure 7.5).

Preliminary work was also undertaken looking at LAIR1 expression in primary human macrophages. Monocytes were derived from donor blood by Percol gradient and plated onto tissue culture plates previously coated with either collagen or diluents. These experiments had to be abandoned as the cells behaved very differently in the presence or absence of collagen; cells cultured on plastic rapidly became adherent; cells cultured on collagen formed colonies in suspension that appeared morphologically to be proliferating and remained in that form for several days before becoming adherent. The planned experiments with these cells were abandoned as the populations were very clearly different, and contained very different numbers of cell despite addition of identical numbers initially. Whilst this observation, suggesting that primary monocytes are far more sensitive to the effects of collagen in the microenvironment that the malignant cell line THP-1, was very interesting there was insufficient time to pursue it further. Differences in behaviour of primary human monocytes cultured on collagen as opposed to glass were described by Kaplan and Gaudernack in 1982, with further work by Kaplan showing that monocytes cultured on glass are cytotoxic to tumour cells while those cultured on collagen are not (Kaplan and Gaudernack, 1982, Kaplan, 1983). In these experiments monocytes cultured on glass showed spontaneous cytotoxic activity against several malignant cell lines suggesting that the glass was activating the monocytes/macrophages to a cytotoxic phenotype as such behaviour is not normally seen when cultured in plastic tissue culture dishes. The effect on morphology and

behaviour however is on interest in the light of the changes seen here in human monocytes exposed to collagen.

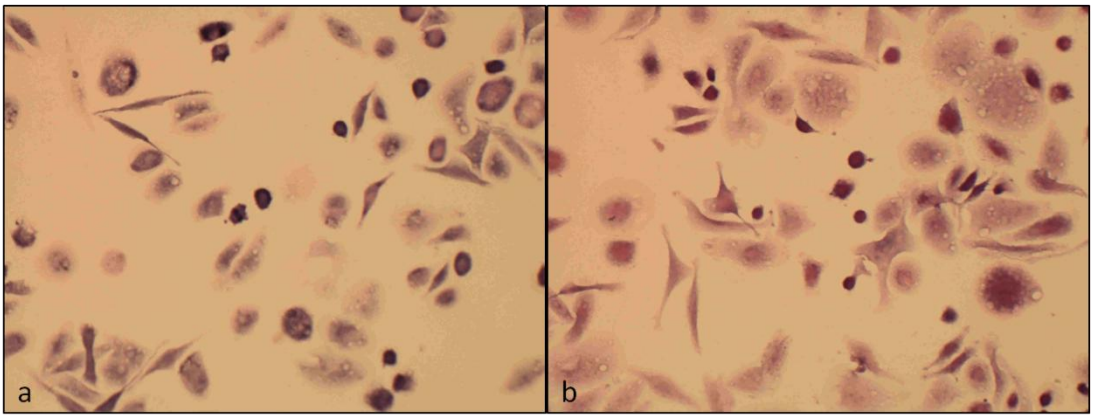


Figure 7.5 THP-1 cells morphology when grown on collagen**.**

THP-1 cells were cultured on chambered slides in the presence (a) or absence (b) of collagen to establish whether collagen induced cell death in THP-1 cells. After 8 days cells were fixed and stained with haematoxylin and eosin. In both culture conditions cells grew well and reached near confluence. Cells showed similar morphology with a proportion of cells remaining small with little cytoplasm but the majority becoming large, adherent and round or dendritic in morphology. (Representative field at x200 magnification from area of low density growth to demonstrate morphology. Photomicrographs from one of three experiments).

7.2 Assessment of tumour cell growth when co-cultured with THP-1 cells grown on collagen.

To assess whether the addition of collagen alters the way THP-1 cells interact with BL2 cells a co-culture system was used in which BL2 cells were cultured in the presence of THP-1 cells grown on collagen or on plastic. Co-culture of tumour cells with macrophages is a well characterised model in the Gregory lab most frequently in the context of bone marrow derived macrophages, and these macrophages have been shown to support or suppress BL2 tumour growth dependent on the exact experimental conditions. In these experiments, BL2 cells were allowed to grow in a favourable environment, i.e. seeded into serum rich medium at a low initial concentration. Cell growth kinetics were calculated by counting cell numbers at 0, 1, 2 and 3 days, after which point there tended to be exhaustion of the medium and consequent high rate of cell death

7.2.1 Presence of collagen does not inhibit BL2 growth.

BL2 cells were grown in tissue culture dishes previously coated with collagen to establish whether the presence of collagen was toxic to the cells or provided a growth stimulus. Cells were counted at day 0, and for the next three days and growth curves plotted. The rate of increase in BL2 cell number was not altered by the addition of collagen implying that collagen itself in the co-culture model does not influence BL2 cell growth. (Figure 7.6) Given the absence of surface expression of LAIR1 on BL2 cells, collagen would not be expected to influence BL2 cell growth, however it was possible that BL2 cells have other surface receptors that interact with collagen.

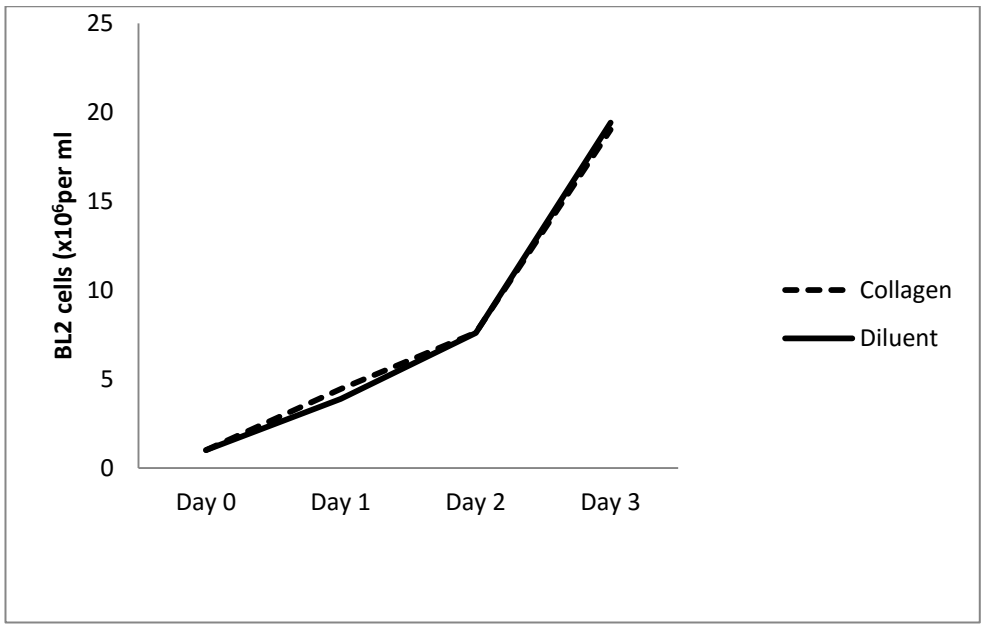


Figure 7.6. BL2 growth kinetics grown on collagen. BL2 cells were plated out at an initial concentration of 1 x10^5 cells per ml in tissue culture plates which were treated with collagen or diluent and grown for 3 days. Cells were counted in triplicate on each day to day 3. (Representative experiment of three).

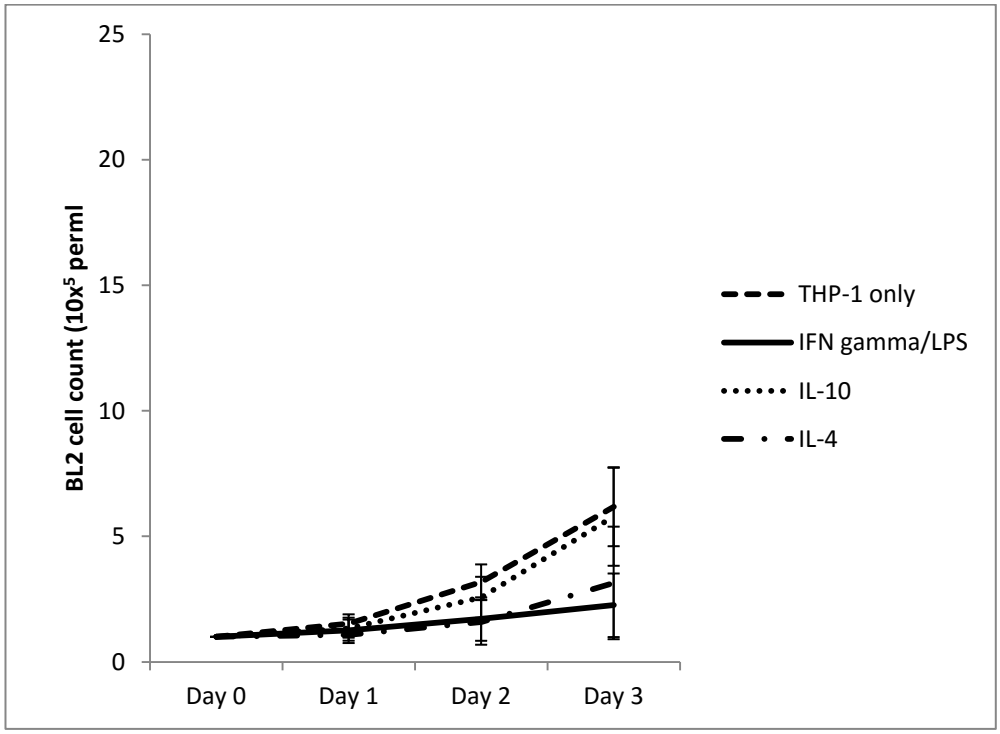


Figure 7.7 BL2 growth kinetics on THP-1 cells. BL2 cells were grown in co-culture with THP-1 cells which were differentiated to macrophages but unpolarised, or cells polarised by the addition of IL-4, IL-10 or IFNγ/LPS following differentiation. Cell counts were obtained at days 0-3. Results calculated in triplicate for each time point. The figure shows means of four experiments from different days (±1sd). Cells grew better on macrophages cultured in

the absence of additional cytokines, or those cultured with IL-10 than those cultured with IL-4 or IFNγ/LPS, but only the IFNγ/LPS culture conditions showed a significant difference at all time points compared to unpolarised macrophages. ($p<0.05$, paired 2-tailed student's t-test).

7.2.2 Collagen influences the growth of BL2 cells where THP-1 ells have been treated with IFNγ/LPS

BL2 cells grown on THP-1 macrophages showed different rates of cell growth dependent on the polarisation condition of the macrophages (Figure 7.7). Those grown on either unpolarised THP-1 cells or cells polarised by the addition of IL-10 grew better than those in which the THP-1 cells were polarised by the addition of IFNγ and LPS or IL-4 but the differences did not reach significance, except when comparing cells treated with IFNγ/LPS with untreated THP-1 cells where growth of BL2 cells was significantly reduced in the IFNγ/LPS conditions. In all cases, BL2 cells grown in the presence of macrophages grew less well than BL2 cells alone. This was supported by a body of work within the group where it was frequently observed in a variety of co-culture systems, that tumour cells in optimum growth conditions are constrained by the addition of macrophages, while those in suboptimal conditions are supported by the addition of macrophages.

When looking at individual culture conditions, only where THP-1 cells were polarised by the addition of IFNγ and LPS did the presence of collagen in the culture conditions alter BL2 cell growth (Figure 7.8). At all time points, BL2 cells grown with the THP-1 cells on collagen grew better than those where the THP-1 cells were on plastic. It could be hypothesised that in the presence of collagen, full activation of the THP-1 macrophages to an M1 type phenotype is inhibited allowing a more supportive/less inhibitory environment for the tumour cells.

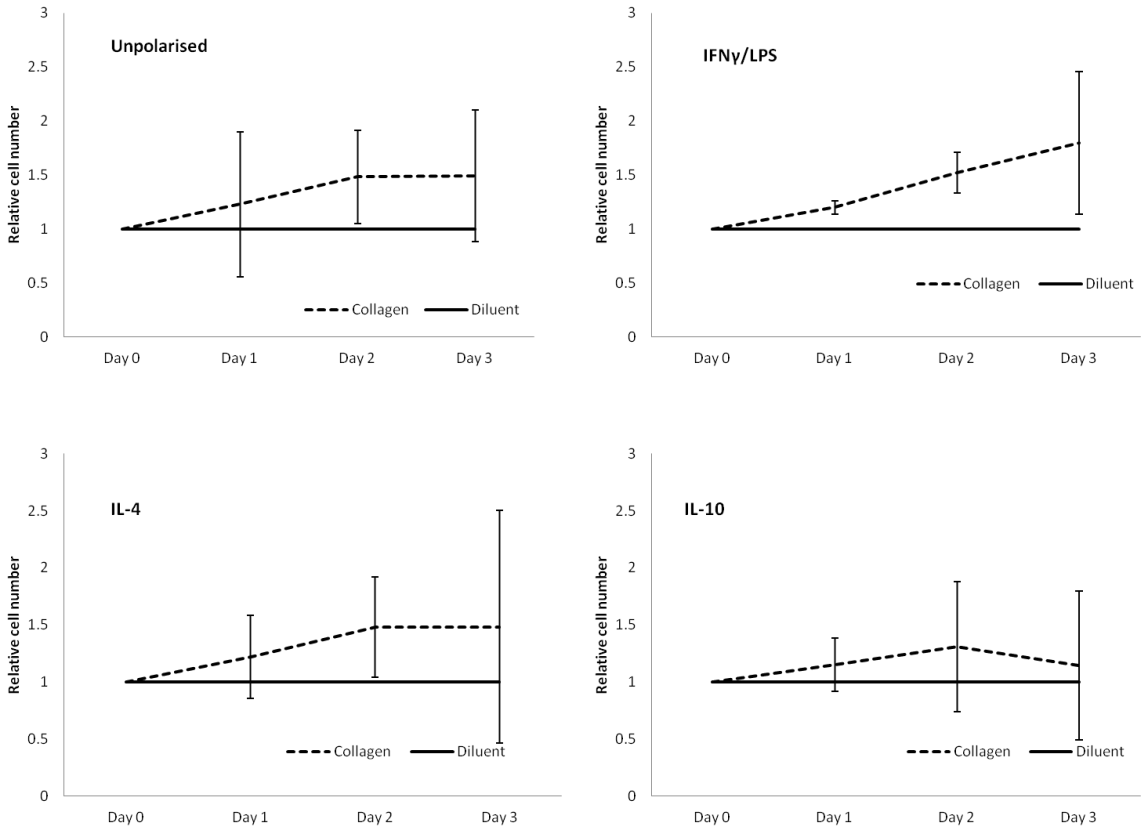


Figure 7.8 BL2 growth kinetics with THP-1 cells grown on collagen

BL2 cells were grown in co-culture with THP-1 cells which were differentiated to macrophage-like cells and then either left unpolarised or polarised by the addition of IL-4, IL-10 or IFNγ/LPS. Cell counts were obtained at days 0-3 of co-culture. Results are plotted relative to growth on THP-1 cells not treated with collagen for each polarisation condition. The results are the means of 4 experiments (±1sd). Only for cells grown on THP-1 cells treated with IFNγ/LPS was there a significant difference in growth on or off collagen and this difference was significant at all time points. ($p<0.05$, paired 2-tailed student's t-test). There was no significant alteration in BL2 cell growth at any time point for any of the other conditions.

7.3 Co-culture of THP-1 cells with apoptotic BL-2 cells does not alter expression of LAIR-1.

Given the known ability of apoptotic cells to modulate macrophage function and activation, experiments were performed to assess whether the presence of apoptotic cells alters the surface expression of LAIR-1 in THP-1 cells. The THP-1 cells were incubated with BL-2 cells that were induced to undergo apoptosis by the addition of staurosporine. Induction of apoptosis in BL2 cells was confirmed by assessment of Annexin-V and propidium iodide staining, confirming that of the BL2 cells added to the culture about 20% of these would be expected to be apoptotic an hour after addition to the THP-1 cells. THP-1 cells were incubated with BL-2 cells over a range of times from 2 to 24 hours and expression of surface LAIR1 assessed by flow cytometry. There was no alteration in the expression of LAIR-1 by flow cytometry for surface LAIR1 by co-culture with apoptotic cells at any of the time points assessed (Figure 7.9). This very preliminary experiment was only performed once due to time constraints.

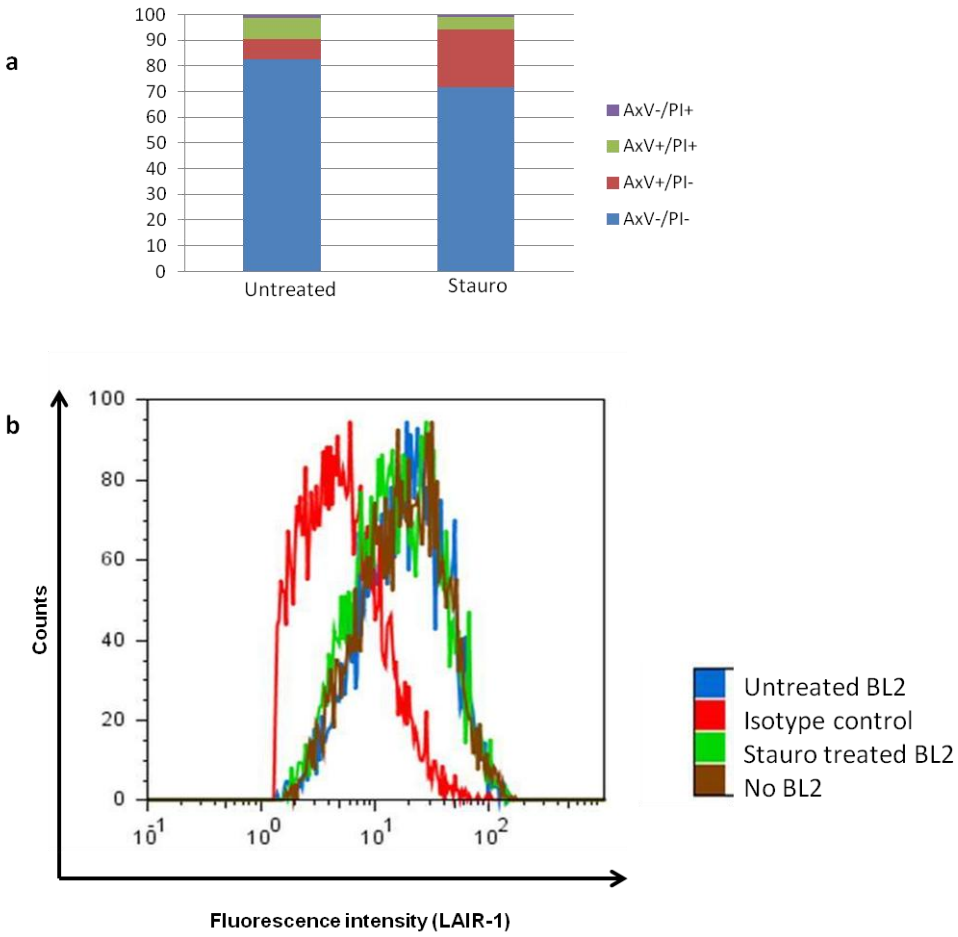


Figure 7.9 Co-culture with BL-2 cells induced to undergo apoptosis does not alter expression of LAIR1.

BL2 cells were induced to undergo apoptosis by exposure to staurosporine at 1µM for five minutes, then thoroughly washed and returned to culture. a) After 75 minutes in culture cells were stained with Annexin V and propidium iodide to assess the percentage of cells in apoptosis. b). Differentiated but unpolarised THP-1 cells were co-cultured with BL2 cells treated with staurosporine, or untreated cells for 2 hours and then BL2 cells washed off and cell surface expression of LAIR1 assessed by flow cytometry. Similar results were obtained after 6 and 24hours of culture. These were very preliminary experiments and only performed once.

7.4 Summary

In this chapter LAIR 1 expression in a tumour associated macrophage model was assessed using THP-1 cells as a source of 'macrophages' and BL2 cells as tumour model. While much of this work is preliminary and would benefit from being repeated and with many questions which have not been addressed, the following conclusions can be reached;

- LAIR-1, an inhibitory receptor that binds collagen as its ligand, is present on naive THP-1 cells.
- Differentiation of THP-1 cells into more mature macrophage-like cells results in a degree of downregulation of expression of LAIR-1 although surface expression remains.
- Surface expression of LAIR-1 does not appear to be affected by the polarisation status of the cells, although intriguingly, there does appear to be a cleaved form of LAIR1 present in differentiated cells particularly those polarised by the addition of IFNγ/LPS.
- LAIR-1 expression in not altered by the addition of collagen to cultured THP-1 cells, and the presence of collagen does not affect THP-1 cell differentiation or BL2 cell growth.
- The addition of collagen does appear to affect how THP-1 cells polarised by IFNγ/LPS support tumour cell growth of BL2 cells with enhanced cell growth in the face of collagen suggesting that while surface expression of LAIR-1 appears unaltered, there may be differences in signalling within the cell.

