Proteomics Analysis of Chronic Lymphocytic Leukaemia Cells

Suliman Abdallah A. Alsagaby

A thesis submitted to Cardiff University for the Degree of Doctor of Philosophy

Department of Haematology
Institute of Cancer & Genetics
School of Medicine
Cardiff University
Cardiff
United Kingdom

Acknowledgments

Foremost, my utmost gratitude is due to Allah The Almighty God whose blessing, help and guidance made this study possible to conduct.

Next, I owe my deepest gratitude to my supervisors: Dr. Paul Brennan, Dr. Christopher Pepper and Dr. Ian Brewis for their supervision. They pleasantly provided the advice, guidance and encouragement necessary for me to proceed through the doctorial programme and complete this thesis.

I also would like to express my special thanks to Dr Christopher Fegan and Dr. Guy Pratt for providing me with patient samples that were used in this project.

The CBS Proteomics facility deserves my sincere thanks for the help they offered in the proteomics part of this study. I am grateful to Dr. Sanjay Khanna for his assistance regarding peptide separation by liquid chromatography and to Dr. Keith Hart for helping me with mass spectrometry for peptide sequencing and protein identification and quantification.

I also would like to show my gratitude to all the members of the CLL group for their friendly assistance and valuable friendship. In particular I would like to thank Dr. Laurence Pearce and Dr. Liam Morgan for their help regarding flow cytometry.

Last but not least, I would like to thank my mother Fatemah Alturky, my father Abdallah Alsagaby and my wife Wafa Almansour for their love, inspiration, support and prayer that, without any doubt, helped throughout this project.

This thesis is dedicated to my mother Fatemah Alturky, my father Abdallah Alsagaby and my wife Wafa Almansour

Abstract

CLL is a malignant disease of B-cells characterised by a heterogeneous clinical outcome. Some patients require an early treatment and have low survival time, while others never need treatment. Many prognostic markers have been established and used to help predict the clinical course of CLL. Despite advances in understanding the biology of CLL, the molecular differences underlying the variable clinical outcome of CLL are not yet fully understood.

The hypothesis of this study was that the heterogeneous outcome of CLL could be driven, in part, by the aberrant expression of proteins in the two forms of CLL. Therefore, this study aimed to identify these proteins using proteomics approaches.

In an attempt to achieve this goal, four steps were performed. Firstly, a cellular fractionation method was developed to extract cellular proteins into two different fractions (NP40 fraction for cytosolic protein enrichment and SDS fraction for nuclear protein enrichment).

Secondly, extracted proteins were subjected to qualitative proteomics analysis using 2D nano-LC and MALDI TOF-TOF mass spectrometry in order to identify CLL proteins. Integrating the identified proteins (n=900) with previously published transcriptome of CLL cells and normal B-cells highlighted

20 proteins with preferential expression in CLL cells - some of which were linked to human cancer.

Thirdly, iTRAQ technology coupled with 2D nano-LC and MALDI TOF-TOF mass spectrometry was used to measure the relative expression of proteins in different CLL samples. This workflow identified 15 altered proteins in the two forms of CLL and detected 14 proteins with variable expression.

Finally, six proteins were selected for investigation in an additional CLL cohort. Of these proteins thyroid hormone receptor-associated protein 3 (TR150), T-cell leukaemia/lymphoma protein 1A (TCL-1) and S100A8 showed association with poor prognosis CLL and early requirement for treatment. Additionally, myosin-9 exhibited reduced expression in poor prognosis CLL samples.

Overall, this study identified proteins with potential importance in CLL prognosis and pathology. These proteins merit investigation in a larger CLL cohort to further confirm their relevance to CLL. In addition, this study showed the usefulness of combining cellular fractionation with proteomics and transcriptomics to identify proteins with potential role in CLL.

Publications

Identification of differentially expressed proteins in chronic lymphocytic leukaemia using qualitative and quantitative proteomics. Suliman Alsagaby, Sanjay Khanna, Keith W Hart, Guy Pratt, Christopher Fegan, Christopher Pepper, Ian A. Brewis, Paul Brennan. Manuscript in preparation.

Differential proteome associate with CD38 expression in chronic lymphocytic leukaemia. Suliman Alsagaby, Sanjay Khanna, Keith W Hart, Christopher Fegan, Christopher Pepper, Ian A. Brewis, Paul Brennan. Abstract and Poster were published in the British Society of Proteomics Research conference 12-14 July 2011: From the visible to the hidden proteome.

Identification of differentially expressed proteins in chronic lymphocytic leukaemia using qualitative and quantitative proteomics. Suliman Alsagaby, Sanjay Khanna, Keith W Hart, Guy Pratt, Christopher Fegan, Christopher Pepper, Ian A. Brewis, Paul Brennan. Abstract and Poster were published in the Saudi scientific international conference 11-14 October 2012.

Presentations

A talk entitled Proteomics analysis of chronic lymphocytic leukaemia was given to the department of Infection, Immunity and Biochemistry in Cardiff University.

A talk entitled Proteomics analysis of chronic lymphocytic leukaemia was given to the department of haematology in Cardiff University.

Talks entitled Proteomics analysis of chronic lymphocytic leukaemia were regularly given to the CLL group in Cardiff University.

Index

List of Figures

List of Tables

Abbreviations

1DE	1 dimensional gel
2DE	2 dimensional gel
AB1IP	amyloid beta A4 precursor protein-binding family B member 1-interacting protein
ACINU	apoptotic chromatin condensation inducer in the nucleus
BCA assay	bicinchoninic acid assay
BCR	B-cell receptor
CHCA	α-cyano-4-hydroxycinnamic acid
cICAT	cleavable isotope-coded affinity tag
CID	collision induced dissociation
CLL	chronic lymphocytic leukaemia
FUS	RNA-binding protein FUS
Glu- Fib	[glu1] – fibrinopeptide B
GNA13	guanine nucleotide-binding protein subunit alpha-13
GPS	global protein server
HSB	high salt buffer
IGHV	immunoglobulin heavy chain variable region genes
IKZF3	zinc finger protein aiolos
iTRAQ	isobaric tags for relative and absolute quantitation
KCl	potassium chloride
LASP-1	LIM and SH3 domain protein 1
M-CLL	mutated-CLL

MALDI	matrix assisted laser desorption/ ionization
$MgCl_2$	magnesium chloride
MMTS	methyl methane-thiosulfonate
PAGE	polyacrylamide gel electrophoresis
PARP	poly (ADP-ribose) polymerase
PBMCs	peripheral blood mononuclear cells
PBS	phosphate buffered saline
PMF	peptide mass fingerprint
PMSF	phenylmethanesulfonylfluoride
PRDX2	peroxiredoxin-2
PRDX5	peroxiredoxin-5, mitochondrial
ProtScore	protein score
RP	reversed phase
SCX	strong cation exchange column
STMN1	stathmin
TCEP	tris-(2-carboxyethyl)phosphine
TCL-1	T-cell leukemia/lymphoma protein 1A
TEAB	triethylammonium bicarbonate
TFA	tetrafluoroacetic acid
THIM	3-ketoacyl-CoA thiolase, mitochondrial
TOF-TOF	time of flight-time of flight
TR150	thyroid hormone receptor-associated protein 3
U-CLL	unmutated-CLL
ZAP-70	zeta associated protein 70

Chapter One

Introduction

1.1 B-cell Chronic Lymphocytic Leukaemia (CLL)

Chronic lymphocytic leukaemia (CLL) is a neoplastic disease of mature-appearing lymphocytes in the peripheral blood, bone marrow, lymph nodes and spleen. The malignant cells in this disease are characterised by their expression of CD19 (B-cell marker) and co-expression of CD5 and CD23 (Dighiero, 2005). The majority of CLL cases are asymptomatic at diagnosis and are discovered by chance in a routine blood check (Redaelli et al., 2004). The clinical course of CLL is heterogeneous ranging from an aggressive form with rapid progression to a stable form with no excess age-adjusted risk of mortality (Pepper et al., 2012).

1.1.1 Epidemiology

CLL is the most common adult leukaemia in western society (Dighiero, 2005). The average incidence of this disease in the UK is 3/100,000 per year and is 5.5/100,000 worldwide (Oscier et al., 2004, Redaelli et al., 2004). The risk of developing this disease increases with age; according to the Surveillance, Epidemiology and End Result (SEER), the median age at diagnosis of male CLL patients is 70 years, while that of female CLL patients is 74 years. However, 20-30% of CLL patients are less than 55 years old at diagnosis (Oscier et al., 2004). This disease is more common in men than women (2:1) and the median survival time of CLL patients is longer among female patients compared to that of male patients (Molica, 2006).

CLL is more prevalent in some ethnic populations than others. American people of Chinese, Japanese and Filipino origin are at 5 times lower risk of developing CLL than Caucasian American people (Weiss, 1979). Likewise, the incidence of CLL in Israel is significantly lower among African and Asian migrants compared to that of European migrants (Bartal et al., 1978).

1.1.2 Aetiology

1.1.2.1 Environmental factors

No single environmental factor has been convincingly established as a risk factor for CLL. Unlike chronic myeloid leukaemia (CML), CLL is not associated with ionising radiation exposure (Boice et al., 1991). However, environmental agents associated with farming have been suggested to account for the high CLL incidence in farmers (Waterhouse et al., 1996). A few studies have reported an increased CLL incidence in people frequently exposed to electromagnetic fields (Floderus et al., 1993, Feychting et al., 1997). While hepatitis C infection is common among B-cell non-Hodgkin's lymphoma patients (42%) compared to the healthy population (1%), it is less common in CLL patients (12%) (Ferri et al., 1996, Luppi et al., 1996).

1.1.2.2 Hereditary factors

CLL shows the highest familial incidence among all major haematological malignancies indicating that some genetic factors may play a role in initiating CLL (Gunz and Veale, 1969). Individuals with first-degree

relatives diagnosed with CLL are more than three times at risk of developing CLL or other lymphoid malignancies than is the general population (Gunz et al., 1975). Moreover, affected individuals in such families develop CLL earlier than patients with sporadic CLL (Yuille et al., 2000). As a result, family history of CLL has been considered as an important risk factor for CLL (Cannon-Albright et al., 1994). Another line of evidence that supports the involvement of genetic risk factors in CLL is the low CLL incidence in some ethnic groups, which does not considerably increase after migration to areas in which CLL incidence is high (Pan et al., 2002).

In spite of the awareness that genetic components play an important role in the increased incidence of CLL in certain families, such factors have not been fully determined. Genome-wide association studies have shown that polymorphisms in some genes are associated with the risk of developing an aggressive form of the disease. For example, a polymorphism in the gene encoding CD5 (Perez-Chacon et al., 2005); CD38 (Jamroziak et al., 2009) or tumour necrosis factor alpha (TNF-α) (Jevtovic-Stoimenov et al., 2008). In addition, certain human leukocyte antigen (HLA) haplotypes, such as HLA-Cw16, have been found to be more common among CLL patients when compared to age-matched controls (Montes-Ares et al., 2006). Single nucleotide polymorphisms (SNPs) in genes encoding for interferon regulatory factor 4, GRAM domain-containing protein 1B, nuclear body protein SP140 and protein kinase D2 have been reported as risk markers of developing CLL (Di Bernardo et al., 2008). Another study has also demonstrated an

association between disease-susceptibility and SNPs in or around genes encoding proteins involved in apoptosis and immune regulatory pathways such as *CCNH*, *APAI*, *IL16*, *CASP8*, *NOS2A and CCR7* (Enjuanes et al., 2008).

1.1.3 Diagnosis of CLL

Most patients are diagnosed with CLL following a routine blood check with an increased lymphocyte count in absence of other CLL symptoms. However, some CLL patients present CLL symptoms, such as fatigue, anaemia, weight loss and lymphadenopathy at diagnosis (Oscier et al., 2004). A lymphocytosis of mature-appearing B-cells ($\geq 5.0 \times 10^9$/L) that lasts for more than one month is the first criterion of CLL diagnosis. Furthermore, CLL cells express specific surface antigens, which can be used to confirm the diagnosis of CLL. These antigens include B-cell markers CD19, CD20 and CD23. In addition, co-expression of CD5, which is a T-cell antigen, and low expression of surface immunoglobulin, CD22 and CD79b. These markers can be determined by performing immunophenotyping (Hallek et al., 2008).

1.1.4 Biology of CLL

The historical view of CLL was that it is an accumulative disease where malignant cells have defects in their programmed cell death system, which results in long-lived cells that accumulate in the peripheral blood, bone marrow, lymph nodes and spleen (Dighiero, 2003). Probably this view has been based on the observation that the majority of CLL cells (>98%) in the

peripheral blood are arrested in G_0 or the early G_1 of the cell cycle and they over express anti-apoptotic proteins (Lanasa, 2010). However, an *in vivo* study using heavy water to measure the birth and death rates of CLL cells showed heterogeneous birth of CLL cells in different patients (0.1 to 1.76% of CLL cells divide per day). Likewise, this study also reported that patients vary in the death rate of their CLL cells (Messmer et al., 2005). It is likely that the ratio between birth and death of CLL cells determines the number of CLL cells and the clinical course of the disease (Lanasa, 2010).

1.1.4.1 CLL cell trafficking

Perhaps the most critical biological process in CLL pathogenesis is trafficking of the malignant cells from peripheral blood to secondary lymphoid tissues and bone marrow. As the vast majority of CLL cells do not divide in the peripheral blood, they need to migrate to secondary lymphoid tissues and bone marrow to receive pro-survival and proliferative signals from accessory cells (Burger, 2012). Migration of CLL cells is a complex process which is regulated by different molecules, such as chemokines, chemokine receptors, integrins and cytoskeleton proteins (Davids and Burger, 2012). Stromal cells in secondary lymphoid tissue and bone marrow secrete chemokines like CXCL12 and CXCL13, which bind to their corresponding receptors (CXCL4 and CXCL5 respectively) on CLL cells in the peripheral blood. This binding induces chemotaxis of CLL cells toward the secondary lymphoid tissue and bone marrow (Deaglio and Malavasi, 2009). Upon migration of CLL cells to these sites, CLL cells attract accessory cells, such as

T-cells and nurse-like cells through active secretion of certain chemokines like CCL2 and CCL3 to create the supportive influence of the microenvironment needed for the survival and expansion of CLL cells (Davids and Burger, 2012).

Given the critical the role of the microenvironment on CLL pathology, blocking migration of CLL cells to secondary lymphoid tissue and bone marrow appears an attractive strategy to prevent CLL cells from receiving the support of the microenvironment. Subsequently, this may facilitate targeting CLL cells using conventional chemotherapy. An *in vivo* study showed that blocking the integrin alpha4 beta1 (also know as very late antigene-4 (VLA-4)) inhibits CLL homing to bone marrow (Binsky et al., 2010). Similarly, targeting CXCR4 using small peptide inhibitors was shown to antagonise chemotaxis of CLL cells (Burger et al., 2005).

1.1.4.2 Accessory cells

In the secondary lymphoid tissues and bone marrow, CLL cells meet with accessory cells such as T-cells, nurse-like cells and bone marrow stromal cells (Burger, 2012). These cells create the optimum microenvironment for CLL cell survival and growth. This occurs through direct contact between CLL cells and the accessory cells as well as through secreted cytokines from the accessory cells (Mainou-Fowler et al., 2001), Burger, 2012). Cytokines, such as IL-4, INFγ and TNFα support CLL survival throug up-regulation of anti-apoptotic proteins like BCL2 (Caligaris-Cappio and Hamblin, 1999, Caligaris-Cappio, 2011). In addition, direct contact between T-cells and

CLL cells *via* CD40L on T-cells and CD40 on CLL cells rescues CLL cells from apoptosis through the up-regulation of survivin and promotion of proliferation of CLL cells (Granziero et al., 2001) Furthermore, ligation of CD100 on CLL cells with plexin B1 receptor on T-cells induces the proliferation of CLL cells as well as prolonging their survival (Granziero et al., 2003). Physical contact between bone marrow stromal cells and CLL cells mediates protection and expansion CLL cells (Panayiotidis et al., 1996). Another type of accessory cells are nurse-like cells (NLCs), which also provide protection to CLL cells upon direct contact (Tsukada et al., 2002, Burger et al., 2000). More precisely, NLCs were shown to highly express CD31 and plexin B1 that bind to their receptors on CLL cells (CD38 and CD100 respectively) and induce growth and survival (Deaglio et al., 2005).

1.1.4.3 Important molecules in the biology of CLL

1.1.4.3.1 B-cell receptor

The B-cell receptor (BCR) complex is made of an assembly of surface immunoglobulin (sIg) and the non-covalently bound heterodimer CD79a/CD79b (Dighiero, 2005). Antigen stimulation of the BCR of CLL cells, that takes place in secondary lymphoid tissues, is a critical event in CLL initiation and progression (Caligaris-Cappio, 2011). This critical event has been implicated in several biological processes that are believed to be important in CLL pathogenesis, such as migration, survival and proliferation (Guarini et al., 2008, Quiroga et al., 2009). *In vitro* analysis showed that ligation of the BCR with an anti-IgM antibody protected CLL cells from

spontaneous apoptosis, enhanced their chemotaxis towards CXCL12 and CXCL13 and induced their proliferation (Guarini et al., 2008, Quiroga et al., 2009).

In CLL cells, expression of the surface Ig, commonly IgM and IgD, is variable but lower than that of normal B-cells (Dighiero, 2005). Different mechanisms accounting for the reduced expression of surface Ig on CLL cells have been proposed. For example, point mutation, deletion and insertion in the gene encoding CD79b were found to be associated with diminished expression of BCR on CLL cells (Thompson et al., 1997). In addition, defective folding and glycosylation of μ and CD79a chains were also related with reduced expression of BCR in CLL cells (Vuillier et al., 2005).

CLL is viewed as two subtypes of disease differing in their clinical outcome; unmutated CLL (U-CLL), where the CLL cells have not undergone somatic hypermutation in their immunoglobulin variable heavy chain genes, and mutated CLL (M-CLL), where the CLL cells have undergone somatic hypermutation in their immunoglobulin variable heavy chain genes (Hamblin et al., 1999, Damle et al., 1999). Although CLL cells express less BCR compared to normal B-cells, U-CLL cells show higher expression of BCR than M-CLL cells (Stevenson et al., 2011). In line with this finding, U-CLL cells were observed to respond better to BCR ligation compared to M-CLL cells (Stevenson and Caligaris-Cappio, 2004). An additional reason for this variation

was suggested to be due to the different affinity of BCR to antigen in U-CLL and M-CLL, where it is less specific in U-CLL (Stevenson et al., 2011).

In regard to the different responses following BCR engagement between U-CLL and M-CLL, a gene expression study showed that genes involved in signal transduction, transcription, cell-cycle regulation and cytoskeleton organisation were up-regulated in U-CLL but not in M-CLL following anti-IgM stimulation. Furthermore, *in vitro* studies showed that IgM cross-linking induced proliferation as well as progression into G_1 phase in U-CLL cells but not in M-CLL cells (Guarini et al., 2008, Quiroga et al., 2009).

1.1.4.3.2 ZAP-70

Zeta associated protein of 70 kDa (ZAP-70) is a protein tyrosine kinase that was first described in T-lymphocytes and shown to be involved in T-cell receptor (TCR) signal transduction. Upon ligation of the TCR, immune receptor tyrosine-based activation motifs (ITAMs) are phosphorylated by a member of the Src family. This event leads to recruitment of ZAP-70 to the phosphorylated ITAMs and becomes activated to participate in the transduction of downstream signalling pathways. Normal B-lymphocytes do not use ZAP-70 for their BCR signalling, but instead use another protein tyrosine kinase, called Syk (Kipps, 2007). Interestingly, in Syk-deficient B-cells, ZAP-70 was shown to compensate for the loss of Syk and facilitate BCR signalling (Kong et al., 1995). Some CLL B-cells express levels of ZAP-70 comparable to that expressed by normal T-lymphocytes (Chen et al., 2002).

High expression of ZAP-70 was implicated in BCR signalling, migration and survival of CLL cells. Enhanced signalling of B-cell receptor (BCR) was reported in CLL cells that expressed high levels of ZAP-70 compared to those that lacked ZAP-70 (Chen et al., 2002, Chen et al., 2005). Moreover, when ZAP-70^{-} CLL cells were forced to express ZAP-70 by using an adenovirus vector encoding ZAP-70, they showed greater BCR signalling following IgM ligation (Chen et al., 2005). In addition, the response of CLL cells to migratory signals, such as CCL21 and CCL19, was shown to be associated with high expression of ZAP-70 (Richardson et al., 2006). Furthermore, enhanced survival of CLL cells, following co-culture with nurse-like cells in the presence of stromal-derived factor 1 (CXCL12), was observed for CLL cells with high levels of ZAP-70 (Richardson et al., 2006).

Targeting ZAP-70 in CLL cells seems an attractive strategy to induce apoptosis. In this regard, activated heat shock protein 90 (Hsp90), which binds to and stabilizes ZAP-70, was found to be highly expressed in ZAP-70^{+} CLL cells compared with normal lymphocytes or ZAP-70^{-} CLL cells. Inhibition of activated Hsp90 was shown to cause degradation of ZAP-70, which in turn led to apoptosis in CLL cells (high ZAP-70) but not in normal lymphocytes (Castro et al., 2005).

1.1.4.3.3 CD38

CD38 is a surface molecule that acts as an enzyme using its extracellular domain to participate in the synthesis of Ca^{2+}-active metabolites

including cyclic ADP-ribose (cADPR) and ADPR and as a receptor, where it regulates the activation of an intracellular signalling pathway (Deaglio et al., 2006). Expression of CD38 is dynamic during B-cell maturation; it is up-regulated in bone marrow immature B-cells, reduced in resting normal B-cells and re-expressed in terminally differentiated plasma cells (Deaglio et al., 2003). CD38 displays different functions during lymphocyte development. For example, it causes bone marrow immature B-cells to undergo apoptosis, whereas it rescues mature B-cells from apoptosis in the germinal centre of the lymph nodes (Kumagai et al., 1995, Zupo et al., 1994).

CD38 expression was linked with important biological processes in CLL, such as proliferation, survival and migration. Expression of CD38 in CLL cells is up-regulated in spleen and lymph node, where they proliferate and survive, suggesting a potential role for CD38 in CLL survival and proliferation (Patten et al., 2008). In line with this observation, the interaction between CD38 and its ligand CD31, which is highly expressed on NLCs, was observed to activate migration and proliferation pathways of CLL cells (Deaglio et al., 2005, Deaglio et al., 2010). Consistently, ligation of CD38 with an agonistic monoclonal antibody (mimicking CD31) in the presence of IL-2 was reported to prolong survival, induce differentiation and proliferation in CLL cells (Deaglio et al., 2003). Importantly, the same study reported that the effect of CD38 ligation on CLL proliferation was more pronounced in $CD38^{+}$ CLL cells. In addition, the expression of CD38 was much higher on CLL cells that showed a response to the migratory signal CXCL12 than those that did not (Vaisitti et al., 2010).

Consistently, the same study showed that the chemotaxis of CLL cells towards CXCL12 was greatly enhanced upon ligation of CD38 with an agonistic monoclonal antibody. Interestingly, this effect was diminished by blocking CD38 with an antagonistic monoclonal antibody (Vaisitti et al., 2010). The role of CD38 in the pathogenesis of CLL, especially in CLL homing, made some authors suggest that blocking CD38 would inhibit CLL trafficking to secondary lymphoid tissues. This in turn would facilitate targeting CLL cells using conventional chemotherapy (Deaglio et al., 2008).

1.1.4.3.4 NF-κB

NF-κB is a group of transcriptional factors, which in human includes c-Rel, Rel B, p50, p52, and Rel A (p65) (Hayden and Ghosh, 2012). In resting, unstimulated cells, IκB, an NF-κB inhibitor, keeps NF-κB inactive by preventing it from translocating to the nucleus (Hayden and Ghosh, 2012). Activation of NF-κB occurs when it is freed from IκB upon phosphorylation and proteasomal degradation of the latter. NF-κB then translocates to the nucleus and binds to the promoters/enhancers of its target genes and starts regulating their expression which influences a wide variety of cellular processes including survival, differentiation, and proliferation (Hayden and Ghosh, 2012). In this regard, NF-κB was reported to inhibit apoptosis and induce survival of different cell types (Chen et al., 1999). Interestingly, NF-κB shows high activity in CLL cells compared to normal B-lymphocytes with heterogeneous basal and inducible NF-κB level in CLL (Furman et al., 2000, Hewamana et al., 2008). The reasons for the high activity of NF-κB seen in CLL are still unclear, but

there are a number of factors that can increase NF-κB activity in CLL cells such as CD40 ligation, IL-4 and IL-13 (Furman et al., 2000, Zaninoni et al., 2003).

High activity of NF-κB was implicated in the pathology of CLL. Rel A DNA binding was reported to correlate negatively with sensitivity to fludarabine and spontaneous apoptosis of freshly isolated CLL cells and positively with CLL cell counts (Hewamana et al., 2008). In agreement with this finding, the authors also showed that Rel A DNA binding was higher in CLL samples from patients with shorter lymphocyte doubling times. The importance of NF-κB in different types of cancer, including CLL, made it an attractive therapeutic target. In this regard, a study used dehydroxymethylepoxyquinomicin (DHMEQ) to diminish NF-κB activity in CLL cells. The results showed that inhibiting the activity of NF-κB induced apoptosis selectively of CLL cells (Horie et al., 2006).

1.1.4.3.5 BCL2

Programmed cell death (apoptosis) is a complex biological process that is required to control cell number and to eliminate infected or damaged cells. Therefore, the balance between anti-apoptotic forces and pro-apoptotic forces is essential for normal cellular development (Adams and Cory, 1998). Dysfunction of this process has been shown to contribute to the development of different diseases including cancer (Hajra and Liu, 2004). Apoptosis is mainly controlled by the BCL2 family, which consists of three

groups. Firstly, there is the anti-apoptotic group, which includes proteins, such as BCL2, BCL-XL and myeloid cell leukaemia sequence 1 (MCL1). Secondly, there is the multi-domain pro-apoptotic group that includes proteins with three BH domains (BH1, BH2 and BH3) like BAX and BAK. Finally, there is the BH3-only group of pro-death proteins which contains BIM, BID, BAD, BIK, and NOXA (Cory and Adams, 2002).

Several studies demonstrated that CLL cells express large amounts of the anti-apoptotic protein BCL2 as compared to normal B-cells or PBMCs from healthy donors (Hanada et al., 1993), (Del Gaizo Moore et al., 2007, McCarthy et al., 2008, Hanada et al., 1993). Unlike follicular lymphoma, *BCL2* translocation is not common in CLL rendering the cause of the high expression of BCL2 in CLL unclear (Dyer et al., 1994). However, different explanations for this observation have been suggested. For example, the expression of cytosolic nucleolin, which stabilises the mRNA of BCL2, was found to positively correlate with the expression of BCL2 in CLL cells (Otake et al., 2007). In addition, miR-15a and miR 16-1, which are natural antisense BCL2 interactors, were shown to inversely correlate with the expression of BCL2 in CLL cells (Cimmino et al., 2005).

High expression of BCL2 is a hallmark of CLL cells and is believed to play a crucial role in the survival of CLL cells and their accumulation in lymph nodes, bone marrow and peripheral blood. In this regard, higher levels of BCL2 are found in progressive CLL compared with

stable CLL (Marschitz et al., 2000). Furthermore, higher expression of BCL2 was shown to positively correlates with high CLL cell counts (>$10^5/\mu$l) (Kitada et al., 1998). Therefore, targeting BCL2 is perhaps an attractive strategy to induce apoptosis of CLL cells. In a lymphoblastic leukaemia model, elimination of BCL2 expression quickly induced apoptosis of the leukaemic cells (Letai et al., 2004). Moreover, apoptosis of CLL cells was induced upon repression of BCL2 by miR 15a and miR16-1 (Cimmino et al., 2005). Similarly, a reduction of BCL2 expression by antisense oligonucleotides was shown to induce apoptosis of CLL cells (Pepper et al., 1999). Furthermore, the small molecule BCL2 inhibitor HA14-1 (ethyl 2-amino-6-bromo-4-(1-cyano-2-ethoxy- 2-oxoethyl)-4H-chromene-3-carboxylate) effectively induced apoptosis of CLL cells (Campas et al., 2006).