Chapter 8

Discussion of the results

8.1 Introduction

This thesis examines some aspects of tumour-associated macrophages in Diffuse Large B cell lymphoma. While there is considerable interest in the microenvironment in this tumour following the large scale gene expression data published in recent years, there is still little data as to the exact role of the tumour associated macrophage in DLBCL and their function and phenotype. The work presented here is based on tissue derived from human DLBCL, with the exception of the preliminary in vitro experiments in the final chapter. While this of necessity means much of the data presented is observational rather than subject to experimental manipulation, the findings represent those of the genuine human disease rather than an animal model.

8.2 Possible role for apoptosis in recruiting macrophages in DLBCL

Initially the possibility that tumour cell apoptosis may be one of the factors involved in the recruitment of macrophages was addressed. Work by others in the group looking at the relationship of proliferation, apoptosis and macrophage infiltration in Burkitt lymphoma has shown an association between these variables. Mechanisms by which apoptotic cells recruit macrophages in these tumours have been described as have some of the ways in which macrophages influence tumour growth (Ford et al 2015, Ogden et al 2005, Truman et al 2008). Such an association between apoptosis, proliferation and macrophage infiltrate has not previously been demonstrated in DLBCL. Presented here in chapter 3, looking at a cohort of 82 cases of DLBCL, an association was demonstrated between apoptosis and proliferation and apoptosis and macrophage infiltration, but not macrophage infiltration and proliferation (Figure 3.3 and Table 3.2). This argues that, as in Burkitt lymphoma, tumour apoptosis may be one of the mechanisms by which

macrophages are recruited into tumours, and influenced by tumour cells. In DLBCL there is no clear association between proliferation and macrophage number to argue that macrophages support tumour growth. These however are heterogeneous tumours with multiple dysregulated pathways and any effect of macrophages on proliferation may very well be too subtle in the face of multiple other influences to be detectable.

Clearly, as in any study looking at human tumour tissue, this analysis represents only a single point in time for each tumour studied and cannot inform as to causal relationships. It could equally be argued that the relationship between macrophages and apoptosis exists because macrophages induce apoptosis and hence large numbers of macrophages induce substantial apoptosis. While this cannot be refuted on the data presented here, a relationship in which apoptotic cells attract macrophages fits better with the data from experimental models in which apoptotic cells release many factors that send 'find me'and'eat me' signals to macrophages and provides a more biologically plausible explanation of the association (Grimsley and Ravichandran, 2003, Truman et al 2008, Ford et al 2015). It should also be noted that, despite some differences in some studies, the weight of data in DLBCL suggests that large numbers of macrophages are associated with poorer outcome which would be at variance with the hypothesis that macrophages in the tumour drive tumour cell apoptosis (Kridel et al, 2015).

Both DLBCL and BL show high rates of proliferation and apoptosis. Although the median rate of apoptosis seen in the DLBCL tumours studies was only 1.4%, given the efficiency with which apoptotic cells are cleared, this is likely to equate to a very high rate of cell death (Figure 3.2) (Gregory and Pound 2011). That there is a relationship between proliferation and apoptosis in lymphoma has been recognised for a long time, with a study in 1993 of a wide variety of lymphomas (classified using the Kiel classification and Working formulation) showing a relationship between these variables (Leoncini et al, 1993). These results are not directly comparable to the data presented here being of a mixture of high and low grade lymphomas and using a now-outdated method of classification, but they support the observations made here that proliferation and apoptosis are related in DLBCL. It

should be noted that apoptosis is much less frequently seen in low grade tumours, and may therefore not have a role to play in recruiting macrophages in these lymphomas. In the example of follicular lymphoma for instance, the tumour appears to use the pre-existing microenvironment of the lymph node and may therefore have less requirement for recruited elements of the microenvironment (Scott and Gascoyne, 2014).

8.3 Capturing the gene expression signature of the TAM by an *in silico* approach

In an effort to gain an unbiased view of the TAM in DLBCL, a gene expression signature of macrophages in situ in human tumours was sought using a bioinformatic approach. This formed part of a wider study looking at the cancer transcriptome in a broad spectrum of human tumours. The analytical approach was based on measuring the correlation between all genes across multiple large datasets in order to provide a statistical measure of the relatedness between expression patterns. This measure was then used to create 3D network visualizations of the graphs using BioLayout *Express*^{3D} a visualization tool that allows exploration of complex networks of a size not previously possible (Freeman et al. 2007). The MCL algorithm (van Dongen 2000) was then used to group nodes (genes) into clusters in a completely unsupervised manner. This algorithm uses the innate structure of the graph together with the correlation values associated with edges to divide the network into groups of co-expressed genes. In this respect MCL has been shown to perform as well or better than other network clustering algorithms (Brohee and van Helden 2006). This approach differs from conventional clustering methods in viewing the data globally i.e. all 'connections' between genes are considered in the placement of genes in the network rather than simply considering data connectivity in a pairwise, hierarchical fashion. Others have also sought to identify gene signatures/modules in cancer data, but have used different analytical approaches, less data, grouped far fewer genes and generally failed to explain the biological significance of their findings (Mosca et al

2009, Rhodes et al 2004, Segal et al 2004, Shaffer et al 2006, Shi et al. 2010). Where correlation networks have been used previously to analyse modularity in gene expression data (Langfelder and Horvath 2008, Stuart et al 2003), available computing frameworks have not permitted the visualization or exploration of the resultant graphs and these studies have focused on only a single dataset. Advances in the ability to generate large graphs allows all the data from a gene expression signature to be considered in a single graph rather than having to apply some supervision to the graph whether that be selection of genes thought to be of interest, selection of differentially expressed genes or correlation with outcome. The approach used here in contrast to many of those discussed above makes no prejudgement about what is important in the data.

The primary focus of this study was to learn more about the functional profile of tumour-associated immune cells. To this end six large datasets of unrelated human primary tumours derived from different populations were examined initially, followed by an unrelated dataset to validate some of the findings from the 'core' datasets. Network analysis of individual cancer datasets demonstrated the correlation graphs to possess a highly structured topology. Detailed examination of these graphs revealed that many of the clusters of genes observed at the micro and macro level of graph structure were highly enriched in functionally related genes. It was clear from the examination of the clusters that this approach was very effective in identifying cohorts of genes expressed in a cell- or pathway- specific manners, as discussed below. Also observed were groups of genes associated with known disease modules (Barabasi et al 2011). These tended not to form a major component of the graph lying mostly as smaller clusters separate from or loosely connected to the main graph. For example, the breast cancer dataset contained a small group of genes whose expression is lower in aggressive tumour subtypes. This included ESR1 (oestrogen receptor alpha), GATA3, FOXA1, and XBP1, and therefore appears to capture a significant proportion of the oestrogen signalling transcriptional network, including modulators and downstream targets (Figure 4.19). Similarly, in the DLBCL dataset, associated with IRF4, one of the markers of the ABC-subtype (Alizadeh et al 2000) were FOXP1, PIM2 and CARD11, all described to be up-

regulated in ABC-subtype of DLBCL (Figure 4.14) (Davis et al 2010). A further example of a disease-specific network comes from the analysis of the nearest neighbours of *SILV* in the skin cancer data (Figure 4.42) (Riker et al 2008). The immediate neighbours of *SILV*, a gene whose product pMel17 (recognised by the HMB-45 antibody) is used clinically in the diagnosis of melanoma (de Vries et al 2001), include *TYR* (tyrosinase) the key enzyme in melanin biosynthesis, *MLPH* (melanophilin), which plays a role in melanosome transport, *GPR 143* expressed on the melanosome membrane, *MLANA* (melan-a), *MITF*, a melanocytic transcription factor and transcriptional regulator of many of these genes (reviewed in (Steingrimsson et al 2004)) and *SNCA* (alpha-synuclein). Whilst there is no literature regarding the expression of *SNCA* in melanoma, given that it plays a role in dopamine metabolism in the CNS, it is not inconceivable that it also functions in melanin metabolism (Sidhu et al 2004), highlighting the potential of this approach to identify novel genes in functional networks. In addition, *CDK2*, a gene downstream of and regulated by MITF and known to be required for melanoma growth and proliferation (Du et al 2004) lies within these 'near neighbours', as does BIRC7, an anti-apoptotic protein contributing to melanoma survival and resistance to therapy (Lazar et al 2010, Vucic et al 2000) which is also regulated by MITF (Dynek et al 2008). In all cases it would appear that the graphs have accurately identified key modules associated with key disease genes and other genes lying in the immediate neighbourhood must surely merit further investigation. The focus of the analysis was not on identifying dysregulated pathways in the individual tumours, but the presence of such networks capturing these pathways known to be of importance in the individual tumours provides some further validation of the method (Barabasi et al 2011).

Much of the recognisable structure of each of the graphs was formed of three major components; cell cycle related clusters, extracellular matrix clusters and immune cell clusters. These were clearly demonstrated to be significantly enriched in genes associated with individual cells or functions and were annotated on that basis. The exact composition of these elements varied, as might be predicted from tumour to tumour with for example a neutrophil signature only being identifiable in tumours in

which histologically neutrophils are prominent. A lymphocyte signature and a macrophage signature were identified in all tumour types. There tended in all tumours to be a relatively tight grouping of immune clusters in one area of the graph with the exact degree of relatedness of these varying between tumours, suggesting a degree of interaction between these elements, reflective of the known interactions between for example lymphocytes and cells of the innate immune system. Similarly, with the exception of the teratoma signatures in the testicular tumour dataset, the extracellular matrix signatures tended to lie together in all graphs and often very closely related to vascular signatures, again suggesting a relationship between these. Similarly cell cycle signatures broke into somewhat different clusters in individual tumours but the cluster tended to group together in every case.

Many of the clusters seen and the patterns of gene expression they showed clearly correlated with histological features of the tumour in question providing further validation for the approach as a method of analysing the cancer transcriptome. For example, the extracellular matrix signature showed a higher level of expression in PMBL than other types of DLBCL, in keeping with the pattern of fibrosis commonly seen in this tumour (Figure 4.12); the cell cycle signature rises in breast cancers of increasing grade and ovarian cancers of increasing malignancy (Figures 4.19 and 4.24); somatic tissue signatures are seen in teratomas in the testicular tumour dataset (Figure 4.35). It was not possible to assign annotation to all clusters or to examine in detail many of the clusters not related to the stromal components identified in each of the individual tumours, although the expression pattern of many of these was very interesting and appeared worthy of further analysis.

In an effort to further define what elements of the signatures were conserved across all tumours a ‘core’ signature was defined by using the mean Pearson correlations between six datasets. This would tend to remove any tumour- or tissue-specific signatures leaving only those signatures common to all tumours (Fig 4.40). To confirm the signatures were truly conserved and not simply present only in the original datasets, this conserved signature was mapped to an independent dataset of skin tumours showing good preservation of the signatures (Fig 4.41 and table 4.78).

As predicted, a number of the clusters generated by this approach were highly similar to those that we had observed in the individual datasets, but with the loss of those that were tumour/tissue-type specific. The conserved gene signatures consisted predominantly of three categories of functionally related genes. By far the largest grouping of genes contained genes whose expression profile over the individual datasets showed little variation and which were poorly characterised. For these reasons they were designated 'house-keeping' (HK) clusters and are not discussed further. A second major structural feature within the 'core' graph was clearly associated with the cell cycle. A central cluster within this area of the graph was highly enriched in genes associated with cell cycle progression when compared to previously derived cell cycle lists (Huang et al 2009, Subramanian et al 2005). It is perhaps unsurprising that a signature relating to proliferation is shared across multiple tumours and when this signature was examined in the individual datasets there was a clear up-regulation of these genes in more aggressive tumours, suggesting that this signature is contributed largely by the malignant cells rather than the stromal elements. This cell cycle cluster is also closely associated with clusters of genes involved in mRNA/protein processing. These presumably represent functional modules whose expression and activity are tied to cellular proliferation. The final class of conserved gene signatures were associated with the stromal components of a tumour i.e. tumour-associated macrophages, other leukocytes, vascular components and mesenchymal cells.

The preservation of specific clusters associated with stromal cells across such a large number of genetically diverse individuals and multiple tumour types argues there is, at least in part, a common tumour microenvironment that controls, and is controlled by, interactions amongst elements of the stroma. One of the larger conserved signatures observed is associated with the extracellular matrix. This was enriched in structural proteins, proteoglycans, modifiers of the extracellular matrix and signalling molecules. Histologically, the presence of a desmoplastic tumour stroma is a well recognised phenomenon occurring in many tumour types. However, like many other elements of the microenvironment the precise role played by this tumour-associated

stroma has been difficult to assess: is the role of the stroma to contain the tumour or to promote the survival of the malignant cells? Recent data from studies of DLBCL suggest that in this tumour different elements of the stroma contribute to both a good and a poor prognosis with elements of a vascular signature particularly being associated with poor prognosis while other stromal elements were associated with a good prognosis (Lenz et al 2008a). The elements they describe as being in the 'good' stomal signature tend to fall into the ECM cluster and to a lesser extent the macrophage cluster in the clustering presented here, while much of the 'bad' stromal signature tends to fall in the endothelial cluster derived here. Work in small cell lung carcinoma has established the role of interactions between the ECM and tumour cells in resisting chemotherapy-induced death highlighting one possible role for the tumour stroma (Hodkinson et al 2006). More recently work by Kim *et al* (Kim et al 2009) in a lung carcinoma model highlighted the role that an ECM component, in this case versican, can play in activating other elements of the microenvironment suggesting that as for other elements of the tumour microenvironment, cross-talk between elements is likely to be of great importance. Similarly, a role has been demonstrated for sequestration of angiogenic signalling molecules in the ECM with these being released by the actions of MMPs (Bergens et al 2000).

The 'core' vasculature signature observed here contains many familiar and well-characterised markers of endothelial cells as well as less well characterised endothelial genes. It contains receptors and co-receptors (*KDR, NRP2*) for VEGF, the major angiogenic factor but also contains elements associated with Notch signalling, another important system in angiogenesis (for a review see (Phng and Gerhardt 2009)). NOTCH3 usually expressed in vascular smooth muscle, lies in the endothelial-related cluster enriched in ECM and basement membrane proteins and thought to represent the structural component of the vascular signature. A recent study investigating the crosstalk between endothelial and mural cells via NOTCH3 signalling showed a reduction in angiogenesis in an *in-vitro* co-culture system when NOTCH3 is knocked out in mural cells (Liu et al 2009).

The macrophage signature in the 'core' dataset is smaller than in other datasets, some elements clearly having been lost as having mean correlation coefficients below the selected threshold. The preserved signature contains gene associated with phagocytosis, MHC class II antigen presentation and T-cell co-stimulation and is therefore suggestive of a macrophage able to act as an antigen-presenting cell. The TAM profile also contains scavenger receptors and genes involved in lipid metabolism suggesting a role in apoptotic cell clearance by TAMs. Notably, the signature does not contain many of the 'classical' markers of macrophage polarisation such as IL-10, IL-12, arginase, etc. While some of these activation markers, derived from mouse data, are not expressed in human macrophages and therefore of little relevance in studying human macrophages, the absence of clear evidence of polarisation is interesting (Martinez et al 2013). When analysed in individual datasets, genes such as IL-10 and IL-12 appear often to be expressed in a very variable manner and frequently do not appear in the network analysis at all, suggesting that these macrophage polarising signals do not represent a uniform response to tumour in any of the datasets we have studied. It must be noted though, that as with all gene expression data generated from more than a single cell, that signatures represent an 'average' impression of cell type within a tissue and cannot provide information about whether, and/or where different subtypes of cells exist in that tissue.

Analysis of the T cell profile demonstrates the presence of an almost completely intact antigen recognition and signalling pathway containing elements of the TCR, co-receptors and downstream signalling molecules. Also in the signature are cytotoxic molecules and markers of activation suggesting these are, at least in part, activated cytotoxic T cells. There is not obviously a T_{reg} signature, but this may reflect in part the relatively limited number of markers of that cell type making the signature difficult to identify if the key markers are not present in a cluster. The preservation across all tumour types of a cluster of genes associated with an interferon response and the presence of *IFNG* in the T cell signature argues that activation of this pathway forms a consistent part of the response to a tumour. This is

in keeping with data derived from murine models in which it was shown that TAMs express many interferon inducible genes (Biswas et al 2006). Taken together, these data do not support the view that TAMs have a so-called M2 (Mantovani et al 2009) phenotype characterised by dominant actions of interleukin 4, rather they appear to represent a cell type with some features in common with alternatively activated macrophage, but areas of difference. Given that the M2 macrophage is an extreme polarisation state derived from a single polarizing signal in vitro, it would be predicted that this signature would not be reproduced completely in the complex microenvironment of the tumour even if certain elements of the signature, of benefit to the tumour, are present.

Of interest is how the different stromal signatures relate to each other. The fact that macrophage, T cell, ECM and endothelial-specific genes form independent clusters, indicate that there is not a tight causal relationship between them. However, it is clear that the functions of these cells may be closely related, with the T cell signature and macrophage signature in particular, demonstrating a similar expression pattern and close spatial relationship in all network graphs. Despite the frequent association of macrophage number with microvessel density in solid tumours (Eerola et al 1999, Lissbrant et al 2000, Orre and Rogers 1999, Sickert et al 2005), these signatures are clearly separate in the tumours studied here suggesting that there are other factors involved in the development of the neovasculature of the tumour in addition to the role played by TAMs. This viewpoint is supported by the fact that the macrophage cluster does not contain any of the known regulators of endothelial proliferation, such as the vascular endothelial growth factors (VEGFs).

The TAM signature from the DLBCL dataset has many features in common with that from the core dataset but contains a larger number of genes. Like the core signature there are genes involved in chemotaxis and phagocytosis, genes involved in the innate recognition of pathogens and defence response (Table 4.3). Macrophage scavenger receptors including MARCO, SCARB2 and CD163 are expressed, a

finding reported from other studies of TAMs (Allavena et al 2010, Ford et al 2015). Complement component C1q, one of the linker molecules involved in clearance of apoptotic cells is expressed, as is LRP1, hinting at the macrophage's role in clearance of dead tumour cells. CXCL12, reported to promote a proangiogenic phenotype in macrophages is in the cluster, as is the angiogenic molecule TYMP, suggestive of a role in tumour angiogenesis (van Overmeire et al 2014). Recent work highlighted a role for NRP1 in recruitment of macrophages to the hypoxic areas of tumour (Casazza et al 2013). This is not present in the DLBCL TAM signature, but the closely related gene NRP2 is included.