1.1.4.3.6 MCL1

Another anti-apoptotic protein, MCL1 was also reported to be important in the pathology of CLL. *In vitro* analysis showed a strong correlation between resistance of CLL to fludarabine and expression of MCL1 (Pepper et al., 2008). Similarly the expression of MCL1 was shown to inversely correlate with chemotherapy response in CLL patients (Kitada et al., 1998). These observations suggest that MCL1 is a critical survival factor in CLL rendering it potentially a good therapeutic target. Different approaches have been employed to decrease MCL1 and thus induce apoptosis of CLL cells. For example, CLL cells were shown to undergo apoptosis upon reduction of MCL1 expression by siRNA. Apoptosis continued even in the presence of sustained

BCR stimulation (Longo et al., 2008). In addition, reduction of MCL1 expression by homoharringtonine (HHT) was shown to induce apoptosis in CLL cells. Interestingly, stromal cell protection of CLL cells via induction of MCL1 expression did not prevent the apoptosis driven by HHT (Chen et al., 2011).

Of other anti-apoptotic proteins, high expression of the anti-apoptotic protein BAG-1 was also reported to be associated with a failure to fulfil complete response to treatment (Kitada et al., 1998). In addition, high levels of BCL-XL mRNA were found in 70% of CLL patients studied. In contrast, mRNA of the pro-apoptotic protein BCL-XS was low in all CLL patients (Gottardi et al., 1996). More importantly, the ratios between anti-apoptotic and pro-apoptotic proteins were shown to play an important role in CLL cell survival. For example, low ratios of BCL2 (anti-apoptotic protein) to BAX (pro-apoptotic protein) were found in CLL cells sensitive to treatment, while high ratios were detected in CLL resistant to treatment (Thomas et al., 1996).

1.1.4.3.7 Chromosomal aberration

The majority of CLL patients show chromosomal aberrations in their CLL cells. A study that used fluorescence *in situ* hybridization to examine chromosomal abnormalities in 325 CLL patients found that 82% of them had chromosomal aberrations in their CLL cells (Dohner et al., 2000). This study also showed that the most common chromosomal change was a deletion in

13q (55%), a deletion in 11q (18%), trisomy 12 (16%), a deletion in 17p (7%), and a deletion in 6q (6%). Importantly, these chromosomal changes were found to have an impact on the pathology of CLL. For example, *miR15* and *miR16* are located in the region deleted in the 13q abnormality. These two genes were either deleted or down-regulated in the majority (68%) of CLL patients (Calin et al., 2002). Interestingly, *miR15* and *miR16* function as antisense *BCL2* integrators, which negatively regulate BCL2 expression at the level of transcriptional level (Cimmino et al., 2005). In addition, expression of *miR29* and *miR181* were shown to be reduced in 11q deleted patients. These two miRs were reported to negatively regulate the expression of the oncogene *TCL1* (Pekarsky et al., 2006). CLL cells with an 11q deletion were also found to highly express ATF5, a member of the activating transcription factor required for cell survival, compared with CLL cells with 13q abnormality or a normal karyotype (Mittal et al., 2007, Persengiev et al., 2002). More importantly, deletion of the 11q includes genes that encode ATM (ataxia telangiectasia mutated) (Stilgenbauer et al., 1996). Upon DNA damage, the tumour suppressor gene product p53 becomes activated by ATM, which in turn promotes cell cycle arrest for either DNA repair or cell death (Hawley and Friend, 1996). CLL patients with an ATM deficiency are associated with an aggressive form of CLL (Starostik et al., 1998). In the same context, the *TP53* tumour suppressor gene, which encodes for p53, is usually deleted in patients with a 17p abnormality (Dohner et al., 1999). CLL patients with deletion or mutation in *TP53* were reported to have an aggressive form of CLL with rapid

lymphocyte doubling times (El Rouby et al., 1993, Shaw and Kronberger, 2000).

1.1.5 Prognosis of CLL

Individual CLL patients have different clinical courses, where some patients have indolent disease and probably will never require treatment, whereas others have an aggressive form of CLL and thus need treatment (Dighiero and Hamblin, 2008). The challenge in CLL is that most patients (both types) are diagnosed with early stage disease (Pepper et al., 2012). Therefore, predicting patients whose CLL will rapidly progress and need treatment can be difficult even with the help of prognostic markers that have been used in CLL.

Some of the molecules that were shown to involve the biology of CLL, which were discussed in section 1.1.4.3, were also reported to serve as prognostic markers in CLL. In this section (1.1.5), the most commonly used prognostic markers in CLL including these molecules are discussed.

1.1.5.1 CLL stages

The first recognised prognsotic tool in CLL was the Rai staging system published in 1975. It was based on clinical observations such as lymphadenopathy, organomegaly and cytopenias (anaemia and thrombocytopenia) (Rai et al., 1975). The Rai staging system includes five different stages as shown in Table 1.

Stage	Clinical features	Median survival
Stage 0	Lymphocytosis in blood in bone marrow	>150 months
Stage I	Lymphocytosis with enlarged lymph nodes	101 months
Stage II	Lymphocytosis with enlarged liver or spleen or both	71 months
Stage III	Lymphocytosis with anaemia	19 months
Stage IV	Lymphocytosis with thrombocytopenia	19 months

Table 1: Original Rai staging system (Adapted from Rai et al., 1975).

The Rai staging system was followed by Binet's staging system (1981) which focused on two parameters; the number of involved lymph nodes and the development of cytopenias, to create three stages A, B and C, Table 2 (Binet et al., 1981).

Stage	Clinical features	Median survival
Stage A	< 3 involved areas	Not reached
Stage B	≥ 3 involved areas	84 months
Stage C	Anaemia and/or thrombocytopenia	24 months

Table 2: Binet staging system. Involved areas include axillary, cervical, inguinal, spleen and liver. (Adapted from Binet et al., 1981)

Due to the simplicity and reproducibility of these staging systems, they have been widely used and their prognostic value has been broadly accepted. Moreover, these systems are effective in identifying subgroups of patients who require treatment. Nevertheless, the major limitation of these two systems is that they do not predict the progression of the disease (Montserrat, 2006). This is very important since the majority of CLL patients are diagnosed at an early stage of the disease (Pepper et al., 2012).

1.1.5.2 *IGHV* gene status

Upon stimulation, B-cells enter germinal centres in the follicles of lymph nodes. There they undergo somatic hypermutation (SHM) of their immunoglobulin variable heavy chain genes (*IGHV*) under the direction of T-helper cells and in the presence of antigen presented by follicular dendritic cells. This process is required to enhance antibody affinity towards an antigen, as B-cells with low affinity antibody undergo apoptosis while those with the highest affinity for antigen are selected to differentiate to plasma cells and memory B-cells (Van Bockstaele et al., 2009). A study by Hamblin *et al* (1999) showed that 45% of CLL cases do not undergo SHM, as defined by ≥98 sequence homology of their *IGHV* genes with the closest germline sequence. The other 55% of CLL cases do undergo *IGHV* SHM as their *IGHV* sequences are >2% different from that of germline.

The mutational status of *IGHV* is regarded as the most accurate, widely applicable, prognostic marker in CLL. U-CLL is associated with more

aggressive disease and shorter survival time, while M-CLL is associated with an indolent disease and longer survival time (Hamblin et al., 1999, Damle et al., 1999, Pepper et al., 2012).

Although mutational status is a good predictor of the clinical outcome of CLL, certain gene segment usage of *IGHV* was shown to associate with a particular form of CLL independent of the mutational status of *IGHV*. For example, patients with V4-34 usage were reported to have a favourable form of CLL and patients with V3-21 usage were found to have an aggressive form of CLL regardless of their *IGHV* mutational status (Parker and Strout, 2011)

1.1.5.3 ZAP-70

Although CLL cells share a common gene expression profile "signature" independent of their *IGHV* mutational status, a group of genes were differentially expressed in U-CLL and M-CLL. Among these genes, ZAP-70 was the most altered gene in the two groups of CLL cases, where it was 4-fold over-expressed in U-CLL (Rosenwald et al., 2001). Subsequent studies confirmed the importance of ZAP-70 in CLL and showed that ZAP-70 is a surrogate marker of *IGHV* mutational status (Chen et al., 2002, Crespo et al., 2003). However, a later study showed that ZAP-70 expression is not always in concordance with the mutational status of *IGHV* (Rassenti et al., 2004). This study reported that 29% of U-CLL cases were negative for ZAP-70 and 17% of M-CLL cases were positive for ZAP-70.

ZAP-70 is a good indicator of the time to first treatment (TTFT) as well as survival of CLL patients. The cut-off value that discriminates between positive and negative ZAP-70 expression, when measured by flow cytometry, has been suggested to be 20% (≥20% = ZAP-70^{+} and <20% = ZAP-70^{-}) (Crespo et al., 2003, Rassenti et al., 2004). The median TTFT in ZAP-70^{+} CLL cases was 3.2 years versus 9 years in ZAP-70^{-} CLL cases (Rassenti et al., 2004). Similarly the median time of survival in ZAP-70^{+} CLL was 7.5 years, whereas the median was not reached in ZAP-70^{-} CLL (Crespo et al., 2003). The same study showed that the expression of ZAP-70 in CLL cells is stable during the course of the disease. T-cells and NK cells express ZAP-70, thus caution should be taken when determining ZAP-70 expression in CLL cells by not including that of T-cells and NK cells (Van Bockstaele et al., 2009).

1.1.5.4 CD38

CD38 expression was the first reported surrogated marker of *IGHV* mutational status (Damle et al., 1999). However, a later study reported 28% discordance between CD38 and *IGHV* mutational status (Hamblin et al., 2002). The same study also showed that CD38 expression in CLL is an independent prognostic marker, where high expression was associated with poor prognosis CLL. In line with this view, another study noticed that high CD38 expression was associated with shorter TTFT and overall survival as well as advanced stage of disease and a poor response to chemotherapy (Durig et al., 2002). The median TTFT in patients with high CD38 expression was 2.7 years versus 10 years in patients with low CD38 expression. In

addition, the median survival time was 10 years in patients with high CD38 expression and not reached in patients with low CD38 expression.

There have been different cut-offs that have been suggested to define $CD38^{+}$ CLL and $CD38^{-}$ CLL e.g. 20% (Durig et al., 2002) and 30% (Hamblin et al., 2002). CD38 expression was reported to change in some patients (24%) during the course of CLL (Hamblin et al., 2002). Nevertheless, Durig et al., 2002 demonstrated that although CD38 expreesion may change over time, it does not change from positive to negative or *vice versa*, this study was based on 20% cut off.

1.1.5.5 Chromosomal abnormalities

Chromosomal aberrations have a prognostic value in CLL. Deletions in 13q are associated with a favourable prognosis of CLL, while deletion in 11q and in 17p are associated with poor prognosis of the disease (Dohner et al., 2000). The same study showed that survival times as well as the TTFT in CLL patients were dramatically different based on the following chromosomal aberrations: 17p, 11q, 6q and 13q (Table 3). Patients with a 17p abnormality show resistance to standard anti-leukaemia agents such as alkylating drugs and/or purine nucleoside analogues. However, they show a better response to alemtuzumab, either alone or in combination with other chemotherapy (Hallek et al., 2008). Furthermore, remarked lymphadenopathy is commonly found in patients with an 11q aberration (Byrd et al., 2004).

	17p	11q	6q	13q
Median TTFT (years)	2.7	6.6	9.5	11.1
Median survival (years)	0.8	1.1	2.8	7.2

Table 3: TTFT and survival in CLL patients on the basis of chromosomal abnormalities. Adapted from (Dohner et al.,2000).

1.1.6 Treatment of CLL

As mentioned earlier CLL follows different clinical courses, where some patients suffer from an aggressive form while others have a very indolent form of the disease. Most patients are diagnosed at an early stage of the disease, where the risk is low and therefore they do not benefit from treatment. Nevertheless, as the disease progresses and becomes symptomatic, patients will start receiving treatments (Gribben, 2010). There have been several factors to make a decision of whether a patients needs treatment or not. For example, advanced clinical staging, symptomatic disease, burden of disease, age, co-morbid illnesses, adverse prognostic factors and the availability of treatments that alter survival (Byrd et al., 2004). It is important to mention that CLL treatment does not cure the disease, yet it can improve patients' health and provide relief of the disease symptoms (Dighiero, 2003). Current CLL treatments include alkylating agents, purine analogues and monoclonal antibodies, used either alone or in combination (Redaelli et al., 2004).

1.1.6.1 Alkylating agents

Alkylating agents such as chlorambucil are the historical mainstay of CLL treatment. Due to its lower toxicity, lower cost and oral route of administration it is often considered as a good first-line therapy for CLL patients. However, patients treated with alkylating agents are at 2.5-fold greater risk of developing a secondary acute myeloid leukaemia (Abbott, 2006). The complete response rates following treatment with chlorambucil are low with no recorded survival benefits, when prescribed alone (Redaelli et al., 2004). Nevertheless, combinations of different alkylating agents, such as CVP (cyclophosphamide, vincristine, prednisone) show a better response but with the same overall survival as chlorambucil alone (Byrd et al., 2004).

1.1.6.2 Purine analogues

Purine analogues, which are potent inhibitors of DNA repair, showed higher activity than alkylating agents in CLL treatment. In this regard, monotherapy of purine analogues including fludarabine, cladribine, or pentostatin demonstrate a greater activity in CLL than alkylating agents (Pinilla-Ibarz and McQuary, 2010). Based on the mode of action of purine nucleoside analogues and akylating agents, where the latter cause DNA damage and the former inhibit DNA repair, the combination of these agents may show a greater activity in CLL. In this regard, the CLL4 clinical trial reported that CLL patients treated with a combination of fludarabine and cyclophosphamide had a higher overall response, complete response and progression-free survival than patients treated with fludarabine did. As a result

the trial concluded that fludarabine and cyclophosphamide should be standard treatment for CLL (Catovsky et al., 2007).

1.1.6.3 Monoclonal Antibodies

The knowledge of the surface antigen expression of CLL cells was used to develop monoclonal antibodies that target these cells using an immunotherapy approach. Examples of these drugs are rituximab and alemtuzumab that target CD20 and CD52 respectively (Abbott, 2006). Possibly due to the low surface expression of CD20 on CLL cells, rituximab showed low activity as a single agent. In contrast, alemtuzumab showed a better activity when prescribed alone and was shown to be beneficial for patients with chromosomal abberations including 11q and 17p or *P53* mutations (Pinilla-Ibarz and McQuary, 2010). Importantly, improvement in CLL patients' overall survival was possible to achieve through introducing rituximab to chemotherapy (fludarabine and cyclophosphamide) as reported in the CLL8 trial (Hallek et al., 2010). In addition, this trial demonstrated that patients treated with chemoimmunotherapy (fludarabine, cyclophosphamide, and rituximab) had a higher progression-free survival than patients who received only fludarabine and cyclophosphamide.

1.2 Introduction to Proteomics

Proteomics is a field that studies proteins expressed by an organism, tissue, or cells on a large scale. This includes protein identification, protein expression profiling, post-translational modifications (PTMs), protein localization, protein-protein interactions and protein activity and function (Twyman, 2004). Proteomics relies on four main pillars: protein/peptide separation technologies, such as electrophoresis and liquid chromatography to reduce sample complexity; mass spectrometry for protein/peptide mass detection; database resources and bioinformatics tools to deal with the large amount of data produced by mass spectrometry analysis (Liebler, 2002).

Before the emergence of proteomics the idea of studying biological molecules globally was introduced by transcriptomic approaches, (Cox and Mann, 2007). However, with proteomics, some of the limitations of transcriptomics can be overcome. For example, protein expression can be more accurately determined by using quantitative proteomics rather than inferring it from mRNA expression (Gygi et al., 1999b, Anderson and Seilhamer, 1997). Moreover, proteomics can detect PTMs of proteins, which are not detectable at the mRNA level (Yates et al., 2009). In addition, transcriptomic output lacks information regarding protein localization, which can be obtained using proteomics approaches when sample subcellular fractionation is employed (Yates et al., 2005). However, proteomics approaches still suffer from other limitations, particularly the difficulty associated with identifying proteins with low abundance (Sriyam et al., 2007).

Therefore, combining these two omic approaches has the potential to compensate for their limitations and thus help to better elucidate the biology of a disease or normal cellular function (Hegde et al., 2003).

1.2.1 Complexity of the human proteome

The human proteome is extremely complex and thus it is impossible, at least presently, to identify the entire proteome of a human cell or body fluid in a single-step (Ahmed, 2009). In this regard, it is estimated that the number of protein-coding genes expressed by a human genome are approximately 20,500 (Clamp et al., 2007). However, the number of proteins is believed to be significantly greater due to alternative splicing and PTMs (Nilsen and Graveley, 2010). Importantly, the complexity of a human proteome is not only caused by the large number of proteins but also by the different properties of these proteins, such as their concentration. In this regard, individual proteins expressed by a human cell greatly vary in their concentration with a dynamic range of at least six-fold (Wu and Han, 2006). This range is even wider in body fluids, such as plasma, where the difference in protein abundance is more than 10 orders of magnitude (Anderson and Anderson, 2002).

1.2.2 Sample preparation

Given the extreme complexity of the human proteome, sample preparation prior to protein identification is a key factor for a successful proteomics study (Ahmed, 2009). There are many methods that can be used

to reduce sample complexity (Dreger, 2003, Ahmed, 2009, Huber et al., 2003). However, in general any sample preparation will lead to the loss of some proteins (Westermeier et al., 2008). Therefore, sample preparation should always be kept to the minimum level as long as it reduces sample complexity sufficiently to facilitate the identification of many proteins.

Although a complete cell lysate can be used for proteomics experiments, performing sample fractionation prior to protein identification reduces sample complexity and allows more coverage of proteome of the sample (Brewis and Brennan, 2010, Dreger, 2003). For example, some studies have used different detergents to generate sample fractions corresponding to different compartments, such as cytosolic proteins, nuclear proteins and membrane proteins from intact cells (Barnidge et al., 2005a, Brennan et al., 2009). These studies have shown that analysing each of these fractions individually by mass spectrometry following further separation by liquid chromatography deepens the indentified proteome and gives a clue to protein localization.

Organelle fractionation methods can be employed when proteins of a particular organelle are targeted for analysis (Huber et al., 2003). This type of sample preparation is commonly performed by homogenizing cells in the absence of detergents followed by isolation of a specific organelle, for example the nucleus, using differential gradient centrifugation (Dreger, 2003). This method reduces sample complexity, enriches specific proteins of

an organelle, for example the nuclear proteins, and also reveals more accurate information about protein localization. However, it is time consuming and labour intensive with low protein recovery (Yates et al., 2005).

1.2.3 Strategies for protein/peptide separation

As mentioned earlier, reducing sample complexity is a very critical step within a successful proteomics workflow. By exploiting the diverse properties of proteins, they can be separated according to their size, such as by SDS-PAGE or gel filtration chromatography; net charge, such as by isoelectric focussing or ion exchange chromatography; hydrophopicity, such as by reverse phase chromatography; and binding characteristics, such as by affinity chromatography (Liebler, 2002, Twyman, 2004, Westermeier et al., 2008).

1.2.3.1 Two-dimensional electrophoresis (2DE)

Using multiple orthogonal separation methods is essential to improving protein separation in order to effectively reduce sample complexity (Sriyam et al., 2007). A typical example of this type of separation is 2DE, where proteins are first separated on the basis of their ioselectric point (1st dimension) using isoelectric focussing and then further separated according to their size (2nd dimension) using SDS-PAGE (Figeys, 2005). 2DE shows a great capacity for resolving a large number of proteins (3000–10000) (Issaq et al., 2002). Nevertheless, it suffers from limitations such as low reproducibility, poor separation of very acidic or basic proteins, poor solubility

of hydrophobic proteins and poor detection of low abundance proteins (Ahmed, 2009).

1.2.3.2 Liquid chromatography (LC) of peptides

As peptides are smaller and more soluble than proteins, they are easier to separate compared with proteins (Issaq, 2001). Therefore, some of limitations of the 2DE, such as the difficulty of resolving membrane proteins, can be overcome when a protein digest is subjected to separation by liquid chromatography (LC) (Issaq, 2001). In LC, the concept of using multiple orthogonal separation methods to improve sample separation can also be applied (Yates et al., 2009). In this type of separation, a sample is subjected to separation by multiple LC columns, such as an ion exchange column followed by a reverse phase column (Westermeier et al., 2008).

1.2.3.2.1 Ion exchange chromatography (IEXC)

Ion exchange chromatography (IEXC) separation of peptides is performed on the basis of peptide pH dependent net charge. Peptides are run through an IEX column containing a matrix of spherical particles covered with positively or negatively charged ionic groups. Peptides will have different ionic strength and thus they will differentially bind to the opposite charge on the solid phase column. Peptides with a net charge of zero or the same charge as the solid phase will not bind to the column and will be eluted first in the breakthrough fraction. However, for bound peptides to be eluted a higher concentration of the elution buffer (mobile phase) is required. Typically salt

ions (Na^+ or Cl^-) of the elution buffer compete with the bound peptides for the charges on the column causing some peptides to begin to elute. Therefore, gradually increasing the salt concentration in the mobile phase elutes peptides with low attraction to the column earlier than those with high attraction (Liebler, 2002, Twyman, 2004, Westermeier et al., 2008).

1.2.3.2.2 Reverse phase chromatography (RPC)

In this kind of separation, peptides are resolved according to their hydrophobicity. The separation occurs in a column packed with matrix covered with hydrophobic alkyl chains. Peptides bind to the hydrophobic solid phase in strength proportional to their hydrophobicity. Hydrophilic peptides will not bind to the column, but instead they will elute in the breakthrough fraction. Elution of bound peptides from the column is a function of increasing the proportion of organic solvent in the mobile phase. Therefore, a linear increase of organic solvent concentration in the mobile phase causes peptides with low hydrophobicity to elute earlier than those with high hydrophobicity (Liebler, 2002, Twyman, 2004, Westermeier et al., 2008).

1.2.3.2.3 On-line two-dimensional liquid chromatography (2DLC) with salt plug

The IEX column and RP column can be on-line with each other (on-line 2DLC with salt plugs), where peptides on the IEX column (1^{st} dimension) are separated according to their net charge using a breakthrough fraction followed by increasing salt fractions. Separated peptides are then

moved to the RP trap where desalting and concentration take place and applied to the RP column (2nd dimension) for further separation on the basis of their hydrophobicity. The power of this configuration is that it is fully automated and requires less material compared with off-line 2DLC (Westermeier et al., 2008, Yates et al., 2009).

1.2.4 Mass spectrometry (MS)

A mass spectrometer is comprised of three main components; an ionization source, an analyser and a detector. Upon the application of sample into the mass spectrometer, peptides are ionised and extracted into the analyser where they are separated on the basis of their mass to charge (m/z) ratio. Separated ions are then recorded by a detector. There are many different mass spectrometry instruments, where different configurations are employed for peptide ionisation and peptide mass analysis (Mann et al., 2001).

1.2.4.1 Peptide ionization and peptide mass fingerprinting

In matrix-assisted laser desorption/ionization (MALDI) mass spectrometry peptides and UV absorbing matrix, such as α-cyano-4-hydroxycinnamic acid (CHCA), are dissolved in a solvent and applied to a metal plate. When the solvent evaporates, it leaves the peptides embedded within the matrix crystals. Upon the application of a focused UV beam on the sample for a short time energy is absorbed by the matrix and then emitted as heat causing sublimation and ionization of peptides (Karas and Hillenkamp,

1988). The ions are extracted into the mass spectrometry (MS) analyzer where they accelerate according to their m/z. Since ions produced by MALDI are predominantly singly charged, the acceleration force is constant on all these ions. However, these ions have different masses, which makes them arrive at the detector at different time points. Ions with low mass arrive at the detector earlier than those with a larger molecular weight (Wysocki et al., 2005, Figeys, 2005).

Soft ionization of peptides can also be conducted using a different method known as electrospray ionization (ESI). In this type of ionization, peptides are dissolved in a solvent and are sprayed through a narrow tube at the end of which a high electric field is held. The solvent in the sprayed droplets starts evaporating, leaving multiple charges on the peptides in the gaseous phase. These peptides enter the MS analyzer and are propelled to the detector at a speed proportional to their m/z (Mann et al., 2001).

The output of such runs does not provide information regarding amino acid sequence of a peptide, but determines the m/z of examined peptides. This approach can be used when a protein of interest is resolved in 2DE, excised, digested by trypsin and analysed by MALDI-TOF MS. The sizes of measured peptides are searched against a database that contains *in silico* digested proteins in order to match the protein that these peptides

belong to with a statistical confidence. This is known as peptide mass fingerprinting (PMF) (Thiede et al., 2005)

1.2.4.2 Tandem mass spectrometry (MS/MS)

To obtain amino acid sequence data of peptides, an additional step is required following the assessment of the peptide masses. This step is peptide fragmentation that occurs in a collision-induced dissociation (CID) chamber. The peptide fragments exit the CID chamber and 'fly' to the detector at a speed proportional to their m/z. The output of this analysis is series of spectra with different m/z as well as different intensities (Steen and Mann, 2004).

The fragmentation of the peptide backbone in the CID chamber occurs predominantly at three possible bonds. Consequently, there are six possible ions that result from such peptide fragmentation; a, b, and c, ions that contain the N-terminal of the parent peptide; and z, y and x ions which contain the C-terminal of the parent peptide (Figure 1.1) (Lim and Elenitoba-Johnson, 2004). The CID of each peptide happens inefficiently and results in multiple fragments of different lengths. The most common fragmentation of the peptide backbone results in either b or y ions. The difference in m/z between two adjacent b or y ions indicates the monoisotopic mass of the amino acid residue (Figure 1.2) (Brewis and Brennan, 2010). It is worth mentioning that the first amino acid to read using y ions is either lysine (K) or arginine (R) when trypsin, the enzyme that is most commonly used

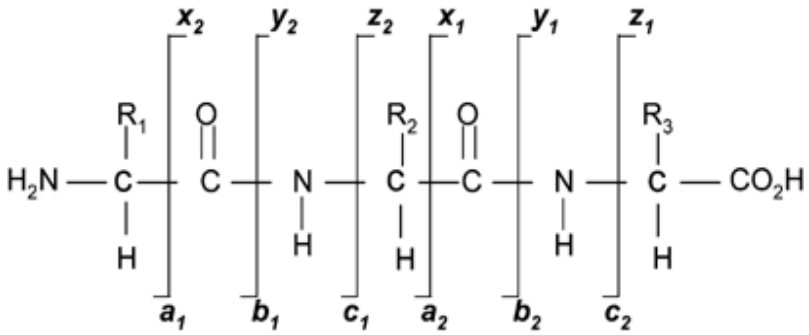


Figure 1.1: Peptide fragmentation for amino acid sequencing using tandem mass spectrometry. Six possible types of ions can be generated from the fragmentation in the peptide backbone. These include a, b, and c, ions that contain the N-terminal of the parent peptide; and z, y and x ions which contain the C-terminal of the parent peptide. Adapted from (Lim and Elenitoba-Johnson, 2004).

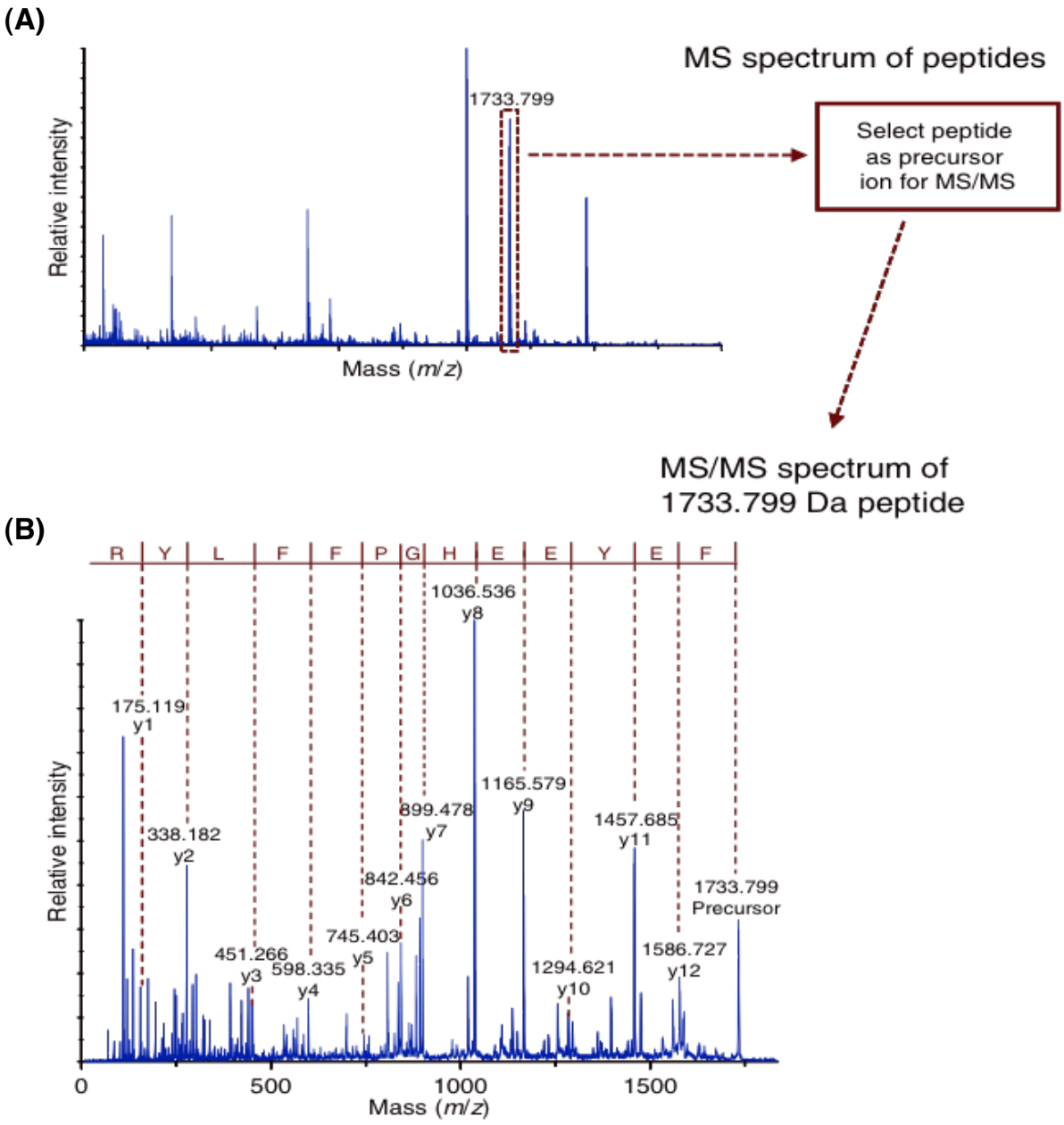


Figure 1.2: Peptide sequencing by tandem mass spectrometry. The first MS determines the masses of different peptides (A). Subsequently, a particular peptide, in this case (1733.799 Da), is selected for fragmentation in the CID chamber to obtain amino acid sequence data (B). Only y ions are shown in this example through which the peptide sequence was deducted. Adapted from (Brewis and Brennan, 2010).

for protein digestion that cleaves on the C-terminal of K or R, is used (Westermeier et al., 2008). The sequence data are then searched against a database to achieve an identification (Steen and Mann, 2004).