To date there are few global gene expression profiles from TAMs and these are often derived from inbred mouse tumour models in which the cells have been separated from their microenvironment and therefore potentially had their gene expression altered by the process of isolation. Previous work looking at a mouse fibrosarcoma model in which TAMS were released from the tissue by physical disruption and enzymatic digestions demonstrated many IFN-inducible genes, despite the presence of the immunosuppressive cytokine IL-10, suggesting that in mouse as in the data presented here the interferon induced signature is important (Biswas et al 2006). They demonstrated that TAMs isolated from tumours maintained an altered response to LPS, with reduced production of proinflammatory genes including IL-12, IL-6, TNF-α and CCL3 suggesting that regardless of whether or not the TAM can be defined as an 'alternatively' activated macrophage, it is resistant to classical inflammatory activation . AIF1, a marker of organ rejection in humans (analysed further in chapters 5 and 6) was also demonstrated as an upregulated gene in TAMs in their analysis (Biswas et al 2006). TAMs from a mouse model of breast cancer in which cells were isolated by tissue mincing, phagocytososis of labelled dextran and fluorescence activated cell sorting showed increased abundance of genes involved in 'immune response', 'tissue development' and 'angiogenesis' among others (Oljavo et al 2010). Similarly to Biswas et al, Oljavo et al found evidence of the pro-inflammatory mediators, in this case the chemokine CCL3 suggesting the TAM profile is more complex than simply an 'alternatively' activated macrophage (Oljavo

et al 2010). Work by Sofia Petrova in our group demonstrated using laser-capture microdissection of macrophages from intact tumours a gene expression profile in a xenograph mouse model in which anti-inflammatory and pro-angiogenic genes were expressed by TAMs and in which the TAM signature was enriched for genes involved in phagocytosis, lipid metabolism and matrix remodelling (Ford et al 2015). There was overlap between the signature generated from mouse, and those generated from human tumours by a bioinformatic approach, but not all genes seen in either setting were seen in the other. As well as differences in analysis this is likely to represent differences in host and tumour type. Allavena et al in one of the few studies in humans and focussing on expression of C-type lectins showed in TAMs from human ovarian carcinoma that c-type lectin receptor expression was similar between TAMs in malignant ascites and macrophages from peritoneal washings, with the exception of a few differentially expressed genes including the mannose receptor (CD206) which showed higher expression in TAMs from malignant ascites than resident peritoneal macrophages (Allavena et al 2010). In the TAM population they studied, on binding of the mannose receptor by tumour mucins there was production of IL-10, but even with IFNγ and LPS stimulation these cells cultured ex vivo could not be induced to express IL-12. Similarly Ford et al showed upregulation of the mannose receptor on TAMs in a mouse xenograft model (Ford et al 2015). The TAM profile from DLBCL generated here did not contain the mannose receptor, which may reflect either a genuine absence or intratumoural heterogeneity of TAMs such that the correlation pattern does not cluster with the rest of the macrophage signature. One of the issues highlighted by the Allavena et al paper is that of comparison with a normal macrophage population (Allavena et al 2010). It is very difficult when comparing macrophage signatures with TAM signatures to know what to consider as a 'normal' macrophage. Is it a resident tissue macrophage? Or a tissue reparative macrophage, a tingible body macrophage? Macrophages represent such intrinsically variable and plastic cells that differences between the normal and tumour associated macrophage unless very carefully selected may simply be the differences between the environment in which they exist rather than any true functional difference due soley to the tumour.

In the bioinformatic data presented here, no equivalent normal signature is present, the datasets consisting of tumour tissue only. However, the macrophages studied would be subject to the multiple influences of the tumour microenvironment and hence the genes that they express are likely to be those they express in the face of multiple apoptotic tumour cells, and/or a hypoxic microenvironment, and/or under the influence of factors secreted by the tumour cells and thus while the analysis cannot inform as to which genes are up- or down-regulated in the face of these multiple influences, it can inform as to which genes are expressed in this context.

It is impossible to know the extent to which the gene expression signatures relate to protein level and localisation in the cells. For example while there is expression at the mRNA level of CD86 in the DLBCL TAM suggesting expression of co-stimulatory molecules, the cell surface expression and availability for signalling of the molecule can only be assumed. Expression at the protein level of some genes of interest was performed in chapter 5, but it was not within the scope of the thesis to explore all possibilities from the signatures. It is also clear that there is considerable intra-tumoral heterogeneity of TAMS (van Overmeire et al 2014). This is unsurprising given the heterogeneity seen within tumours, with areas of differential oxygenation frequently seen histologically as viable areas surrounding blood vessels, sub-clonal evolution giving rise to areas of morphological variation, differences in abundance of stromal components etc. For this reason it was important to study the spatial distribution of the molecules of interest in the tissue. A signature derived from whole tissue can only ever supply an 'average' of all the components contributing to the signature which may mask considerable spatial variation in protein expression. This was explored to a limited extent in chapter 5.

8.4 Validation of the TAM signature on an independent cohort of DLBCL

Markers of TAMs from the *in silico* analysis of DLBCL and other tumour types were examined in a large cohort of cases of reactive lymph nodes and DLBCL by

immunohistochemistry. This carried the *in silico* analysis back to human tumour tissue and allowed assessment of the spatial expression of the studied proteins. Gene expression analysis in any setting, as with many other techniques that look at homogenized tissue, represents an 'average' of the tissue studied with loss of any spatial heterogeneity. Examination of the selected markers at the protein level by immunohistochemistry did not provide a direct validation of the gene expression signature, looking at protein rather than mRNA expression, but does provide additional information about where and by what cells the proteins are being expressed.

All the selected markers identified in the signature and for which acceptable immunohistochemical staining could be identified were present in macrophages in tumour tissue and reactive lymph nodes (Figures 5.2 – 5.8), but several showed marked variation in expression patterns with LGALS3BP in particular showing only weak and focal expression in tumours.

Defining an equivalent 'normal' tissue in which to look at expression outwith the tumour is difficult and a reactive lymph node population was used as a non-tumour tissue. The normal lymph node is a very changeable environment and given the plasticity of the macrophage lineage, there are multiple possible interactions of the macrophage and its environment all likely to represent subtly different activation states. The tissue that was used was from lymph nodes that did not contain lymphoma and crucially were not derived from the drainage field of epithelial malignancies (which numerically represent the vast majority of non-lymphomatous lymph nodes in the pathology archives), and therefore not susceptible to soluble influences from tumours. The nodes used were lymph nodes removed in the investigation of lymphadenopathy, not containing lymphoma and representing a variety of patterns of reactive lymphadenopathy secondary probably to multiple infectious/inflammatory/reactive conditions.

Macrophages in the non-tumour lymph nodes were considered in three, broad, morphologically identifiable groups; sinus histiocytes, stromal macrophages and tingible body macrophages. CD68 and CD163 were included in the panel for analysis as both are recognised markers of the monocyte/macrophage lineage and allowed comparison of the expression pattern of these to the other less 'classical' macrophage markers. Limited double staining was undertaken where the markers were expressed in different cellular compartments. Due to the high autofluorescence of formalin-fixed paraffin embedded tissue, only chromogenic labelling of cells was carried out which limited the combinations of antibodies studied. CD68, AIF1, LAIR 1 and TYMP were seen in all compartments in the reactive lymph node although AIF1 showed some variability in intensity of staining with much more intense staining seen in sinus histiocytes than other macrophages (Figures 5.2 -5.8). Interestingly, CD163 staining was seen mostly in sinus histiocytes with only occasional weak staining in tingible body macrophages and absence in stromal macrophages. There was variable staining in tumours ranging from very occasional entirely negative cases, through cases with occasional positive cells to cases with many positive cells. This did not appear to relate to macrophage content as assessed by CD68 staining, as there were cases with abundant CD68 positive cells and sparse CD163 positive cells, neither did it relate to phagocytosis of apoptotic cells as despite the relative absence in tingible body macrophages, expression was seen in tumour with high rate of apoptosis where the macrophages contained abundant apoptotic debris. CD163 is widely quoted in the literature as a marker of M2 macrophages, as a direct extrapolation from mouse data. A concensus statement however from a group of 'macrophage scientists' in an effort to bring clarity to the confused terminology of macrophage activation/polarisation highlights the fact that on activation by IL-4 i.e. the originally described M2 activation, CD163 is not expressed by human macrophages (Murray et al 2014). CD163 is a member of the scavenger receptor cystine-rich family and expressed exclusively on cells of the monocyte lineage and used here as a marker of that lineage (Buechler et al 2000). Expression is regulated by cytokine expression, tending to be upregulated by IL-10 and IL-6 and downregulated by IFNγ, LPS, TNFα and IL-4 (Buechler et al 2000). It is clear from the non-tumour lymph node TMA that it is only expressed in certain

subpopulations of macrophages being seen more frequently in sinus histiocytes and being seen also in subpopulations of macrophages in tumours. This is supported by the double-labelling immunohistochemistry (Figure 5.13) where expression of TYMP and CD163 are seen together in a proportion of cells but single positive populations are identified for each marker.

The marker from the TAM signature which showed most variation in expression was LGALS3BP which showed expression almost exclusively in sinus histiocytes in the 'normal' lymph node and limited expression in TAMs.

The genes selected from the *in silico* signature of TAMs did all appear to be expressed by TAMs in DLBCL, but none of these appeared to be specific to TAMs all also being expressed by macrophages in the non-lymphomatous lymph node. The absence of specificity for TAMs does not however exclude a role for any of these molecules in support of the tumour as functions seen in the reactive lymph node may also be functions that provide support to the tumour. The germinal centre, in particular, showing high rates of proliferation and death and a rapid expansion in size could be argued to share many features seen in malignancy.

8.5 Relationship of expression of AIF1, LAIR1, LGALS3BP and TYMP with outcome in a patient cohort treated with CHOP

Four genes of interest from the macrophage signature were studied further. These were all selected on the basis that their known function suggested that they may have a role to play in the macrophage/tumour interaction, and their expression in DLBCL had not previously been studied. Of these, three of the genes, AIF1, TYMP and LGALS3BP did not show any association with outcome when their expression pattern was studied in a large cohort of DLBCL treated with CHOP. This may reflect a genuine absence of function of these molecules in the tumour microenvironment, or perhaps more likely, a minor effect, not identifiable by the relatively crude measure of patient outcome. It is likely that the tumour microenvironment represents a complex balance of multiple influences on tumour growth and survival and is likely to be subtly different in every tumour representing

both the individual genetic landscape of the patient and the mutational landscape of the tumour. Unless any effect exerted by a single molecule has a profound effect on tumour cell growth then it will be difficult to detect on the basis of patient outcome alone. It is of note from the large scale gene expression analyses that these have highlighted the importance of a signature made up of large numbers of genes in predicting prognosis and that examination of isolated genes/their protein products have proved far less successful in predicting prognosis emphasising the multiple elements that contribute to the prognostic signature (Alizadeh et al, 2000, Lenz et al 2008c, Adida et al 2000, Barrans et al 2004, Gascoyne et al 1997, Iqbal et al 2006, Uherova et al 2001, Hsi 2001, Ohshima et al 2001)

The only marker that showed some relationship with outcome was LAIR1 which showed a significantly better overall survival in cases where the TAMs showed a low intensity of staining than where there was high intensity of staining, but no significant difference was seen in disease free survival suggesting that this may represent a spurious association rather than a genuine effect of TAMs on outcome. Examination of a larger cohort of cases may help to address this question as despite the non-significant result, the Kaplan-Meier curves for disease free survival for this parameter remained separate. LAIR1 contains two ITIM motifs and functions as an inhibitory receptor (Meyaard et al 1997). The pattern of expression noted in this cohort of cases is suggestive that lower expression of LAIR1 results in more readily activatable macrophages and hence better outcome. As such it suggests that if the association seen is genuine, then the role of LAIR1 may either be via activation to a tumouricidal role in the face of therapy or activation to a tumour supportive role in the developing tumour. Again, as human archival tissue represents a single timepoint in the life of the lymphoma, this question is not answerable on the current data.

The ligand for LAIR1 is part of the collagen molecule and hence it has a plausible role in the microenvironment of the tumour (Lebbink et al 2006). Collagen is a

common component of almost every tissue, the aberrant tissue of tumours included. As such it is possible that binding of macrophages to collagen in almost any setting outwith the blood stream will increase the activation threshold of the cell. While this has been studied in T cells and NK cells it has not been studied in macrophages and it is unclear if it plays a role in macrophage activation and whether the role it plays is affected by the nature of the activating signal, i.e. binding via Fc receptors, toll-like receptor signalling pathways, cytokine signalling or others (Meyaard et al 1997, Verbrugge et al 2003, Maasho et al 2005). The final chapter of preliminary experiments looks at the effects of macrophage activation on expression of LAIR1 in an *in vitro* cell line model using THP-1 cells.

8.6 Preliminary data on the effect signalling through LAIR1 has on macrophage activation

Given the possibility that LAIR1 expression by TAMs in human DLBCL may impact on outcome, some further studies were undertaken looking at the expression of LAIR-1 in various activation states of macrophage-like cells differentiated from THP-1 cells and the effect the addition of collagen had on malignant cell growth. Whilst this is very preliminary work and requires repeated and expanded, certain points can be drawn from it. In THP-1 cells, there appears to be a downregulation of LAIR1 surface expression as the cells differentiate towards macrophage-like cells (Figure 7.2). A similar pattern of downregulated expression has been seen in human B cell and T cells as they mature from naive to memory cells (Jansen et al 2007, Maasho et al 2005, Merlo et al 2005). This alters the activation threshold in naive as opposed to antigen-experienced lymphocytes and could be considered as one of the mechanisms contributing to appropriate lymphocyte activation. The role of downregulation in cells of the innate immune system is less clear, but it could be argued that as monocytes are recruited out of circulation to become macrophages in inflammatory/regenerating/apoptotic sites there is a need for them to be more activatable than when in circulation. Expression has not previously been studied in human monocytes differentiating to macrophages.

The data as to the role macrophage polarisation has on expression is somewhat unclear; the does not appear to be differences in expression levels on the cell surface by flow cytometry and there is no difference in levels of LAIR1 mRNA, but western blotting using an antibody against an intracellular component of the molecule suggests a cleavage product is differentially present, being seen more when cells are activated with IFNγ and LPS (Figure 7.3). It is difficult to equate the cell surface expression and whole cell lysate data, but it must be noted that the flow cytometry data is confounded by the high levels of cell death induced by detaching the cells from the culture dishes, an effect not seen in the Western blotting where lysis was performed on the cell culture plates. Secretion of either the extracellular component of LAIR1 or of LAIR2 molecule is reported in various conditions and it would be of interest to see if there was shedding of LAIR1 in the supernatants of the THP-1 cells (Ouyang et al 2004, Lebbink et al 2008). These experiments should also be repeated with primary human monocytes as while THP-1 cells provide a convenient and reproducible source of macrophage-like cells, clearly they may not have a normal LAIR1 signalling pathway/response.

Co-culture of THP-1 macrophage-like cells showed differences in cell growth where IFNγ/LPS activated cells were cultured with or without collagen, BL2 cells growing with collagen showing an enhanced growth compared to those without collagen (Figure 7.8). No effect was seen in any of the other activation states. This suggests that LAIR1 may function to downregulate 'classical' inflammatory activation of macrophages. As such this could play a physiological role in dampening down the inflammatory response in wound healing situations but may also provide a growth advantage to tumour cells preventing activation of a cytotoxic phenotype in the presence of collagen. Mechanistically, this would suggest that the desmoplastic stroma seen in many tumours represents another route of immune suppression in the tumour microenvironment. Again these experiments require repetition in primary cells as well as in different culture situations. They were performed here in optimum growth conditions for BL2 cells, and the relationship of cell growth with

macrophages has been seen to alter dependent on the conditions in which the cells are maintained (C Ford, unpublished observations).

This thesis examines some aspects of the tumour associated macrophage in diffuse large B cell lymphoma. Much further work could be done to expand this study; in the macrophage signature derived from the bioinformatic analysis genes were selected for further study on the basis that their function was not entirely unknown, however the genes in the signature about which very little is known also provide intriguing subjects for further study. Their presence within the TAM signature strongly argues that they are genuinely macrophage-expressed genes on a 'guilt-by association' principle and that they are probably relatively specifically expressed by macrophages or their signature would be 'diluted out' by expression in other cell types and they would not fall within the signature. The presence of macrophage genes in TAMs about which little or nothing is known is an interesting gap in our current knowledge.

Not directly related to the TAM signature, this analysis approach also highlighted genes of interest in the tumour signatures, beyond the scope of this thesis to explore. These suggest several interesting avenues to explore in relation to deregulated pathways in DLBCL. Further work could also expand the data looking at outcome in relation to expression of additional selected molecules of interest. Although all cases of DLBCL over a long time period were considered for analysis, given the relative rarity of the tumour the number of cases available for study is still relatively small and more subtle influences on outcome may be detectable with a larger cohort of cases. The influence of the selected genes on outcome could also be studied in a CHOP-Rituximab treated cohort of patients as the available data suggests that the role of the TAM may be altered by the addition of Rituximab. Finally the preliminary data presented in the final chapter needs repeated and expanded to a more physiological macrophage-like cell having been performed so far only in a malignant 'macrophage-like' cell line.

8.7 Future directions

Some further study of the molecules selected from the DLBCL TAM signature should be undertaken. In the first instance, the work performed looking at LAIR 1 in a co-culture model should be extended from the use of a malignant macrophage-like cell line to primary human monocyte derived macrophages to give a better impression of whether LAIR1 signalling is likely to play a role in macrophage activation in a more physiological model of macrophages. These experiments would look both at activation of macrophages in the presence and absence of collagen as well as by cross-linking of LAIR1 on the surface of macrophages by antibodies. As well as looking at the growth of tumour cells, measurable outcomes would include levels of secreted cytokines, surface expression of markers of activation and migratory and phagocytic activity of macrophages. Given the intriguing findings of a possible cleaved product of LAIR1 in the presence of 'classical' activation of the THP-1 cells, this finding also could be explored further. The presence of secreted LAIR1 or LAIR 2 in supernatants could be sought in different activation states to establish whether this cleaved form represented a way of modulating LAIR1 availability for signalling.

Other molecules from the TAM profile should also be studied. TYMP in particular, having a reported role in both protection of cells from apoptosis and promotion of angiogenesis, is a molecule of interest. Initially there are relatively simple in vitro experiments that should be performed to establish whether this may play a role in protection of lymphoma cells from apoptosis. Protection from apoptosis in many settings is shown to be due to the actions of the metabolite 2-deoxy-D-ribose, by a poorly understood mechanism (Brown and Bicknell, 1998, Ikeda et al 2006). Addition of this molecule to lymphoma cells grown in culture in a variety of conditions such as hypoxia and serum starvation may help establish whether this molecule by itself has a direct role to play in modulating rates of apoptosis. Similarly addition of the molecule in a co-culture setting could help establish whether any effect observed of 2-deoxy- D-ribose, is due only to action on tumour cells or whether it alters the way tumour cells interact with macrophages. Clearly

any effect seen in vitro may not be reflected in an effect in vivo, as is seen in the modulation of apoptosis in an animal model of Burkitt lymphoma reported in Ford et al (Ford et al 2015). The role, if any, of TYMP in modulating tumour growth in the in vivo situation is less straightforward but could be assessed by a variety of approaches; firstly by looking in real human lymphomas for any relationship between TYMP-expressing tumour associated macrophages and the microvasculature by immunohistochemistry for these two elements of the tumour stroma. Modulation of TYMP expression in an in vivo model of tumour development is less straightforward as, given the molecule is expressed in the 'normal' cells of the tumour host, manipulation of this would require manipulation either in the whole animal with potentially widespread effects beyond that of the tumour/macrophage interaction or as a conditional knockout in macrophages only, were such a model to be used. A TYMP knockout mouse does exist, which may represent a possible avenue to explore. A further possibility to explore looking only at the action of the 2-deoxy-D-ribose metabolite of TYMP, is the use of the stereoisomer 2-deoxy-L-ribose which has been used in both the co-culture and xenograft settings to inhibit the action of the active metabolite (Liekens et al 2007).

There are other markers in the TAM signature that are of interest for further study. These include several markers that were not assessed in this thesis at all in terms of protein expression and localisation and for these molecules that would represent the first step in further investigations of their role in TAMs with additional work dependent on the confirmation of the presence of these in TAMs and the relative patterns of expression between TAMs and tumours in benign reactive lymph nodes. Of particular interest, from the little that is known about them, are myoferlin and chitinase 3-like 1. Myoferlin is a protein implicated in abnormalities of muscle function and which is described in endothelial cells, where it is implicated in VEGF signal transduction, and more recently in tumour cells where is seems to play a role in membrane repair (Davis et al 2000, Bernatchez et al 2007, Bernatchez et al 2009, Leung et al 2013). It does appear however to be expressed by macrophages in the Human Protein Atlas (Uhlen et al 2010, Uhlen et al 2015) and is one of the genes

upregulated in M2-polarised macrophages in the gene expression study by Martinez et al (Martinez et al 2013). The function of this gene in macrophages is currently unknown. Chitinase 3-like 1 was present in the TAM signature in DLBCL but not the 'core' TAM signature. It is a secreted protein, found in activated macrophages, neutrophils, chondrocytes and synovial cells (Coffman 2008). It is said to be a marker of alternative macrophage activation (Lee et al 2009). What little mechanistic data there is suggests a role for the molecule in promoting tumor angiogenesis as well as roles in tissue remodelling and development of fibrosis, and given that it forms part of the DLBCL TAM signature but is not seen more widely in TAM signatures may inform as to remodelling of the microenvironment in DLBCL (Shao R, 2013, Fillippeli et al 2015).

References

Adida C, Haioun C, Gaulard P, Lepage E, Morel P, Briere J, Dombret H, Reyes F, Diebold J, Gisselbrecht C, Salles G, Altieri DC, Molina TJ. (2000). Prognostic significance of survivin expression in diffuse large B-cell lymphomas. *Blood*. 96(5):1921-1925.