Although PMF is high throughput and less expensive, it fails to identify proteins in mixtures as it relies on the peptide mass rather than peptide sequence (Thiede et al., 2005, He et al., 2008). This challenge can be overcome by using tandem mass spectrometry as it identifies proteins in mixtures on the basis of their peptide sequences. In other words, even if two proteins have identical tryptic peptides masses yet different sequence, tandem mass spectrometry is capable of distinguishing between these proteins. Therefore, tandem mass spectrometry has largely superseded PMF (Coon et al., 2005).

1.2.5 Options for the proteomics workflows

There are different options forßß proteomics workflows that suit different aims. Proteins can be separated by 2DE, protein(s) of interest is excised, digested with trypsin and subjected to MS analysis (Voss et al., 2001, Perrot et al., 2011, Boyd et al., 2003, Cochran et al., 2003). Another option is to resolve proteins by 1DE, excise gel bands followed by trypsin digestion, separate the peptide mixtures on LC and carry out MS/MS analysis (Perrot et al., 2011, Boyd et al., 2003). This workflow is called (GeLC-MS or 1DE-LC and MS/MS) which is likely to achieve more proteome coverage in a given sample compared with the 2DE-MS approach (Brewis and Brennan, 2010). In

a different workflow, which does not include the use of gels, solubilised proteins are trypsin-digested and resolved by LC and subjected to MS/MS analysis (Barnidge et al., 2005a). In the proteomics field 2DLC using IEXC and RPC has become more popular as it more effectively separates peptide mixtures than 1DLC using RP alone (Westermeier et al., 2008, Yates et al., 2009). These different workflows are complementary as each one of them may reveal unique proteins (Brewis and Brennan, 2010). Figure 1.3 shows a schematic of a gel-free proteomics workflow.

1.2.6 Quantitative proteomics

Assessment of the relative abundance of proteins on a large scale has the potential to gain insights into normal cellular function and disease pathology (Twyman, 2004). There are a number of strategies that may be used to perform relative quantification of proteins on a global scale with mass spectrometry being the central platform. They may be divided into label-based quantitative proteomics and label free quantitative proteomics (Nakamura and Oda, 2007). In the latter, relative quantification of proteins is assessed using the peak intensity or spectral counting of peptide(s) that belong to one protein (Zhu et al., 2010). In principle peptide concentration strongly correlates with the intensity of its spectral peak intensity (Chelius and Bondarenko, 2002). Likewise, the more protein present the more peptide can result from trypsin digestion, which in turn increases the spectral count (Liu et al., 2004).

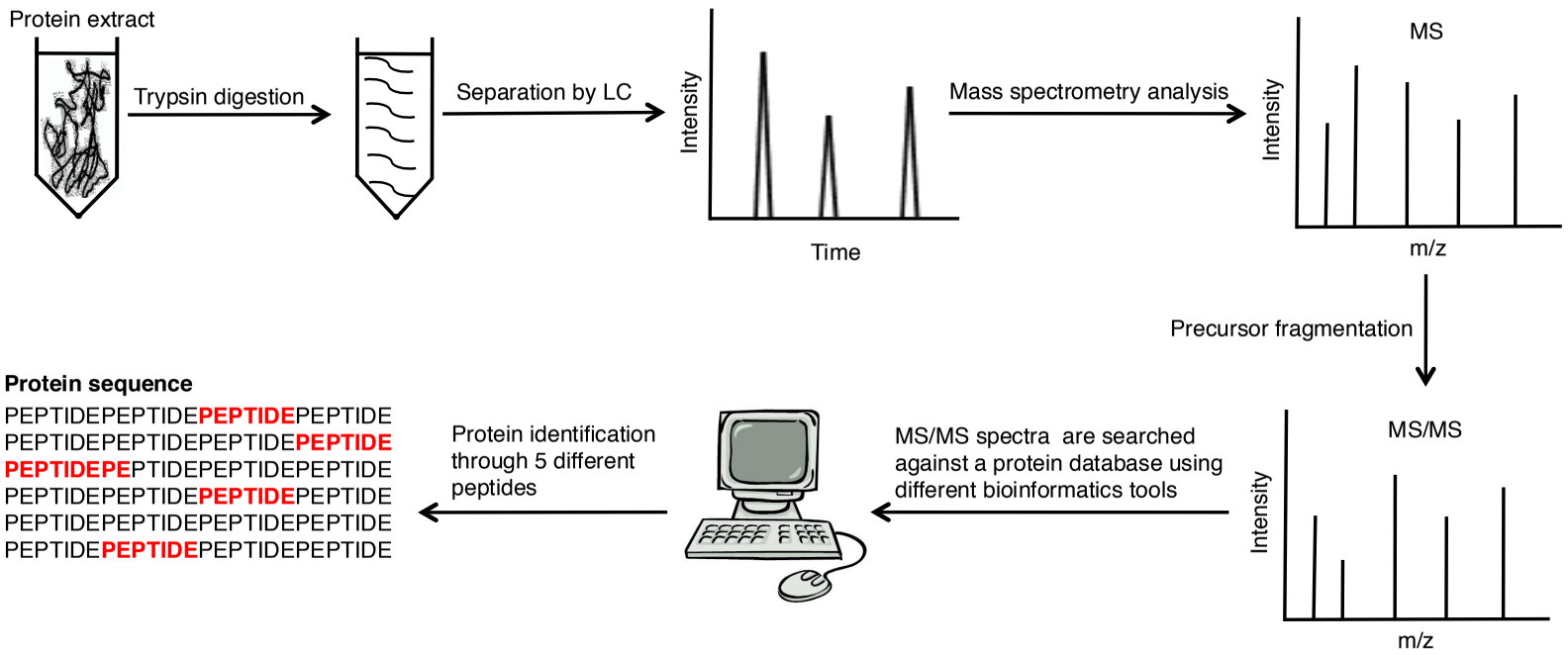


Figure 1.3: Schematic of gel-free proteomics workflow. Fraction of cellular proteins (e.g. cytosolic proteins) are extracted from the sample and subjected to digestion by trypsin. The peptide soup is then separated by LC and subjected to mass spectrometry. Firstly, the m/z of precursors is assessed followed by precursor fragmentation to generate MS/MS spectra. The latter are searched against a protein database (e.g. the Swiss-Prot database) using different bioinformatics tools, such as the MASCOT search engine, to determine the peptide sequence in order to identify the protein to which these peptides belong.

The other type of quantitative proteomics is based on the use of a mass label. While label free quantitative proteomics reveal more proteome coverage compared with label-based quantitative proteomics, the latter provide more accurate quantitative data than the former (Bantscheff et al., 2007). Of the mass label-based quantitative proteomics, iostope-coded affinity tags (ICAT) was introduced by Gygi et al., (1999a). ICAT comprises two tags that specifically bind to cysteine residues. One sample is labelled with the heavy tag and the other is labelled with the light tag. Labelled samples are pooled and digested. In the MS mode, peptides of the sample labelled with the heavy tag are distinguishable from peptides of the sample labelled with light tag by an increase in their mass by 8 Da (Gygi et al., 1999a). With ICAT, two samples can be labelled and it is only peptides containing cysteine that can be used to measure protein relative abundance (Gygi et al., 1999a).

Another approach of label-based quantitative proteomics is stable isotope labeling by amino acids in cell culture (SILAC) (Ong et al., 2002). In SILAC cells are cultured in a medium that contains isotopically labeled amino acids such as deuterated leucine. Cultured cells incorporate the deuterated leucine in the newly synthesised proteins. Cell lysate of the labelled sample is mixed with cell lystate of a control sample and subjected to protein digestion. This leads to an increase in the peptide masses of the labelled sample over non-labelled sample at the MS level. Other more commonly used labelled amino acids include lysine and arginine as one of them should be in the C-terminal of peptides that result from protein digestion

with trypsin (Yan and Chen, 2005). SILAC is suitable to label proteins of cells that can be readily grown in cell culture. The maximum number of samples that can be spontaneously studied using this approach is three (unlabeled, $^{13}C_6$, and $^{13}C_6{}^{15}N_4$-labeled amino acids) (Bantscheff et al., 2007).

Another elegant strategy of mass label-based quantitative proteomics is Isobaric tags for relative and absolute quantitation (iTRAQ) demonstrated by Ross et al, (2004). The iTRAQ reagents consist of three components; a peptide reactive group which binds to the N-terminal of a peptide and to the side amino group of lysine, a balance group with mass ranges from 28 to 31 Da and a reporter group with mass ranges from 114 to 117 Da (Figure 1.4A). There are four different iTRAQ reagents, which are named after their reporter mass; 114, 115, 116 and 117. These four reagents were designed so that they have the same mass and thus their influence on the mass of labelled peptides is the same. However, upon MS/MS fragmentation of pooled labelled samples the reporter group is fragmented from the iTRAQ reagents and detected at low m/z values; 114, 115, 116 and 117. The intensity of the reporter represents the relative abundance of the peptide in the sample (Figure 1.4B, C, D and E) (Ross et al., 2004). The iTRAQ strategy is the approach most widely used, particularly for samples that are difficult or can not be grown in cell culture. The method can label four samples spontaneously using the 4-plex iTRAQ reagents, or up to eight samples using the 8-plex iTRAQ reagents (114-121) (Nakamura and Oda, 2007).

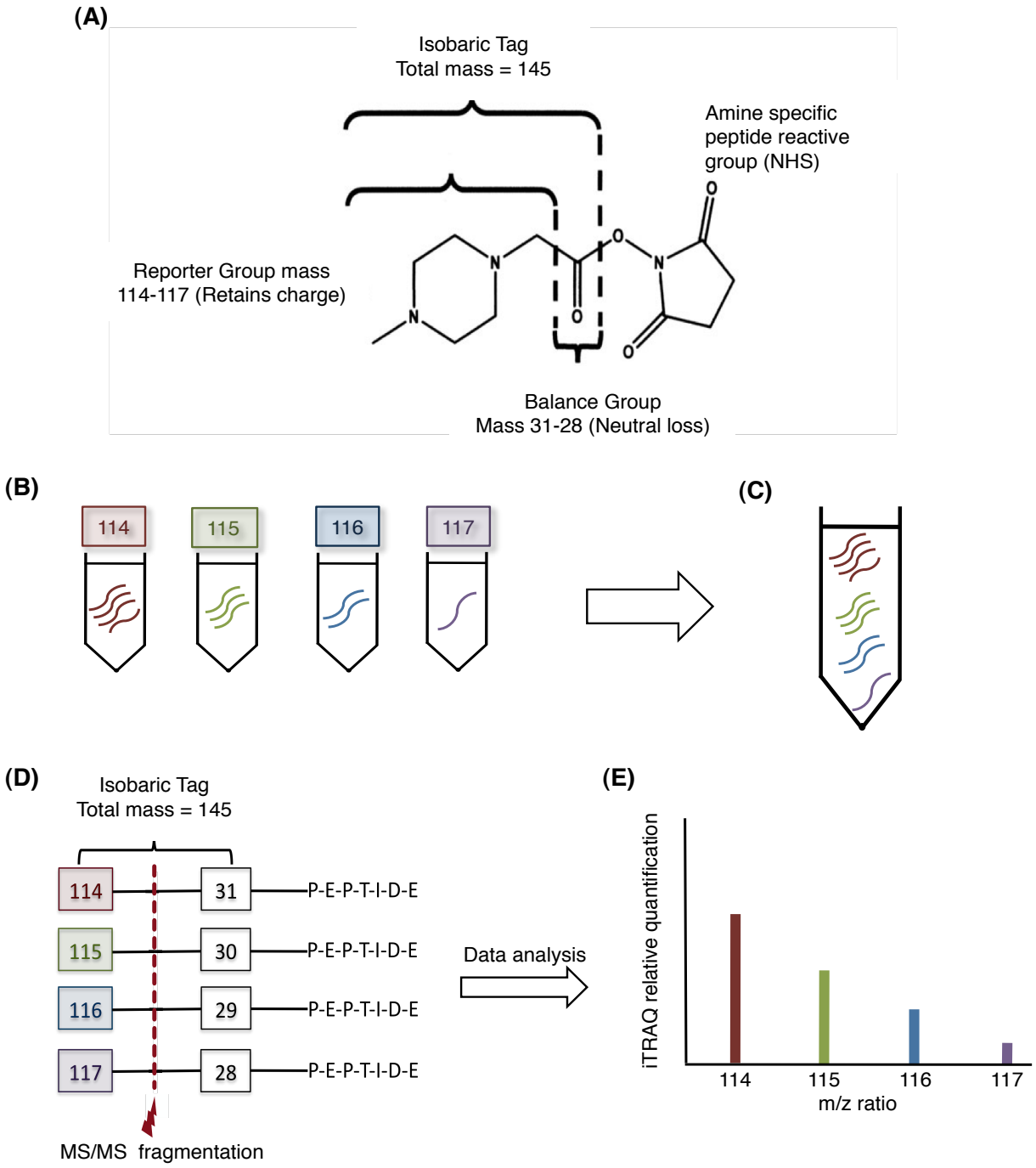


Figure 1.4: Quantitative proteomics using the iTRAQ approach. The structure of iTRAQ reagents is shown in (A). Labelling of four biological samples with 4-plex iTRAQ reagents (B) followed by multiplexing them into one tube (C). Following LC separation and MS analysis the iTRAQ reporter is cleaved upon MS/MS fragmentation (D) and recorded in the MS/MS spectra at the low m/z ratio (E). The intensity of the reporter represents the relative abundance of the peptide in the sample. Adapted from (Ross et al., 2004).

1.3 CLL proteomics

Seven studies using different proteomics approaches have been conducted on CLL samples (Table 1.4). Each of these studies will be discussed in turn. One group used 2DE-MS to study protein expression in a whole cell lysate of primary CLL cells from 24 patients with poor prognosis or with good prognosis, based on chromosomal aberrations (Voss et al., 2001). This study identified 17 proteins with altered expression on the basis of chromosomal abnormalities or patient survival rates. For instance, it was observed that high expression of heat shock protein 27 (HSP27), which is an apoptosis inhibitor, was associated with CLL cells from patients with chromosomal abnormality 11q and 17p compared to 13q. In addition, it was shown that the reduced expression of thioredoxin peroxidase 2 and protein disulfide isomerase might predict low survival in CLL patients. The altered expression of HSP27 was validated using western blotting and antibody detection in four CLL samples with 11q, six samples with 13q and five samples with 17p. The analysis confirmed high expression of this protein in 11q and 17p CLL samples compared to 13q CLL samples.

Cochran et al., (2003) used 2DE-MS to investigate protein expression in complete cell lysates from six U-CLL versus six M-CLL samples. This study reported an altered expression of proteins linked to organisation and activation of the cytoskeleton in U-CLL compared to M-CLL, for example, F-actin-capping protein beta subunit and laminin-binding protein precursor. Furthermore, nucleophosmin, which enhances p53 stability and function, was

Table1.4: List of published CLL proteomics studies.

Cell type	Comparison	Sample preparation	Quantitative approach	Proteomics workflow	Number of peptide required for protein identification	Number of detected spots or identified proteins	Number of identified altered proteins	Number of validated altered proteins	Ref
Primary CLL cells	Poor prognosis versus good prognosis based on chromosomal aberrations and survival rate	Complete cell lysate	2DE/Silver stain	2DE-MS	≥4 peptides	17 proteins	17	1	Voss et al., 2001
Primary CLL cells	N/A	Plasma membrane fraction	N/A	1DE-MS and 1DE/1DLC-MS/MS	≥1 peptide	500 proteins*	N/A	N/A	Boyd et al., 2003
Primary CLL cells	U-CLL versus M-CLL and $CD38^{+}$ CLL versus $CD38^{-}$ CLL	Complete cell lysate	2DE/colloidal Coomassie Blue	2DE-MS	≥5 peptides	800 spots 60 of them were identified	4	0	Cochran et al., 2003
Primary CLL cells	UU-CLL versus M-CLL	Cytolsolic fraction and membrane fraction	cICAT	1DLC-MS/MS	Not specified	538 proteins*	13	2	Barnidge et al., 2005a
Primary CLL cells	N/A	Membrane fraction	N/A	2DLC-MS/MS	≥1 peptide	695 proteins*	N/A	N/A	Barnidge et al., 2005b
Primary MCL, SLL and CLL cells	MCL versus SLL and CLL	Membrane proteins using MPs	N/A	1DE/1DLC-MS/MS	≥1 peptide	371 proteins	33	1	Miguet et al., 2009
Primary CLL cells	U-CLL versus M-CLL and stimulated CLL versus un-stimulated CLL	Complete cell lysate	DIGE	2DE-MS 2DE-MS/MS	≥2 peptides	1873 spots 31 of them were identified	31	2	Perrot et al., 2011

MCL: mantle cell lymphoma, SLL: small cell lymphoma, N/A: not applicable.

* No FDR data were provided in these studies and hence there is uncertainty regarding the total number of proteins.

not detected in U-CLL samples but was present in M-CLL samples. The same study also used the same samples to compare the proteome of five $CD38^+$ CLL samples to that of seven $CD38^-$ CLL samples. However, this comparison did not show statistically significant differences between $CD38^+$ CLL and $CD38^-$ CLL in terms of protein expression.

Barnidge *et al*., (2005) used cleavable ICAT (cICAT) coupled with 1DLC-MS/MS to seek deferentially expressed proteins in cytosolic and membrane fractions generated form one U-CLL sample and one M-CLL sample. This study reported 326 proteins in cytosolic extracts and 212 proteins in membrane extracts. However, it was not specified whether the total number of proteins (538) included redundant proteins that were detected in both fractions. The identification of these proteins was based on one or more peptides, yet the authors did not mention how many proteins were detected with multiple or single peptides and did not conduct false discovery rate (FDR) analysis to indicate proteins that might have been falsely identified. Hence it is not possible to fully ascertain the quality of the dataset. Of the reported proteins, 13 were shown with altered expression in the two types of CLL samples. Of these proteins, acidic leucine-rich nuclear phosphoprotein 32 family member A (A32A) and cytochrome c oxidise subunit 6B1 (COXG) were selected for validation in six U-CLL and six M-CLL using western blotting and antibody detection. The validation data were consistent with the quantitative proteomics data only for COXG, which was reduced in U-CLL compared to M-

CLL. This study did not link the different expression of COXG with the different disease outcomes associated with U-CLL and M-CLL.

Perrot *et al*., (2011) used 2D differential gel electrophoresis (2D-DIGE) and MALDI-TOF MS as well as 1D LC and tandem mass spectrometry to explore the different protein expression in complete protein extracts from three U-CLL samples versus three M-CLL samples. In addition, this study investigated the effect of BCR stimulation on the proteome of the six CLL samples. Thirty one proteins showed altered expression in this study. Of these proteins five exhibited differential expression in U-CLL compared with M-CLL without BCR stimulation. Consistent with gene expression profile studies, which showed that the BCR stimulation response is almost limited to U-CLL cells (Guarini et al., 2008, Quiroga et al., 2009), this study reported that 25 proteins demonstrated altered expression in U-CLL samples following BCR stimulation. However, only six proteins showed a change in their expression in M-CLL samples upon BCR stimulation. Of the 25 proteins, the altered expression of programmed cell death protein 4 (PDCD4), which is a tumor suppressor protein, and UV excision repair protein (RAD23B), which is a protein involved in DNA repair and activation and function of p53, were validated in 11 CLL samples using western blotting and antibody detection. The validation data consistently with the quantitative proteomics data showed that the expression of these two proteins was dramatically reduced upon stimulation of BCR in U-CLL samples.

In an attempt to identify surface proteins that may improve the diagnosis of mantle cell lymphoma (MCL), small cell lymphoma (SLL) and CLL, Miguet *et al*., (2009) used 1DE-1DLC and MS/MS to analyse plasma membrane proteins of the three types of B-cell malignancy (one sample for each type of the malignant B-cells). This study identified 579 non-redundant proteins in these three types of B-cell malignancy, of which 371 were detected in the CLL sample. The identification of these proteins was based on multiple or single peptides. The FDR was less than 1% for both types of identification, whilst the number of proteins that were identified with multiple or single peptides was not shown. Of the 597 proteins, 33 showed possible preferential expression in MCL sample. Of these proteins, CD148 was further studied in 158 patient samples and 30 control samples from healthy donors using flow cytometry. The analysis showed higher expression of CD148 in MCL compared to CLL and SLL and the authors suggested that high expression of CD184 may exclude the diagnosis of CLL and SLL and probably indicate a diagnosis of MCL.

Based on the potential importance of surface proteins as prognostic or/and therapeutic targets in CLL, Boyd et al., (2003) used 1DE-MS as well as 1DE, 1DLC and MS/MS to analyse plasma membrane protein fractions from CLL samples. This study identified 500 proteins using multiple or single peptides. Data about FDR analysis and the number of proteins that were identified with either multiple or single peptides were not reported in this study. Of the 500 proteins, 48% were membrane-associated proteins,

reflecting the quality of sample preparation for membrane proteins. Importantly, this study described novel proteins such as MIG2B and B-cell Novel Protein1 (BCNP1). The transcripts of these two proteins were found to be highly expressed in B-cell malignancies and were restricted to B-cells containing tissues such as lymph node and spleen.

Barnidge et al., (2005b) used 2DLC-MS/MS to characterise the proteome of crude membrane extracts of CLL samples. This study reported the largest number of identified proteins, 695, in CLL samples. The identification of these proteins was carried out using multiple or single peptides. However, no data about FDR analysis or the number of proteins that were identified with multiple or single peptides were shown in this study.

1.4 Justifications of this study

The majority of the CLL proteomics studies, which sought altered protein expression in poor prognosis CLL compared with good prognosis CLL, used gel-based proteomics approaches (Voss et al., 2001, Cochran et al., 2003, Perrot et al., 2011). In contrast, only one CLL proteomics study, which explored protein expression in one U-CLL sample versus one M-CLL sample, used a gel free proteomics approach (Barnidge et al., 2005a). Although these proteomics approaches have their advantages and disadvantages, it is easier to identify larger numbers of proteins using gel free proteomics approachs (Monteoliva and Albar, 2004). This was the justification to use a gel-free proteomics approach, in my project, to study 12 primary CLL samples in order to identify proteins with potential relevance to CLL.

As mentioned earlier, reducing sample complexity, such as by performing cellular fractionation, is key to a successful proteomics study (section 1.2.2). Of the CLL proteomics studies, which were performed on poor and good prognosis CLL, the majority (3/4) used complete cell lysates (Voss et al., 2001, Cochran et al., 2003, Perrot et al., 2011). In fact, only one study used two different protein extracts (cytosol and membrane) to explore protein expression in two CLL samples (U-CLL versus M-CLL) (Barnidge et al., 2005a). In this project an effort was made to subject as many proteins as possible from CLL samples to proteomics analyses by attempting to minimise the sample complexity. This was done by using a fractionation method that was developed in my study to generate two different fractions from CLL

samples. As this method was developed on the basis of using different detergents it had the great potential to solubilise more CLL proteins.

Several prognostic makers have been used to help predict the outcome of CLL (section 1.1.5). However, the majority of CLL proteomics studies, which compared the proteome the two forms of CLL, mainly used the mutational status of *IGVH* to discriminate between poor prognosis and good prognosis of CLL (Cochran et al., 2003, Barnidge et al., 2005a, Perrot et al., 2011). In fact, only one CLL proteomics study used chromosomal aberration and patient survival to distinguish between poor prognosis CLL and good prognosis CLL (Voss, et al., 2001). My project was designed to study protein expression in poor prognosis and good prognosis CLL on the basis of CD38 expression. However, further analyses of protein expression in poor prognosis and good prognosis CLL based on other commonly used prognostic markers were also conducted.

1.5 Hypothesis and aims of this project

Given that proteins are the main functional molecules in a living cell, the rationale of this project was that the heterogeneous outcome of CLL may be driven by aberrant expression of proteins in poor prognosis CLL compared with good prognosis CLL cells. Therefore, this project aimed to identify these proteins using different proteomics apporaches. To achieve this four major steps were required to be conducted:

1- Develop a cellular fractionation method that enables the extraction of as many cellular proteins from primary CLL cells as possible and keeps sample complexity reduced.
2- Perform a gel-free qualitative proteomics analysis to allow the identification of proteins in the protein fractions generated from CLL patient samples.
3- Perform gel-free quantitative proteomics studies that enable the relative quantification of proteins in the generated protein extracts from primary CLL samples in order to compare the proteome of poor prognosis CLL samples with the proteome of good prognosis CLL samples.
4- Use an antibody-based approach to validate and further investigate proteins identified by proteomics with potential importance in CLL in an additional cohort of CLL patient samples.

Chapter Two

Materials and Methods

2.1 Cell isolation

2.1.1 Isolation of mononuclear cells from peripheral blood samples of CLL patients

Peripheral blood samples from CLL patients were obtained in accordance with the ethical approval obtained from South East Wales Research Ethics Committee (02/4806). Separation of peripheral mononuclear blood cells was conducted using a previously described method (Pepper et al., 1999). Blood samples were gently shaken, transferred to 15ml tubes and diluted with an equal volume of phosphate-buffered saline (PBS; 0.01M phosphate buffer, 2.7mM KCl, 137mM NaCl). Syringes were used to slowly add 4ml Ficoll to the bottom of the tubes of each sample. Samples were carefully put in a centrifuge for a spin at 280xg for 20 min without use of the brake at the end of the run. Following centrifugation, each sample was separated into 4 layers organized from the top to the bottom of the tube as follows: plasma, mononuclear cells, Ficoll and finally red blood cells with granulocytes. A plastic pipette was used to transfer the mononuclear cell layer to a new 15ml tube, which was topped up with PBS. Samples were then spun at 280xg for 5 min. The supernatant (wash) was discarded and the cell pellet was re-suspended for 15-30 seconds with 3 ml sterile water to lyse contaminating red blood cells. The sample was topped up with PBS followed by centrifugation at 280xg for 5 min. The supernatant (wash) was then discarded and the cellular pellet was re-suspended with an amount of PBS proportional to its size. To assess the cell count, a small aliquot (60μl) from each sample was diluted with 540μl PBS and applied to a Vi-cell XR cell

counter (Beckman Coulter). Figure 2.1 shows the schematic of mononuclear cell isolation from peripheral blood samples of CLL patients.