Afshar-Sterle S, Zotos D, Bernard NJ, Scherger AK, Rödling L, Alsop AE, Walker J, Masson F, Belz GT, Corcoran LM, O'Reilly LA, Strasser A, Smyth MJ, Johnstone R, Tarlinton DM, Nutt SL, Kallies A. (2014). Fas ligand-mediated immune surveillance by T cells is essential for the control of spontaneous B cell lymphomas. *Nat Med.* 20(3):283-90

Ai WZ, Hou JZ, Zeiser R, Czerwinski D, Negrin RS, Levy R. (2009). Follicular lymphoma B cells induce the conversion of conventional CD4+ T cells to T-regulatory cells. *Int J Cancer*. 124(1):239-44.

Akiyama S, Furukawa T, Sumizawa T, Takebayashi Y, Nakajima Y, Shimaoka S, Haraguchi M. (2004). The role of thymidine phosphorylase, an angiogenic enzyme, in tumor progression. *Cancer Sci.* 95(11):851-7.

Allavena P, Sica A, Solinas G, Porta C, Mantovani A. (2008). The inflammatory micro-environment in tumor progression: the role of tumor-associated macrophages. *Crit Rev Oncol Hematol*. 66(1):1-9.

Allavena P, Chieppa M, Bianchi G, Solinas G, Fabbri M, Laskarin G, Mantovani A. (2010). Engagement of the mannose receptor by tumoral mucins activates an immune suppressive phenotype in human tumor-associated macrophages. *Clin Dev Immunol.* 2010:547179.

Allavena P, Mantovani A. (2012). Immunology in the clinic review series; focus on cancer: tumour-associated macrophages: undisputed stars of the inflammatory tumour microenvironment. *Clin Exp Immunol.* 167(2):195-205.

Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J Jr, Lu L,Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO, and Staudt LM. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*. 403(6769):503-511.

Alizadeh AA, Gentles AJ, Alencar AJ, Liu CL, Kohrt HE, Houot R, Goldstein MJ, Zhao S, Natkunam Y, Advani RH, Gascoyne RD, Briones J, Tibshirani RJ, Myklebust JH, Plevritis SK, Lossos IS, Levy R. (2011). Prediction of survival in diffuse large B-cell lymphoma based on the expression of 2 genes reflecting tumor and microenvironment. *Blood.* 118(5):1350-8.

Alkassab F, Gourh P, Tan FK, McNearney T, Fischbach M, Ahn C, Arnett FC, Mayes MD. (2007). An allograft inflammatory factor 1 (AIF1) single nucleotide polymorphism (SNP) is associated with anticentromere antibody positive systemic sclerosis. *Rheumatology (Oxford).* 46(8):1248-51.

Alvaro T, Lejeune M, Salvadó MT, Bosch R, García JF, Jaén J, Banham AH, Roncador G, Montalbán C, Piris MA. (2005). Outcome in Hodgkin's lymphoma can be predicted from the presence of accompanying cytotoxic and regulatory T cells. *Clin Cancer Res.* 11(4):1467-73.

Andjelic B, Mihaljevic B, Todorovic M, Bila J, Jakovic L, Jovanovic MP. (2012). The number of lymphoma-associated macrophages in tumor tissue is an independent prognostic factor in patients with follicular lymphoma. *Appl Immunohistochem Mol Morphol*. 20(1):41-6.

Aoukaty A, Lee IF, Wu J, Tan R. (2003). Chronic active Epstein-Barr virus infection associated with low expression of leukocyte-associated immunoglobulin-like receptor-1 (LAIR-1) on natural killer cells. *J Clin Immunol.* 23(2):141-5.

Arima J, Imazono Y, Takebayashi Y, Nishiyama K, Shirahama T, Akiba S, Furukawa T, Akiyama S, Ohi Y. (2000). Expression of thymidine phosphorylase as an indicator of poor prognosis for patients with transitional cell carcinoma of the bladder. *Cancer.* 88(5):1131-8.

Asai K, Hirano T, Kaneko S, Moriyama A, Nakanishi K, Isobe I, Eksioglu YZ, Kato T. (1992). A novel glial growth inhibitory factor, gliostatin, derived from neurofibroma. *J Neurochem.* 59(1):307-17.

Asselin-Labat ML, Sutherland KD, Barker H, Thomas R, Shackleton M, Forrest NC, Hartley L, Robb L, Grosveld FG, van der Wees J, Lindeman GJ, Visvader JE. (2007). Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. *Nat Cell Biol*. 9(2):201-9.

Bai M, Skyrlas A, Agnantis NJ, Kamina S, Tsanou E, Grepi C, Galani V, Kanavaros P. (2004). Diffuse large B-cell lymphomas with germinal center B-cell-like differentiation immunophenotypic profile are associated with high apoptotic index, high expression of the proapoptotic proteins bax, bak and bid and low expression of the antiapoptotic protein bcl-xl. *Mod Pathol.* 17(7):847-56

Barabási AL, Gulbahce N, Loscalzo J. (2011). Network medicine: a network-based approach to human disease. *Nat Rev Genet.* 12(1):56-68.

Barrans SL, Fenton JA, Banham A, Owen RG, Jack AS. (2004). Strong expression of FOXP1 identifies a distinct subset of diffuse large B-cell lymphoma (DLBCL) patients with poor outcome. *Blood*. 104(9):2933-2935.

Barrans S, Crouch S, Smith A, Turner K, Owen R, Patmore R, Roman E, Jack A. (2010). Rearrangement of MYC is associated with poor prognosis in patients with diffuse large B-cell lymphoma treated in the era of rituximab. *J Clin Oncol.* 28(20):3360-5

Bastard C, Deweindt C, Kerckaert JP, Lenormand B, Rossi A, Pezzella F, Fruchart C, Duval C, Monconduit M, Tilly H. (1994). LAZ3 rearrangements in non-Hodgkin's lymphoma: correlation with histology, immunophenotype, karyotype, and clinical outcome in 217 patients. *Blood.* 83(9):2423-7.

Becker R, Lenter M, Vollkommer T, Boos A, Pfaff D, Augustin H and Christian S. (2008). Tumor stroma marker endosialin (Tem1) is a binding partner of metastasis-related protein Mac-2 BP/90K. *The FASEB J.* 22, 3059-6067.

Bergers G, Brekken R, McMahon G, Vu TH, Itoh T, Tamaki K, Tanzawa K, Thorpe P, Itohara S, Werb Z, Hanahan D. (2000). Matrix metalloproteinase-9 triggers the angiogenic switch during carcinogenesis. *Nat Cell Biol*. 2(10):737-44

Bergers G, Benjamin LE. Tumorigenesis and the angiogenic switch. (2003*).* *Nat Rev Cancer*. 3(6):401-10.

Berglund M, Thunberg U, Amini RM, Book M, Roos G, Erlanson M, Linderoth J, Dictor M, Jerkeman M, Cavallin-Ståhl E, Sundström C, Rehn-Eriksson S, Backlin C, Hagberg H, Rosenquist R, Enblad G. (2005). Evaluation of immunophenotype in diffuse large B-cell lymphoma and its impact on prognosis. *Mod Pathol*. 18(8):1113-1120.

Bernatchez PN, Acevedo L, Fernandez-Hernando C, Murata T, Chaloni C, Kim J, Erdjument-Bromage H, Shah V, Gratton JP, McNally EM, Tempst P, Sessa WC. (2007). Myoferlin regulates vascular endothelial growth factor receptor-2 stability and function. *J Biol Chem.* 282(42):30745-53.

Bernatchez PN, Sharma A, Kodoman P, Sessa WC. (2009). Myoferlin is critical for endocytosis in endothelial cells. *Am J Physiol Cell Physiol.* 297(3):C484-492.

Biswas SK, Gangi L, Paul S, Schioppa T, Saccani A, Sironi M, Bottazzi B, Doni A, Vincenzo B, Pasqualini F, Vago L, Nebuloni M, Mantovani A, Sica A. (2006). A distinct and unique transcriptional program expressed by tumor-associated macrophages (defective NF-kappaB and enhanced IRF-3/STAT1 activation). *Blood*. 107(5):2112-22.

Biswas SK, Mantovani A. (2010). Macrophage plasticity and interaction with lymphocyte subsets: cancer as a paradigm. *Nat Immunol.* 2010 Oct;11(10):889-96.

Blanquicett C, Gillespie GY, Nabors LB, Miller CR, Bharara S, Buchsbaum DJ, Diasio RB, Johnson MR. (2002). Induction of thymidine phosphorylase in both irradiated and shielded, contralateral human U87MG glioma xenografts: implications for a dual modality treatment using capecitabine and irradiation. *Mol Cancer Ther*. 1(12):1139-45.

Bosco MC, Puppo M, Santangelo C, Anfosso L, Pfeffer U, Fardin P, Battaglia F, Varesio L. (2006). Hypoxia modifies the transcriptome of primary human monocytes: modulation of novel immune-related genes and identification of CC-chemokine ligand 20 as a new hypoxia-inducible gene. *J Immunol.* 177(3):1941-55.

Bournazou I, Pound JD, Duffin R, Bournazos S, Melville LA, Brown SB, Rossi AG, Gregory CD. (2009). Apoptotic human cells inhibit migration of granulocytes via release of lactoferrin. *J Clin Invest*. 119(1):20-32.

Brandt S, Montagna C, Georgis A, Schüffler PJ, Bühler MM, Seifert B, Thiesler T, Curioni-Fontecedro A, Hegyi I, Dehler S, Martin V, Tinguely M, Soldini D. (2013). The combined expression of the stromal markers fibronectin and SPARC improves the prediction of survival in diffuse large B-cell lymphoma. *Exp Hematol Oncol*. 2(1):27.

Brohée S, van Helden J. (2006). Evaluation of clustering algorithms for protein-protein interaction networks. *BMC Bioinformatics*. 7:488.

Bronckaers A, Gago F, Balzarini J, Liekens S. (2009). The dual role of thymidine phosphorylase in cancer development and chemotherapy. *Med Res Rev.* 29(6):903-53.

Brown LF, Harrist TJ, Yeo KT, Ståhle-Bäckdahl M, Jackman RW, Berse B, Tognazzi K, Dvorak HF, Detmar M. (1995). Increased expression of vascular permeability factor (vascular endothelial growth factor) in bullous pemphigoid, dermatitis herpetiformis, and erythema multiforme. *J Invest Dermatol.* 104(5):744-9.

Brown NS, Bicknell R. (1998). Thymidine phosphorylase, 2-deoxy-D-ribose and angiogenesis. *Biochem J*. 334 (Pt 1):1-8.

Buechler C, Ritter M, Orsó E, Langmann T, Klucken J, Schmitz G. (2000). Regulation of scavenger receptor CD163 expression in human monocytes and macrophages by pro- and antiinflammatory stimuli. *J Leukoc Biol.* 67(1):97-103.

Burisch J, Munkholm P. (2015). The epidemiology of inflammatory bowel disease. *Scand J Gastroenterol*. 50(8):942-51.

Calvo KR, Dabir B, Kovach A, Devor C, Bandle R, Bond A, Shih JH, Jaffe ES. (2008). IL-4 protein expression and basal activation of Erk in vivo in follicular lymphoma. *Blood.* 112(9):3818-26.

Carmeliet P, Jain RK. (2011). Molecular mechanisms and clinical applications of angiogenesis. *Nature*. 473(7347):298-307.

Casazza A, Laoui D, Wenes M, Rizzolio S, Bassani N, Mambretti M, Deschoemaeker S, Van Ginderachter JA, Tamagnone L, Mazzone M. (2013). Impeding macrophage entry into hypoxic tumor areas by Sema3A/Nrp1 signaling blockade inhibits angiogenesis and restores antitumor immunity. *Cancer Cell*. 24(6):695-709

Challa-Malladi M, Lieu YK, Califano O, Holmes AB, Bhagat G, Murty VV, Dominguez-Sola D, Pasqualucci L, Dalla-Favera R. (2011). Combined genetic inactivation of β2 Microglobulin and CD58 reveals frequent escape from immune recognition in diffuse large B cell lymphoma. *Cancer Cell*. 20(6):728-40.

Cheah CY, Herbert KE, O'Rourke K, Kennedy GA, George A, Fedele PL, Gilbertson M, Tan SY, Ritchie DS, Opat SS, Prince HM, Dickinson M, Burbury K, Wolf M, Januszewicz EH, Tam CS, Westerman DA, Carney DA, Harrison SJ, Seymour JF. (2014). A multicentre retrospective comparison of central nervous system prophylaxis strategies among patients with high-risk diffuse large B-cell lymphoma. *Br J Cancer*. 111(6):1072-9.

Chetaille B, Bertucci F, Finetti P, Esterni B, Stamatoullas A, Picquenot JM, Copin MC, Morschhauser F, Casasnovas O, Petrella T, Molina T, Vekhoff A, Feugier P, Bouabdallah R, Birnbaum D, Olive D, Xerri L. (2009). Molecular profiling of classical Hodgkin lymphoma tissues uncovers variations in the tumor microenvironment and correlations with EBV infection and outcome. *Blood.* (12):2765-3775.

Ch'ng ES, Tuan Sharif SE, Jaafar H. (2013). In human invasive breast ductal carcinoma, tumor stromal macrophages and tumor nest macrophages have distinct relationships with clinicopathological parameters and tumor angiogenesis. *Virchows Arch*. 462(3):257-67.

Choi WW, Weisenburger DD, Greiner TC, Piris MA, Banham AH, Delabie J, Braziel RM, Geng H, Iqbal J, Lenz G, Vose JM, Hans CP, Fu K, Smith LM, Li M, Liu Z, Gascoyne RD, Rosenwald A, Ott G, Rimsza LM, Campo E, Jaffe ES, Jaye DL, Staudt LM, Chan WC. (2009). A new immunostain algorithm classifies diffuse large B-cell lymphoma into molecular subtypes with high accuracy. *Clin Cancer Res*. 15(17):5494-502

Ci W, Polo JM, Cerchietti L, Shaknovich R, Wang L, Yang SN, Ye K, Farinha P, Horsman DE, Gascoyne RD, Elemento O, Melnick A. (2009). The BCL6 transcriptional program features repression of multiple oncogenes in primary B cells and is deregulated in DLBCL. *Blood*. 113(22):5536-48.

Coelho V, Krysov S, Ghaemmaghami AM, Emara M, Potter KN, Johnson P, Packham G, Martinez-Pomares L, Stevenson FK. (2010). Glycosylation of surface Ig creates a functional bridge between human follicular lymphoma and microenvironmental lectins. *Proc Natl Acad Sci U S A.* 107(43):18587-92.

Coffman FD. (2008). Chitinase 3-Like 1 (CHI3L1): a putative disease marker at the interface of proteomics and glycomics. *Crit Rev Clin Lab Sci.* 45(6):531-62.

Colomo L, López-Guillermo A, Perales M, Rives S, Martínez A, Bosch F, Colomer D, Falini B, Montserrat E, Campo E. (2003). Clinical impact of the differentiation profile assessed by immunophenotyping in patients with diffuse large B-cell lymphoma. *Blood*. 101(1):78-84

Conti A, Rodriguez GC, Chiechi A, Blazquez RM, Barbado V, Krènacs T, Novello C, Pazzaglia L, Quattrini I, Zanella L, Picci P, De Alava E, Benassi MS. (2011). Identification of potential biomarkers for giant cell tumor of bone using comparative proteomics analysis. *Am J Pathol.* 178(1):88-97.

Coutinho R, Clear AJ, Mazzola E, Owen A, Greaves P, Wilson A, Matthews J, Lee A, Alvarez R, da Silva MG, Cabeçadas J, Neuberg D, Calaminici M, Gribben JG. (2015). Revisiting the immune microenvironment of diffuse large B-cell lymphoma using a tissue microarray and immunohistochemistry: robust semi-automated analysis reveals CD3 and FoxP3 as potential predictors of response to R-CHOP. *Haematologica.* 100(3):363-9.

Curiel TJ, Coukos G, Zou L, Alvarez X, Cheng P, Mottram P, Evdemon-Hogan M, Conejo-Garcia JR, Zhang L, Burow M, Zhu Y, Wei S, Kryczek I, Daniel B, Gordon A, Myers L, Lackner A, Disis ML, Knutson KL, Chen L, Zou W. (2004). Specific recruitment of regulatory T cells in ovarian carcinoma fosters immune privilege and predicts reduced survival. *Nat Med.* 10(9):942-9.

Dakappagari N, Ho SN, Gascoyne RD, Ranuio J, Weng AP, Tangri S. (2012). CD80 (B7.1) is expressed on both malignant B cells and nonmalignant stromal cells in non-Hodgkin lymphoma. *Cytometry B Clin Cytom.* 82(2):112-9.

Davis DB, Delmonte AJ, Ly CT, McNally EM. (2000). Myoferlin, a candidate gene and potential modifier of muscular dystrophy. *Hum Mol Genet.* 9(2):217-226.

Davis RE, Brown KD, Siebenlist U, Staudt LM. (2001). Constitutive nuclear factor kappaB activity is required for survival of activated B cell-like diffuse large B cell lymphoma cells. *J Exp Med.* 194(12):1861-74.

Davis RE, Ngo VN, Lenz G, Tolar P, Young RM, Romesser PB, Kohlhammer H, Lamy L, Zhao H, Yang Y, Xu W, Shaffer AL, Wright G, Xiao W, Powell J, Jiang JK, Thomas CJ, Rosenwald A, Ott G, Muller-Hermelink HK, Gascoyne RD, Connors JM, Johnson NA, Rimsza LM, Campo E, Jaffe ES, Wilson WH, Delabie J, Smeland EB, Fisher RI, Braziel RM, Tubbs RR, Cook JR, Weisenburger DD, Chan WC, Pierce SK, Staudt LM. (2010). Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. *Nature.* 463(7277):88-92.

Deininger MH, Seid K, Engel S, Meyermann R, Schluesener HJ. (2000). Allograft inflammatory factor-1 defines a distinct subset of infiltrating macrophages/microglial cells in rat and human gliomas. *Acta Neuropathol*. 100(6):673-80.

Deininger MH, Meyermann R, Schluesener HJ. (2002). The allograft inflammatory factor 1 family of proteins. *FEBS Lett*. 514(2-3):115-21.

Del Galdo F, Maul GG, Jiménez SA, Artlett CM. (2006). Expression of allograft inflammatory factor 1 in tissues from patients with systemic sclerosis and in vitro differential

expression of its isoforms in response to transforming growth factor beta. *Arthritis Rheum*. 54(8):2616-25.

Deng L, Zhou JF, Sellers RS, Li JF, Nguyen AV, Wang Y, Orlofsky A, Liu Q, Hume DA, Pollard JW, Augenlicht L, Lin EY. (2010). A novel mouse model of inflammatory bowel disease links mammalian target of rapamycin-dependent hyperproliferation of colonic epithelium to inflammation-associated tumorigenesis. *Am J Pathol.* 176(2):952-67.

De Paepe P, Achten R, Verhoef G, Wlodarska I, Stul M, Vanhentenrijk V, Praet M, De Wolf-Peeters C. (2005). Large cleaved and immunoblastic lymphoma may represent two distinct clinicopathologic entities within the group of diffuse large B-cell lymphomas. *J Clin Oncol*. 23(28):7060-7068.

Diehl V, Kirchner HH, Schaadt M, Fonatsch C, Stein H, Gerdes J, Boie C. (1981). Hodgkin's disease: establishment and characterization of four in vitro cell lies. *J Cancer Res Clin Oncol*. 101(1):111-24.

Ding M, Fu X, Tan H, Wang R, Chen Z, Ding S. (2012). The effect of vascular endothelial growth factor C expression in tumor-associated macrophages on lymphangiogenesis and lymphatic metastasis in breast cancer. *Mol Med Rep.* 6(5):1023-9.

Doig TN, Hume DA, Theocharidis T, Goodlad JR, Gregory CD, Freeman TC. Coexpression analysis of large cancer datasets provides insight into the cellular phenotypes of the tumour microenvironment. BMC Genomics. 2013 Jul 11;14:469.

Du J, Widlund HR, Horstmann MA, Ramaswamy S, Ross K, Huber WE, Nishimura EK, Golub TR, Fisher DE. (2004). Critical role of CDK2 for melanoma growth linked to its melanocyte-specific transcriptional regulation by MITF. *Cancer Cell.* 6(6):565-76.