2.1.2 Depletion of T-cells from the isolated peripheral blood mononuclear cells of CLL patients

To increase the purity of CLL cells, T-cells were depleted using the method and reagent provided with the CD3-magnetic beads (Invitrogen: 111-51D) from CLL samples with CD19 <95%. Peripheral blood mononuclear cells of CLL patients in 15ml tubes were re-suspended in 5ml PBS and were mixed with an appropriate amount of magnetic beads (75µl beads/1×10^7 cells). The beads were coated with a primary monoclonal antibody specific for the CD3 membrane antigen and thus they specifically bind to T-cells. Each sample was then incubated on a roller for 20 min at 4°C. After that, the sample was placed next to a magnet for 2 min at room temperature. A plastic pipette was used to transfer the cell suspension to a new 15ml tube without touching the bead-bound cells that were stuck to the wall of the tube. A small aliquot of sample from which T-cells were depleted were examined for CD19 expression using flow cytometry following staining with anti-CD19 antibody conjugated to allophycocyanin (Invitrogen: MHCF1905). Figure 2.2 shows a schematic of T-cell depletion from CLL samples.

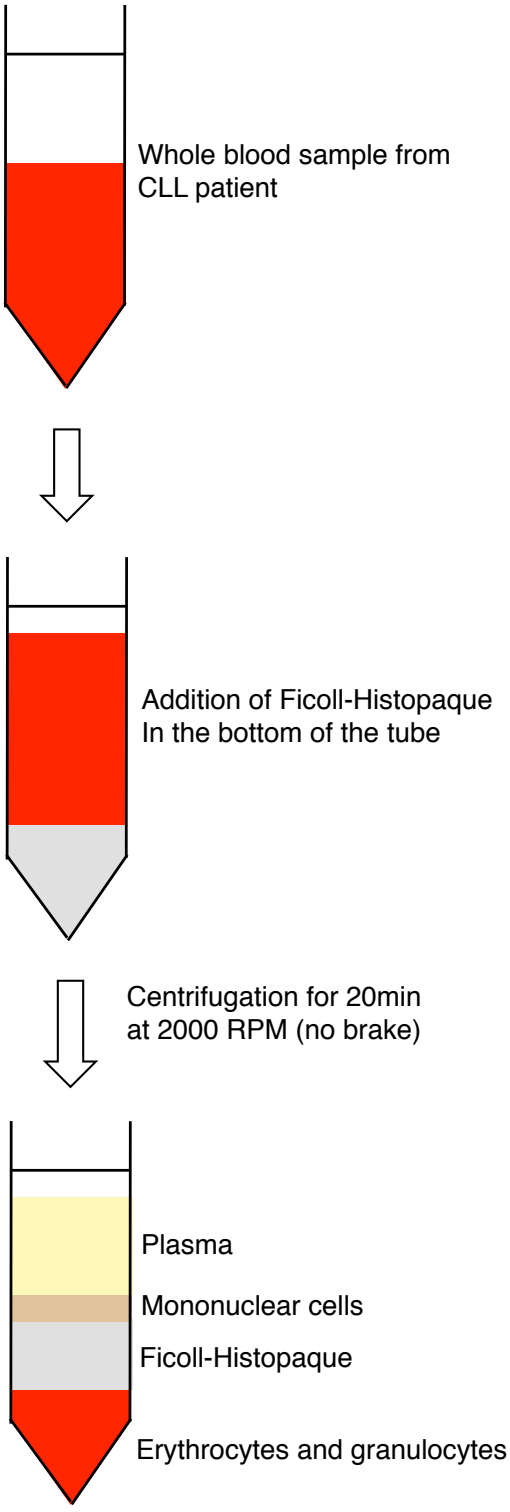


Figure 2.1: Schematic of the isolation of mononuclear cells from a whole blood sample of CLL patient using Ficoll.

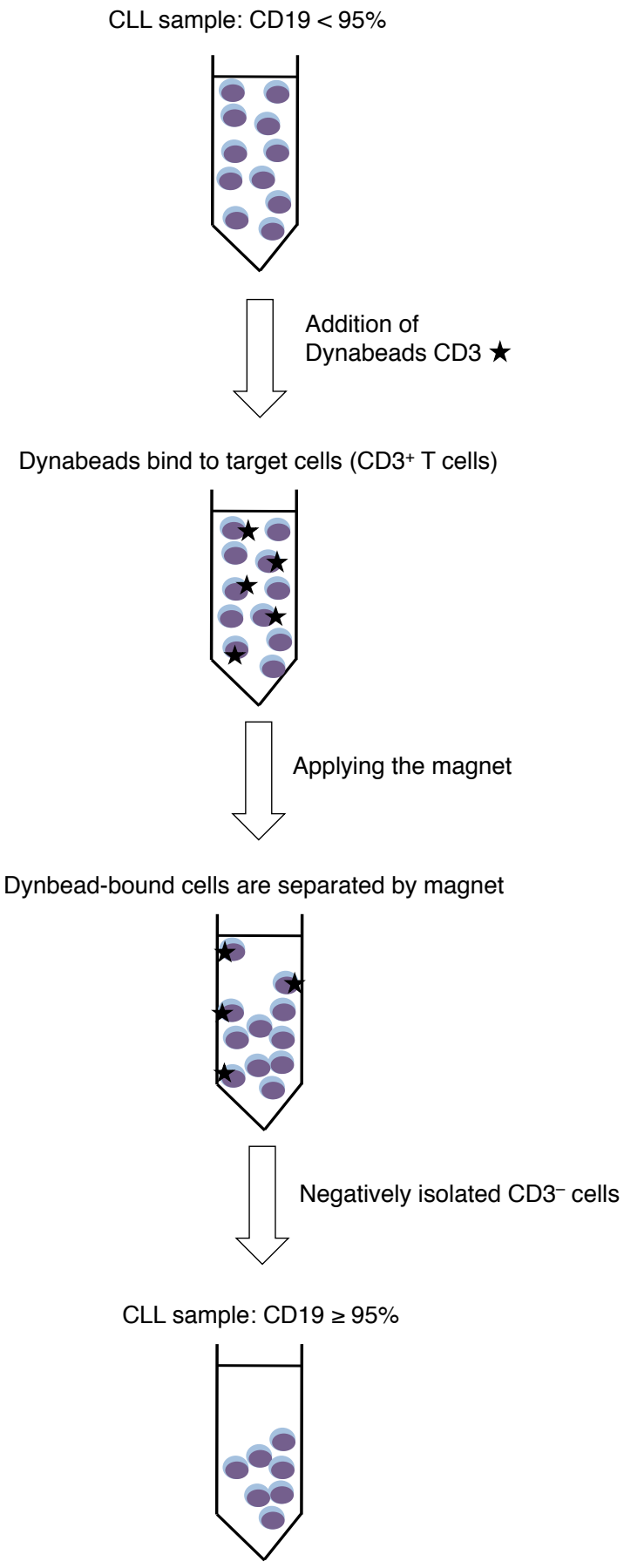


Figure 2.2: Schematic of T-cell depletion from CLL samples.

2.1.3 Isolation of mononuclear cells from buffy coat samples of healthy donors

Buffy coat samples from healthy volunteers were obtained from the Welsh Blood Service. 70ml buffy coat was transferred to a 200ml flask and diluted with 50ml PBS followed by gentle mixing. The sample (120ml) was then transferred into three 50ml tubes (35ml for each tube). Syringes were used to slowly dispense 15ml Ficoll into the bottom of each tube, which were spun at 280xg for 20 min without use of the brake at the end of the run. A plastic pipette was used to transfer the mononuclear cells layer from each tube to one new 50ml tube, which was topped up with PBS and spun at 280xg for 5 min. The supernatant was discarded and the pellet was re-suspended in 45ml PBS followed by centrifugation at 280xg for 5 min. The pellet was re-suspended in 3ml sterile water for 15-30 seconds and topped up with PBS followed by centrifugation at 280xg for 5 min. This step was performed twice. The cell pellet was then re-suspended in PBS. To assess the cell count, a small aliquot (60 µl) from the cell suspension was diluted with 540 µl PBS and applied to a Vi-cell XR cell counter (Beckman Coulter).

2.1.4 Positive isolation of B-cells from peripheral blood mononuclear cells of healthy individuals

Positive isolation of B-cells from the separated peripheral blood mononuclear cells of healthy individuals was performed according to the instructions provided with the CD19-magnetic beads (Invitrogen: 111.34D). These beads are coated with a primary monoclonal antibody specific to the

CD19 membrane-bound antigen and thus they specifically bind to B-cells. Peripheral blood mononuclear cells in a 15ml tube were re-suspended with 5ml PBS and mixed with an appropriate amount of the beads (100µl beads/1×10^7 target cells). The sample was then incubated on a roller for 20 min at 4°C. After that, the sample was placed on a magnet for 2 min at room temperature. A plastic pipette was used to transfer the cell suspension containing the PBMCs to a new 15ml tube. This was done without touching the bead-bound cells that were stuck to the wall of the tube. To wash the bead-bound cells, the tube containing them was removed from the magnet and they were re-suspended with PBS (1ml/1×10^7 beads) and placed back on the magnet for 2 min. Then the supernatant was discarded using a plastic pipette. The washing step was performed 4 times. The bead-bound cells were kept and denoted B-cell. Figure 2.3 shows a schematic of B-cells isolation from peripheral blood mononuclear cells.

2.1.5 Detaching the magnetic beads from positively isolated B-cells

The magnetic beads with the primary antibody were detached from the positively isolated B-cells using the protocol and reagent provided with the detaching reagent, DETACHaBEADS CD19 (Invitrogen: 125.06D). Bead-bound cells in a 15ml tube, were re-suspended with an appropriate amount of PBS (1ml PBS/100µl beads) and were mixed with the detaching reagent (40µl detaching reagent per 1ml sample). The sample was then incubated on a roller for 45 min at 4°C. After that, the sample was mixed with

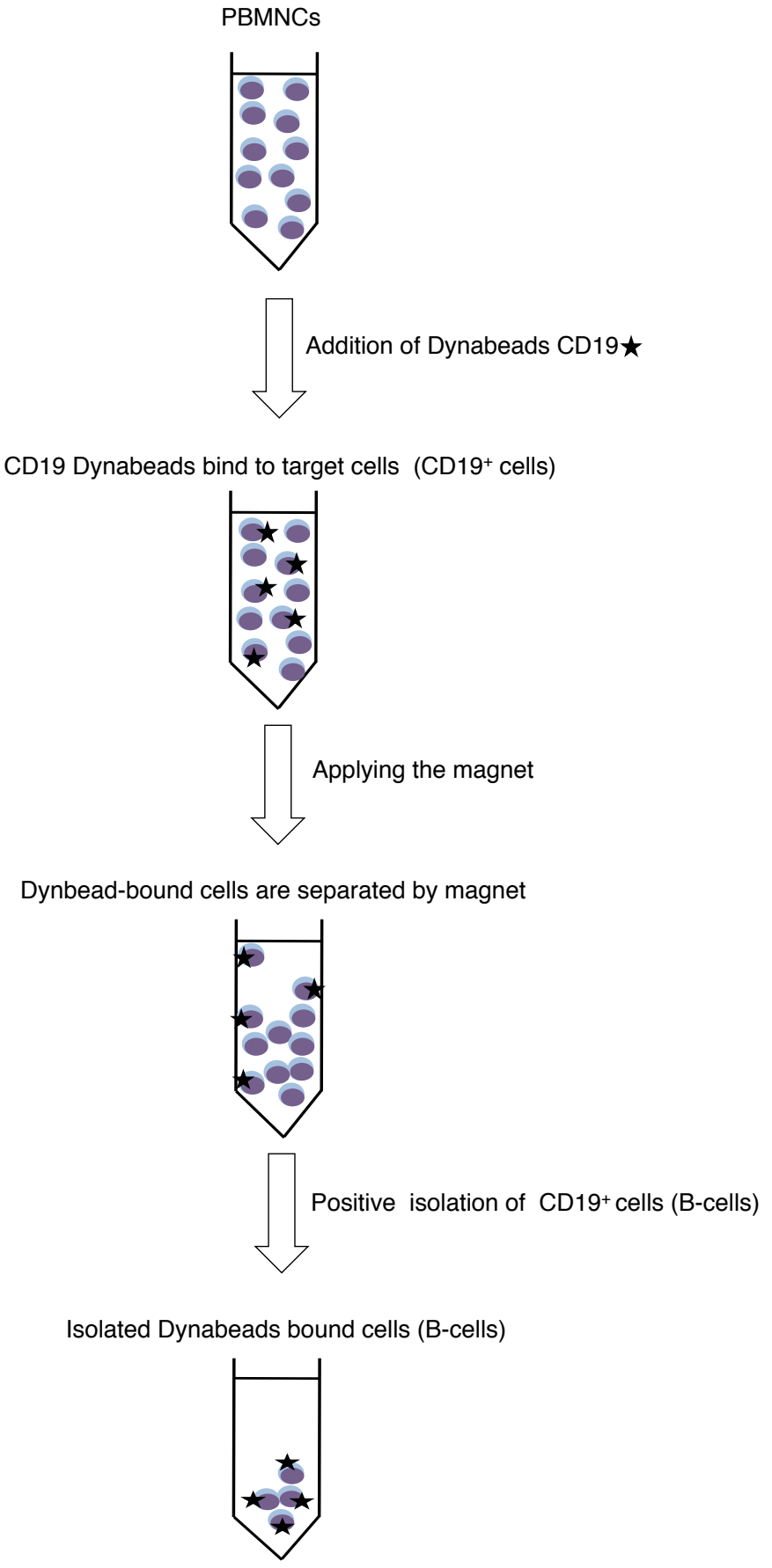


Figure 2.3: Schematic of B-cell isolation from peripheral blood mononuclear cells.

a pipette and placed on a magnet for 2 min at room temperature. The supernatant containing the released cells was transferred using a plastic pipette to a new 15ml tube and mixed with 10ml PBS followed by centrifugation at 280xg for 5 min. A schematic of detaching the magnetic beads from positively isolated B-cells is shown in Figure 2.4.

Following separation of mononuclear cells from peripheral blood samples of CLL patients with or without T-cell depletion, CLL samples (CD19 ≥95%) were pelleted and frosen on dry ice for 10-20 min and stored at -80°C for subsequent experiments. Likewise, isolated B-cells and PBMCs from buffy coat samples of healthy donors were pelleted and frosen on dry ice for 10-20 min and stored at –80°C for subsequent experiments.

2.2 Flow cytometry analysis

2.2.1 Surface staining of the isolated mononuclear cells

To assess the surface expression of CD19 on the mononuclear cells as well as the surface expression of CD38 on CD19 positive cells, two aliquots (3x 10^5 cells) from each sample were transferred to two FACS tubes and topped up to 100µl with PBS. One aliquot was stained with 4µl anti-CD19 antibody conjugated to allophycocyanin (APC) (Invitrogen: MHCF1905) and 4µl anti-CD38 antibody conjugated to Phycoerythrin (PE) (R&D system: FAB2404P). Samples were kept in the dark at room temperature for 10 min. Then 3ml PBS was added to each tube followed by centrifugation at 280xg for 5 min. The supernatant was tipped off and the pellets were re-suspended with

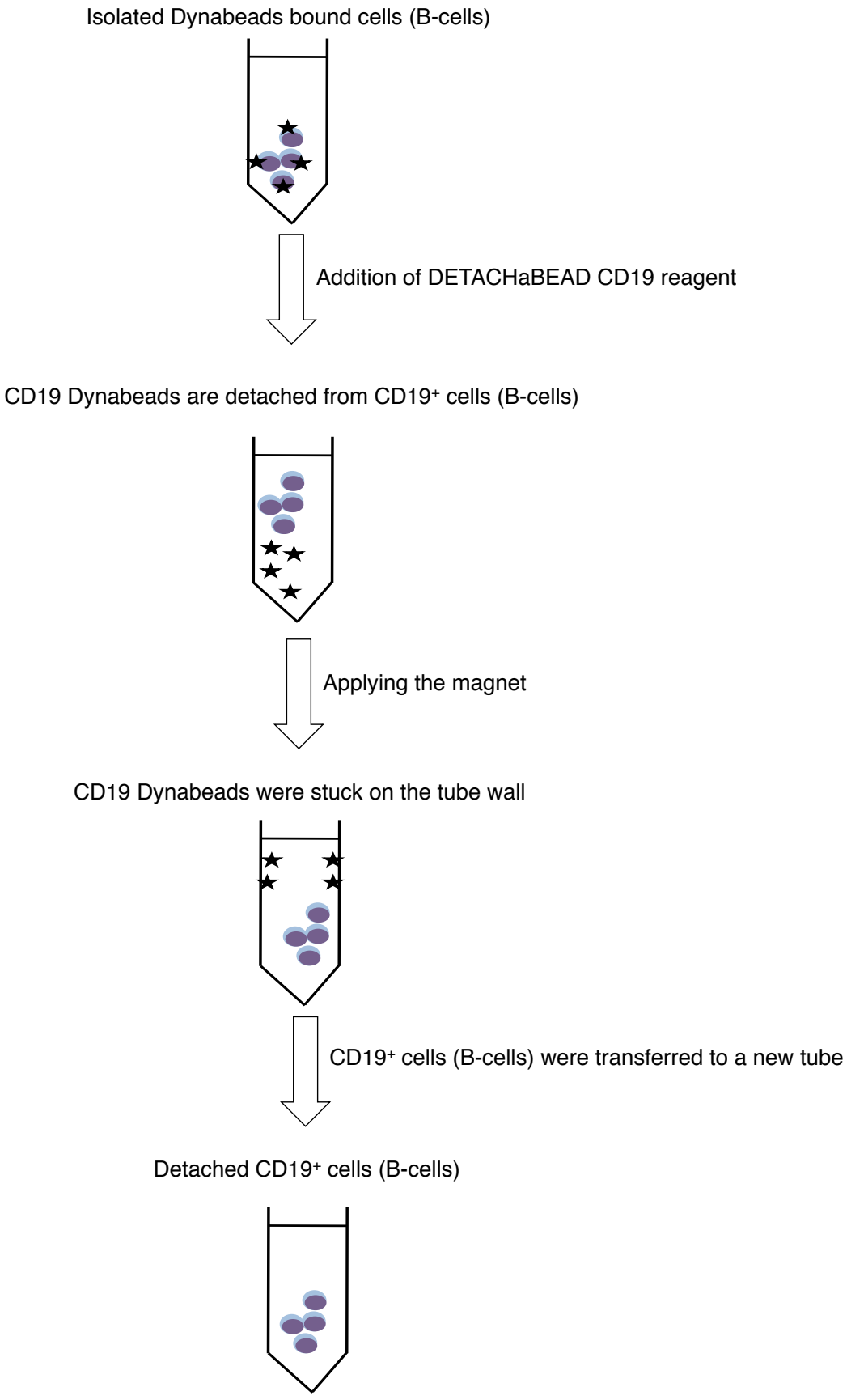


Figure 2.4: Schematic of detaching the magnet beads from positively isolated B-cells.

200µl 1% (w/v) paraformaldehyde. For each patient there were two tubes: a test tube containing the antibodies and a control tube containing no antibodies. Samples were analysed by an Accuri C6 cytometer. CD19 expression was measured through fluorescence channel 4, while CD38 expression was assessed through fluorescence channel 2.

2.2.2 Apoptosis assay

The viability of isolated cells was assessed as described by the Annexin V apoptosis assay kit (eBiosceience, BMS500FI/300CE). 5 x 10^5 cells from each sample were re-suspended with 200µl binding buffer (1X). For early apoptosis detection, 4µl Annexin V-FITC was added to each sample followed by gentle shaking and incubation at room temperature for 10 min. For late apoptosis detection 10µl Propidium Iodide (PI) was added to each sample prior to analysis by an Accuri C6 cytometer. Annexin V-FITC was measured through fluorescence channel 1, whereas PI was determined via fluorescence channel 2.

2.3 Cellular fractionation

Initially, a cellular fractionation method described by Brennan *et al*., (2008) was developed to generate three different protein fractions from primary CLL cells. Frozen CLL cells were thawed and then incubated with 0.5ml of hypotonic buffer (10mM triethylammonium bicarbonate, 1.5mM $MgCl_2$ and 10mM KCl) supplemented with 1mM phenylmethanesulfonylfluoride and 0.1% (v/v) Nonidet P40 detergent for 15 min on ice. Samples were then

centrifuged (10,000xg) for 5 min at 4°C. The supernatant was removed and denoted the “NP40 fraction”. It typically contained cytosolic proteins. The resulting pellet was washed twice with 1ml hypotonic buffer followed by centrifugation. The washed pellet was re-suspended in 0.2ml of High Salt Buffer (20mM triethylammonium bicarbonate, 420mM NaCl, 1.5mM $MgCl_2$, 25% glycerol (v/v)) supplemented with 1mM phenylmethanesulfonylfluoride for 30 min on ice. Samples were then centrifuged (10,000xg) for 5 min at 4°C. The supernatant was retrieved and labelled as the “High Salt Buffer (HSB) fraction”. The High Salt buffer fraction typically contained nuclear proteins, including some transcription factors. The resultant pellet was washed twice with 1 ml hypotonic buffer followed by centrifugation. As a final step the cellular pellet was then solubilised in 0.2ml 1% (w/v) Sodium dodecyl sulphate (SDS) solution, heated for 20 min at 90°C and then sonicated; this fraction was labelled as the “SDS fraction”. It also contained nuclear proteins and other relatively insoluble proteins. As part of the development of our fractionation method, the High Salt buffer step was subsequently omitted from the workflow.

2.4 Protein assay

Protein concentrations were determined using the instructions and reagents provided with the bicinchoninic acid protein assay kit (Sigma, BCA1 and B 9643). Using a 96-well plate, 25µl triplicates of a serial dilution of bovine serum albumin (BSA: 0.17-1.00 mg/ml) and 25µl duplicate of protein extracts from CLL cells were mixed with 200µl of the BCA working reagent (50

parts of reagent A with 1 part of reagent B). The plate was covered with a plastic sheet and incubated at 37°C for 30 min and read by an ELISA plate reader at 570nm. Then the average absorbance of each triplicate BSA serial dilution and sample duplicate was calculated. After this, a standard curve was generated by plotting the concentrations of the BSA against their absorbance at 570nm and was used to calculate the protein concentration of samples tested based on their absorbance.

2.5 Protein detection

2.5.1 One dimensional-electrophoresis (1DE)

Separation of proteins by 1DE was performed using the instructions and reagents of the NuPAGE electrophoresis system (Invitrogen, IM-1001). Protein extracted from CLL cells was mixed with NuPAGE LDS (Lithium Dodecyl Sulfate) sample buffer (4X) and NuPAGE reducing agent (10X) using the following formula: 6.5µl protein + 2.5µl NuPAGE LDS sample buffer (4X) + 1µl NuPAGE reducing agent (10X). The mixture was heated at 70°C for 10 min and spun for 30 seconds. An XCell SureLock Mini-Gel (Invitrogen: EI0001) was set up followed by loading the prepared samples into pre-made NuPAGE 4-12% Bistros Zoom gels (Invitrogen, NP0321BOX). Samples were separated using NuPAGE MOPS (3-(N-morpholino) propanesulfonic acid) SDS running buffer for 55 min at a constant 200 Volts. Pre-stained molecular weight markers (Seeblue, Invitrogen: LC5925) were also resolved in the same gel to indicate the size of the separated proteins. Some samples were resolved on a gel for protein visualisation by colloidal

Coomassie blue stain, while other samples were separated in a gel for specific protein detection using western blotting and antibody detection.

2.5.2 Staining with colloidal coomassie blue

Extracted proteins from CLL cells that were resolved in a gel were visualised using the instructions and reagents provided in the colloidal coomassie Bbue staining kit (Invitrogen: LC6025). The gel was fixed in a fixing solution (40ml deionised water, 50ml methanol and 10ml acetic acid) for 10 min on a shaker at room temperature. Then the gel was transferred and immersed in staining solution (55ml deionised water, 20ml methanol and 20ml strainer A) for 10 min on a shaker at room temperature. Then 5ml strainer B was added to the staining solution. The gel was left overnight in the staining solution on a shaker at room temperature. After that, the gel was washed with 200ml deionized water until the background of the gel was clear. The gel was scanned for further analysis.

2.5.3 Western blotting and antibody detection

Using the instructions and the equipment provided with the XCell II Blot module (Invitrogen, EI9051), proteins resolved in a gel were transferred to Polyvinylidene fluoride membrane (PVDF, GE Healthcare: RPN303FP) in NuPAGE transfer buffer (1X) (Invitrogen: NP0006) with 10% (v/v) methanol by electroblotting for 90 min at 30 V and 170 A. Upon completion of the transfer step, the membrane was quickly washed with PBS-Tween20 and blocked with a blocking buffer (0.2% (w/v) I-Block, 0.1% (v/v) Tween-20 and 0.4% (w/v)

sodium azide) for 2 hours on a shaker at room temperature. The blocked membrane was incubated with diluted primary antibodies of interest overnight on a shaker at 4°C. Next, the membrane was washed three times with PBS-Tween20 for 15 min on a shaker at room temperature. Then the membrane was incubated with diluted secondary antibody for one hour on a shaker at room temperature. The membrane was then washed 3 times with PBS-Tween20 as before and was incubated with Alkaline Phosphatase (AP) buffer for 5 minutes. After that, the membrane was placed on a plastic sheet and was covered with 500µl chemiluminscence substrate (TROPIX, CDP-Star-Manufacture). Then the membrane was exposed to light sensitive photographic film (Kodak Scientific Imaging Film) for different periods of time. Finally films were developed using a developer machine. The developed film was then scanned for further analysis.

2.5.4 Antibodies

Antibodies against nine proteins were purchased and were used as shown in Table 2.1. Secondary antibodies were used at dilution of 1/10000.

Table 1.2: The primary antibodies that were used for protein detection in this project.

Protein	Concentration or dilution	Catalogue Number	Company
Poly-(ADP)ribose polymerase	200ng/ml	sc-7150	Santa Cruz Biotechnology

Protein S100A8	1/1000 dilution	sc-7150	Santa Cruz Biotechnology
Tubulin	1/1000 dilution	T-9026	Sigma
Actin*	1/1000 dilution	A-2066	Sigma
Myosin-9	1/1000 dilution	10341-2B3	Sigma
T-cell leukaemia/lymphoma protein 1A	1/1000 dilution	39-4800	Invitrogen
Thyroid hormone receptor-associated protein 3*	1/1000 dilution	ab71985	Abcam
Heterochromatin protein 1-binding protein 3*	1/1000 dilution	ab98894	Abcam
Histone H4	0.5/1000 dilution	ab31830	Abcam

The antibodies shown in this table were either mouse or rabit monoclonal antibodies. Antibody labled with (*) were polyclonal antibodies. Dilution 1/1000: 1μg/ml

2.6 Proteomics analysis

2.6.1 Preparation of reference samples for quantitative proteomics analysis

An NP40 fraction of one CLL sample was used as a reference sample for the relative quantification of proteins in the NP40 fractions of the examined CLL samples. In the course of optimising the quantitative proteomics workflow, the SDS reference sample was prepared by mixing an

equivalent amount of protein (20μg) from the SDS fraction of the 12 CLL samples (six poor prognosis CLL samples and six good prognosis CLL samples). This SDS reference sample was used to assess the relative quantification of proteins in the SDS fractions of the CLL samples.

2.6.2 Protein precipitation

Protein extracted from CLL cells (NP40 and SDS fractions) were precipitated separately using the method and reagents provided in the 2D Clean Up kit (GE Healthcare). Equivalent amounts of protein (20μg) from CLL samples were transferred to new 1.5ml tubes and mixed with 300µl precipitant and incubated on ice for 15 min. Then 300µl co-precipitant was added to each sample followed by mixing and centrifugation at 12000xg for 5 min. The supernatant was removed and the pellet was covered with 40µl co-precipitant and incubated on ice for 5 min. Then samples were centrifuged as before and the supernatant was discarded. The pellet in each sample was then mixed and vortexed with 25µl deionised water. Then 1ml pre-chilled wash buffer and 5µl wash adaptive were added to each sample. Samples were incubated in ice for 30 min and vortexed every 10 min. After this, samples were centrifuged as before and the supernatant was removed and the protein precipitant was re-suspended in 20μml of 20mM triethylammonium bicarbonate (TEAB) and kept for downstream experiments.

2.6.3 Protein digestion and iTRAQ labelling

Following protein precipitation, proteins were digested with trypsin (for qualitative proteomics analyses) or trypsin digested and labelled with iTRAQ reagents (for quantitative proteomics analyses) as described in the Applied Biosystems iTRAQ reagent multiplex kit. Samples were spun at 12000xg for 5 min and the supernatant was removed. Protein precipitant was resuspended in 20µl of 0.5mM TEAB, denaturated with 1µl 2% (w/v) SDS and reduced with 2µl 50mM tris-(2- carboxyethyl) phosphate (TCEP) at 60°C for 1 hour. Then proteins were alkylated with 1µl 200mM methyl methane-thiosulfonate (MMTS) in isopropanol for 10 min at room temperature. For protein digestion, samples were incubated with 2µl of sequencing grade porcine trypsin (Promega) overnight at 37°C. For iTRAQ labelling, 70µl ethanol was added to each iTRAQ vial and mixed for 1 min. The content of each iTRAQ vial was transferred to one sample. Samples were then incubated for 1 hour at room temperature. iTRAQ-labelled samples were combined in a new 1.5ml tube and dried down using a vacuum centrifuge dryer. Next, the combined sample was re-suspended in 60µl LC loading buffer; 2% (v/v) acetonitrile ACN in water with 0.05% (v/v) trifluoroacetic acid TFA and an aliquot (15µl) was immediately subjected to 2D nano-LC separation while the remaining (45µl) was stored at –20°C.