Duffield JS, Ware CF, Ryffel B, Savill J. (2001). Suppression by apoptotic cells defines tumor necrosis factor-mediated induction of glomerular mesangial cell apoptosis by activated macrophages. *Am J Pathol.* 159(4):1397-404.

Dynek JN, Chan SM, Liu J, Zha J, Fairbrother WJ, Vucic D. (2008). Microphthalmia-associated transcription factor is a critical transcriptional regulator of melanoma inhibitor of apoptosis in melanomas. *Cancer Res*. 68(9):3124-32.

Eeckhoute J, Keeton EK, Lupien M, Krum SA, Carroll JS, Brown M. (2007). Positive cross-regulatory loop ties GATA-3 to estrogen receptor alpha expression in breast cancer. *Cancer Res.* 67(13):6477-83.

Eerola AK, Soini Y, Pääkkö P. (1999). Tumour infiltrating lymphocytes in relation to tumour angiogenesis, apoptosis and prognosis in patients with large cell lung carcinoma. *Lung Cancer*. 26(2):73-83.

Eike MC, Olsson M, Undlien DE, Dahl-Jørgensen K, Joner G, Rønningen KS, Thorsby E, Lie BA. (2009). Genetic variants of the HLA-A, HLA-B and AIF1 loci show independent associations with type 1 diabetes in Norwegian families. *Genes Immun.* 10(2):141-50.

Elliott MR, Chekeni FB, Trampont PC, Lazarowski ER, Kadl A, Walk SF, Park D, Woodson RI, Ostankovich M, Sharma P, Lysiak JJ, Harden TK, Leitinger N, Ravichandran KS. (2009). Nucleotides released by apoptotic cells act as a find-me signal to promote phagocytic clearance. *Nature.* 461(7261):282-6.

Elliott MR, Ravichandran KS. (2010). Clearance of apoptotic cells: implications in health and disease. *J Cell Biol.* 189(7):1059-70.

Enzler T, Gillessen S, Manis JP, Ferguson D, Fleming J, Alt FW, Mihm M, Dranoff G. (2003). Deficiencies of GM-CSF and interferon gamma link inflammation and cancer. *J Exp Med*. 197(9):1213-9.

Espinosa I, José Carnicer M, Catasus L, Canet B, D'angelo E, Zannoni GF, Prat J. (2010). Myometrial invasion and lymph node metastasis in endometrioid carcinomas: tumor-associated macrophages, microvessel density, and HIF1A have a crucial role. *Am J Surg Pathol*. 34(11):1708-14.

Evan GI, Wyllie AH, Gilbert CS, Littlewood TD, Land H, Brooks M, Waters CM, Penn LZ, Hancock DC. (1992). Induction of apoptosis in fibroblasts by c-myc protein. *Cell.* 69(1):119-28.

Fadok VA, Bratton DL, Konowal A, Freed PW, Westcott JY, Henson PM. (1998). Macrophages that have ingested apoptotic cells in vitro inhibit proinflammatory cytokine production through autocrine/paracrine mechanisms involving TGF-beta, PGE2, and PAF. *J Clin Invest.* 101(4):890-8.

Fang HY, Hughes R, Murdoch C, Coffelt SB, Biswas SK, Harris AL, Johnson RS, Imityaz HZ, Simon MC, Fredlund E, Greten FR, Rius J, Lewis CE. (2009). Hypoxia-inducible factors 1 and 2 are important transcriptional effectors in primary macrophages experiencing hypoxia. *Blood.* 114(4):844-59.

Ferreri AJ, Bruno-Ventre M, Donadoni G, Ponzoni M, Citterio G, Foppoli M, Vignati A, Scarfò L, Sassone M, Govi S, Caligaris-Cappio F. (2015). Risk-tailored CNS prophylaxis in a mono-institutional series of 200 patients with diffuse large B-cell lymphoma treated in the rituximab era. *Br J Haematol*. 168(5):654-62.

Filippelli M, Cuppari C, Giacchi V, Lanzafame A, Rotolo N, Garozzo MT, Capizzi A, Musumeci M, Musumeci S, Leonardi S. (2015). Serum and BAL YKL-40 levels in different inflammatory lung diseases: an update. *J Biol Regul Homeost Agents.* 29 (2 suppl 1):130-6.

Ford CA, Petrova S, Pound JD, Voss JJ, Melville L, Paterson M, Farnworth SL, Gallimore AM, Cuff S, Wheadon H, Dobbin E, Ogden CA, Dumitriu IE, Dunbar DR, Murray PG, Ruckerl D, Allen JE, Hume DA, van Rooijen N, Goodlad JR, Freeman TC, Gregory CD. (2015). Oncogenic properties of apoptotic tumor cells in aggressive B cell lymphoma. *Curr Biol*. 25(5):577-88.

Fornarini B, D'Ambrosio C, Natoli C, Tinari N, Silingardi V and Iacobelli S. (2000). Adhesion to 90K (Mac-2 BP) as a mechanism for lymphoma drug resistance in vivo. *Blood.* 96 (6), 3282-3285.

Freeman TC, Goldovsky L, Brosch M, van Dongen S, Mazière P, Grocock RJ, Freilich S, Thornton J, Enright AJ. (2007). Construction, visualisation, and clustering of transcription networks from microarray expression data. *PLoS Comput Biol.* 3(10):2032-42.

Fried B, Reddy A, Mayer D. (2011). Helminths in human carcinogenesis. *Cancer Lett.* 305(2):239-49.

Friedkin M, Roberts D. (1954). The enzymatic synthesis of nucleosides. I. Thymidine phosphorylase in mammalian tissue. *J Biol Chem*. 207(1):245-56.

Fujimoto J, Ichigo S, Sakaguchi H, Hirose R, Tamaya T. (1998). Expression of platelet-derived endothelial cell growth factor and its mRNA in uterine endometrium during the menstrual cycle. *Mol Hum Reprod*. 4(5):509-13.

Fujimoto J, Sakaguchi H, Hirose R, Wen H, Tamaya T. (1999). Clinical implication of expression of platelet-derived endothelial cell growth factor (PD-ECGF) in metastatic lesions of uterine cervical cancers. *Cancer Res*. 59(13):3041-4.

Fujita N, Nishie A, Aishima S, Kubo Y, Asayama Y, Ishigami K, Kakihara D, Ushijima Y, Takayama Y, Shirabe K, Oda Y, Honda H. (2014). Role of tumor-associated macrophages in the angiogenesis of well-differentiated hepatocellular carcinoma: pathological-radiological correlation. *Oncol Rep.* 31(6):2499-505.

Fukaya Y, Shimada H, Wang L, Zandi E, and DeClerck Y. (2008). Identification of Galectin-3-binding Protein as a Factor Secreted by Tumor Cells That Stimulates Interleukin-6 Expression in the Bone Marrow Stroma. *J. Biol. Chem.* 283(27), 18573-18581.

Fukushima H, Ohsawa M, Ikura Y, Naruko T, Sugama Y, Suekane T, Kitabayashi C, Inoue T, Hino M, Ueda M. (2006). Mast cells in diffuse large B-cell lymphoma; their role in fibrosis. Histopathology. 49(5):498-505

Fusco O, Querzoli P, Nenci I, Natoli C, Brakebush C, Ullrich A and Iacobelli S. (1998). 90K (Mac-2 BP) gene expression in breast cancer and evidence for the production of 90K by peripheral-blood mononuclear cells. *Int J Cancer*. 79 (1), 23-26.

Gascoyne RD, Adomat SA, Krajewski S, Krajewska M, Horsman DE, Tolcher AW, O'Reilly SE, Hoskins P, Coldman AJ, Reed JC, Connors JM. (1997). Prognostic significance of Bcl-

2 protein expression and Bcl-2 gene rearrangement in diffuse aggressive non-Hodgkin's lymphoma. *Blood*. 90(1):244-251.

Ghilardi C, Chiorino G, Dossi R, Nagy Z, Giavazzi R, Bani M. (2008). Identification of novel vascular markers through gene expression profiling of tumor-derived endothelium. *BMC Genomics*. 9:201.

Ghose A, Elias HK, Guha G, Yellu M, Kundu R, Latif T. (2015). Influence of Rituximab on Central Nervous System Relapse in Diffuse Large B-Cell Lymphoma and Role of Prophylaxis-A Systematic Review of Prospective Studies. *Clin Lymphoma Myeloma Leuk*. S2152-2650(15)00070-1.

Goede V, Klein C, Stilgenbauer S. (2015). Obinutuzumab (GA101) for the Treatment of Chronic Lymphocytic Leukemia and Other B-Cell Non-Hodgkin's Lymphomas: A Glycoengineered Type II CD20 Antibody. *Oncol Res Treat*. 38(4):185-92.

Golay J, Zaffaroni L, Vaccari T, Lazzari M, Borleri GM, Bernasconi S, Tedesco F, Rambaldi A, Introna M. (2000). Biologic response of B lymphoma cells to anti-CD20 monoclonal antibody rituximab in vitro: CD55 and CD59 regulate complement-mediated cell lysis. *Blood*. 95(12):3900-8.

Golpon HA, Fadok VA, Taraseviciene-Stewart L, Scerbavicius R, Sauer C, Welte T, Henson PM, Voelkel NF. (2004). Life after corpse engulfment: phagocytosis of apoptotic cells leads to VEGF secretion and cell growth. *FASEB J*. 18(14):1716-8.

Gonzalez FJ, Vicioso L, Alvarez M, Sevilla I, Marques E, Gallego E, Alonso L, Matilla A, Alba E. (2007). Association between VEGF expression in tumour-associated macrophages and elevated serum VEGF levels in primary colorectal cancer patients. *Cancer Biomark*. 3(6):325-33.

Gregory CD, Pound JD. (2011). Cell death in the neighbourhood: direct microenvironmental effects of apoptosis in normal and neoplastic tissues. *J Pathol*. 223(2):177-94.

Griffiths L, Dachs GU, Bicknell R, Harris AL, Stratford IJ. (1997). The influence of oxygen tension and pH on the expression of platelet-derived endothelial cell growth factor/thymidine phosphorylase in human breast tumor cells grown in vitro and in vivo. *Cancer Res*. 57(4):570-2.

Grimsley C, Ravichandran KS. (2003). Cues for apoptotic cell engulfment: eat-me, don't eat-me and come-get-me signals. *Trends Cell Biol*. 13(12):648-56.

Hagemann T, Wilson J, Kulbe H, Li NF, Leinster DA, Charles K, Klemm F, Pukrop T, Binder C, Balkwill FR. (2005) Macrophages induce invasiveness of epithelial cancer cells via NF-kappa B and JNK. *J Immunol*. 175(2):1197-205

Hagemann T, Lawrence T, McNeish I, Charles KA, Kulbe H, Thompson RG, Robinson SC, Balkwill FR. (2008). "Re-educating" tumor-associated macrophages by targeting NF-kappaB. *J Exp Med.* 205(6):1261-8

Hans CP, Weisenburger DD, Greiner TC, Gascoyne RD, Delabie J, Ott G, Müller-Hermelink HK, Campo E, Braziel RM, Jaffe ES, Pan Z, Farinha P, Smith LM, Falini B, Banham AH, Rosenwald A, Staudt LM, Connors JM, Armitage JO, Chan WC. (2004). Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. *Blood*. 103(1):275-82.

Hans CP, Weisenburger DD, Greiner TC, Chan WC, Aoun P, Cochran GT, Pan Z, Smith LM, Lynch JC, Bociek RG, Bierman PJ, Vose JM, Armitage JO. (2005). Expression of PKC-beta or cyclin D2 predicts for inferior survival in diffuse large B-cell lymphoma. *Mod Pathol.* 18(10):1377-1384.

Hasselblom S, Sigurdadottir M, Hansson U, Nilsson-Ehle H, Ridell B, Andersson PO. (2007). The number of tumour-infiltrating TIA-1+ cytotoxic T cells but not FOXP3+ regulatory T cells predicts outcome in diffuse large B-cell lymphoma. *Br J Haematol.* 137(4):364-73.

Hedvat CV, Hegde A, Chaganti RS, Chen B, Qin J, Filippa DA, Nimer SD, Teruya-Feldstein J. (2002). Application of tissue microarray technology to the study ofnon-Hodgkin's and Hodgkin's lymphoma. *Hum Pathol*. 33(10):968-74.

Herbert JM, Stekel D, Sanderson S, Heath VL, Bicknell R. (2008). A novel method of differential gene expression analysis using multiple cDNA libraries applied to the identification of tumour endothelial genes. *BMC Genomics*. 9:153.

Hermine O, Haioun C, Lepage E, d'Agay MF, Briere J, Lavignac C, Fillet G, Salles G, Marolleau JP, Diebold J, Reyas F, Gaulard P.(1996). Prognostic significance of bcl-2 protein expression in aggressive non-Hodgkin's lymphoma. Groupe d'Etude des Lymphomes de l'Adulte (GELA). *Blood.* 87(1):265-272.

Hirayama S, Ishii G, Nagai K, Ono S, Kojima M, Yamauchi C, Aokage K, Hishida T, Yoshida J, Suzuki K, Ochiai A. (2012). Prognostic impact of CD204-positive macrophages in lung squamous cell carcinoma: possible contribution of Cd204-positive macrophages to the tumor-promoting microenvironment. *J Thorac Oncol*. 7(12):1790-7.

Hodkinson PS, Elliott T, Wong WS, Rintoul RC, Mackinnon AC, Haslett C, Sethi T. (2006). ECM overrides DNA damage-induced cell cycle arrest and apoptosis in small-cell lung cancer cells through beta1 integrin-dependent activation of PI3-kinase. *CellDeath Differ*. 13(10):1776-88.

Honeychurch J, Alduaij W, Azizyan M, Cheadle EJ, Pelicano H, Ivanov A, Huang P, Cragg MS, Illidge TM. (2012). Antibody-induced nonapoptotic cell death in human lymphoma

and leukemia cells is mediated through a novel reactive oxygen species-dependent pathway. *Blood*. 119(15):3523-33.

Hotchkiss KA, Ashton AW, Klein RS, Lenzi ML, Zhu GH, Schwartz EL. (2003). Mechanisms by which tumor cells and monocytes expressing the angiogenic factor thymidine phosphorylase mediate human endothelial cell migration. *Cancer Res.* 63(2):527-33.

Hsi ED. (2001). The search for meaningful prognostic markers in diffuse large B-cell lymphoma. *Am J Clin Pathol.* 115(4):481-3.

Huang X, Bai X, Cao Y, Wu J, Huang M, Tang D, Tao S, Zhu T, Liu Y, Yang Y, Zhou X, Zhao Y, Wu M, Wei J, Wang D, Xu G, Wang S, Ma D, Zhou J. (2010). Lymphoma endothelium preferentially expresses Tim-3 and facilitates the progression of lymphoma by mediating immune evasion. *J Exp Med.* 207(3):505-20.

Huang da W, Sherman BT, Lempicki RA. (2009a)Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 4(1):44-57.

Huang da W, Sherman BT, Lempicki RA. (2009b). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 37(1):1-13.

Huber W, Carey VJ, Long L, Falcon S, Gentleman R. (2007). Graphs in molecular biology. *BMC Bioinformatics.* 8 Suppl 6:S8.

Huminiecki L, Bicknell R. (2000). In silico cloning of novel endothelial-specific genes. *Genome Res*. 10(11):1796-806.

Hummel M, Bentink S, Berger H, Klapper W, Wessendorf S, Barth TF, Bernd HW, Cogliatti SB, Dierlamm J, Feller AC, Hansmann ML, Haralambieva E, Harder L, Hasenclever D, Kühn M, Lenze D, Lichter P, Martin-Subero JI, Möller P, Müller-Hermelink HK, Ott G, Parwaresch RM, Pott C, Rosenwald A, Rosolowski M, Schwaenen C, Stürzenhofecker B, Szczepanowski M, Trautmann H, Wacker HH, Spang R, Loeffler M, Trümper L, Stein H, Siebert R; Molecular Mechanisms in Malignant Lymphomas Network Project of the Deutsche Krebshilfe. (2006). A biologic definition of Burkitt's lymphoma from transcriptional and genomic profiling. *N Engl J Med.* 354(23):2419-30.

Huynh ML, Fadok VA, Henson PM. (2002). Phosphatidylserine-dependent ingestion of apoptotic cells promotes TGF-beta1 secretion and the resolution of inflammation. *J Clin Invest*. 109(1):41-50.

Iacobelli S, Arno E, D'Orazio A and Coletti G. (1986). Detection of antigens recognised by a novel monoclonal antibody in tissue and serum of patients with breast cancer. *Cancer Res.* 46 (6), 3005-3010.

Iacobelli S, Natoli C, D'Egidio M, Tanburrini E, Antinori A and Ortona L. (1991). Lipoprotein 90K in human immunodeficiency virus-infected patients: a further serological marker of progression. *J Infect Dis.* 164 (4), 819.

Iacovazzi P, Trisolini A, Barletta D, Elba S, Manghisi O and Correale M. (2001). Serum 90K/MAC-2BP glycoprotein in patients with liver cirrhosis and hepatocellular carcinoma: a comparison with alpha-fetoprotein. *Clin Chem Lab Med.* 39 (10), 961-965.

Ikeda R, Che XF, Ushiyama M, Yamaguchi T, Okumura H, Nakajima Y, Takeda Y, Shibayama Y, Furukawa T, Yamamoto M, Haraguchi M, Sumizawa T, Yamada K, Akiyama S. (2006). 2-Deoxy-D-ribose inhibits hypoxia-induced apoptosis by suppressing the phosphorylation of p38 MAPK. *Biochem Biophys Res Commun.* 342(1):280-5

Illidge TM, Cheadle EJ, Honeychurch J. (2014). New opportunities for anti-CD20 monoclonal antibody to give a direct punch to the tumor. *Leuk Lymphoma*. 55(1):3-4.

Inohara H, Akahni S, Koths K, and Raz A. (1996). Interactions between Galectin-3 and Mac-2 binding protein mediate cell-cell adhesion. *Cancer Res.* 56, 4530-4534.

The International Non-Hodgkin's Lymphoma Prognostic Factors Project. (1993). A predictive model for aggressive non-Hodgkin's lymphoma. *N Engl J Med.* 329(14):987-94.

Iqbal J, Sanger WG, Horsman DE, Rosenwald A, Pickering DL, Dave B, Dave S, Xiao L, Cao K, Zhu Q, Sherman S, Hans CP, Weisenburger DD, Greiner TC, Gascoyne RD, Ott G, Müller-Hermelink HK, Delabie J, Braziel RM, Jaffe ES, Campo E, Lynch JC, Connors JM, Vose JM, Armitage JO, Grogan TM, Staudt LM, Chan WC. (2004). BCL2 translocation defines a unique tumor subset within the germinal center B-cell-like diffuse large B-cell lymphoma. *Am J Pathol.* 165(1):159-66.

Iqbal J, Neppalli VT, Wright G, Dave BJ, Horsman DE, Rosenwald A, Lynch J, Hans CP, Weisenburger DD, Greiner TC, Gascoyne RD, Campo E, Ott G, Müller-Hermelink HK, Delabie J, Jaffe ES, Grogan TM, Connors JM, Vose JM, Armitage JO, Staudt LM, Chan WC. (2006). BCL2 expression is a prognostic marker for the activated B-cell-like type of diffuse large B-cell lymphoma. *J Clin Oncol*. 24(6):961-968.

Iqbal J, Greiner TC, Patel K, Dave BJ, Smith L, Ji J, Wright G, Sanger WG, Pickering DL, Jain S, Horsman DE, Shen Y, Fu K, Weisenburger DD, Hans CP, Campo E, Gascoyne RD, Rosenwald A, Jaffe ES, Delabie J, Rimsza L, Ott G, Müller-Hermelink HK, Connors JM, Vose JM, McKeithan T, Staudt LM, Chan WC; Leukemia/Lymphoma Molecular Profiling Project. (2007). Distinctive patterns of BCL6 molecular alterations and their functional consequences in different subgroups of diffuse large B-cell lymphoma. *Leukemia.* 21(11):2332-43.

Ishikawa F, Miyazono K, Hellman U, Drexler H, Wernstedt C, Hagiwara K, Usuki K, Takaku F, Risau W, Heldin CH. (1989. Identification of angiogenic activity and the cloning and expression of platelet-derived endothelial cell growth factor. *Nature*. 338(6216):557-62.