2.6.4 Liquid chromatography (LC) separation of digested proteins

Separation of peptides by LC was performed using a two-dimensional salt plug method that was previously described by (Brennan et al., 2009, Welton et al., 2010). Following protein digestion and labelling with iTRAQ, 15µl of the prepared samples (containing approximately 20µg of digested proteins) were loaded into the 2D nano-LC system (UltiMate 30000, Dionex). The machine was set to separate 10µl (approximately 13.3µg of digested proteins). Peptides were firstly separated on a strong cation exchange column (Bio-SCX, 500mm, 15mm, 5mm, Dionex) in a step-wise elution using increasing concentrations of NaCl (breakthrough, 100mM, 200mM, 400mM, 800mM, and 1M). Each salt fraction was desalted on a reverse phase desalting column and subsequently further separated on reverse phase column (PepMap 75mm id, 15cm, 3mm, 100Å, Dionex) at a flow rate of 300nl/ min. Two buffers were used A = 2% (v/v) CAN (Sigma) in water with 0.05% (v/v) TFA (sigma) and B = 90% (v/v) ACN in water with 0.01% (v/v) TFA. Separation of peptides was performed using a three-step gradient; the first step was from 5 to 20% solvent B for the first 34 min, the second step was from 20% to 50% solvent B for 21 min and the third step was from 50% to 90% solvent B for 4 min. A chromatogram was recorded at 214nm. Separated peptides were spotted onto an LC-MALDI plate (one spot/8 sec) using a Probot microfraction collector (Dionex). The MALDI matrix α-Cyano-4-hydroxycinnamic acid (CHCA) (Sigma) was used (2mg/ml in 70% (v/v) ACN in 0.1% (v/v) TFA containing 10fmol/ml Glu-Fib (Sigma)) which was

continuously added to the column effluent via a μ-tee mixing piece at a flow rate of 1.4 ml/min. For each salt fraction there were 255 eluted spots starting from the 25th min until the 59th min. The Probot microfraction collector left gaps between each salt fraction. There were 109 empty spots between the eluted spots of the breakthrough and the eluted spots of the 100mM salt fraction. Also upon completion of the 100mM salt run the Probot left 5 empty spots and started eluting spots of the next salt fraction. This was the case with the rest of the salt fractions. In total there were 1530 spots eluted on the MALDI plate from the sample. The plate also contained an additional 8 calibration spots (matrix, Cal Mix 1 and Cal Mix 2, the volume for each spot was 0.5μl). The nano-LC machine was programmed and operated by Dr. Sanjay Khanna.

2.6.5 Mass Spectrometry

The mass spectrometry analyses were conducted using a previously described method (Brennan et al., 2009, Welton et al., 2010). Separated peptides on the LC-MALDI plate were subjected to mass spectrometry for MS and MS/MS analysis. MS was performed using an Applied Biosystems 4800 MALDI TOF/TOF mass spectrometer with a 200Hz solid-state laser operating at a wavelength of 355nm. After screening of all LC-MALDI sample positions in MS positive reflector mode using 800 laser shots (mass range 700–3000Da; focus mass 1400), the fragmentation of up to six automatically selected precursors was performed (the most intense ion signals per spot position with signal to noise above 30 and the strongest was

analysed first). Internal calibration of each spot in the MS was achieved against the Glu-Fib added to the matrix. Common trypsin autolysis peaks and matrix ion signals and precursors within 300 resolution of each other were excluded from the selection. In MS/MS positive ion mode 4000 spectra were averaged with 1 kV collision energy (collision gas was air at a pressure of 1.661026 Torr) and default calibration. The MALDI TOF/TOF MS machine was operated by Dr. Keith W Hart.

Chapter Three

Preparation of primary CLL samples for Proteomics Analysis

3.1 Introduction

To perform a good quality proteomics study, good quality samples and sample preparation should be utilized. The aim of this chapter was to select CLL samples and develop fractionation method to generate different protein fractions from CLL samples for proteomics analysis. This chapter explains the criteria for selecting good quality CLL samples as well as the processes of isolating CLL cells from the peripheral blood of CLL patients. In addition, it describes the processes of developing a good quality and reproducible cellular fractionation method that solubilises cellular proteins and leaves no obvious cellular debris.

3.1.1 Staining with annexin V and propidium iodide

The phospholipid phosphatidylserine (PS) is normally located on the inner side of eukaryotic plasma membrane. However, in the early stage of apoptosis PS translocates to the outer leaflet of plasma membrane. PS acts as a flag for phagocytes to engulf apoptotic cells (Martin et al., 1995). Annexin V is a ligand for PS so when annexin V is labelled with a fluorescent molecule such as fluorescein isothiocyanate (FITC), it can be used to detect early apoptosis. Dead cells or cells in a late stage of apoptosis lose the integrity of their plasma membrane allowing a DNA binding dye such as propidium iodide (PI) to enter the cell and stain DNA (Ormerod et al., 1993). As a result, mixing cells with annexin V and propidium iodide followed by flow cytometry analysis can be a powerful tool to discriminate between viable cells and cells undergoing early apoptosis or late apoptosis/necrosis.

3.1.2 CD38 expression discriminates between CLL patients with bad or good prognosis

The outcome of CLL is heterogeneous; it can be an aggressive disease with rapid progression and short survival time in some patients and stable with long survival time in other patients (Hamblin et al, 1999, Rassenti et al., 2004, Damle et al., 1999). There are numerous prognostic markers that can be used to help predict the outcome of CLL. One example of these is the surface antigen CD38 that is expressed at low levels on the CLL cells of patients with a stable form of CLL and longer patient survival. In contrast, it is often more highly expressed on the CLL cells of patients with an aggressive form of the disease (Durig et al., 2002, Pepper et al., 2012).

3.2 Results

3.2.1 CLL cells

To separate CLL cells, Ficoll was used to isolate the layer of mononuclear cells containing CLL cells. Although the low-density layer of mononuclear cells contains cells other than lymphocytes such as monocytes, CLL cells usually predominate since CLL patients often have a profound lymphocytosis.

3.2.1.1 Purity of CLL cells for proteomics analysis

For accurate proteomics analysis, samples must contain a high purity of CLL cells. Although all samples were taken from patients diagnosed with CLL, not all the CLL samples collected contained a sufficiently high

percentage of CLL cells to warrant inclusion in this study. To determine the percentage of CLL cells, cells were stained with an anti-CD19 antibody conjugated to allophycocyanin (APC) followed by flow cytometric analysis. A gate was drawn around the lymphocyte population as shown in Figure 3.1A and CD19 expression was measured on the gated cells using fluorescence channel FL4 as demonstrated in Figure 3.1B and 3.1C. If samples contained less than 95% CD19 positive B-cells, for example the sample shown in Figure 3.1C, then negative isolation of CLL cells by depleting T-cells using magnetic beads was performed. Figure 3.1D shows an example of the resulting increase in purity of CLL cells (the percentage of CD19 positive cells increased from 92.3 to 96.5). Only samples with 95% or greater of CD19 positive B-cells were used for proteomics analyses.

3.2.1.2 Viability of CLL cells

After being satisfied with the purity of CLL samples, the next criterion to investigate was the viability of CLL cells. In an attempt to eliminate irrelevant factors that could contribute to differences in the proteome of the poor prognosis and good prognosis CLL samples, all samples utilized had to show high cell viability. CLL samples vary in terms of their viability from patient to patient. To assess viability, cells were stained with annexin V and propidium iodide followed by flow cytometric analysis. A large gate was drawn around the lymphocyte population and annexin V on gated cells was measured in fluorescence channel FL1, while PI in gated cells was measured in fluorescence channel FL2. All CLL samples that were used for proteomics analyses were ≥95% viable (Figure 3.2).

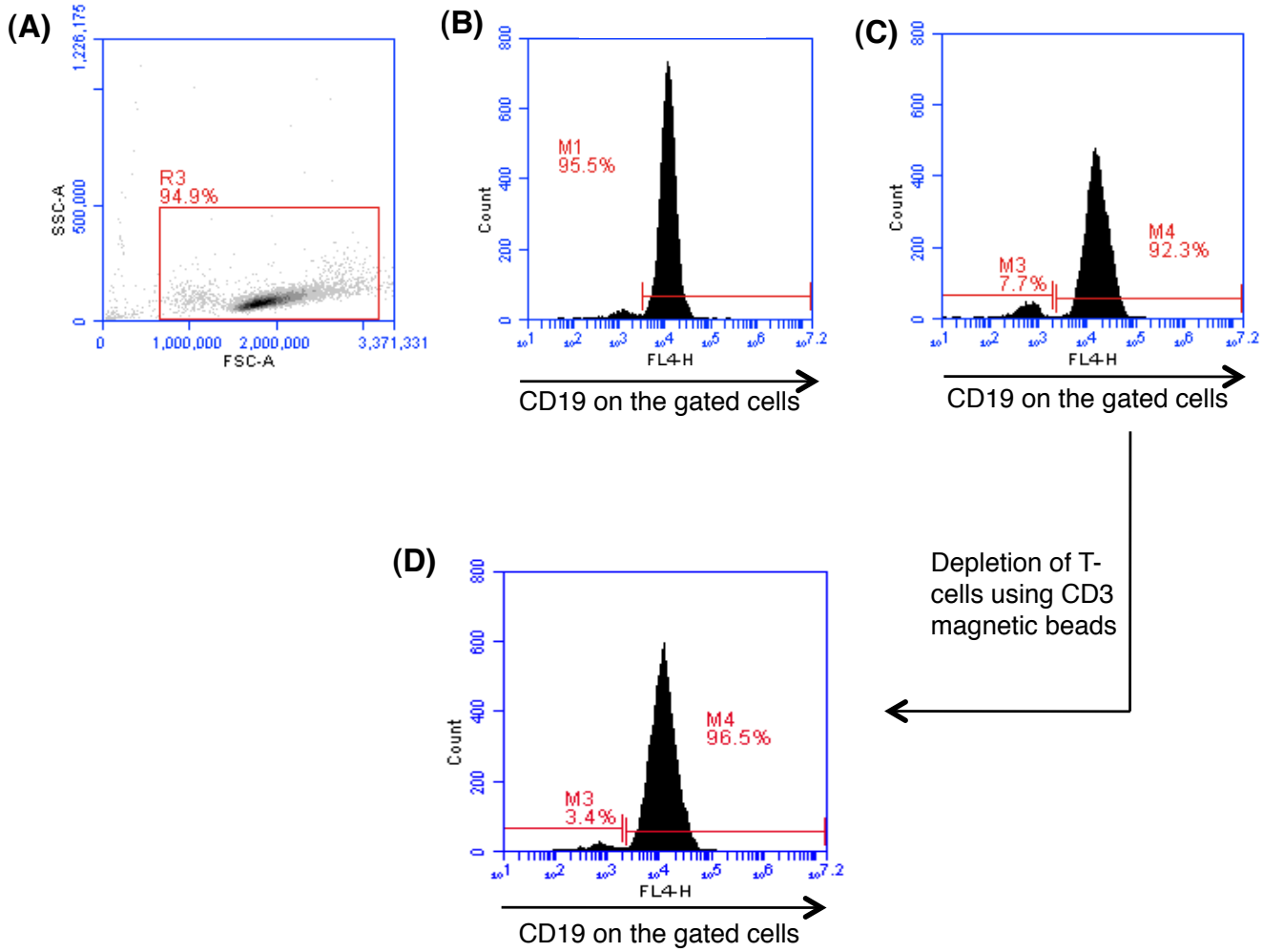


Figure 3.1: Purity of CLL samples. Following staining of CLL cells with an anti CD19 antibody, cells were analysed by flow cytomery. A gate was drawn around lymphocytes which appeared on the screen of side scatter and foreward scatter (A). Percent of CD19 expression was assessed on the gated cells (B and C). A sample with 92.3% CD19 positive cells was subjected to negative isolation by depleting T cells using magnetic beads. Negatively isolated cells were stained with an anti CD19 antibody and analysed by flow cytometry. Percent of CD19 positive cells increased from 92.3 to 96.5 in the negatively isolated cells (D).

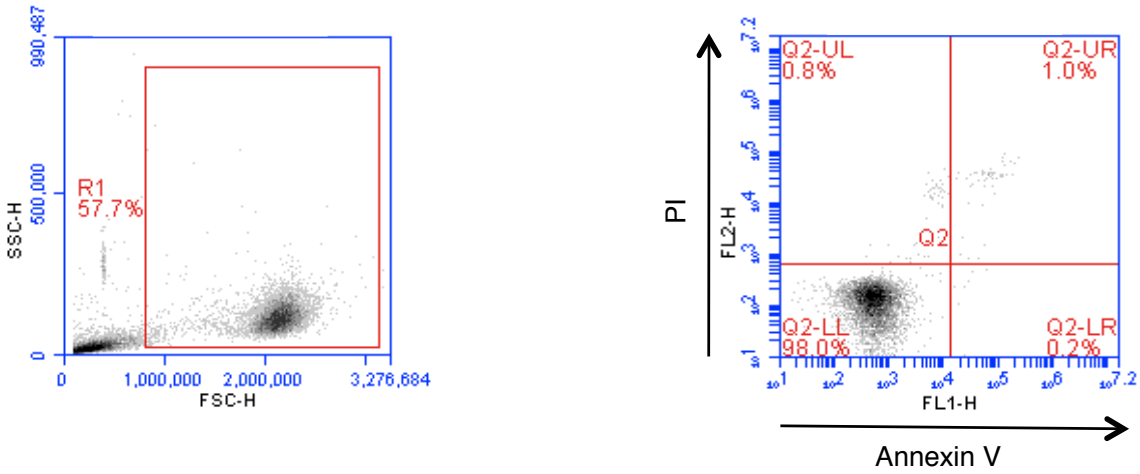


Figure 3.2: An example of the viability of CLL cells that were used for proteomics analyses. Following isolation of CLL cells from peripheral blood samples of CLL patients, cells were stained with with annexin V and PI and analysed by flow cytometry. As samples at this stage had high purity (≥ 95% CD19 positive cells) a large gate was drawn around all event (dots) excluding debris (A). Annexin V and PI were then measured on the gated cells. All CLL samples that were used for proteomics analysis were 95% or more viable. This figure shows that this sample had 98% viable cells.

3.2.1.3 Two groups of CLL samples were collected

One of the main questions of this PhD project was: what is the difference between the favourable form and the unfavourable form of CLL with regard to protein expression? CD38 expression was used to discriminate between the two forms of CLL. To determine the percentage of CD38 expression on CLL cells, CLL samples were stained with an APC-labelled anti-CD19 antibody and a phycoerythrin (PE)-labelled anti-CD38 antibody followed by flow cytometric analysis. Normally T-cells express CD38, so to avoid the false result that could be generated from measuring CD38 expression on T-cells plus B-cells, CD38 expression was measured only on the B-cells in CLL patient samples. This was achieved by firstly measuring CD19 expression on the lymphocyte population using fluorescence channel FL4. Then CD38 expression was specifically measured on CD19 positive cells in channel FL2. Two groups of CLL samples were collected: $CD38^-$ CLL samples (CD38 expression ≤5%) representing the good prognosis form of CLL; and $CD38^+$ CLL (CD38 expression ≥40%), representing the aggressive form of CLL. Figure 3.3 shows an example of two different CLL samples.

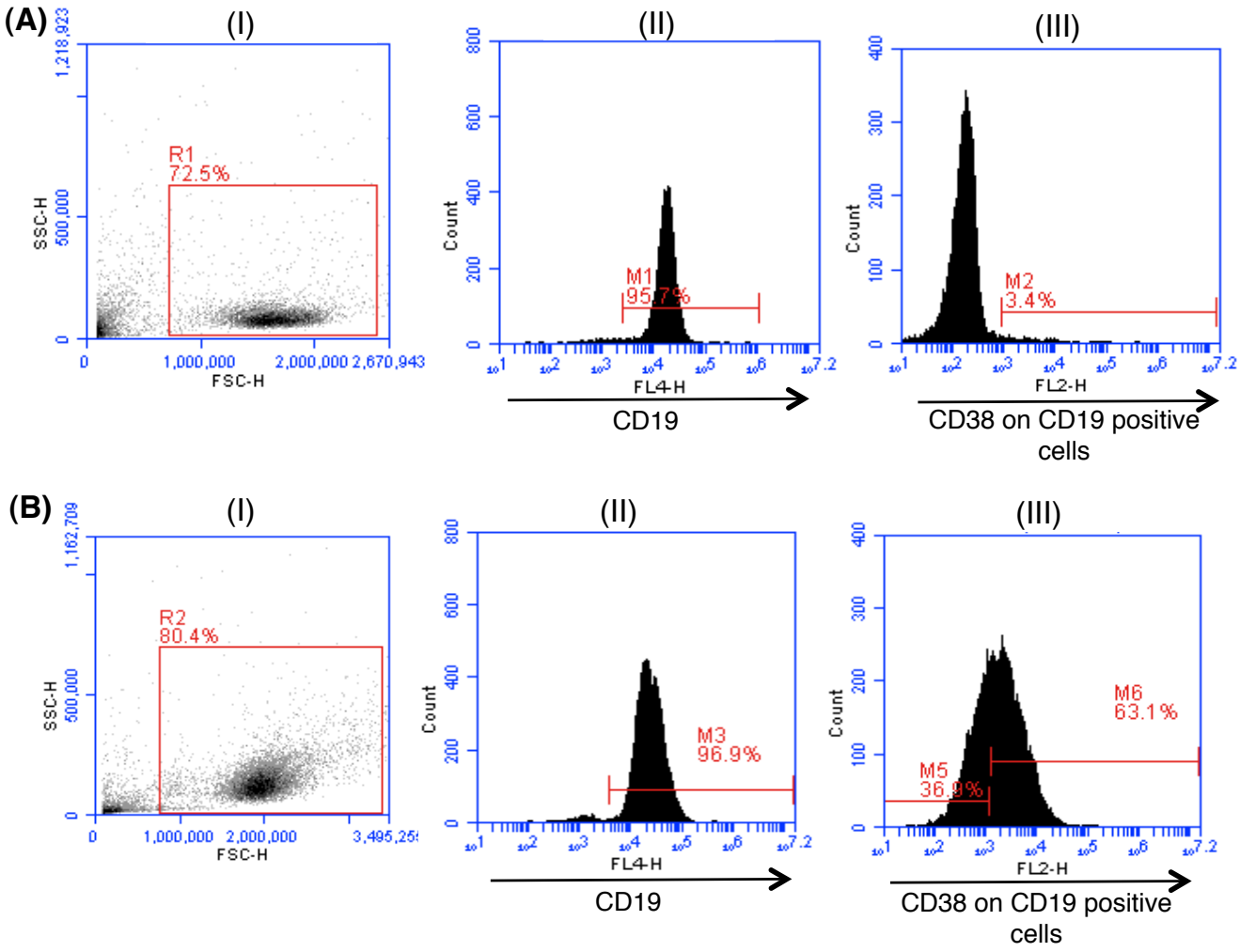


Figure 3.3: CD38 expression on CLL samples. Two groups of CLL samples were collected; CLL samples with low CD38 expression (≤ 5%) and CLL samples with high expression of CD38 (≥ 40%). CLL cells were stained with an anti CD19 antibody and an anti CD38 antibody followed by flow cytometric analysis. Firstly, CD19 expression was determined (A and B; II) and then CD38 expression was measured only on CD19 positive cells (B-CLL cells) (A and B; III). Both samples were > 95% CD19 positive cells (B cells). Nevertheless, one sample had low CD38 expression(≤ 5%) (A; III) and was thus considered $CD38^-$ CLL sample. The other sample had high CD38 expression (≥ 40%) (B; III) and was thus considered $CD38^+$ CLL sample.

3.2.2 Cellular fractionation of CLL samples

One of the most challenging issues in proteomics analysis is the difficulty associated with identifying low abundant proteins, which are masked by high abundance proteins. Therefore, reducing sample complexity is a key to successful proteomics studies. One way of reducing sample complexity is by performing fractionation of the sample (Ahmed, 2009, Huber et al., 2003, Yates et al., 2005).

3.2.2.1 Development of three-step cellular fractionation method

A two-step cellular fractionation procedure using 0.1% (w/v) NP40 to isolate the cytosolic fraction followed by high salt buffer (HSB) to extract nuclear proteins from B-cells has been described but this approach left an obvious pellet of cellular debris (Brennan et al., 2009). Given that the pellet remaining might contain interesting proteins, in this study the pellet was further solubilised with 1% (w/v) SDS enhanced with heating and this was sufficient to completely solubilise the pellet. Figure 3.4 shows a schematic for how the three-step cellular fractionation was carried out.

To evaluate the importance of the SDS fraction in terms of protein content, the bicinchoninic acid protein BCA assay was performed to determine the amount of protein in the three fractions: NP40, high salt buffer (HSB) and SDS. The average content of protein from $1x10^7$ CLL cells in the three fractions was NP40: 250 ± 27 µg (65%); HSB: 21 ± 5 µg (5%); SDS: 115 ± 37 µg (30%) (mean ± SD for six individual samples; Figures 3.5).

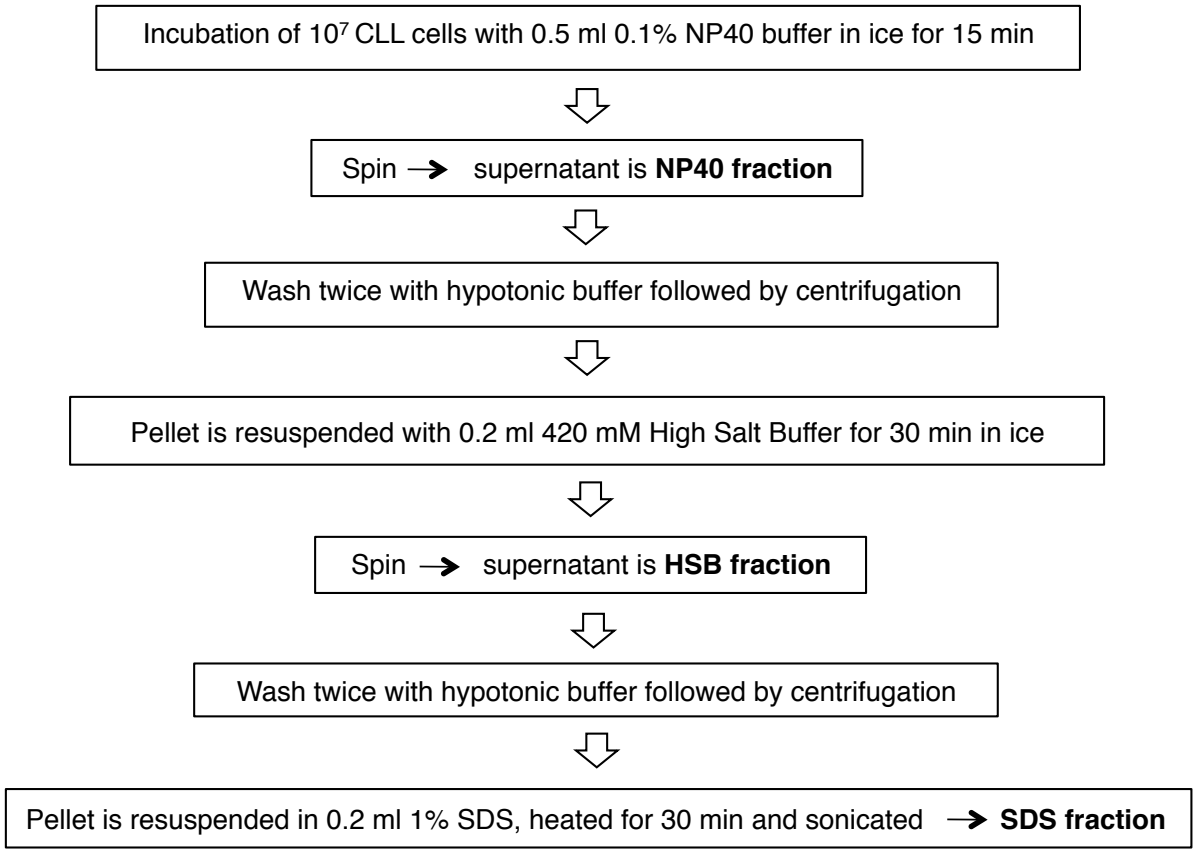


Figure 3.4: Schematic workflow of the three-step cellular fractionation.

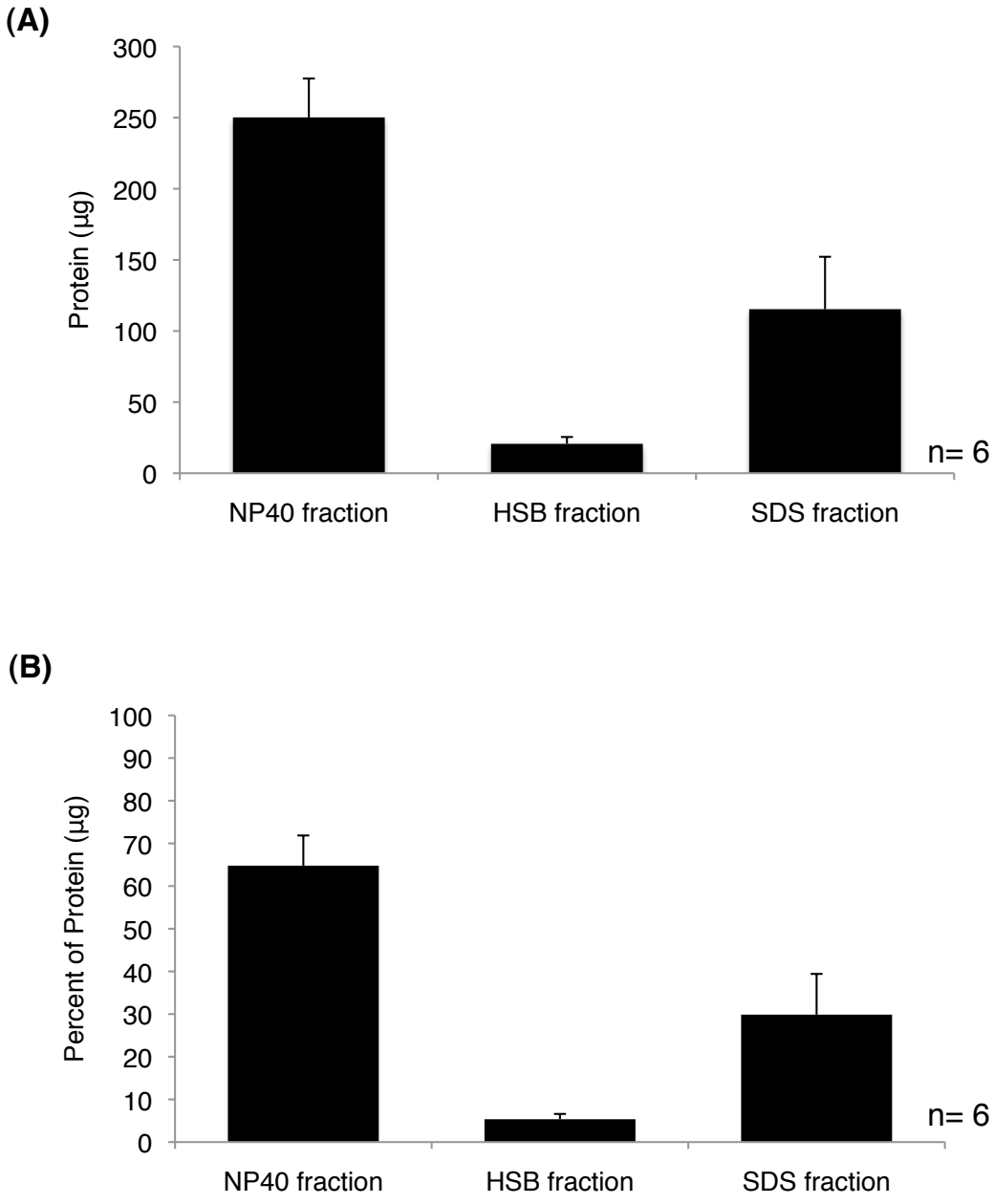


Figure 3.5: Protein yield in the three fractions (NP40, HSB and SDS). Three fractions were generated from primary CLL cells (NP40, HSB and SDS). Protein concentration in each fraction was determined using BCA assay (A and B).