Iris FJ, Bougueleret L, Prieur S, Caterina D, Primas G, Perrot V, Jurka J, Rodriguez-Tome P, Claverie JM, Dausset J, et al. (1993). Dense Alu clustering and a potential new member of the NF kappa B family within a 90 kilobase HLA class III segment. *Nat Genet.* 3(2):137-45.

Iwamoto T, Pusztai L. (2010). Predicting prognosis of breast cancer with gene signatures: are we lost in a sea of data? *Genome Med*. 2(11):81

Jansen CA, Cruijsen CW, de Ruiter T, Nanlohy N, Willems N, Janssens-Korpela PL, Meyaard L. (2007). Regulated expression of the inhibitory receptor LAIR-1 on human peripheral T cells during T cell activation and differentiation. *Eur J Immunol.* 37(4):914-24.

Johnson PW, Watt SM, Betts DR, Davies D, Jordan S, Norton AJ, Lister TA. (1993). Isolated follicular lymphoma cells are resistant to apoptosis and can be grown in vitro in the CD40/stromal cell system. *Blood.* 82(6):1848-57.

Jorissen RN, Lipton L, Gibbs P, Chapman M, Desai J, Jones IT, Yeatman TJ, East P, Tomlinson IP, Verspaget HW, Aaltonen LA, Kruhøffer M, Orntoft TF, Andersen CL, Sieber OM. (2008). DNA copy-number alterations underlie gene expression differences between microsatellite stable and unstable colorectal cancers. *Clin Cancer Res.* 14(24):8061-9.

Kaplan G, Gaudernack G. (1982). In vitro differentiation of human monocytes. Differences in monocyte phenotypes induced by cultivation on glass or on Collagen. *J Exp Med.* 156(4):1101-14.

Kaplan G. (1983). In vitro differentiation of human monocytes. Monocytes cultured on glass are cytotoxic to tumor cells but monocytes cultured on collagen are not. *J Exp Med*. 157(6):2061-72.

Kaplan RN, Riba RD, Zacharoulis S, Bramley AH, Vincent L, Costa C, MacDonald DD, Jin DK, Shido K, Kerns SA, Zhu Z, Hicklin D, Wu Y, Port JL, Altorki N, Port ER, Ruggero D, Shmelkov SV, Jensen KK, Rafii S, Lyden D. (2005). VEGFR1-positive haematopoietic bone marrow progenitors initiate the pre-metastatic niche. *Nature.* 438(7069):820-7.

Kawahara A, Hattori S, Akiba J, Nakashima K, Taira T, Watari K, Hosoi F, Uba M, Basaki Y, Koufuji K, Shirouzu K, Akiyama S, Kuwano M, Kage M, Ono M. (2010). Infiltration of thymidine phosphorylase-positive macrophages is closely associated with tumor angiogenesis and survival in intestinal type gastric cancer. *Oncol Rep.* 24(2):405-15.

Keane C, Gill D, Vari F, Cross D, Griffiths L, Gandhi M. (2013). CD4(+) tumor infiltrating lymphocytes are prognostic and independent of R-IPI in patients with DLBCL receiving R-CHOP chemo-immunotherapy. *Am J Hematol*. 88(4):273-6.

Kelley TW, Pohlman B, Elson P, Hsi ED. (2007). The ratio of FOXP3+ regulatory T cells to granzyme B+ cytotoxic T/NK cells predicts prognosis in classical Hodgkin lymphoma and is independent of bcl-2 and MAL expression. *Am J Clin Pathol*. 128(6):958-65.

Kerr JF, Wyllie AH, Currie AR. (1972). Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. *Br J Cancer*. 26(4):239-57.

Kessenbrock K, Plaks V, Werb Z. (2010). Matrix metalloproteinases: regulators of the tumor microenvironment. *Cell*. 141(1):52-67

Kim S, Takahashi H, Lin WW, Descargues P, Grivennikov S, Kim Y, Luo JL, Karin M. (2009). Carcinoma-produced factors activate myeloid cells through TLR2 to stimulate metastasis. *Nature*. 457(7225):102-6

Kim TD, Li G, Song KS, Kim JM, Kim JS, Kim JS, Yun EJ, Park JI, Park HD, Hwang BD, Lim K, Yoon WH. (2009). Radiation-induced thymidine phosphorylase upregulation in rectal cancer is mediated by tumor-associated macrophages by monocyte chemoattractant protein-1 from cancer cells. *Int J Radiat Oncol Biol Phys*. 73(3):853-60.

Kimura M, Kawahito Y, Obayashi H, Ohta M, Hara H, Adachi T, Tokunaga D, Hojo T, Hamaguchi M, Omoto A, Ishino H, Wada M, Kohno M, Tsubouchi Y, Yoshikawa T. (2007). A critical role for allograft inflammatory factor-1 in the pathogenesis of rheumatoid arthritis. *J Immunol*. 178(5):3316-22.

Kitamura T, Qian BZ, Soong D, Cassetta L, Noy R, Sugano G, Kato Y, Li J, Pollard JW. (2015). CCL2-induced chemokine cascade promotes breast cancer metastasis by enhancing retention of metastasis-associated macrophages. *J Exp Med*. 212(7):1043-59.

Kobara M, Sunagawa N, Abe M, Tanaka N, Toba H, Hayashi H, Keira N, Tatsumi T, Matsubara H, Nakata T. (2008). Apoptotic myocytes generate monocyte chemoattractant protein-1 and mediate macrophage recruitment. J Appl Physiol (1985). 104(3):601-9.

Koh YW, Park CS, Yoon DH, Suh C, Huh J. (2014). CD163 expression was associated with angiogenesis and shortened survival in patients with uniformly treated classical Hodgkin lymphoma. *PLoS One*. 9(1):e87066.

Köhler C. (2007). Allograft inflammatory factor-1/Ionized calcium-binding adapter molecule 1 is specifically expressed by most subpopulations of macrophages and spermatids in testis. *Cell Tissue Res*. 330(2):291-302.

Korkola JE, Houldsworth J, Chadalavada RS, Olshen AB, Dobrzynski D, Reuter VE, Bosl GJ, Chaganti RS. (2006). Down-regulation of stem cell genes, including those in a 200-kb gene cluster at 12p13.31, is associated with in vivo differentiation of human male germ cell tumors. *Cancer Res*. 66(2):820-7.

Koths K, Taylor E, Halenbeck R, Casipit C and Wang A. (1993). Cloning and characterisation of a human Mac-2-binding protein, a new member of the superfamily defined by the macrophage scavenger receptor cysteine-rich domain. *J Biol Chem.* 286 (19), 14245-14249.

Kridel R, Steidl C, Gascoyne RD. (2015). Tumor-associated macrophages in diffuse large B-cell lymphoma. *Haematologica*. 100(2):143-5.

Kurahara H, Takao S, Kuwahata T, Nagai T, Ding Q, Maeda K, Shinchi H, Mataki Y, Maemura K, Matsuyama T, Natsugoe S. (2012). Clinical significance of folate receptor β-expressing tumor-associated macrophages in pancreatic cancer. *Ann Surg Oncol*. 19(7):2264-71.

Langfelder P, Horvath S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 9:559.

Lauber K, Bohn E, Kröber SM, Xiao YJ, Blumenthal SG, Lindemann RK, Marini P, Wiedig C, Zobywalski A, Baksh S, Xu Y, Autenrieth IB, Schulze-Osthoff K, Belka C, Stuhler G, Wesselborg S. (2003). Apoptotic cells induce migration of phagocytes via caspase-3-mediated release of a lipid attraction signal. *Cell.* 113(6):717-30.

Lauber K, Blumenthal SG, Waibel M, Wesselborg S. (2004). Clearance of apoptotic cells: getting rid of the corpses. *Mol Cell.* 14(3):277-87.

Lazar I, Yaacov B, Shiloach T, Eliahoo E, Kadouri L, Lotem M, Perlman R, Zakay-Rones Z, Panet A, Ben-Yehuda D. (2010). The oncolytic activity of Newcastle disease virus NDV HUJ on chemoresistant primary melanoma cells is dependent on the proapoptotic activity of the inhibitor of apoptosis protein Livin. *J Virol.* 84(1):639-46.

Lebbink RJ, de Ruiter T, Kaptijn GJ, Meyaard L. (2005). Identification and characterization of the rat homologue of LAIR-1. *Immunogenetics.* 57(5):344-51.

Lebbink RJ, de Ruiter T, Adelmeijer J, Brenkman AB, van Helvoort JM, Koch M, Farndale RW, Lisman T, Sonnenberg A, Lenting PJ, Meyaard L. (20060. Collagens are functional, high affinity ligands for the inhibitory immune receptor LAIR-1. *J Exp Med.* 203(6):1419-25.

Lebbink RJ, de Ruiter T, Kaptijn GJ, Bihan DG, Jansen CA, Lenting PJ, Meyaard L. (2007). Mouse leukocyte-associated Ig-like receptor-1 (mLAIR-1) functions as an inhibitory collagen-binding receptor on immune cells. *Int Immunol.* 19(8):1011-9.

Lebbink RJ, van den Berg MC, de Ruiter T, Raynal N, van Roon JA, Lenting PJ, Jin B, Meyaard L. (2008). The soluble leukocyte-associated Ig-like receptor (LAIR)-2 antagonizes the collagen/LAIR-1 inhibitory immune interaction. *J Immunol.* 180(3):1662-9.

Lee CG, Hartl D, Lee GR, Koller B, Matsuura H, Da Silva CA, Sohn MH, Cohn L, Homer RJ, Kozhich AA, Humbles A, Kearley J, Coyle A, Chupp G, Reed J, Flavell RA, Elias JA. (2009). Role of breast regression protein 39 (BRP_39)/chitinase 3-like 1in Th2 and IL-13 induced tissue responses and apoptosis. *J Exp Med.* 206(5):1149-66.

Leitinger B, Hohenester E. (2007). Mammalian collagen receptors. *Matrix Biol.* 26(3):146-55.

Lenz G, Wright GW, Emre NC, Kohlhammer H, Dave SS, Davis RE, Carty S, Lam LT, Shaffer AL, Xiao W, Powell J, Rosenwald A, Ott G, Muller-Hermelink HK, Gascoyne RD, Connors JM, Campo E, Jaffe ES, Delabie J, Smeland EB, Rimsza LM, Fisher RI, Weisenburger DD, Chan WC and Staudt LM. (2008a). Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. *Proc Natl Acad Sci*. 105(36):13520 13525.

Lenz G, Davis RE, Ngo VN, Lam L, George TC, Wright GW, Dave SS, Zhao H, Xu W, Rosenwald A, Ott G, Muller-Hermelink HK, Gascoyne RD, Connors JM, Rimsza LM, Campo E, Jaffe ES, Delabie J, Smeland EB, Fisher RI, Chan WC, Staudt LM. (2008b) Oncogenic CARD11 mutations in human diffuse large B cell lymphoma. *Science.* 319(5870):1676-9.

Lenz G, Wright G, Dave SS, Xiao W, Powell J, Zhao H, Xu W, Tan B, Goldschmidt N, Iqbal J, Vose J, Bast M, Fu K, Weisenburger DD, Greiner TC, Armitage JO, Kyle A, May L, Gascoyne RD, Connors JM, Troen G, Holte H, Kvaloy S, Dierickx D, Verhoef G, Delabie J, Smeland EB, Jares P, Martinez A, Lopez-Guillermo A, Montserrat E, Campo E, Braziel RM, Miller TP, Rimsza LM, Cook JR, Pohlman B, Sweetenham J, Tubbs RR, Fisher RI, Hartmann E, Rosenwald A, Ott G, Muller-Hermelink HK, Wrench D, Lister TA, Jaffe ES, Wilson WH, Chan WC, Staudt LM; Lymphoma/Leukemia Molecular Profiling Project. (2008c). Stromal gene signatures in large-B-cell lymphomas. *N Engl J Med.* 359(22):2313-23

Leoncini L, Del Vecchio MT, Megha T, Barbini P, Galieni P, Pileri S, Sabattini E, Gherlinzoni F, Tosi P, Kraft R, et al. (1993) Correlations between apoptotic and proliferative indices in malignant non-Hodgkin's lymphomas. *Am J Pathol.* 142(3):755-63

Leung C, Yu C, Lin MI, Tognon C, Bernatchez P. (2013). Expression of myoferlin in human and murine carcinoma tumors: role in membrane repair, cell proliferation, and tumorigenesis. *Am J Pathol.* 182(5): 1900-9.

Lewis JS, Landers RJ, Underwood JC, Harris AL, Lewis CE. (2000). Expression of vascular endothelial growth factor by macrophages is up-regulated in poorly vascularized areas of breast carcinomas . *J Pathol*. 192(2):150-8

Liekens S, Bronckaers A, Pérez-Pérez MJ, Balzarini J. (2007) Targeting platelet-derived endothelial cell growth factor/thymidine phosphorylase for cancer therapy. *Biochem Pharmacol.* 74(11):1555-67.

Liguori M, Solinas G, Germano G, Mantovani A, Allavena P. (2011). Tumor-associated macrophages as incessant builders and destroyers of the cancer stroma. *Cancers.* 3(4):3740-61

Lin EY, Li JF, Gnatovskiy L, Deng Y, Zhu L, Grzesik DA, Qian H, Xue XN, Pollard JW. (2006). Macrophages regulate the angiogenic switch in a mouse model of breast cancer. *Cancer Res.* 66(23):11238-46.

Linderoth J, Edén P, Ehinger M, Valcich J, Jerkeman M, Bendahl PO, Berglund M, Enblad G, Erlanson M, Roos G, Cavallin-Ståhl E. (2008). Genes associated with the tumour microenvironment are differentially expressed in cured versus primary chemotherapy-refractory diffuse large B-cell lymphoma. *Br J Haematol*. 141(4):423-32.

Lipford E, Wright JJ, Urba W, Whang-Peng J, Kirsch IR, Raffeld M, Cossman J, Longo DL, Bakhshi A, Korsmeyer SJ. (1987). Refinement of lymphoma cytogenetics by the chromosome 18q21 major breakpoint region. *Blood*. 70(6):1816-23.

Lissbrant IF, Stattin P, Wikstrom P, Damber JE, Egevad L, Bergh A. (2000). Tumor associated macrophages in human prostate cancer: relation to clinicopathological variables and survival. *Int J Oncol*. 17(3):445-51

Liu H, Kennard S, Lilly B. (2009). NOTCH3 expression is induced in mural cells through an autoregulatory loop that requires endothelial-expressed JAGGED1. *Circ Res.* 104(4):466-75.

Liu S, Tan WY, Chen QR, Chen XP, Fu K, Zhao YY, Chen ZW. (2008). Daintain/AIF-1 promotes breast cancer proliferation via activation of the NF-kappaB/cyclin D1 pathway and facilitates tumor growth. *Cancer Sci.* 99(5):952-7.

Lodermeyer V, Suhr K, Schrott N, Kolbe C, Stürzel CM, Krnavek D, Münch J, Dietz C, Waldmann T, Kirchhoff F, Goffinet C. (2013). 90K, an interferon-stimulated gene product, reduces the infectivity of HIV-1. *Retrovirology.* 10:111.

Maasho K, Masilamani M, Valas R, Basu S, Coligan JE, Borrego F. (2005). The inhibitory leukocyte-associated Ig-like receptor-1 (LAIR-1) is expressed at high levels by human naive T cells and inhibits TCR mediated activation. *Mol Immunol*. 42(12):1521-30.

Mainou-Fowler T, Angus B, Miller S, Proctor SJ, Taylor PR, Wood KM. (2006). Micro-vessel density and the expression of vascular endothelial growth factor (VEGF) and platelet-derived endothelial cell growth factor (PdEGF) in classical Hodgkin lymphoma (HL). *Leuk Lymphoma.* 47(2):223-30.

Maloney DG, Smith B, Rose A. (2002). Rituximab: mechanism of action and resistance. *Semin Oncol.* 29(1 Suppl 2):2-9.

Mandelbaum J, Bhagat G, Tang H, Mo T, Brahmachary M, Shen Q, Chadburn A, Rajewsky K, Tarakhovsky A, Pasqualucci L, Dalla-Favera R. (2010). BLIMP1 is a tumor suppressor gene frequently disrupted in activated B cell-like diffuse large B cell lymphoma. *Cancer Cell*. 18(6):568-79.

Mantovani A, Sica A, Allavena P, Garlanda C, Locati M. (2009). Tumor-associated macrophages and the related myeloid-derived suppressor cells as a paradigm of the diversity of macrophage activation. *Hum Immunol.* 70(5):325-30.

Mantovani A, Sica A. (2010). Macrophages, innate immunity and cancer: balance, tolerance, and diversity. *Curr Opin Immunol*. 22(2):231-7.

Mantovani A, Biswas SK, Galdiero MR, Sica A, Locati M. (2013). Macrophage plasticity and polarization in tissue repair and remodelling. *J Pathol.* 229(2):176-85

Marchetti A, Tinari N, Buttitta F, Chella A, Angeletti C, Sacco R, Mucilli F, Ullrich A and Iacobelli S. (2002). Expression of 90K (Mac-2 BP) correlates with distant metastasis and predicts survival in stage I non-small cell lung cancer patients. *Cancer Res.* 62 (9), 2535-2539.

Marinaccio C, Ingravallo G, Gaudio F, Perrone T, Nico B, Maoirano E, Specchia G, Ribatti D. (2014). Microvascular density, CD68 and tryptase expression in human diffuse large B cell lymphoma. *Leuk Res*. 38(11):1374-7.

Marinaccio C, Ingravallo G, Gaudio F, Perrone T, Ruggieri S, Opinto G, Nico B, Maiorano E, Specchia G, Ribatti D. (2015). T cells, mast cells and microvascular density in diffuse large B cell lymphoma. *Clin Exp Med.* 2015 May 10. [Epub ahead of print]

Martelli M, Ferreri AJ, Agostinelli C, Di Rocco A, Pfreundschuh M, Pileri SA. (2013). Diffuse large B-cell lymphoma. *Crit Rev Oncol Hematol.* 87(2):146-71.

Martinez FO, Helming L, Milde R, Varin A, Melgert BN, Draijer C, Thomas B, Fabbri M, Crawshaw A, Ho LP, Ten Hacken NH, Cobos Jiménez V, Kootstra NA, Hamann J, Greaves DR, Locati M, Mantovani A, Gordon S. (2013). Genetic programs expressed in resting and IL-4 alternatively activated mouse and human macrophages: similarities and differences. *Blood.* 121(9):e57-69.

McDaniel DO, Rigney DA, McDaniel KY, Windham WJ, Redmond P, Williams B, Zhou X, Hawxby A, Butt F. (2013). Early expression profile of inflammatory markers and kidney allograft status. *Transplant Proc.* 45(4):1520-3.

Merlo A, Tenca C, Fais F, Battini L, Ciccone E, Grossi CE, Saverino D. (2005). Inhibitory receptors CD85j, LAIR-1, and CD152 down-regulate immunoglobulin and cytokine production by human B lymphocytes. *Clin Diagn Lab Immunol*. 12(6):705-12.

Meyaard L, Adema GJ, Chang C, Woollatt E, Sutherland GR, Lanier LL, Phillips JH. (1997). LAIR-1, a novel inhibitory receptor expressed on human mononuclear leukocytes. *Immunity*. 7(2):283-90.

Meyaard L. (2008). The inhibitory collagen receptor LAIR-1 (CD305). *J Leukoc Biol.* 83(4):799-803.

Meyer PN, Fu K, Greiner T, Smith L, Delabie J, Gascoyne R, Ott G, Rosenwald A, Braziel R, Campo E, Vose J, Lenz G, Staudt L, Chan W, Weisenburger DD. (2011). The stromal cell marker SPARC predicts for survival in patients with diffuse large B-cell lymphoma treated with rituximab. *Am J Clin Pathol.* 135(1):54-61

Miller TP, Lippman SM, Spier CM, Slymen DJ, Grogan TM. (1988. HLA-DR (Ia) immune phenotype predicts outcome for patients with diffuse large cell lymphoma. *J Clin Invest*. 82(1):370-372.

Minard-Colin V, Xiu Y, Poe JC, Horikawa M, Magro CM, Hamaguchi Y, Haas KM, Tedder TF. (2008). Lymphoma depletion during CD20 immunotherapy in mice is mediated by macrophage FcgammaRI, FcgammaRIII, and FcgammaRIV. *Blood.* 112(4):1205-1213.