Interestingly, this analysis demonstrated that the SDS fractionation contained almost a third of the total cellular protein.

3.2.2.2 Quality of the three-step cellular fractionation method

The rationale behind developing this cellular fractionation method was to extract relatively different proteins in each fraction, in order to reduce the complexity of the proteome in each fraction and hopefully identify more proteins in the subsequent proteomics analyses. To evaluate the quality of the cellular fractionation method, an equivalent amount of proteins from each of the three fractions was resolved by SDS-PAGE followed by colloidal Coomassie Blue staining (Figure 3.6). This analysis showed that all three fractions contained proteins across the full range of molecular weights. Focussing on the lower molecular weights, the pattern of the three fractions was relatively different suggesting different proteins were extracted in each fraction.

3.2.2.3 Reproducibility of three-steps cellular fractionation method

To investigate the variation from cell extraction to cell extraction, three groups of protein bands were selected to perform densitometric analyses of four different CLL samples. This was performed using ImageJ software version 1.44o. The three groups of bands included bands present in all of the three fractions, bands found only in two fractions and bands that were predominantly present in only one fraction (Figure 3.7A I, II, III). The

small error bars indicate that the banding pattern was consistent suggesting good reproducibility of the cellular fractionation (Figure 3.7B I, II, III).

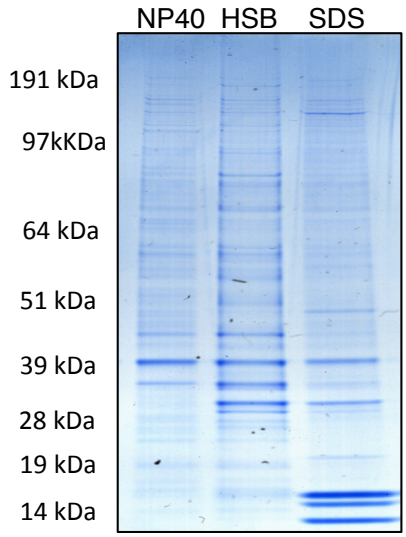


Figure 3.6: Visualisation of the proteins in the three fractions (NP40, HSB and SDS). 1.5 μg protein from the NP40 fraction, HSB fraction and SDS fraction were respectively loaded onto a 4-12% gradient polyacrylamide gel. The gel was then stained with colloidal Coomassie blue. The pattern of the three fractions was relatively different. This suggested that different proteins were extracted in each fraction.

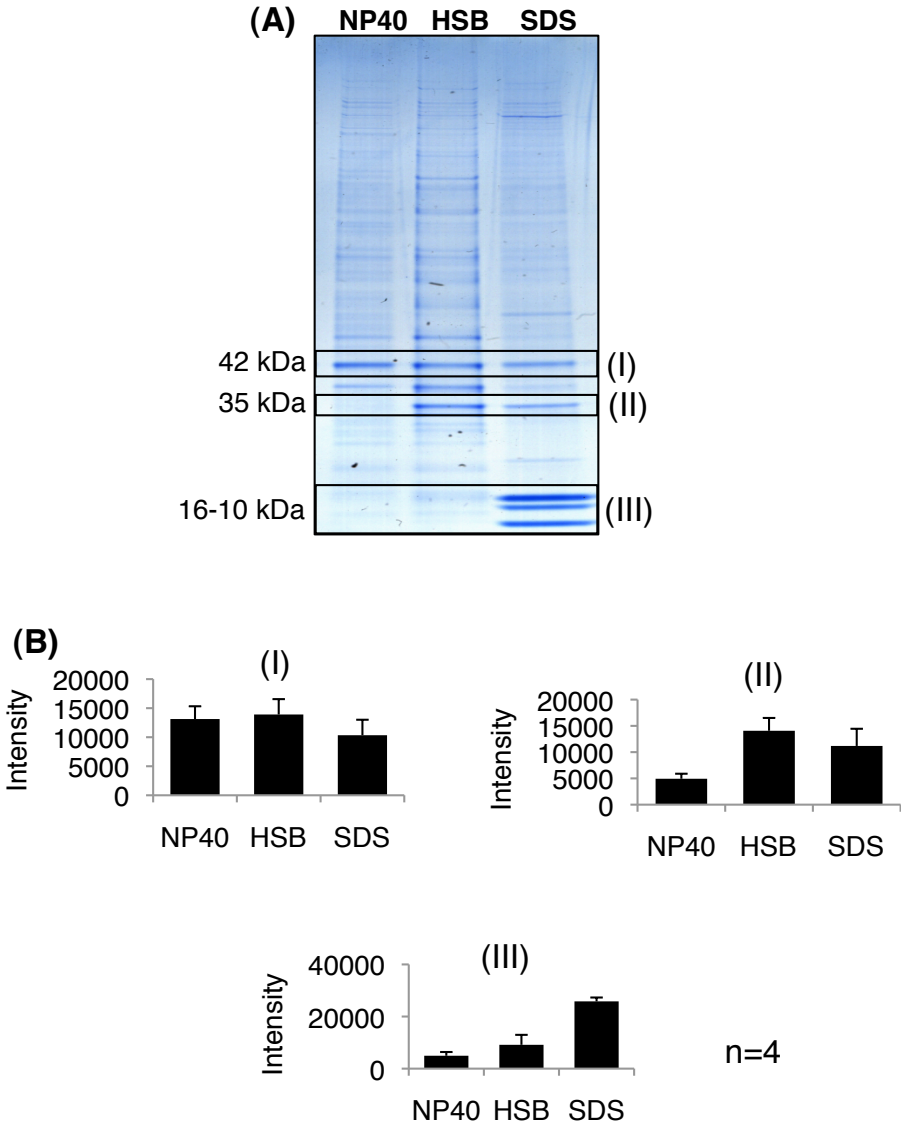


Figure 3.7: Reproducibility of the three-step cellular fractionation method. Three groups of protein bands were chosen for densitometric analysis. These bands include bands present in all three fractions (A-I), bands appeared only in two fractions (A-II) bands found predominantly in only one fraction (A-III). These analyses were performed on four different CLL samples (B-I, II, III). The similar intensity of the measured bands in four different CLL samples (as indicated by the small error bars) suggested that this method was reproducible.

The BCA assay and the analysis by SDS-PAGE followed by colloidal Coomassie staining demonstrated that the SDS fraction is worthy of including for proteomics analysis. In this regard, the SDS fraction contained 30% of the total protein recovered. Furthermore, the SDS fraction contained proteins with low molecular sizes (approximately 10-16 kDa) that were not extracted in the NP40 fraction or in the HSB fraction.

3.2.2.4 Purity of the three fractions

To check the purity of the fractions, equivalent amounts of protein from each fraction were used to carry out SDS-PAGE followed by western blotting and antibody detection. Specific protein detection was carried out for poly(ADP-ribose) polymerase (PARP), as a nuclear protein marker, and tubulin, as a cytosolic protein marker (Brennan et al., 2009). Figure 3.8 showed that PARP was not found in the NP40 fraction suggesting that the NP40 fraction was not contaminated by the HSB fraction or the SDS fraction. Likewise, tubulin was not present in the HSB or the SDS fractions, which indicates that they were not contaminated by the NP40 fraction. However, PARP was found in the HSB and SDS fractions suggesting an overlap between HSB and SDS fractions.

3.2.2.5 Development of two-step cellular fractionation method

Following the analysis described in 3.2.2.4, it was decided not to continue with the HSB step but rather to combine the HSB and SDS fractions to give a two-step cellular fractionation method for two reasons. The first was

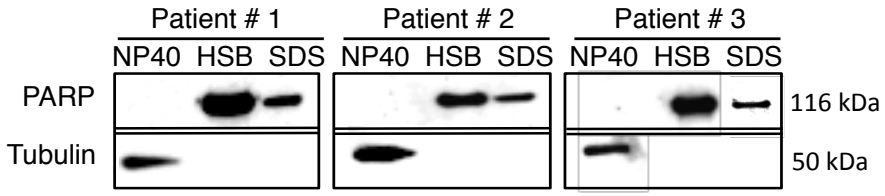


Figure 3.8: Analysis of PARP and tubulin in NP40 fraction, HSB fraction and SDS fraction. 2µg of protein from each fraction were separated by SDS-PAGE followed by western blotting and antibody detection using specific antibody to poly (ADP-ribose) polymerase (PARP), a nuclear protein, and an antibody to tubulin, a cytosolic protein. The analysis suggested that there was no cross contamination between NP40 fraction and other two fractions. However, HSB and SDS fractions found to have different amount of PARP. This analysis shows the result driven from three different CLL samples.

that the HSB fraction contained only 5% of the total protein. Secondly, there was an obvious overlap between the HSB fraction and the SDS fraction as represented by the presence of PARP in both fractions. Figure 3.9 illustrates how this method was performed to generate two different fractions, an NP40 fraction and an SDS fraction.

The protein yield in the NP40 fraction and the SDS fraction that were generated using the two-step cellular fractionation method was determined using the BCA assay. The average content of protein from $1x10^7$ CLL cells in the two fractions was: NP40: 250 ± 44 µg (65%) and SDS: 131 ± 19 µg (35%) (data represent the mean ± SD for 20 samples; Figure 3.10).

3.2.2.6 Quality of two-step cellular fractionation method

To evaluate the quality of the two-step cellular fractionation method, equal amounts of protein from the NP40 and SDS fractions were separated by SDS-PAGE followed by colloidal Coomassie Blue staining. This analysis demonstrated that the pattern of the two fractions was relatively different suggesting different proteins were extracted in each fraction. Figure 3.11 shows analysis by SDS-PAGE/ colloidal Coomassie Blue staining of the NP40 fraction and SDS fraction of three different CLL samples.

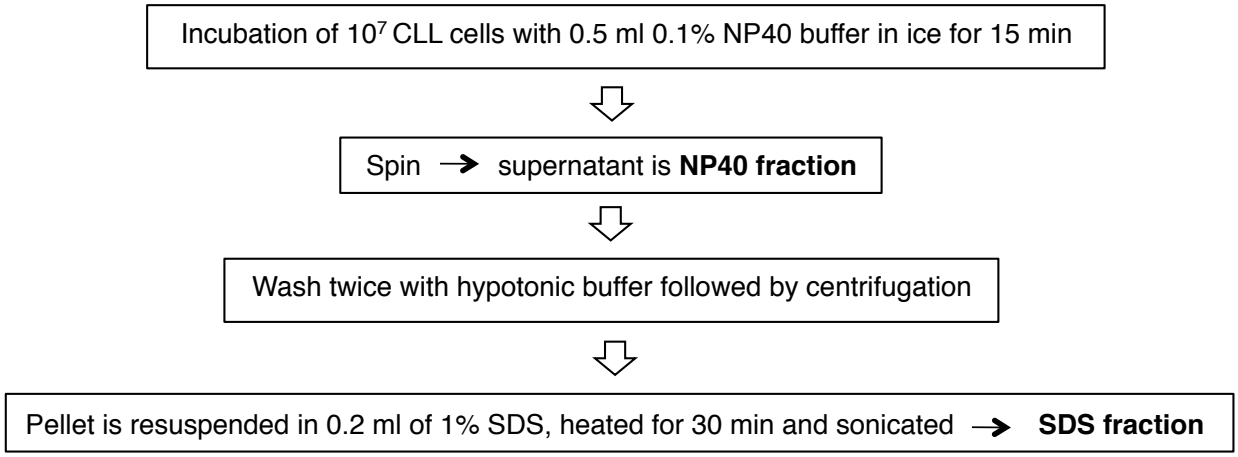


Figure 3.9: Schematic workflow of the two step cellular fractionation method.

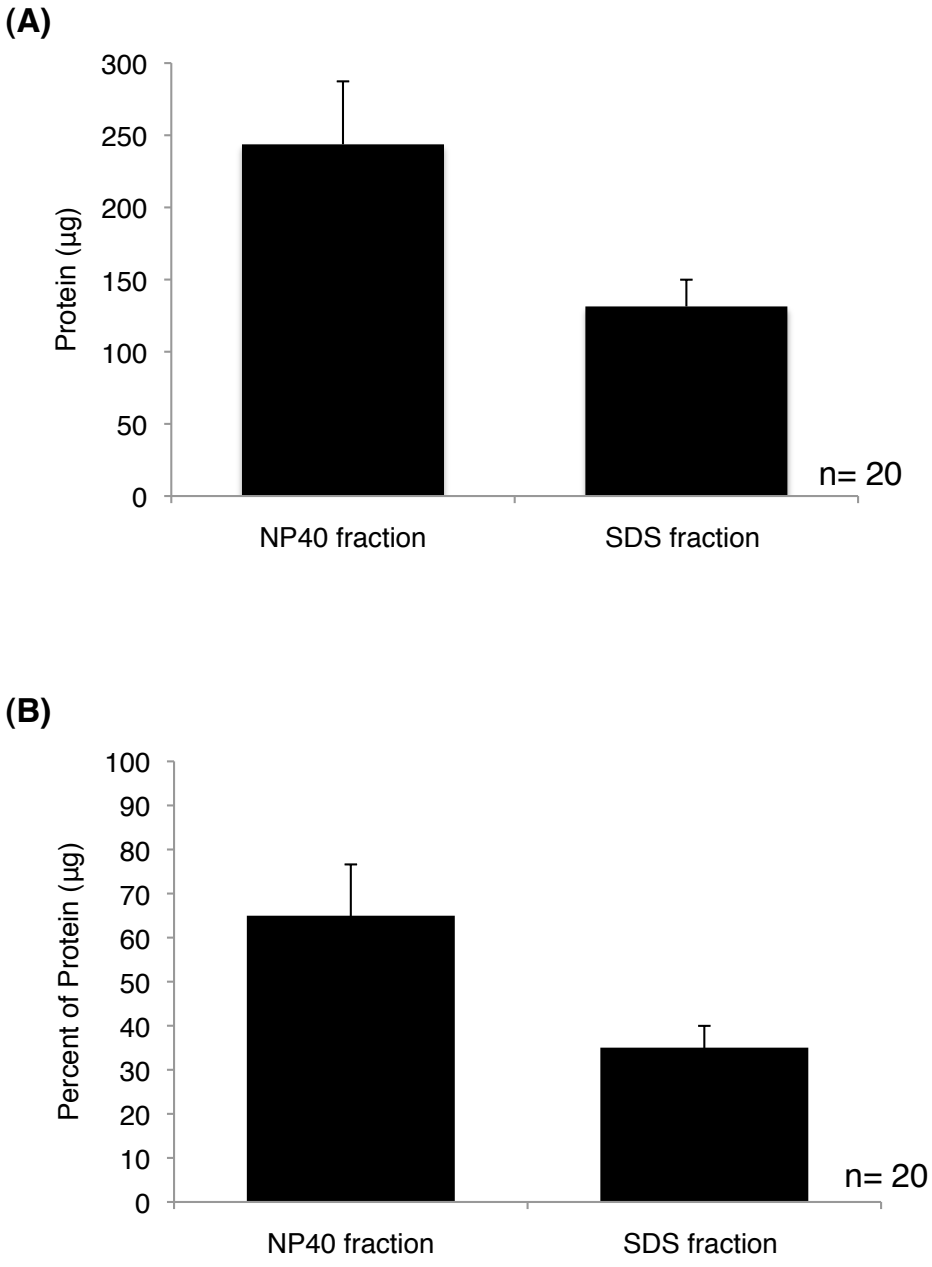


Figure 3.10: Protein yield from the two fractions (NP40 and SDS). Two fractions were generated from primary CLL cells (NP40 and SDS). Protein amount in each fraction was determined using BCA assay (A and B).

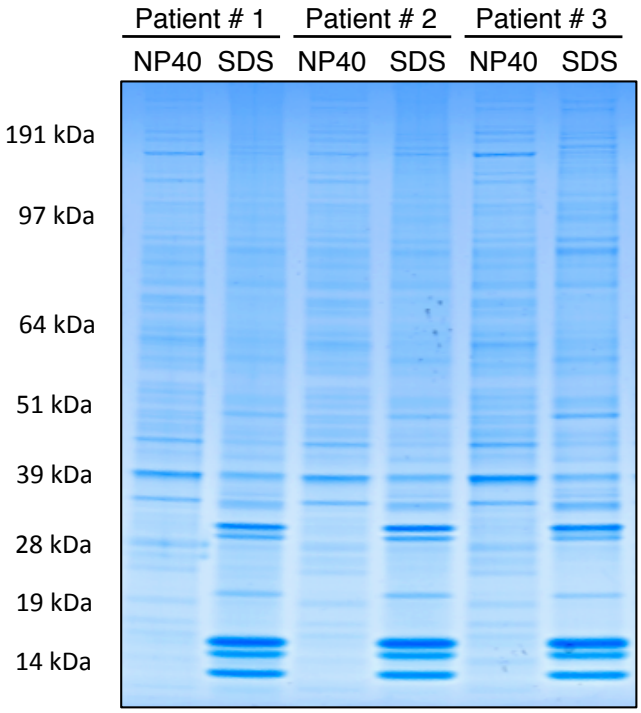


Figure 3.11: Visualisation of the proteins from the NP40 fraction and the SDS fraction. 5 μg protein of NP40 and SDS fractions were respectively loaded onto a 4-12% gradient polyacrylamide gel. The gel was then stained with Coomassie blue. The pattern of the two fractions was relatively different. This figure shows visualized proteins present in the NP40 fraction and SDS fraction from three different CLL samples.

3.2.2.7 Reproducibility of two-step cellular fractionation method

To investigate the variation from extraction to extraction, three groups of protein bands were selected to perform densitometric analyses for them in four different CLL samples. This was performed using ImageJ software version 1.44o. The three groups of bands included bands present in all of the two fractions and bands that were predominantly present in only one fraction (Figure 3.12A I, II, III). The small error bars indicate that the banding pattern was consistent suggesting good reproducibility of the cellular fractionation (Figure 3.12B I, II, III)

3.2.2.8 Purity of the two fractions

To check the purity of the fractions, equivalent amounts of protein from each fraction was used to carry out SDS-PAGE followed by western blotting and antibody detection using an anti-poly(ADP-ribose) polymerase (PARP) antibody, which was used as a nuclear protein marker, and an anti-tubulin antibody, which was used as a cytosolic protein marker. PARP was not found in the NP40 fraction suggesting that the NP40 fraction was not contaminated by the SDS fraction. Likewise, tubulin was not present in the SDS fractions, which probably indicates that it was not contaminated by the NP40 fraction (Figure 3.13). All samples used for proteomics analyses showed similar results.

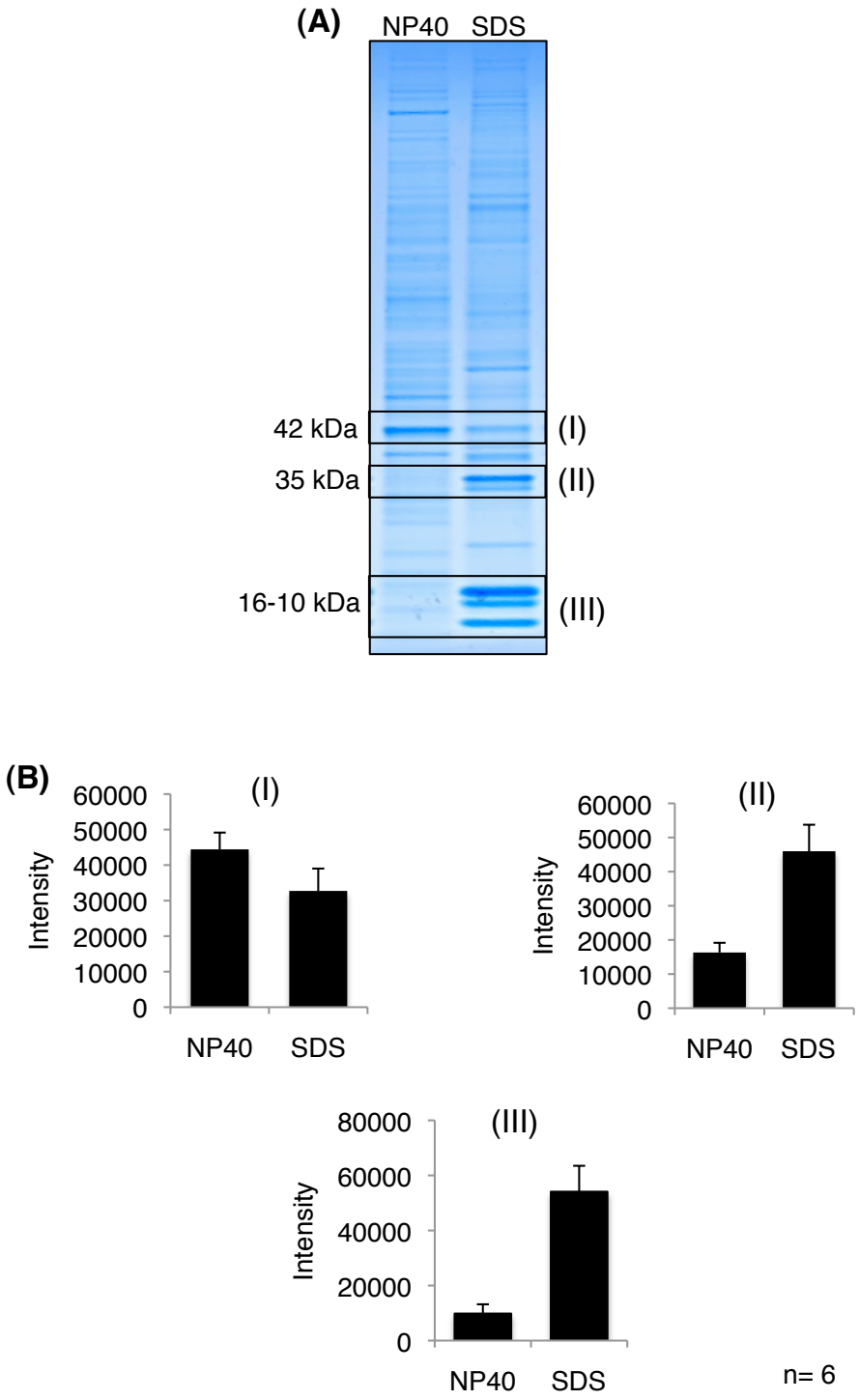


Figure 3.12: Reproducibility of the two-step cellular fractionation method. Three groups of protein bands were chosen for densitometric analysis. These bands include bands present in the two fractions (A-I) and bands that appeared only in only one fraction (A-II and III). This analysis was performed on six different CLL samples (B-I, II, III). The similar intensity of the measured bands in six different CLL samples (as indicated by the small error bars) suggested that this method was reproducible.

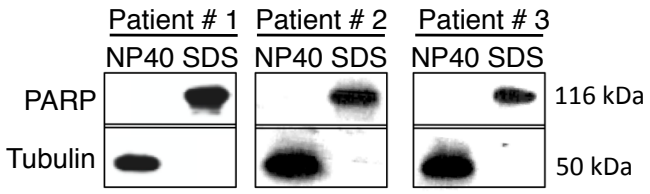


Figure 3.13: Analysis of PARP and tubulin in the NP40 fraction and SDS fraction. 5 µg protein from the NP40 fraction and SDS fraction were resolved by SDS-PAGE followed by western blotting and antibody detection using specific antibodies to poly (ADP-ribose) polymerase (PARP), as a nuclear protein marker, and to tubulin, as a cytosolic protein marker. The analysis suggested that there was no cross contamination between the two fractions. This figure shows results from three different CLL samples.

3.3 Discussion

In order to perform good quality CLL proteomics analysis, considerable effort was made to ensure the quality of the CLL samples utilized. In this regard, all CLL samples were separated from peripheral blood samples derived from CLL patients within 4 hours of collection. This was very important to avoid *ex vivo* manipulation of CLL cells that might result in pathologically irrelevant changes in the two forms of CLL in terms of protein expression.

The purity of CLL cells was taken into consideration; only CLL samples with ≥95% CD19 positivity were used. This was not the case with some other CLL proteomics studies where the purification step of CLL cells was not included in their sample preparation workflow (Voss et al., 2001, Cochran et al., 2003). Restricting the selection of CLL samples to those with $CD19^+$ ≥95% is an essential quality control step in order to avoid including the proteome of other mononuclear cells such as monocytes or T-lymphocytes in the analysis of the CLL proteome. This quality control step becomes even more critical when using CLL samples derived from treated patients as treatment kills CLL cells, which in turn makes their percentage lower among other isolated blood mononuclear cells.

CD19 is not a specific marker for CLL cells as it is expressed on both normal and malignant B cells (Nadler et al., 1983). The reason why CD19 was used to determine the purity of the CLL cells is that all of the CLL samples were separated from whole peripheral blood obtained from patients

who had already been diagnosed with CLL. Given the lymphocytosis exhibited by these patients, the majority of B-cells in the peripheral blood samples of CLL patients were CLL cells.

The second criterion applied on selection of CLL samples was viability of CLL cells. This was necessary because some of the CLL samples were obtained from patients who had previously received treatment and from patients with different clinical courses. This, combined with the known propensity of some CLL cells to apoptosis *ex vivo* made it important to isolate only the viable CLL cells (Collins et al., 1989). All of the seven published CLL proteomics studies (discussed in section: 1.3) used CLL samples regardless of their viability (Voss et al., 2001, Cochran et al., 2003, Boyd et al., 2003, Barnidge et al., 2005a, Barnidge et al., 2005b, Miguet et al., 2009, Perrot et al., 2011). The risk of using CLL samples with low viability is that it may contribute to changes in protein expression that are irrelevant to the pathology of the disease. As a result, this study limited the use of CLL samples to those with high viability.

Although mutational status of *IGHV* is regarded as the most accurate and widely applicable prognostic marker in CLL (Damle et al., 1999, Hamblin et al, 1999), CD38 has been used as a prognostic marker to discriminate between CLL patients with bad or good prognosis (Durig et al., 2002, Pepper et al., 2012). This decision was made because the majority (3/4) of the published CLL proteomics studies that explored the protein

expression of the two forms of CLL ware based on the mutational status of *IGHV* genes (Cochran et al., 2003, Barnidge et al., 2005a, Perrot et al., 2011). In fact, only one CLL proteomics study was partially conducted to compare the proteome of $CD38^+$ CLL to that of $CD38^-$ CLL and showed no significant changes (Cochran et al, 2003). Furthermore, transcriptomic studies that compared $CD38^+$ CLL to $CD38^-$ CLL showed a number of significantly altered mRNA between the two sets of samples (Pepper et al, 2007, Huttmann et al, 2006). In addition, CD38 expression is relatively easy to determine by flow cytometry following staining with anti-CD19 antibody and an anti-CD38 antibody when CLL samples were received.

Several studies have demonstrated different cut-offs for CD38 expression that discriminate poor prognosis from good prognosis CLL. These cut-offs include 30% (Damle et al., 1999), 20% (Durig et al, 2002, Pepper et al., 2012), and 7% (Krober et al., 2002). Different studies that focused on the impact of CD38 expression on CLL pathology and prognosis used 20% as a cut-off to differentiate between $CD38^+$ CLL and $CD38^-$ CLL (Deaglio et al., 2007, Vaisitti et al., 2010, Huttmann et al., 2006). In this study, CLL samples that were collected had CD38 expression of 40% or more representing the bad prognosis of CLL and 5% or less representing the good prognosis of CLL. Such a distinct percentage of CD38 expression was used in the hope that differences between these groups of CLL samples would be highly pronounced at the protein expression level.

As mentioned earlier, one of the keys to a successful proteomics study is the quality of the protein extract to be analysed (Dreger, 2003, Yates et al., 2005, Ahmed, 2009, Brewis and Brennan, 2010). Of the published CLL proteomics studies, 3/7 used whole cell lysate (Voss et al., 2001, Cochran et al., 2003, Perrot et al., 2011). While this had the potential to subject all cellular proteins to analysis, it did not reduce sample's complexity, which in turn limited the potential to identify a large number of proteins. In addition, these approaches did not provide information regarding protein localization (Brewis and Brennan, 2010).

Other published CLL proteomics studies (3/7) specifically focused on one fraction namely membrane proteins (Boyd et al., 2003, Barnidge et al., 2005b, Miguet et al., 2009). The advantage of using one protein extract is that sample complexity can be reduced and information about protein localization can be obtained. However, it leaves other protein fractions, such as the nuclear fraction and cytosolic fraction, unanalyzed.