Mishima T, Iwabuchi K, Fujii S, Tanaka SY, Ogura H, Watano-Miyata K, Ishimori N, Andoh Y, Nakai Y, Iwabuchi C, Ato M, Kitabatake A, Tsutsui H, Onoé K. (2008). Allograft inflammatory factor-1 augments macrophage phagocytotic activity and accelerates the progression of atherosclerosis in ApoE-/- mice. *Int J Mol Med*. 21(2):181-7.

Mitselou A, Ioachim E, Skoufi U, Tsironis C, Tsimogiannis KE, Skoufi C, Vougiouklakis T, Briasoulis E. (2012). Predictive role of thymidine phosphorylase expression in patients with colorectal cancer and its association with angiogenesis-related proteins and extracellular matrix components. *In Vivo.* 26(6):1057-67.

Miyadera K, Sumizawa T, Haraguchi M, Yoshida H, Konstanty W, Yamada Y, Akiyama S. (1995). Role of thymidine phosphorylase activity in the angiogenic effect of platelet derived endothelial cell growth factor/thymidine phosphorylase. *Cancer Res.* 55(8):1687-90.

Miyazono K, Okabe T, Urabe A, Takaku F, Heldin CH. (1987). Purification and properties of an endothelial cell growth factor from human platelets. *J Biol Chem.* 262(9):4098-103.

Moghaddam A, Bicknell R. (1992). Expression of platelet-derived endothelial cell growth factor in Escherichia coli and confirmation of its thymidine phosphorylase activity. *Biochemistry*. 31(48):12141-6.

Moghaddam A, Zhang HT, Fan TP, Hu DE, Lees VC, Turley H, Fox SB, Gatter KC, Harris AL, Bicknell R. (1995), Thymidine phosphorylase is angiogenic and promotes tumor growth. *Proc Natl Acad Sci U S A.* 92(4):998-1002.

Monti S, Savage KJ, Kutok JL, Feuerhake F, Kurtin P, Mihm M, Wu B, Pasqualucci L, Neuberg D, Aguiar RC, Dal Cin P, Ladd C, Pinkus GS, Salles G, Harris NL, Dalla-Favera R, Habermann TM, Aster JC, Golub TR, Shipp MA. (2005). Molecular profiling of diffuse large B-cell lymphoma identifies robust subtypes including one characterized by host inflammatory response. *Blood.* 105(5):1851-61.

Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J,Puigserver P, Carlsson E, Ridderstråle M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, Groop LC. (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet.* 34(3):267-73.

Morita T, Matsuzaki A, Suzuki K, Tokue A. (2001). Role of thymidine phosphorylase in biomodulation of fluoropyrimidines. *Curr Pharm Biotechnol.* 2(3):257-67.

Mosca E, Bertoli G, Piscitelli E, Vilardo L, Reinbold RA, Zucchi I, MilanesiL. (2009). Identification of functionally related genes using data mining and data integration: a breast cancer case study. *BMC Bioinformatics*. 10 Suppl 12:S8.

Müller WE, Krasko A, Skorokhod A, Bünz C, Grebenjuk VA, Steffen R, Batel R, Schröder HC. (2002). Histocompatibility reaction in tissue and cells of the marine sponge Suberites domuncula in vitro and in vivo: central role of the allograft inflammatory factor 1. *Immunogenetics*. 54(1):48-58.

Muris JJ, Meijer CJ, Vos W, van Krieken JH, Jiwa NM, Ossenkoppele GJ, Oudejans JJ. (2006). Immunohistochemical profiling based on Bcl-2, CD10 and MUM1 expression improves risk stratification in patients with primary nodal diffuse large B cell lymphoma. *J Pathol*. 208(5):714-723.

Muris JJ, Ylstra B, Cillessen SA, Ossenkoppele GJ, Kluin-Nelemans JC, Eijk PP, Nota B, Tijssen M, de Boer WP, van de Wiel M, van den Ijssel PR, Jansen P, de Bruin PC, van Krieken JH, Meijer GA, Meijer CJ, Oudejans JJ. (2007). Profiling ofapoptosis genes allows for clinical stratification of primary nodal diffuse large B-cell lymphomas. *Br J Haematol.* 136(1):38-47.

Murray PJ, Allen JE, Biswas SK, Fisher EA, Gilroy DW, Goerdt S, Gordon S, Hamilton JA, Ivashkiv LB, Lawrence T, Locati M, Mantovani A, Martinez FO, Mege JL, Mosser DM, Natoli G, Saeij JP, Schultze JL, Shirey KA, Sica A, Suttles J, Udalova I, van Ginderachter JA, Vogel SN, Wynn TA. (2014). Macrophage activation and polarization: nomenclature and experimental guidelines. *Immunity.* 41(1):14-20.

Murri AM, Hilmy M, Bell J, Wilson C, McNicol AM, Lannigan A, Doughty JC, McMillan DC. (2008). The relationship between the systemic inflammatory response, tumour proliferative activity, T-lymphocytic and macrophage infiltration, microvessel density and survival in patients with primary operable breast cancer. *Br J Cancer*. 99(7):1013-9.

Nagakawa Y, Nomoto S, Kato Y, Montgomery RA, Williams GM, Klein AS, Sun Z. (2004). Over-expression of AIF-1 in liver allografts and peripheral blood correlates with acute rejection after transplantation in rats. *Am J Transplant.* 4(12):1949-57.

Nakshatri H, Badve S. (2009). FOXA1 in breast cancer. *Expert Rev Mol Med.* 11:e8.

Natoli C, Dianzani F, Mazzotta F, Balocchini E, Pierotti P, Antonelli G and Iacobelli S. (1993). 90K protein: a new predictor of disease progression in human immunodeficiency virus infection. *J Acquir Immune Defic Syndr*. 6 (4), 370-375.

Ngo VN, Young RM, Schmitz R, Jhavar S, Xiao W, Lim KH, Kohlhammer H, Xu W, Yang Y, Zhao H, Shaffer AL, Romesser P, Wright G, Powell J, Rosenwald A, Muller-Hermelink HK, Ott G, Gascoyne RD, Connors JM, Rimsza LM, Campo E, Jaffe ES, Delabie J, Smeland EB, Fisher RI, Braziel RM, Tubbs RR, Cook JR, Weisenburger DD, Chan WC, Staudt LM. (2011). Oncogenically active MYD88 mutations in human lymphoma. *Nature*. 470(7332):115-9

Novak ML, Koh TJ. (2013). Phenotypic transitions of macrophages orchestrate tissue repair. *Am J Pathol*. 183(5):1352-63.

O'Brien TS, Fox SB, Dickinson AJ, Turley H, Westwood M, Moghaddam A, Gatter KC, Bicknell R, Harris AL. (1996). Expression of the angiogenic factor thymidine phosphorylase/platelet-derived endothelial cell growth factor in primary bladder cancers. *Cancer Res.* 56(20):4799-804

O'Byrne KJ, Koukourakis MI, Giatromanolaki A, Cox G, Turley H, Steward WP, Gatter K, Harris AL. (2000). Vascular endothelial growth factor, platelet-derived endothelial cell growth factor and angiogenesis in non-small-cell lung cancer. *Br J Cancer*. 82(8):1427-32.

Offit K, Lo Coco F, Louie DC, Parsa NZ, Leung D, Portlock C, Ye BH, Lista F, Filippa DA, Rosenbaum A, et al. (1994). Rearrangement of the bcl-6 gene as a prognostic marker in diffuse large-cell lymphoma. *N Engl J Med.* 331(2):74-80.

Ogden CA, Pound JD, Batth BK, Owens S, Johannessen I, Wood K, Gregory CD. (2005). Enhanced apoptotic cell clearance capacity and B cell survival factor production by IL-10 activated macrophages: implications for Burkitt's lymphoma. *J Immunol.* 174(5):3015-23.

Oh JK, Weiderpass E. (2014). Infection and cancer: global distribution and burden of diseases. *Ann Glob Health*. 80(5):384-92.

Ohno H, Fukuhara S. (1997). Significance of rearrangement of the BCL6 gene in B-cell lymphoid neoplasms. *Leuk Lymphoma.* 27(1-2):53-63.

Ohshima K, Kawasaki C, Muta H, Muta K, Deyev V, Haraoka S, Suzumiya J, Podack ER, Kikuchi M. (2001). CD10 and Bcl10 expression in diffuse large B-cell lymphoma: CD10 is a marker of improved prognosis. *Histopathology*. 39(2):156-62.

Ojalvo LS, Whittaker CA, Condeelis JS, Pollard JW. (2010). Gene expression analysis of macrophages that facilitate tumor invasion supports a role for Wnt-signaling in mediating their activity in primary mammary tumors. *J Immunol.* 184(2):702-12.

Okada H, Tsuzuki T, Shindoh H, Nishigaki A, Yasuda K, Kanzaki H. (2014). Regulation of decidualization and angiogenesis in the human endometrium: mini review. *J Obstet Gynaecol Res.* 40(5):1180-7.

Orre M, Rogers PA. (1999). Macrophages and microvessel density in tumors of the ovary. *Gynecol Oncol*. 73(1):47-50.

Ouyang W, Xue J, Liu J, Jia W, Li Z, Xie X, Liu X, Jian J, Li Q, Zhu Y, Yang A, Jin B. (2004). Establishment of an ELISA system for determining soluble LAIR-1 levels in sera of patients with HFRS and kidney transplant. *J Immunol Methods*. 292(1-2):109-17.

Palanca-Wessels MC, Czuczman M, Salles G, Assouline S, Sehn LH, Flinn I, Patel MR, Sangha R, Hagenbeek A, Advani R, Tilly H, Casasnovas O, Press OW, Yalamanchili S, Kahn R, Dere RC, Lu D, Jones S, Jones C, Chu YW, Morschhauser F. (2015). Safety and activity of the anti-CD79B antibody-drug conjugate polatuzumab vedotin in relapsed or refractory B-cell non-Hodgkin lymphoma and chronic lymphocytic leukaemia: a phase 1 study. *Lancet Oncol*. S1470-2045(15)70128-2.

Pangault C, Amé-Thomas P, Ruminy P, Rossille D, Caron G, Baia M, De Vos J, Roussel M, Monvoisin C, Lamy T, Tilly H, Gaulard P, Tarte K, Fest T. (2010). Follicular lymphoma cell niche: identification of a preeminent IL-4-dependent T(FH)-B cell axis. *Leukemia*. 24(12):2080-9

Pasqualucci L, Neumeister P, Goossens T, Nanjangud G, Chaganti RS, Küppers R, Dalla-Favera R. (2001). Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature.* 412(6844):341-6.

Pasqualucci L, Bereshchenko O, Niu H, Klein U, Basso K, Guglielmino R, Cattoretti G, Dalla-Favera R. (2003). Molecular pathogenesis of non-Hodgkin's lymphoma: the role of Bcl-6. *Leuk Lymphoma*. 44 Suppl 3:S5-12.

Pasqualucci L, Compagno M, Houldsworth J, Monti S, Grunn A, Nandula SV, Aster JC, Murty VV, Shipp MA, Dalla-Favera R. (2006). Inactivation of the PRDM1/BLIMP1 gene in diffuse large B cell lymphoma. *J Exp Med.* 203(2):311-7.

Pasqualucci L, Dominguez-Sola D, Chiarenza A, Fabbri G, Grunn A, Trifonov V, Kasper LH, Lerach S, Tang H, Ma J, Rossi D, Chadburn A, Murty VV, Mullighan CG, Gaidano G,

Rabadan R, Brindle PK, Dalla-Favera R. (2011a). Inactivating mutations of acetyltransferase genes in B-cell lymphoma. *Nature*. 471(7337):189-195.

Ruminy P, Etancelin P, Couronné L, Parmentier F, Rainville V, Mareschal S, Bohers E, Burgot C, Cornic M, Bertrand P, Lenormand B, Picquenot JM, Jardin F, Tilly H, Bastard C. (2011). The isotype of the BCR as a surrogate for the GCB and ABC molecular subtypes in diffuse large B-cell lymphoma. *Leukemia*. 25(4):681-8.

Pasqualucci L, Dominguez-Sola D, Chiarenza A, Fabbri G, Grunn A, Trifonov V,Kasper LH, Lerach S, Tang H, Ma J, Rossi D, Chadburn A, Murty VV, Mullighan CG, Gaidano G, Rabadan R, Brindle PK, Dalla-Favera R. (2011b). Inactivating mutations of acetyltransferase genes in B-cell lymphoma. *Nature*. 471(7337):189-95

Pavlopoulos GA, Wegener AL, Schneider R. (2008). A survey of visualization tools for biological network analysis. *BioData Min*. 1:12.

Pawitan Y, Bjöhle J, Amler L, Borg AL, Egyhazi S, Hall P, Han X, Holmberg L, Huang F, Klaar S, Liu ET, Miller L, Nordgren H, Ploner A, Sandelin K, Shaw PM, Smeds J, Skoog L, Wedrén S, Bergh J. (2005). Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. *Breast Cancer Res.* 7(6):R953-64.

Phng LK, Gerhardt H. (2009). Angiogenesis: a team effort coordinated by notch. *Dev Cell*. 16(2):196-208.

Piccolo E, Tinari N, Semeraro D, Traini S, Fichera I, Cumashi A, La Sorda R, Spinella F, Bagnato A, Lattanzio R, D'Egidio M, Di Risio A, Stampolidis P, Piantelli M, Natoli C, Ullrich A and Iacobelli S. (2012). GALS3BP, lectin galactosidase-binding soluble 3 binding protein, induces vascular endothelial growth factor in human breast cancer cells and promotes angiogenesis. *J. Mol. Med.* 91, 83-94.

Poggi A, Tomasello E, Ferrero E, Zocchi MR, Moretta L. (1998). p40/LAIR-1 regulates the differentiation of peripheral blood precursors to dendritic cells induced by granulocyte-monocyte colony-stimulating factor. *Eur J Immunol.* 28(7):2086-91.

Poggi A, Pellegatta F, Leone BE, Moretta L, Zocchi MR. (2000). Engagement of the leukocyte-associated Ig-like receptor-1 induces programmed cell death and prevents NF kappaB nuclear translocation in human myeloid leukemias. *Eur J Immunol.* 30(10):2751-8.

Poon IK, Lucas CD, Rossi AG, Ravichandran KS. (2014). Apoptotic cell clearance: basic biology and therapeutic potential. *Nat Rev Immunol*. 14(3):166-80.

Qian BZ, Pollard JW. (2010). Macrophage diversity enhances tumor progression and metastasis. *Cell.* 141(1):39-51

Qian BZ, Li J, Zhang H, Kitamura T, Zhang J, Campion LR, Kaiser EA, Snyder LA, Pollard JW. (2011). CCL2 recruits inflammatory monocytes to facilitate breast-tumour metastasis. *Nature*. 475(7355):222-5.

Raes G, Van den Bergh R, De Baetselier P, Ghassabeh GH, Scotton C, Locati M, Mantovani A, Sozzani S. (2005). Arginase-1 and Ym1 are markers for murine, but not human, alternatively activated myeloid cells. *J Immunol*. 174(11):6561

Rassidakis GZ, Jones D, Thomaides A, Sen F, Lai R, Cabanillas F, McDonnell TJ, Medeiros LJ. (2002). Apoptotic rate in peripheral T-cell lymphomas. A study using a tissue microarray with validation on full tissue sections. Am J Clin Pathol. 118(3):328-34.

Reiter I, Krammer B, Schwamberger G. (1999). Cutting edge: differential effect of apoptotic versus necrotic tumor cells on macrophage antitumor activities. *J Immunol*. 163(4):1730-2.

Relf M, LeJeune S, Scott PA, Fox S, Smith K, Leek R, Moghaddam A, Whitehouse R, Bicknell R, Harris AL. (1997). Expression of the angiogenic factors vascular endothelial cell growth factor, acidic and basic fibroblast growth factor, tumor growth factor beta-1, platelet-derived endothelial cell growth factor, placenta growth factor, and pleiotrophin in human primary breast cancer and its relation to angiogenesis. *Cancer Res.* 57(5):963-9.

Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan AM. (2004). Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. *Proc Natl Acad Sci U S A*. 101(25):9309-14.

Riker AI, Enkemann SA, Fodstad O, Liu S, Ren S, Morris C, Xi Y, Howell P, Metge B, Samant RS, Shevde LA, Li W, Eschrich S, Daud A, Ju J, Matta J. (2008). The gene expression profiles of primary and metastatic melanoma yields a transition point of tumor progression and metastasis. *BMC Med Genomics*. 1:13.

Riemersma SA, Jordanova ES, Schop RF, Philippo K, Looijenga LH, Schuuring E, Kluin PM. (2000). Extensive genetic alterations of the HLA region, including homozygous deletions of HLA class II genes in B-cell lymphomas arising in immune-privileged sites. *Blood*. 96(10):3569-77.

Rimsza LM, Roberts RA, Miller TP, Unger JM, LeBlanc M, Braziel RM, Weisenberger DD, Chan WC, Muller-Hermelink HK, Jaffe ES, Gascoyne RD, Campo E, Fuchs DA, Spier CM, Fisher RI, Delabie J, Rosenwald A, Staudt LM, Grogan TM. (2004). Loss of MHC class II gene and protein expression in diffuse large B-cell lymphoma is related to decreased tumor immunosurveillance and poor patient survival regardless of other prognostic factors: a follow-up study from the Leukemia and Lymphoma Molecular Profiling Project. *Blood*. 103(11):4251-8.

Rius J, Guma M, Schachtrup C, Akassoglou K, Zinkernagel AS, Nizet V, Johnson RS, Haddad GG, Karin M. (2008). NF-kappaB links innate immunity to the hypoxic response through transcriptional regulation of HIF-1alpha. *Nature*. 453(7196):807-11.

Romero S, Szafranska J, Cabrera E, Gonzalez A, Peiró A, Llauger J, Ortega L, Bague S, Canet B, Espinosa I, Prat J. (2012). Role of tumor-associated macrophages and angiogenesis in desmoid-type fibromatosis. *Virchows Arch*. 461(2):117-22.

Roschewski M, Staudt LM, Wilson WH. (2014). Diffuse large B-cell lymphoma-treatment approaches in the molecular era. *Nat Rev Clin Oncol*. 11(1):12-23.

Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, Gascoyne RD, Muller-Hermelink HK, Smeland EB, Giltnane JM, Hurt EM, Zhao H, Averett L, Yang L, Wilson WH, Jaffe ES, Simon R, Klausner RD, Powell J, Duffey PL, Longo DL, Greiner TC, Weisenburger DD, Sanger WG, Dave BJ, Lynch JC, Vose J, Armitage JO, Montserrat E, López-Guillermo A, Grogan TM, Miller TP, LeBlanc M, Ott G, Kvaloy S, Delabie J, Holte H, Krajci P, Stokke T, Staudt LM; Lymphoma/Leukemia Molecular Profiling Project. (2002). The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med*. 346(25):1937-47.

Rowe M, Rooney CM, Edwards CF, Lenoir GM and Rickinson AB. (1986). Epstein-Barr virus status and tumour cell phenotype in sporadic Burkitt's lymphoma. *Int J Cancer*. 37(3):367-373.

Sasaki T, Brakebush C, Engel J and Timpl R. (1998). Mac-2 binding protein is a cell-adhesive protein of the extracellular matrix which self-assembles into ring-like structures and binds β1 integrins, collagens and fibronectin. *The EMBO Journal*. 17 (6), 1606-1613.

Sathish JG, Johnson KG, Fuller KJ, LeRoy FG, Meyaard L, Sims MJ, Matthews RJ. (2001). Constitutive association of SHP-1 with leukocyte-associated Ig-like receptor-1 in human T cells. *J Immunol*. 166(3):1763-70

Savill J, Dransfield I, Gregory C, Haslett C. (2002). A blast from the past: clearance of apoptotic cells regulates immune responses. *Nat Rev Immunol*. 2(12):965-75.

Saverino D, Fabbi M, Merlo A, Ravera G, Grossi CE, Ciccone E. (2002). Surface density expression of the leukocyte-associated Ig-like receptor-1 is directly related to inhibition of human T-cell functions. *Hum Immunol*. 63(7):534-46.

Sawada N, Ishikawa T, Fukase Y, Nishida M, Yoshikubo T, Ishitsuka H. (1998). Induction of thymidine phosphorylase activity and enhancement of capecitabine efficacy by taxol/taxotere in human cancer xenografts. *Clin Cancer Res*. 4(4):1013-9.

Sawada N, Ishikawa T, Sekiguchi F, Tanaka Y, Ishitsuka H. (1999). X-ray irradiation induces thymidine phosphorylase and enhances the efficacy of capecitabine (Xeloda) in human cancer xenografts. *Clin Cancer Res*. 5(10):2948-53.