Only one published CLL proteomics study analysed cellular proteins in both cytosolic and membrane fractions (Barnidge et al., 2005a). In the current project, a fractionation method that was previously described by Brennan et al. (2009) was developed to fractionate cellular proteins into two different fractions termed the NP40 fraction and the SDS fraction. The cellular fractionation methods shown by Barnidge et al. (2005a) and those described in this study had the potential to subject more cellular proteins to analysis,

whilst at the same time reducing the sample complexity. This should facilitate the identification of many proteins by proteomics and provide information about protein localization. Nevertheless, the feature of the method demonstrated in this project is that it is less time consuming compared to the method described by Barnidge *et al*., (2005a) which involves two long centrifugation steps of 45 min each.

Using 1DE followed by silver staining Barnidge *et al*., (2005a) showed that the cytosolic fraction and the membrane fraction contained relatively different proteins. Similarly, the analysis of the NP40 fraction and the SDS fraction by 1DE, followed by colloidal Coomassie Blue staining, has also demonstrated that relatively different proteins were extracted in the two fractions. In fact, four intense bands with low molecular weight (<39 kDa) were only present in the SDS fraction. This reflects the quality of the fractionation method that was developed in this project.

While cellular fractionation methods can be used for enrichment of particular proteins, such as cytosolic proteins, it is very difficult to generate a completely pure protein fraction (Huber et al., 2003). However, relative purity of a protein extract can be monitored using protein markers (Brennan et al., 2009). In the two-step cellular fractionation method developed in this chapter, tubulin, a cytosolic protein marker, was detected only in the NP40 fraction, whereas PARP, a nuclear protein marker, was detected only in the SDS fraction. In contrast, in the fractionation method demonstrated by

Barnidge et al. (2005a), tubulin was detected in the cytosolic fraction as well as the membrane fraction suggesting that the latter was contaminated by the former. This may indicate that the two-step cellular fractionation method described in this chapter generates relatively more pure fractions compared with the fractionation method demonstrated by Barnidge et al. (2005a).

In this chapter, good quality NP40 fractions and SDS fractions were generated from poor prognosis and good prognosis CLL samples. The next step was to be subject these protein extracts to different proteomics approaches in an attempt to identify proteins with potential involvement in CLL.

Chapter Four

Qualitative Proteomics Analysis of Primary CLL Samples

4.1 Introduction

Different proteomics approaches have been used to study the proteome of CLL (discussed in section 1.3). Three studies used 2DE as a tool for protein separation (Voss et al., 2001, Cochran et al., 2003, Perrot et al., 2011). While Cochran et al., (2003) and Perrot et al., (2011) detected a large number of proteins (spots) by 2DE (800 spots and 1873 spots respectively), they identified a small number proteins (≤ 60 proteins) by mass spectrometry (MS). Other CLL proteomics studies used LC either alone or in combination with 1DE for sample separation prior to mass spectrometry analysis (Boyd et al., 2003, Barnidge et al., 2005a, Barndige et al., 2005b, Miguet et al., 2009). These studies identified relatively large numbers of proteins (≥ 371 proteins).

This chapter aimed to identify many proteins in the NP40 fractions and the SDS fractions that were generated from CLL samples. This was achieved by subjecting CLL protein extracts to a qualitative proteomics workflow.

In the previous chapter the processes of isolating good quality CLL samples and developing a good quality cellular fractionation method were explained in detail. This chapter explains the qualitative proteomics workflow. Briefly, the protein content of the NP40 fraction or SDS fraction was precipitated and digested by trypsin. The resulting peptide 'soup' was then subjected to separation by 2D nano-LC; separated peptides were subjected to MALDI TOF-TOF mass spectrometry for MS and MS/MS analysis. Resulting

MS/MS spectra were searched against the Swiss-Prot database using GPS/MASCOT for peptide sequencing and protein identification (Figure 4.1).

4.2 Results

4.2.1 Protein precipitation and digestion

In addition to proteins, the NP40 fractions and SDS fractions also contained other molecules such nucleic acids, lipids and detergents. For a good quality mass spectrometry experiment molecules other than proteins should be removed from the samples (Ahmed, 2009). To do that, protein precipitation was carried out using the 2D clean Up kit reagent and method.

The strategy that was used for protein identification was a bottom up approach (shotgun proteomics), where sequencing data of some peptides are used to identify their 'parent' proteins (Yates et al., 2009). Consequently, protein digestion was essential to produce peptides that would be interrogated by MALDI mass spectrometry. Therefore, following protein precipitation, proteins were incubated overnight with 2μg of trypsin at 37°C for protein digestion. Trypsin cleaves the C-terminal of lysine (K) and arginine (R) amino acid residues (Westermeier et al., 2008).

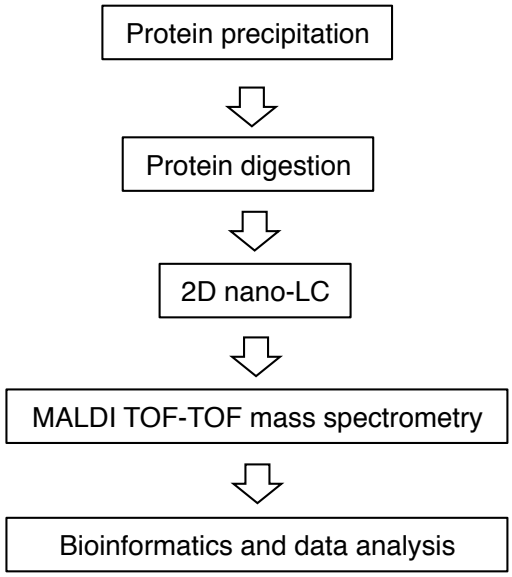


Figure 4.1: Workflow of the qualitative proteomics analysis. Proteins from the NP40 fraction and SDS fraction were separately subjected to protein precipitation using the 2D clean up kit followed by trypsin digestion. Resulting peptides were separated by 2D nano-LC on an SCX column using stepwise elution with increasing concentrations of NaCl (breakthrough, 100mM, 200mM, 400mM, 800mM, and 1M). Each salt fraction was subsequently further separated on an RP column. Separated peptides were then spotted out on a MALDI plate for analysis by MALDI TOF-TOF mass spectrometry. Generated MS/MS spectra were searched against the Swiss-Prot database by using GPS explorer software/MASCOT.

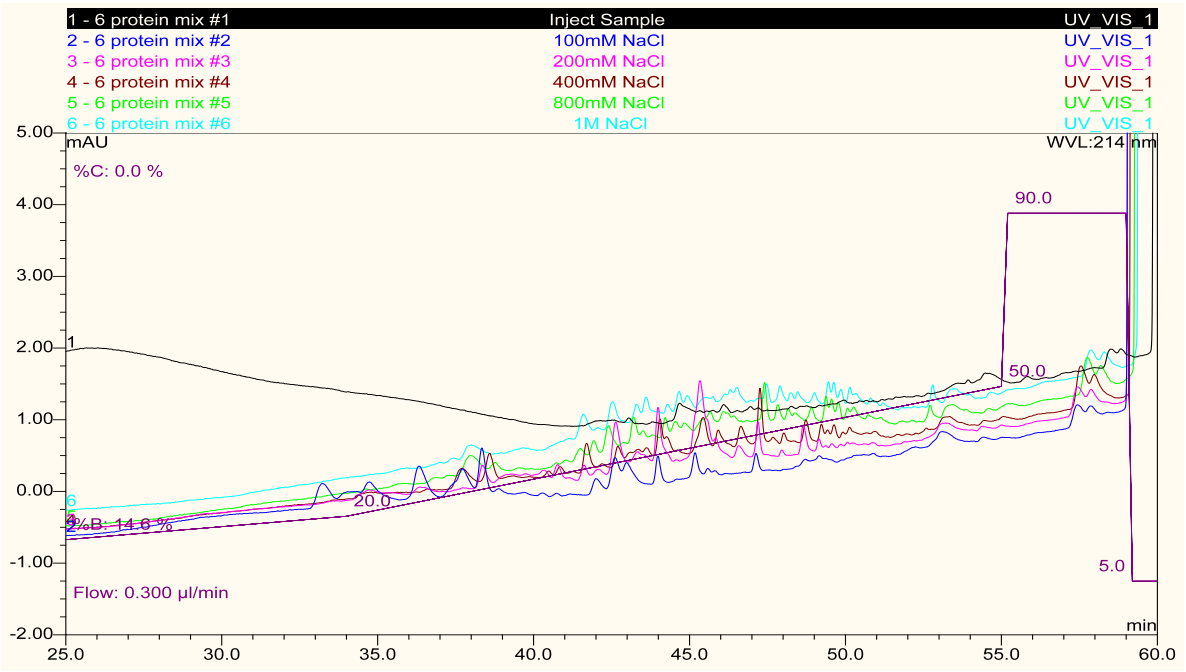
4.2.2 Two-dimensional nano liquid chromatography (2D nano-LC)

4.2.2.1 Quality control of 2D nano-LC

Although the sample complexity was partially reduced by performing the two-step cellular fractionation method (described in the previous chapter), separation by 2D nano-LC of the protein digests would facilitate identification of many proteins by MALDI mass spectrometry. In the course of optimising the qualitative proteomics workflow, in particular separation by 2D nano-LC, quality control runs were employed to avoid using the 2D nano-LC with poorly functioning properties to separate CLL peptides extracts. Examples of poorly separated CLL protein extracts on the 2D nano-LC before introducing the use of quality control checks to the 2D nano-LC are shown in Appendix 1.

The quality control checks included examining the SCX column using a protein mixture digest (P/N 161088; Dionex) and the RP column using Cytochrome digest (P/N 161088; Dionex). Figure 4.2A shows an example of a quality control run for SCX column using protein mixture digest, where many peaks, which probably represent peptides, were recorded at different times with various intensities in every salt fraction (breakthrough, 100mM, 200mM, 400mM, 800mM, and 1M). This indicated that the SCX column was performing well. In addition, the chromatogram of cytochrome digest on an RP column shows different peaks, which probably represent Cytrochrome peptides. These were eluted at different times with different intensities (Figure 4.2B). This indicated that the RP column was also performing well.

(A)



(B)

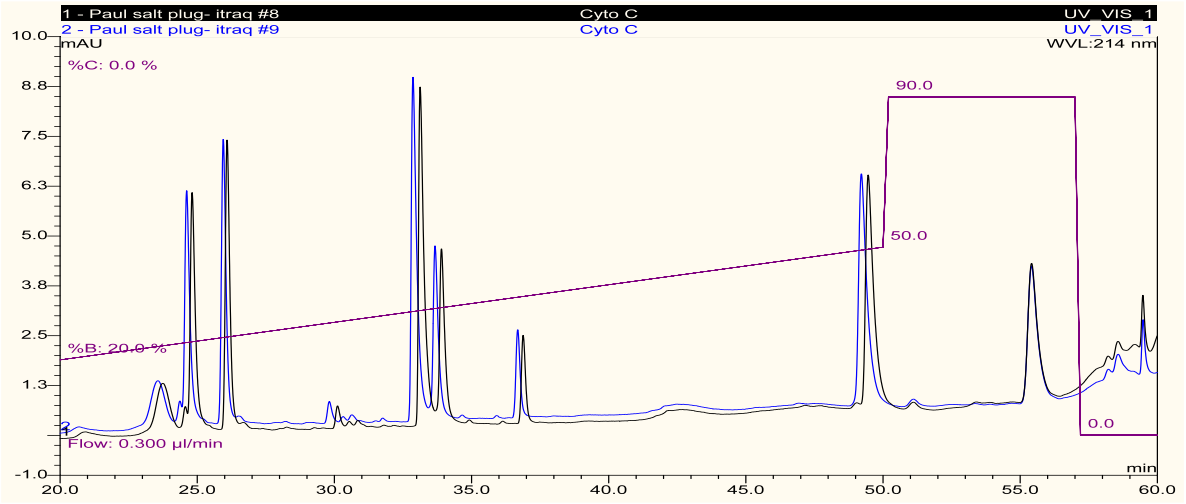


Figure 4.2: Monitoring the performance of the 2D nano-LC. A commercially available protein mixture digest was used to examine the quality of the SCX column (A). Likewise, a commercially available cytochrome digest was used to check the quality of the RP column (B). This step was done every time when biological samples were to be separated on the 2D nano-LC. This analysis suggested that both columns were performing well and were ready to be used for separating the biological samples.

4.2.2.2 Separation of CLL protein digest by 2D nano-LC

After being satisfied with the performance of the nano-LC, peptides from either the NP40 fraction or SDS fraction were subjected to separation by 2D nano-LC. First, the peptides were separated on an SCX column in stepwise elution using increasing concentrations of NaCl (breakthrough, 100mM, 200mM, 400mM, 800mM, and 1M). Each salt fraction was then desalted on an RP desalting column and subsequently further separated on an RP column. Figure 4.3 shows examples of a UV trace record of peptides separated by 2D nano LC. Separated peptides were mixed with a peptide Glu-Fib and CHCA and then spotted out onto a MALDI plate (one spot/8 sec). The peptide Glu-Fib will be used as quality control for MALDI mass spectrometry performance and for internal calibration, while CHCA was used for peptide soft ionisation in the MALDI mass spectrometry.

4.2.3 MALDI TOF-TOF mass spectrometry

Separated peptides were subjected to analysis by MALDI TOF-TOF mass spectrometry to generate MS and MS/MS spectra. First, the intensity and mass of peptides in the 1530 eluted spots across the MALDI plate were measured (MS mode). Then the MS/MS mode started where the most abundant six peptides in each eluted spot were selected for precursor fragmentation in the CID chamber to generate MS/MS spectra, which would be used later for peptide sequencing and protein identification.

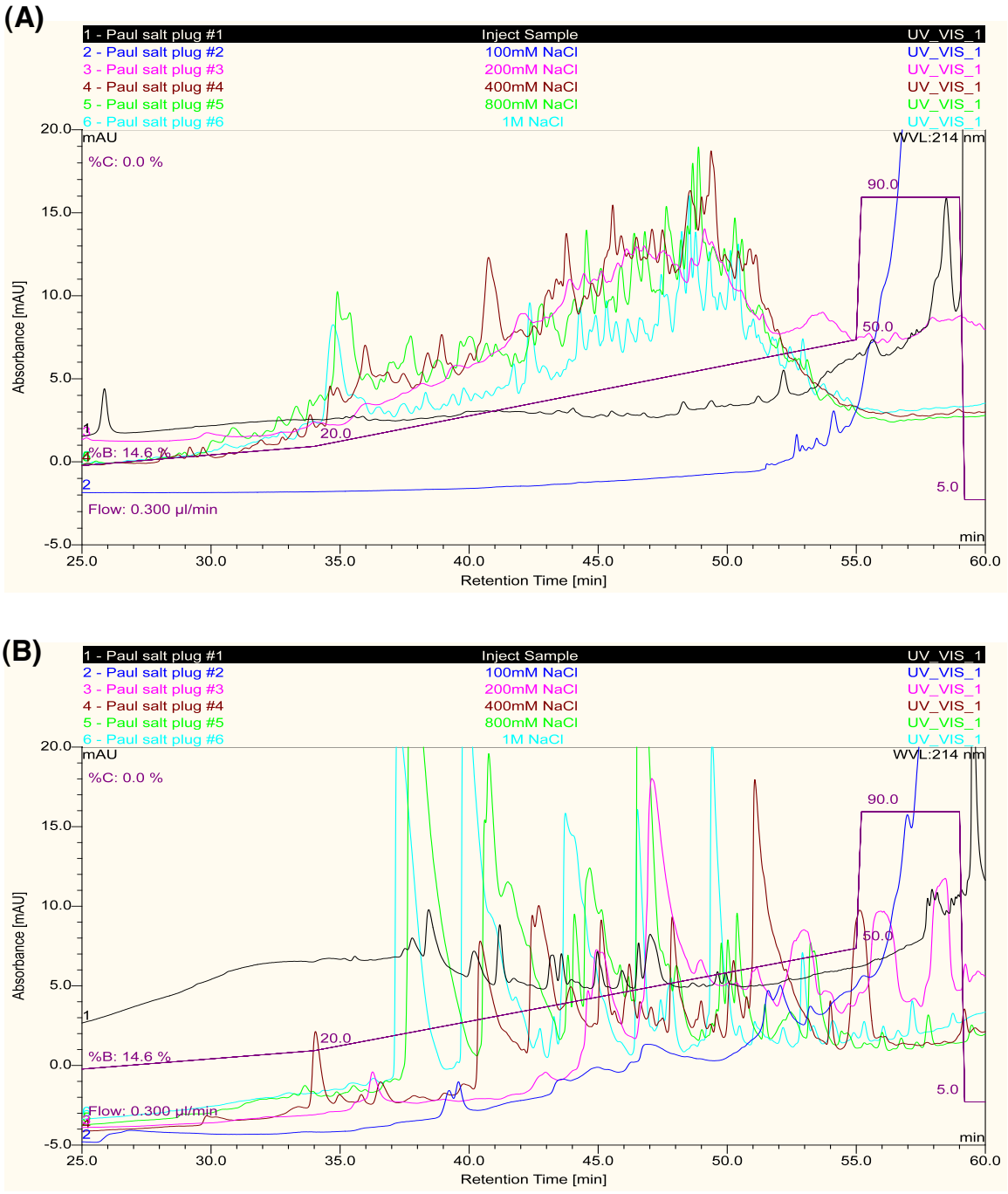


Figure 4.3: Chromatogram of peptides separated by 2D nano-LC. Protein digest (13.3 µg) from the NP40 fraction (A) and from the SDS fraction (B) were separated by 2D nano-LC. Firstly, peptides were separated on the SCX column using six salt fractions with an increasing concentration. Each fraction was then further separated on an RP column using a three-step gradient. The first step was from 5 to 20% solvent B for the first 34 mins, the second step was from 20% to 50% solvent B for 21 mins and the third step was from 50% to 90% solvent B for 4 mins. Peptides were collected from the 25th-59th mins for each salt fraction. A total of 1530 spots were eluted onto the MALDI target plate.

4.2.3.1 Monitoring the quality of the MALDI mass spectrometry analysis

The quality of the MALDI mass spectrometry runs were monitored by performing extracted ion chromatography (XIC) analysis, which can be used to check whether a particular precursor (peptide) is detected by MALDI mass spectrometry in MS mode. XIC analysis was performed for the Glu-Fib peptide (1570.677 Da), which was spiked into each eluted spot (1530) across the MALDI plate. This peptide should be detected in each of these spots. Figure 4.4A shows an example of an XIC analysis for the Glu-Fib peptide on a MALDI mass spectrometry run, which demonstrates that the Glu-Fib peptide was detected in all of the eluted spots. As elucidated in the materials and methods (section: 2.6.4), there were gaps between the eluted spots of different salt fractions. For example, there were 109 empty spots between the eluted spots from the breakthrough fraction and the eluted spots from the 100mM salt fraction. In addition, there were 5 empty spots in the remaining salt fractions. In these gaps Glu-Fib peptides was not detected. Figure 4.4B and 4.4C show examples of Glu-Fib precursor (1570.68 Da) in two different spots.

The quality of the MALDI mass spectrometry runs was also estimated by carrying out a Base Peak Chromatogram (BPC) analysis, which shows the summed intensity of peptides that were detected in every eluted spot (1530) across the MALDI plate. Figure 4.5 demonstrates that peptides with different intensities were detected in each eluted spot on the MALDI plate.

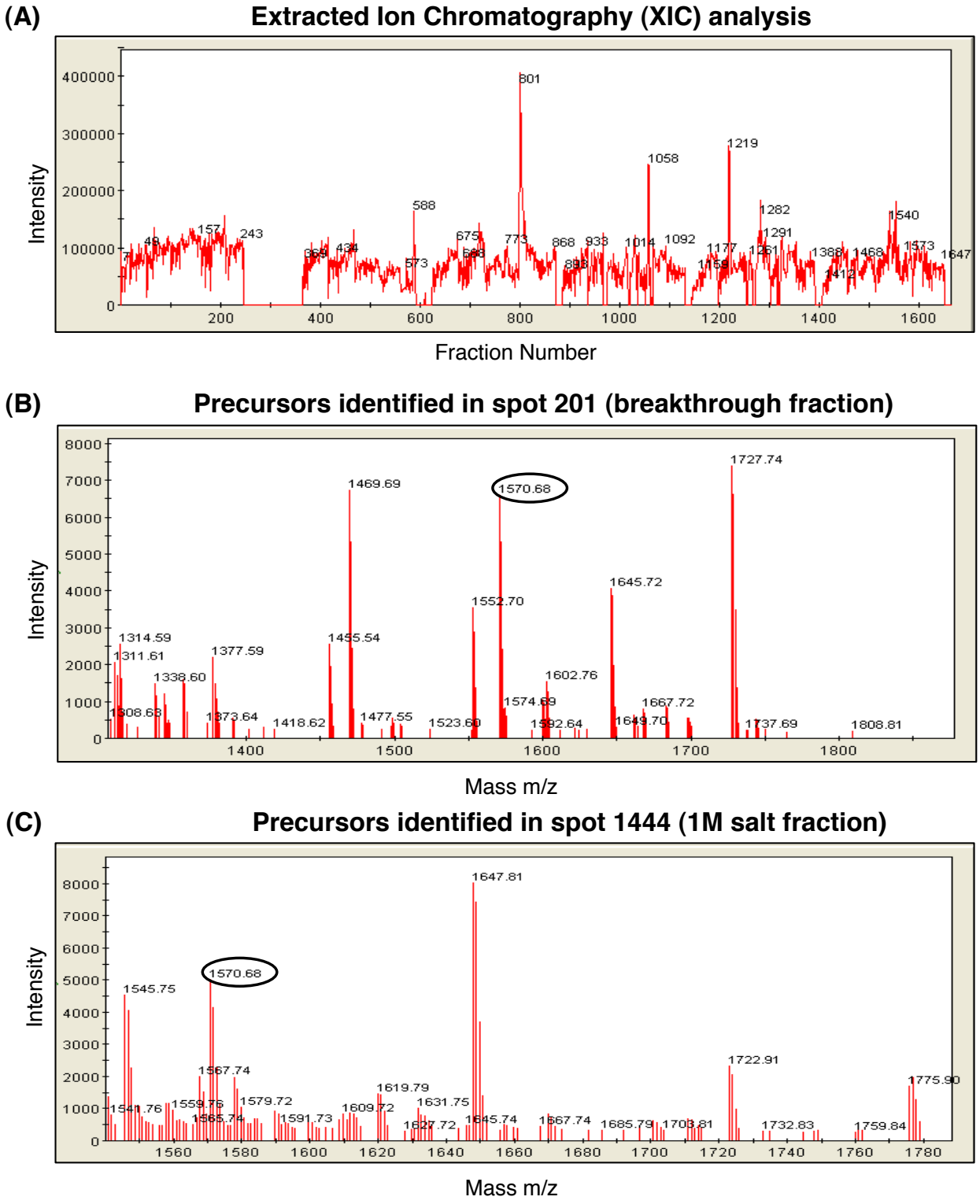


Figure 4.4: Glu-Fib detection by MALDI mass spectrometry. XIC can be used to check whether a particular precursor (peptide) was detected by MALDI mass spectrometry in MS mode. The molecular weight of Glu-Fib peptide is 1570.677 Da. Therefore, the XIC search was done with a mass of 1570.677± 0.2. The analysis showed that a peptide (1570.677± 0.2 Da), which was probably Glu-Fib, was present in all the eluted spots (1530) on the MALDI plate (A). An example of detected Glu-Fib precursor (1570.68 Da) is shown in two different spots; 201 (B) and 1444 (C).

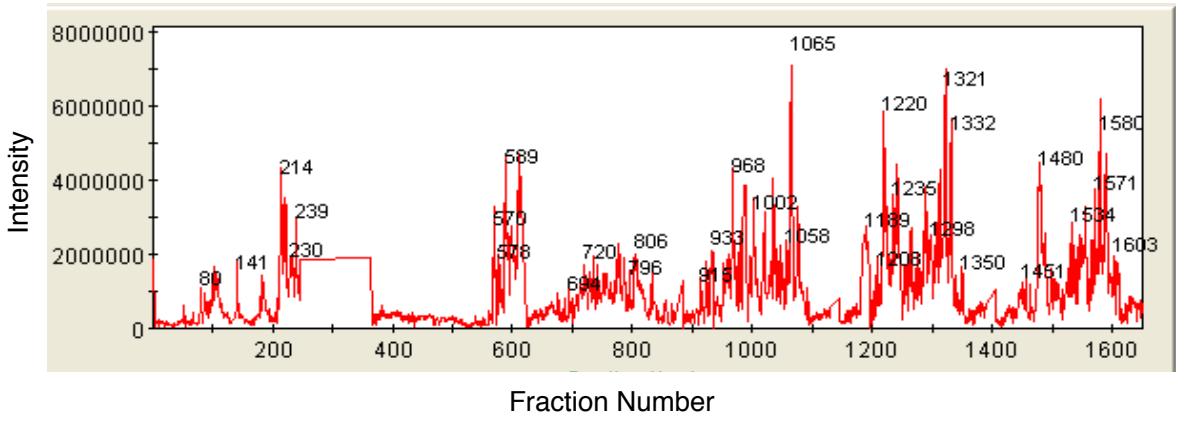


Figure 4.5: Base Peak Chromatography (BPC) analysis of an MS analysisby MALDI mass spectrometry. BPC measures the summed identified precursors (peptides) in each eluted spot and plots them against their retention time across the MALDI plate. Therefore, it can give an indication of the quality of the MS run by MALDI mass spectrometry. This figure shows an example of MS scan of peptides spotted on a MALDI plate, where in almost all eluted spots (1530) there were detected peptides. In addition it reflected the quality of the peptides separation by 2D nano-LC, where in each salt fraction (6x) there were detected peptides. This BPC analysis was for peptides separated on the 2D nano-LC shown previously in Figure 4.3A.

4.2.3.2 Data analysis

Generated MS/MS spectra were searched against the Swiss-Prot database using the MASCOT Database search engine v2.1 (Matrix Science) embedded into GPS Explorer software v3.6 (Applied Biosystems). The search was performed specifically using human taxonomy with trypsin digestion (only one missed cleavage was allowed); tolerance of MS/MS spectra was 0.3Da. Cysteine modification by methyl methanethiosulfonate (MMTS) was employed as a fixed modification. Variable modifications were oxidation (M), pyro-glu (N-term E) and pyro-glu (N-term Q). For a protein to be reported as a detected protein in the sample it must be identified with at least one peptide (ion score ≥95% confidence interval C.I.).

4.2.4 Characterisation of the qualitative proteome of primary human CLL cells

In total 27 LC-MALDI mass spectrometry runs were performed on 12 different CLL samples. Nine LC-MALDI mass spectrometry runs were performed using the NP40 fractions. Of these nine, two were performed on a single sample (one patient), two runs were done on two compined samples (four patients) and five runs were performed on four pooled samples (12 patients). In addition, ten LC-MALDI mass spectrometry runs were performed using the SDS fractions. Of these ten runs, two were done on a single sample (one patient), two runs were performed on two combined samples (four patients) and six runs were done on four pooled samples (12 patients). The other eight LC-MALDI mass spectrometry runs were performed on a mixture

of the cellular fractions (NP40 and SDS fractions). These eight runs consist of five runs done on single samples (two patients) and three runs performed on two pooled samples (four patients).

In total, 900 proteins were identified (Appendix 2). Of these proteins, 625 (69%) were identified with 2 or more peptides (Ion Score ≥95% C.I., False Discovery Rate (FDR) = 0%) and 275 (31%) were identified with one peptide (Ion Score >99.9% C.I., FDR = 3.2%). Figure 4.6 shows the peptide count of proteins detected with multiple peptides.