Scambia G, Pancini P, Baiocchi G, Perrone L, Iacobelli S and Mancuso S. (1988). Measurement of a monoclonal antibody-defined antigen (90K) in the sera of patients with ovarian cancer. *Anticancer Res.* 8 (4), 761-764.

Schmidlin H, Diehl SA, Nagasawa M, Scheeren FA, Schotte R, Uittenbogaart CH, Spits H, Blom B. (2008). Spi-B inhibits human plasma cell differentiation by repressing BLIMP1 and XBP-1 expression. *Blood.* 112(5):1804-12.

Schwaller J, Schneider P, Mhawech-Fauceglia P, McKee T, Myit S, Matthes T, Tschopp J, Donze O, Le Gal FA, Huard B. (2007). Neutrophil-derived APRIL concentrated in tumor lesions by proteoglycans correlates with human B-cell lymphoma aggressiveness. *Blood*. 109(1):331-8.

Schwartz EL, Wan E, Wang FS, Baptiste N. (1998). Regulation of expression of thymidine phosphorylase/platelet-derived endothelial cell growth factor in human colon carcinoma cells. *Cancer Res*. 58(7):1551-7.

Scott DW, Gascoyne RD. (2014). The tumour microenvironment in B cell lymphomas. *NatRev Cancer.* 14(8):517-34.

Segal E, Friedman N, Koller D, Regev A. (2004). A module map showing conditional activity of expression modules in cancer. *Nat Genet.* 36(10):1090-8.

Sehn LH, Gascoyne RD. (2015). Diffuse large B-cell lymphoma: optimizing outcome in the context of clinical and biologic heterogeneity. *Blood.* 125(1):22-32.

Shaffer AL, Yu X, He Y, Boldrick J, Chan EP, Staudt LM. (2000). BCL-6 represses genes that function in lymphocyte differentiation, inflammation, and cell cycle control. *Immunity.* 13(2):199-212.

Shaffer AL, Wright G, Yang L, Powell J, Ngo V, Lamy L, Lam LT, Davis RE, Staudt LM. (2006). A library of gene expression signatures to illuminate normal and pathological lymphoid biology. *Immunol Rev.* 210:67-85.

Shaffer AL, Emre NC, Romesser PB, Staudt LM. (2009). IRF4: Immunity. Malignancy! Therapy? *Clin Cancer Res.* 15(9):2954-61.

Shain KH, Dalton WS, Tao J. (2015). The tumor microenvironment shapes hallmarks of mature B-cell malignancies. *Oncogene*. 2015 Feb 2. doi: 10.1038/onc.2014.403. [Epub ahead of print]

Shao R. (2013). YKL-40 acts as an angiogenic factor to promote tumor angiogenesis. *Front Physiol.* 4:122

Shi Z, Derow CK, Zhang B. (2010). Co-expression module analysis reveals biological processes, genomic gain, and regulatory mechanisms associated with breast cancer progression. *BMC Syst Biol*. 4:74.

Shigeoka M, Urakawa N, Nakamura T, Nishio M, Watajima T, Kuroda D, Komori T, Kakeji Y, Semba S, Yokozaki H. (2013). Tumor associated macrophage expressing CD204 is associated with tumor aggressiveness of esophageal squamous cell carcinoma. *Cancer Sci*. 104(8):1112-9.

Sierra JR, Corso S, Caione L, Cepero V, Conrotto P, Cignetti A, Piacibello W, Kumanogoh A, Kikutani H, Comoglio PM, Tamagnone L, Giordano S. (2008). Tumor angiogenesis and progression are enhanced by Sema4D produced by tumor-associated macrophages. *J Exp Med*. 205(7):1673-85.

Sickert D, Aust DE, Langer S, Haupt I, Baretton GB, Dieter P. (2005). Characterization of macrophage subpopulations in colon cancer using tissue microarrays. *Histopathology*. 46(5):515-21.

Sidhu A, Wersinger C, Moussa CE, Vernier P. (2004). The role of alpha-synuclein in both neuroprotection and neurodegeneration. *Ann N Y Acad Sci*. 1035:250-70.

Sivridis E, Giatromanolaki A, Gatter KC, Harris AL, Koukourakis MI; Tumor and Angiogenesis Research Group. (2002). Association of hypoxia-inducible factors 1alpha and 2alpha with activated angiogenic pathways and prognosis in patients with endometrial carcinoma. *Cancer*. 95(5):1055-63.

Sommerville LJ, Kelemen SE, Ellison SP, England RN, Autieri MV. (2012). Increased atherosclerosis and vascular smooth muscle cell activation in AIF-1 transgenic mice fed a high-fat diet. *Atherosclerosis*. 220(1):45-52.

Stampolidis P, Ullrich A and Iacobelli S. LGALS3BP, lectin galactoside-binding soluble 3 binding protein, promotes oncogenic cellular events impeded by antibody intervention. (2015). *Oncogene*. 34, 39–52.

Steidl C, Lee T, Shah SP, Farinha P, Han G, Nayar T, Delaney A, Jones SJ, Iqbal J, Weisenburger DD, Bast MA, Rosenwald A, Muller-Hermelink HK, Rimsza LM, Campo E, Delabie J, Braziel RM, Cook JR, Tubbs RR, Jaffe ES, Lenz G, Connors JM, Staudt LM, Chan WC, Gascoyne RD. (2010). Tumor-associated macrophages and survival in classic Hodgkin's lymphoma. *N Engl J Med*. 362(10):875-85.

Steidl C, Connors JM, Gascoyne RD. (2011). Molecular pathogenesis of Hodgkin's lymphoma: increasing evidence of the importance of the microenvironment. *J Clin Oncol*. 29(14):1812-26.

Steingrímsson E, Copeland NG, Jenkins NA. (2004). Melanocytes and the microphthalmia transcription factor network. *Annu Rev Genet*. 38:365-411.

Strickland LA, Jubb AM, Hongo JA, Zhong F, Burwick J, Fu L, Frantz GD, Koeppen H. (2005) Plasmalemmal vesicle-associated protein (PLVAP) is expressed by tumour endothelium and is upregulated by vascular endothelial growth factor-A (VEGF). *J Pathol.* 206(4):466-75.

Strizzi L, Muraro R, Vianale G, Natoli C, Talone L, Catalano A, Mutti L, Tassi G and Procopio A. (2002). Expression of glycoprotein 90K in human malignant pleural mesothelioma: correlation with patient survival. *J Pathol.* 197 (2), 218-223.

Stuart JM, Segal E, Koller D, Kim SK. (2003). A gene-coexpression network for global discovery of conserved genetic modules. *Science.* 302(5643):249-55.

Swerdlow SH, Campo E, Harris NL, Jaffe E, Pileri SA, Stein H, Thiele J and Vardiman J (Eds). (2008) . WHO Classification of Tumours of Haematopoitic and Lymphoid Tissues, 4th edition. Lyon, IARC

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 102(43):15545-50.

Suyanı E, Sucak GT, Akyürek N, Sahin S, Baysal NA, Yağcı M, Haznedar R. (2013). Tumor-associated macrophages as a prognostic parameter in multiple myeloma. *Ann Hematol*. 92(5):669-77.

Takebayashi Y, Miyadera K, Akiyama S, Hokita S, Yamada K, Akiba S, Yamada Y, Sumizawa T, Aikou T. (1996a). Expression of thymidine phosphorylase in human gastric carcinoma. *Jpn J Cancer Res.* 87(3):288-95.

Takebayashi Y, Yamada K, Miyadera K, Sumizawa T, Furukawa T, Kinoshita F, Aoki D, Okumura H, Yamada Y, Akiyama S, Aikou T. (1996b). The activity and expression of thymidine phosphorylase in human solid tumours. *Eur J Cancer*. 32A(7):1227-32

Takebayashi Y, Natsugoe S, Baba M, Akiba S, Fukumoto T, Miyadera K, Yamada Y, Takao S, Akiyama S, Aikou T. (1999). Thymidine phosphorylase in human esophageal squamous cell carcinoma. *Cancer*. 85(2):282-9.

Tam W, Gomez M, Chadburn A, Lee JW, Chan WC, Knowles DM. (2006). Mutational analysis of PRDM1 indicates a tumor-suppressor role in diffuse large B-cell lymphomas. *Blood*. 107(10):4090-100.

Tan KL, Scott DW, Hong F, Kahl BS, Fisher RI, Bartlett NL, Advani RH, Buckstein R, Rimsza LM, Connors JM, Steidl C, Gordon LI, Horning SJ, Gascoyne RD. (2012). Tumor associated macrophages predict inferior outcomes in classic Hodgkin lymphoma: a correlative study from the E2496 Intergroup trial. *Blood.* 120(16):3280-7.

Theocharidis A, van Dongen S, Enright AJ, Freeman TC. (2009). Network visualization and analysis of gene expression data using BioLayout Express(3D). *Nat Protoc.* 4(10):1535-50.

Tian Y, Kelemen SE, Autieri MV. (2006). Inhibition of AIF-1 expression by constitutive siRNA expression reduces macrophage migration, proliferation, and signal transduction initiated by atherogenic stimuli *Am J Physiol Cell Physiol*. 290(4):C1083-91.

Toi M, Yamamoto Y, Taniguchi T, Saji S, Hayashi K, Tominaga T. (1996). Regulation of endothelial growth factor expressions in breast cancer. *Gan To Kagaku Ryoho.* 23 Suppl 1:75-9.

Torroella-Kouri M, Silvera R, Rodriguez D, Caso R, Shatry A, Opiela S, Ilkovitch D, Schwendener RA, Iragavarapu-Charyulu V, Cardentey Y, Strbo N, Lopez DM. (2009). Identification of a subpopulation of macrophages in mammary tumor-bearing mice that are neither M1 nor M2 and are less differentiated. *Cancer Res.* 69(11):4800-9.

Tothill RW, Tinker AV, George J, Brown R, Fox SB, Lade S, Johnson DS, Trivett MK, Etemadmoghadam D, Locandro B, Traficante N, Fereday S, Hung JA, Chiew YE, Haviv I; Australian Ovarian Cancer Study Group, Gertig D, DeFazio A, Bowtell DD. (2008). Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clin Cancer Res.* 14(16):5198-208.

Truman LA, Ford CA, Pasikowska M, Pound JD, Wilkinson SJ, Dumitriu IE, Melville L, Melrose LA, Ogden CA, Nibbs R, Graham G, Combadiere C, Gregory CD. (2008). CX3CL1/fractalkine is released from apoptotic lymphocytes to stimulate macrophage chemotaxis. *Blood.* 112(13):5026-36.

Tweeddale ME, Lim B, Jamal N, Robinson J, Zalcberg J, Lockwood G, Minden MD, Messner HA. (1987). The presence of clonogenic cells in high-grade malignant lymphoma: a prognostic factor. *Blood.* 69(5):1307-14.

Tzankov A, Zimpfer A, Pehrs AC, Lugli A, Went P, Maurer R, Pileri S, Dirnhofer S. (2003). Expression of B-cell markers in classical hodgkin lymphoma: a tissue microarray analysis of 330 cases. *Mod Pathol.* 16(11):1141-7.

Tzankov A, Meier C, Hirschmann P, Went P, Pileri SA, Dirnhofer S. (2008). Correlation of high numbers of intratumoral FOXP3+ regulatory T cells with improved survival in germinal center-like diffuse large B-cell lymphoma, follicular lymphoma and classical Hodgkin's lymphoma. *Haematologica.* 93(2):193-200.

Uherova P, Ross CW, Schnitzer B, Singleton TP, Finn WG. (2001). The clinical significance of CD10 antigen expression in diffuse large B-cell lymphoma. *Am J Clin Pathol.* 115(4):582-8.

Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, Zwahlen M, Kampf C, Wester K, Hober S, Wernerus H, Björling L, Ponten F. (2010). Towards a knowledge-based Human Protein Atlas. *Nat Biotechnol.* 28(12):1248-50.

Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szigyarto CA, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist PH, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, von Heijne G, Nielsen J, Pontén F. (2015). Proteomics. Tissue-based map of the human proteome. *Science.* 347(6220):1260419.

Ullrich A, Sures I, D'Egidio M, Jallal B, Powell T, Herbst R, Dreps A, Azam M, Rubenstein M, Natoli C, Shawver L, Schlessinger J and Iacobelli S. (1994). The secreted Tumor associated antigen 90K is a potent immune stimulator. *J Biol Chem.* 269 (28), 18401-1840

Ulmer T, Keeler V, Loh L, Chibbar R, Torlakovic E, Andre S, Gabius H, Laferte S. (2006). Tumor-associated antigen 90K/Mac-2-binding protein: Possible role in Colon Cancer. *J. Cell. Biochem.* 98, 1351-1366.

Utans U, Arceci RJ, Yamashita Y, Russell ME. (1995). Cloning and characterization of allograft inflammatory factor-1: a novel macrophage factor identified in rat cardiac allografts with chronic rejection. *J Clin Invest.* 95(6):2954-62.

Utans U, Quist WC, McManus BM, Wilson JE, Arceci RJ, Wallace AF, Russell ME. (1996). Allograft inflammatory factory-1. A cytokine-responsive macrophage molecule expressed in transplanted human hearts. *Transplantation.* 61(9):1387-92.

Van Dongen, Stijn, *Graph Clustering by Flow Simulation*. PhD thesis, University of Utrecht, May 2000.

Van Overmeire E, Laoui D, Keirsse J, Van Ginderachter JA, Sarukhan A. (2014). Mechanisms driving macrophage diversity and specialization in distinct tumor microenvironments and parallelisms with other tissues. *Front Immunol.* 5:127.

Veelken H, Vik Dannheim S, Schulte Moenting J, Martens UM, Finke J, Schmitt-Graeff A. (2007). Immunophenotype as prognostic factor for diffuse large B-cell lymphoma in patients undergoing clinical risk-adapted therapy. *Ann Oncol.* 18(5):931-939.

Verbrugge A, Ruiter Td Td, Clevers H, Meyaard L. (2003). Differential contribution of the immunoreceptor tyrosine-based inhibitory motifs of human leukocyte-associated Ig-like receptor-1 to inhibitory function and phosphatase recruitment. *Int Immunol*. 15(11):1349 58.

Voll RE, Herrmann M, Roth EA, Stach C, Kalden JR, Girkontaite I. (1997). Immunosuppressive effects of apoptotic cells. *Nature.* 390(6658):350-1.

Vucic D, Stennicke HR, Pisabarro MT, Salvesen GS, Dixit VM. (2000). ML-IAP, a novel inhibitor of apoptosis that is preferentially expressed in human melanomas. *Curr Biol.* 10(21):1359-66

Wallgard E, Larsson E, He L, Hellström M, Armulik A, Nisancioglu MH, Genove G, Lindahl P, Betsholtz C. (2008). Identification of a core set of 58 gene transcripts with broad and specific expression in the microvasculature. *Arterioscler Thromb Vasc Biol.* 28(8):1469-76.

Watano K, Iwabuchi K, Fujii S, Ishimori N, Mitsuhashi S, Ato M, Kitabatake A, Onoé K. (2001). Allograft inflammatory factor-1 augments production of interleukin-6, -10 and -12 by a mouse macrophage line. *Immunology.* 104(3):307-16.

Weiss LM, Warnke RA, Sklar J, Cleary ML. (1987). Molecular analysis of the t(14;18) chromosomal translocation in malignant lymphomas. *N Engl J Med.* 317(19):1185-9.

Wilkinson ST, Vanpatten KA, Fernandez DR, Brunhoeber P, Garsha KE, Glinsmann-Gibson BJ, Grogan TM, Teruya-Feldstein J, Rimsza LM. (2011). Partial plasma cell differentiation as a mechanism of lost major histocompatibility complex class II expression in diffuse large B-cell lymphoma. *Blood.* 119(6):1459-67.

Wu H, Xu JB, He YL, Peng JJ, Zhang XH, Chen CQ, Li W, Cai SR. (2012). Tumor associated macrophages promote angiogenesis and lymphangiogenesis of gastric cancer. *J Surg Oncol.* 106(4):462-8.

Wyckoff JB, Wang Y, Lin EY, Li JF, Goswami S, Stanley ER, Segall JE, Pollard JW, Condeelis J. (2007). Direct visualization of macrophage-assisted tumor cell intravasation in mammary tumors. *Cancer Res*. 67(6):2649-56.

Yamamoto A, Dhar DK, El-Assal ON, Igarashi M, Tabara H, Nagasue N. (1998). Thymidine phosphorylase (platelet-derived endothelial cell growth factor), microvessel density and clinical outcome in hepatocellular carcinoma. *J Hepatol.* 29(2):290-9.

Yan DJ, Chen ZW. (2010). 17β-Estradiol increased the expression of daintain/AIF-1 in RAW264.7 macrophages. *Biosci Biotechnol Biochem*. 74(10):2103-5.

Yang ZF, Ho DW, Lau CK, Lam CT, Lum CT, Poon RT, Fan ST. (2005). Allograft inflammatory factor-1 (AIF-1) is crucial for the survival and pro-inflammatory activity of macrophages. *Int Immunol.* 17(11):1391-7.

Yang ZZ, Novak AJ, Stenson MJ, Witzig TE, Ansell SM. (2006). Intratumoral CD4+CD25+ regulatory T-cell-mediated suppression of infiltrating CD4+ T cells in B-cell non-Hodgkin lymphoma. *Blood.* 107(9):3639-46

Yang ZZ, Novak AJ, Ziesmer SC, Witzig TE, Ansell SM. (2007). CD70+ non-Hodgkin lymphoma B cells induce Foxp3 expression and regulatory function in intratumoral CD4+CD25 T cells. *Blood.* 110(7):2537-44

Yang ZZ, Novak AJ, Ziesmer SC, Witzig TE, Ansell SM. (2009). Malignant B cells skew the balance of regulatory T cells and TH17 cells in B-cell non-Hodgkin's lymphoma. *Cancer Res.* 69(13):5522-30.

Ye Y, Miao S, Lu R, Xia X, Chen Y, Zhang J, Wu X, He S, Qiang F, Zhou J. (2014). Allograft inflammatory factor-1 is an independent prognostic indicator that regulates β catenin in gastric cancer. *Oncol Rep.* 31(2):828-34.

Yoshikawa T, Suzuki K, Kobayashi O, Sairenji M, Motohashi H, Tsuburaya A, Nakamura Y, Shimizu A, Yanoma S, Noguchi Y. (1999). Thymidine phosphorylase/platelet-derived endothelial cell growth factor is upregulated in advanced solid types of gastric cancer. *Br J Cancer*. 79(7-8):1145-50.

Yunis JJ, Mayer MG, Arnesen MA, Aeppli DP, Oken MM, Frizzera G. (1989). Bcl-2 and other genomic alterations in the prognosis of large-cell lymphoma. *N Engl J Med.* 320(16):1047-54.

Zeisberger SM, Odermatt B, Marty C, Zehnder-Fjällman AH, Ballmer-Hofer K, Schwendener RA. (2006). Clodronate-liposome-mediated depletion of tumour-associated macrophages: a new and highly effective antiangiogenic therapy approach. *Br J Cancer.* 95(3):272-81.

Zhang L, Mackenzie IZ, Rees MC, Bicknell R. (1997). Regulation of the expression of the angiogenic enzyme platelet-derived endothelial cell growth factor/thymidine phosphorylase in endometrial isolates by ovarian steroids and cytokines. *Endocrinology.* 138(11):4921-30.

Zhao YY, Yan DJ, Chen ZW. (2013). Role of AIF-1 in the regulation of inflammatory activation and diverse disease processes. *Cell Immunol.* 284(1-2):75-83.

Zhu D, McCarthy H, Ottensmeier CH, Johnson P, Hamblin TJ, Stevenson FK. (2002). Acquisition of potential N-glycosylation sites in the immunoglobulin variable region by somatic mutation is a distinctive feature of follicular lymphoma. *Blood.* 99(7):2562-8.

Zocchi MR, Pellegatta F, Pierri I, Gobbi M, Poggi A. (2001). Leukocyte-associated Ig-like receptor-1 prevents granulocyte-monocyte colony stimulating factor-dependent proliferation and Akt1/PKB alpha activation in primary acute myeloid leukemia cells. *Eur J Immunol.* 31(12):3667-75.

Appendix 1

Selected gene lists from bioinformatic analysis.

See enclosed CD-ROM in back of thesis.

Appendix 2.

Publication arising.