Of the proteins identified with multiple peptides (625), 568 (90%) were identified in multiple MALDI mass spectrometry analyses. Table 4.1 presents 10 examples of the most and least frequently identified proteins based on multiple peptides. Similarly, of the proteins that were detected with a single peptide (275), 167 (61%) were found in multiple MALDI mass spectrometry analyses. Ten examples of these proteins are shown in Table 4.2. These data were encouraging to investigate whether there was a relation between peptide count or total ion score (TIS: total ion score of peptides either distinct or identical that were identified by mass spectrometry) and the chance of identifying the same protein by MALDI mass spectrometry in technical or biological replicates. Figure 4.7 demonstrates a significant correlation between the feasibility of identifying the same protein in multiple MALDI mass spectrometry runs and peptide count (A) or TIS (B). This analysis suggested

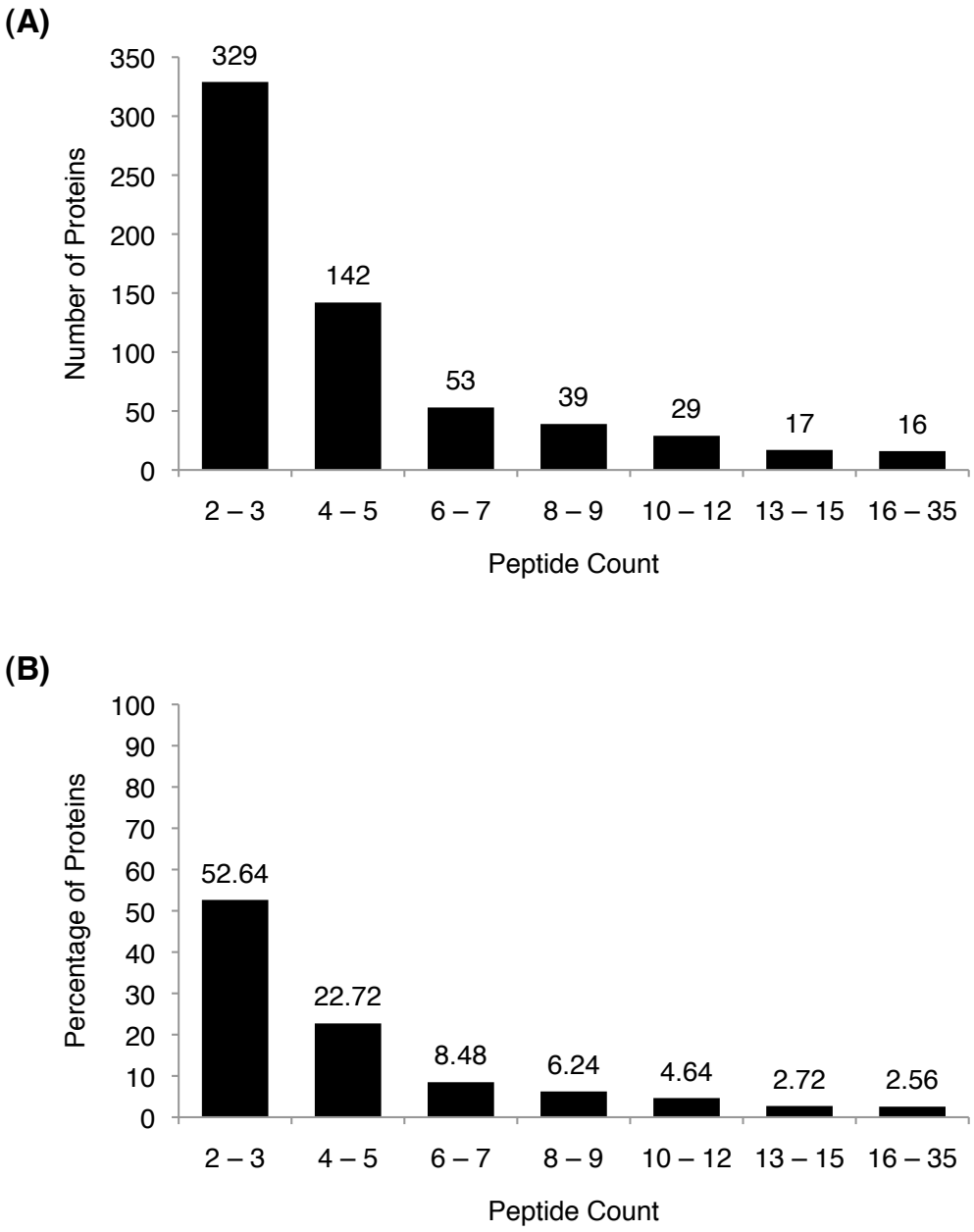


Figure 4.6: Protein identifications based on multiple peptides. Of the 900 proteins, 625 (69%) were identified with multiple peptides. This figure shows the number (A) and percentage (B) of proteins assigned to the number of peptides used for identification.

Table 4.1: Examples of the most and the least frequently identified proteins based on multiple peptides in MALDI mass spectrometry analyses

Protein Name	Accession Number	Peptide Count	TIS	BIS	BIS C.I. %	Number of MALDI MS analyses
Actin, cytoplasmic 2	ACTG_HUMAN	17	1439	186	100	20
Heterogeneous nuclear ribonucleoproteins A2/B1	ROA2_HUMAN	15	1296	229	100	20
Histone H4	H4_HUMAN	10	1016	150	100	20
Nucleophosmin	NPM_HUMAN	7	377	72	100	20
Keratin, type II cytoskeletal 1	K2C1_HUMAN	32	2936	201	100	19
Ubiquitin carboxyl-terminal hydrolase 7	UBP7_HUMAN	2	76	40	99.5	1
Lamin-A/C	LMNA_HUMAN	2	76	39	99.4	1
Phenylalanyl-tRNA synthetase alpha chain	SYFA_HUMAN	2	73	38	99.2	1
26S protease regulatory subunit 6B	PRS6B_HUMAN	2	71	38	99.1	1
Guanine nucleotide-binding protein G(k) subunit alpha	GNAI3_HUMAN	2	73	37	99	1

Protein extracts from CLL cells (the NP40 fractions and the SDS fractions) were digested, separated by 2D nano-LC and analysed by MALDI mass spectrometry. These 10 proteins were chosen after the list of proteins detected with multiple peptides was sorted according to the number of MALDI mass spectrometry analyses through which proteins were detected, peptide count and BIS C.I%. TIS: total ion score, BIS: best ion score, BIS C.I.%: best ion score confidence interval percentage.

Table 4.2: Examples of the most and the least frequently identified proteins based on a single peptide in MALDI mass spectrometry analyses

Protein Name	Accession Number	Peptide Count	TIS	BIS	BIS C.I. %	Number of MALDI MS analyses
Small nuclear ribonucleoprotein Sm D3	SMD3_HUMAN	1	83	83	100	13
Cytochrome c oxidase subunit 2	COX2_HUMAN	1	66	66	100	10
Histone H4-like protein type G	H4G_HUMAN	1	89	89	100	9
Small ubiquitin-related modifier 3	SUMO3_HUMAN	1	85	85	100	9
Splicing factor, arginine/serine-rich 7	SFRS7_HUMAN	1	62	62	99.98	9
Kinectin	KTN1_HUMAN	1	48	48	99.92	1
Putative double homeobox protein 3	DUX3_HUMAN	1	51	52	99.92	1
Microsomal glutathione S-transferase 3	MGST3_HUMAN	1	48	48	99.92	1
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 6	NDUA6_HUMAN	1	48	48	99.91	1
RNA-binding protein 4	RBM4_HUMAN	1	47	47	99.91	1

Protein extracts from CLL cells (The NP40 fractions and the SDS fractions) were digested, separated by 2D nano-LC and analysed by MALDI mass spectrometry. These 10 proteins were selected after the list of proteins detected with a single peptide was sorted according to the number of MALDI mass spectrometry analyses through which proteins were detected, peptide count and BIS C.I%. TIS: total ion score, BIS: best ion score, BIS C.I.%: best ion score confidence interval percentage.

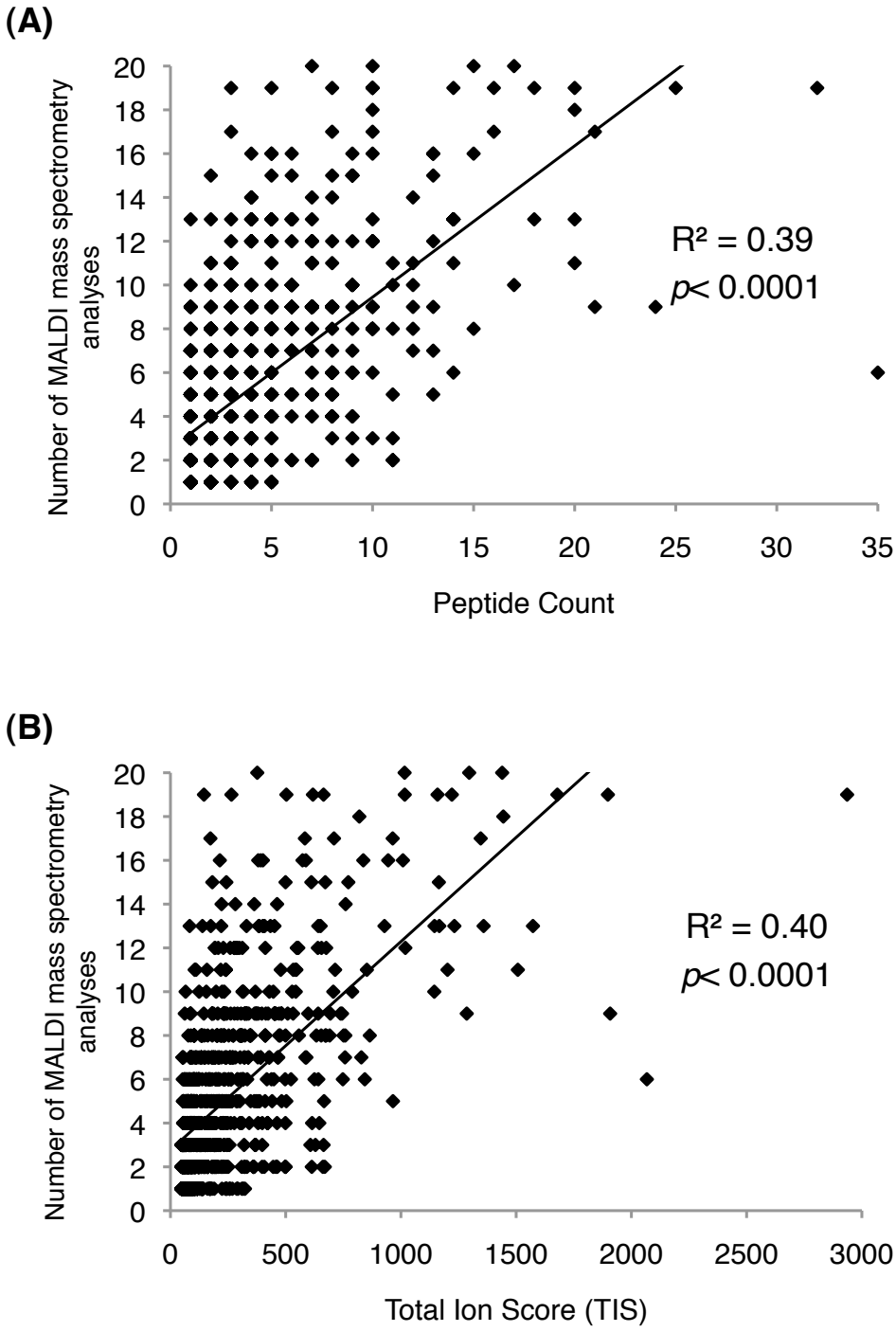


Figure 4.7: Relationship between detecting the same proteins in multiple MALDI mass spectrometry analyses and peptide count or TIS. This analysis showed a significant correlation between the possibility of identifying a protein frequently in MALDI mass spectrometry analyses and its peptide count (A) or TIS (B).

that the larger the peptide count or TIS of a protein, the higher the chance to identify this protein in technical or biological replicates.

Of the 900 proteins, 729 proteins were identified in the NP40 fraction, while 326 were identified in the SDS fraction; 188 proteins were common to both fractions (Figure 4.8). An example of a protein that was identified in the NP40 fractions was tubulin beta chain, which was detected with 13 peptides (Table 4.3) covering 37% of the sequence (Figure 4.9). An example of MS and MS/MS spectra of one of these peptides are shown in Figure 4.10. In addition, one protein that was identified in the SDS fractions was PARP, which was detected with 12 peptides (Table 4.4) covering 16% of the protein sequence (Figure 4.11). An example of MS and MS/MS spectra of one of these peptides is shown in Figure 4.12. Proteins from the NP40 fractions and/or the SDS fractions that were identified with the highest peptide count (in all experiments) are listed in Table 4.5.

4.2.5 Transcriptomic data support single peptide-based protein identification

Of the identified 900 proteins, 108 proteins (12%) were identified in a single mass spectrometry analysis with a single peptide ID (Ion Score >99.9% C.I.). To add more confidence to this type of identification, independent published transcriptomic data derived from six CLL samples (Huttmann et al., 2006) were used to check whether CLL samples expressed the cognate transcript encoding the 108 proteins. Only Affymetrix signals with

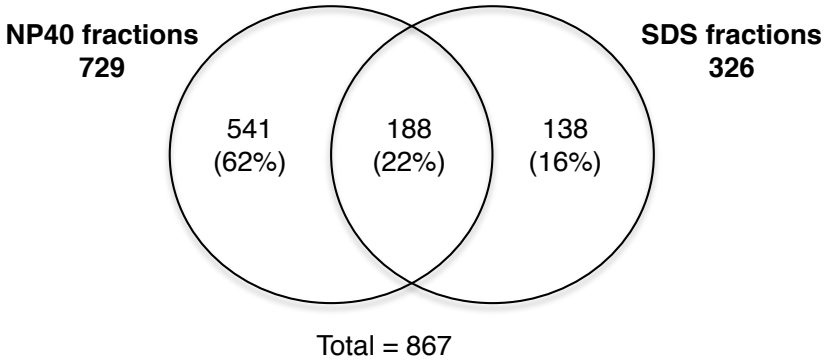


Figure 4.8: Venn diagram of proteins identified in the NP40 fraction and in the SDS fraction. These data represents 9 LC-MALDI runs performed on the NP40 fractions and 10 LC-MALDI runs on the SDS fractions.

Table 4.3: Examples of peptides that were identified in the NP40 fractions and assigned to tubulin beta chain (TBB5_HUMAN).

Peptide sequence	IS	IS C.I. %
EIVHIQAGQCGNQIGAK	131	100
ISVYYNEATGGK	120	100
AILVDLEPGTMDSVR	103	100
FPGQLNADLR	102	100
GHYTEGAELVDSVLDVVR	96	100
EVDEQMLNVQNK	87	100
YLTVAAVFR	82	100
IMNTFSVVPSPK	81	100
ISEQFTAMFR	77	100
LAVNMVPFPR	76	100
NSSYFVEWIPNNVK	72	100
ALTVPELTQQVFDAK	55	100
NMMAACDPR	31	96

Peptides from the NP40 fractions were separated by 2D nano-LC and introduced into the MALDI mass spectrometry for MS and MS/MS analysis. This table shows peptide sequences of 13 different peptides with a minimum ion score 96% C.I. that were mapped to Tubulin beta chain (TBB5_HUMAN). These data were extracted from four MALDI mass spectrometry analyses. These peptides were sorted according to their ion score (IS). Data were extracted from four MALDI mass spectrometry analyses that were performed on the NP40 fractions.

37 % coverage of the Tubulin beta chain sequence

MR**EIVHIQAGQCGNQIGAK**FWEVISDEHGIDPTGTYHGDSDLQLDR**ISVYYNEATGGK**YVPR**AILVDLEPGTMDSVR**S
GPFGQIFRPDNFVFGQSGAGNNWAK**GHYTEGAELVDSVLDVVR**KEAESCDCLQGFQLTHSLGGGTGSGMGTLLISKI
REEYPDR**IMNTFSVVPSPK**VSDTVVEPYNATLSVHQLVENTDETYCIDNEALYDICFRTLKLTTPTYGDLNHLVSATMSG
VTTCLR**FPGQLNADLRKLAVNMVPFPR**LHFFMPGFAPLTSRGSQQYR**ALTVPELTQQVFDAKNMMAACDPR**HGR**YL
TVAAVFR**GRMSMK**EVDEQMLNVQNKNSSYFVEWIPNNVK**TAVCDIPPRGLKMAVTFIGNSTAIQELFKR**ISEQFTAMFR**
RKAFLHWYTGEGMDEMEFTEAESNMNDLVSEYQQYQDATAEEEEDFGEEAEEEA

Figure 4.9: Sequence coverage of Tubulin beta chain (TBB5_HUMAN). Data generated from four MALDI mass spectrometry analyses that were performed on the NP40 fractions were used to investigate the sequence coverage of Tubulin beta chain. Of the full sequence of Tubulin beta chain (444 amino acids), 36.7% (136 amino acids) were identified (**Red Bold**). Underlining was used to discriminate between adjacent peptides that were identified.

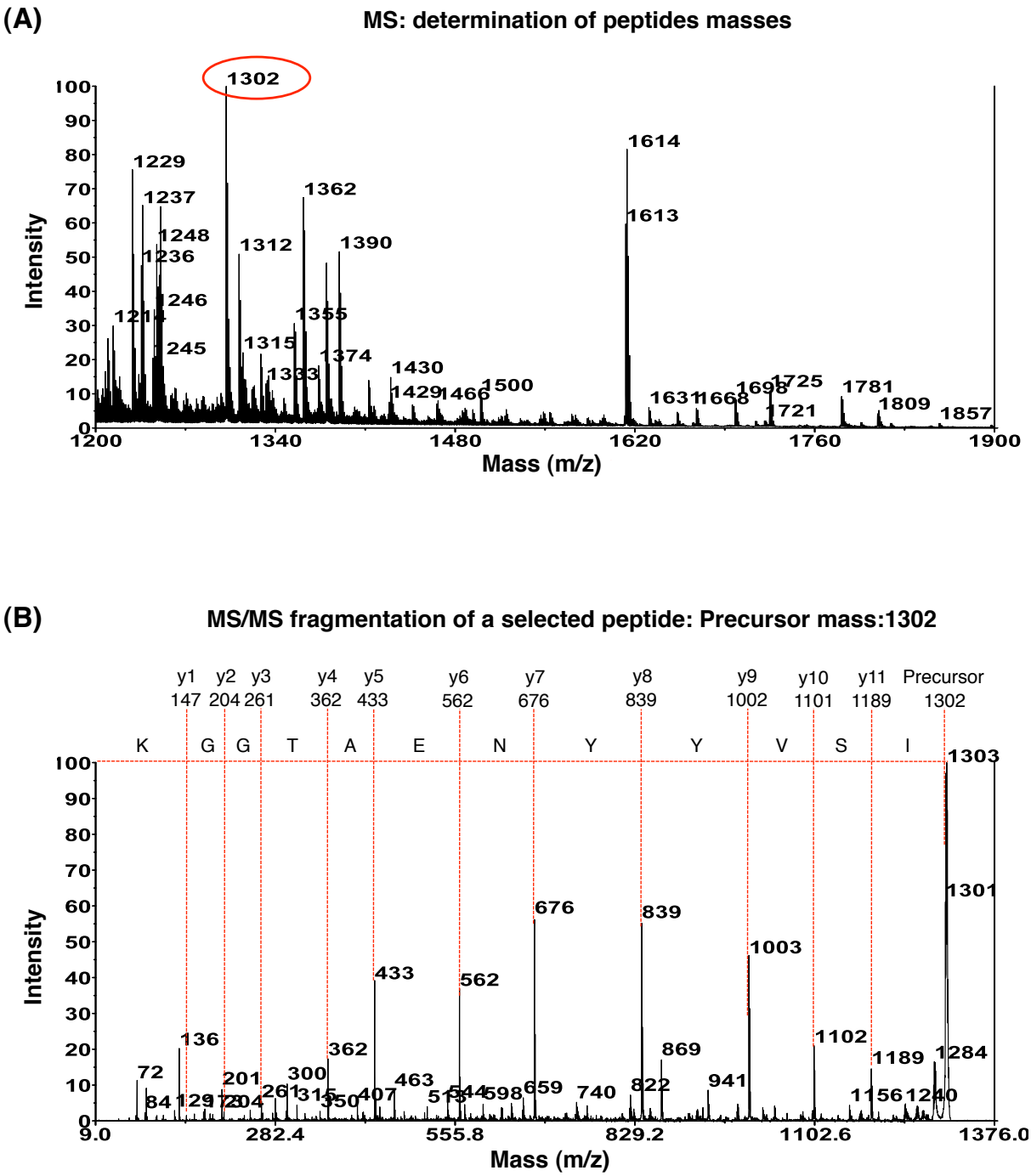


Figure 4.10: Example of MS and MS/MS spectra of Tubulin beta chain (TBB5_HUMAN) that was identified in the NP40 fractions. Separated peptides were applied to MALDI mass spectrometry to measure their masses and to perform peptide fragmentation in order to identify their sequences. A peptide with mass of 1302 (A) was subjected to MS/MS analysis to identify its sequence (B). The peptide sequence on the MS/MS spectra is read from the C terminus (**ISVYYNEATGGK)** as only y ions were shown in this figure.

Table 4.4: Examples of peptides that were identified in the SDS fractions and assigned to Poly [ADP-ribose] polymerase 1 (PARP1_HUMAN).

Peptide sequence	IS	IS C.I. %
HPDVEVDGFSELR	140	100
TTNFAGILSQGLR	118	100
QQVPSGESAILDR	103	100
GGAAVDPDSGLEHSAHVLEK	85	100
AEPVEVVAPR	68	100
VVSEDFLQDVSASTK	66	100
MAIMVQSPMFDGK	61	100
VVDRDSEEAEIIR	50	100
NREELGFRPEYSASQLK	47	100
KPPLLNNADSVQAK	37	99
TLGDFAAEYAK	37	99
GIYFADMVSK	32	97

Peptides from the SDS fractions were separated by 2D nano-LC and introduced into the MALDI mass spectrometry for MS and MS/MS analysis. This table shows the peptide sequence of 12 different peptides with a minimum ion score 97% C.I. that were mapped to PARP. These data were extracted from four MALDI mass spectrometry analyses. These peptides were sorted according to their ion score (IS). Data were extracted from four MALDI mass spectrometry analyses that were performed on the SDS fractions.

16% coverage of Poly [ADP-ribose] polymerase 1 sequence

MAESSDKLYRVEYAKSGRASCKKCSESIPKDSLR**MAIMVQSPMFDGK**VPHWYHFSCFWKVGHSIR**HPDVEVDGFS**
ELRWDDQQKVKKTAEAGGVTGKGQDGIGSKAEK**TLGDFAAEYAK**SNRSTCKGCMEKIEKGQVRLSKKMVDPEKP
QLGMIDRWYHPGCFVK**NREELGFRPEYSASQLK**GFSLLATEDKEALKKQLPGVKSEGKRKGDEVDGVDEVAKKKS
KKEKDKDSKLEKALKAQNDLIWNIKDELKKVCSTNDLKELLIFNK**QQVPSGESAILDR**VADGMVFGALLPCEECSGQ
LVFKSDAYYCTGDVTAWTKCMVKTQTPNRKEWVTPKEFREISYLKKLKVKKQDRIFPPETSASVAATPPPSTASAPA
AVNSSASADKPLSNMKILTLGKLSRNKDEVKAMIEKLGGKLTGTANKASLCISTKKEVEKMNKKMEEVKEANIR**VVSE**
DFLQDVSASTKSLQELFLAHILSPWGAEVK**AEPVEVVAPR**GKSGAALSKKSKGQVKEEGINKSEKRMKLTLK**GGAA**
VDPDSGLEHSAHVLEKGGKVFSATLGLVDIVKGTNSYYKLQLLEDDKENRYWIFRSWGRVGTVIGSNKLEQMPSKE
DAIEHFMKLYEEKTGNAWHSKNFTKYPKKFYPLEIDYGQDEEAVKKLTVNPGTKSKLPKPVQDLIKMIFDVESMKKA
MVEYEIDLQKMPLGKLSKRQIQAAYSILSEVQQAVSQGSSDSQILDLSNRFYTLIPHDFGMK**KPPLLNNADSVQAK**V
EMLDNLLDIEVAYSLLRGGSDDSSKDPIDVNYEKLKTDIK**VVDRDSEEAEIIR**KYVKNTHATTHNAYDLEVIDIFKIERE
GECQRYKPFKQLHNRRLLWHGSR**TTNFAGILSQGLR**IAPPEAPVTGYMFGK**GIYFADMVSK**SANYCHTSQGDPIGL
ILLGEVALGNMYELKHASHISKLPKGKHSVKGLGKTTPDPSANISLDGVDVPLGTGISSGVNDTSLLYNEYIVYDIAQV
NLKYLLKLKFNFKTSLW

Figure 4.11: Sequence coverage of Poly [ADP-ribose] polymerase 1 (PARP1_HUMAN). Data generated from four MALDI mass spectrometry analyses that were performed on the SDS fractions were used to investigate the sequence coverage of PARP. Of the full sequence of PARP (1014 amino acids), 16% (162 amino acids) were identified (**Red Bold**).

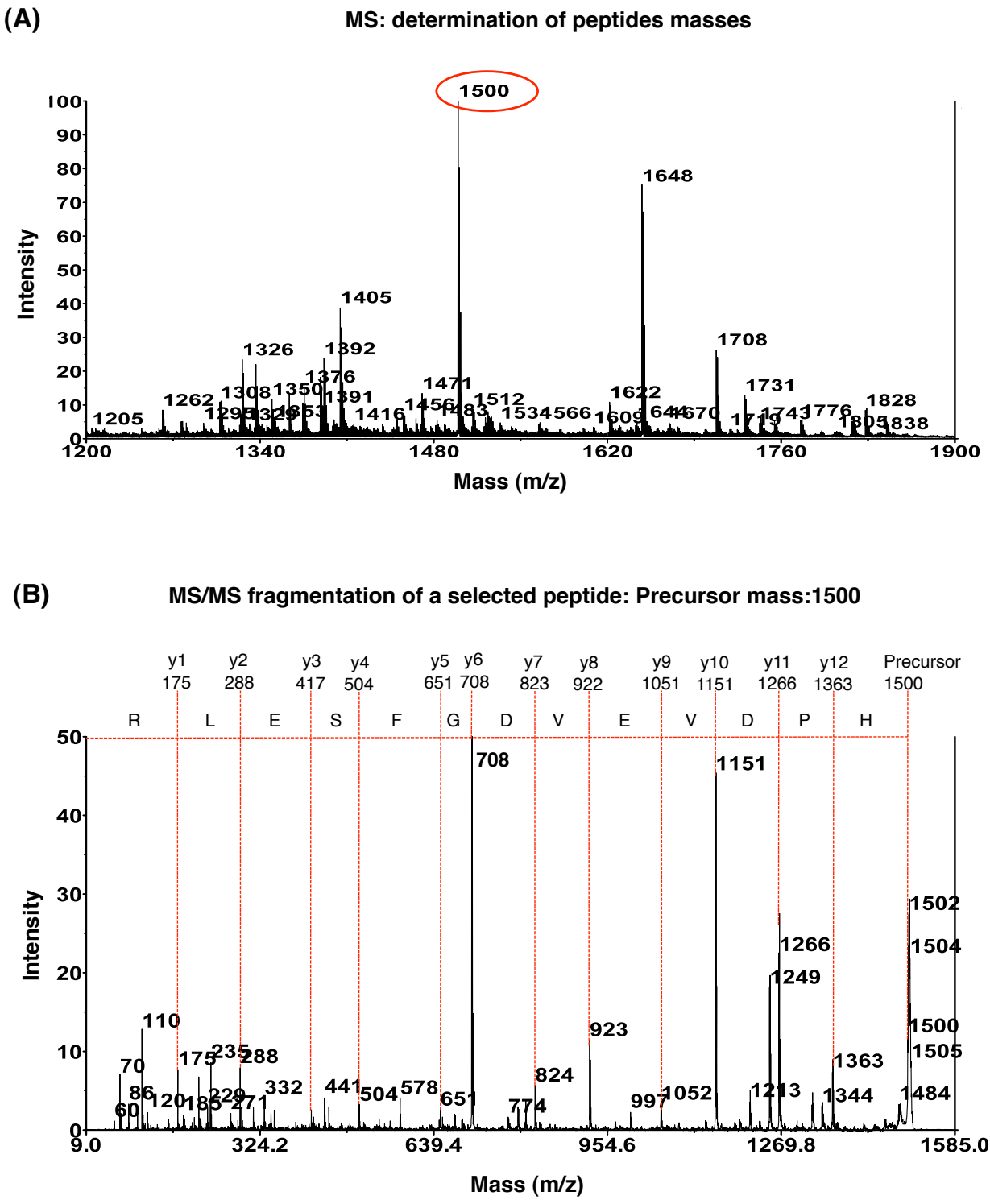


Figure 4.12: Example of MS and MS/MS spectra of Poly [ADP-ribose] polymerase 1 (PARP1_HUMAN) that was identified in the SDS fractions. Separated peptides were applied to MALDI mass spectrometry to measure their masses and to perform peptide fragmentation in order to identify their sequences. A peptide with mass of 1500 (A) was subjected to MS/MS analysis to identify its sequence (B). The peptide sequence based on the MS/MS spectra is read from the C terminus (**HPDVEVDGFSELR)** as only y ions were shown in this figure.

Table 4.5: A list of the ten proteins with the highest peptide counts that were identified in the NP40 fraction and/or in the SDS fraction.

Proteins identified in the NP40 fraction					
Protein name	**Accession Number**	**Peptide Count**	**TIS**	**BIS**	**BIS C.I. %**
Moesin	MOES_HUMAN	21	1346	101	100
Annexin A6	ANXA6_HUMAN	21	1286	97	100
Plastin-2	PLSL_HUMAN	20	1507	167	100
Vimentin	VIME_HUMAN	18	1559	192	100
Filamin-A	FLNA_HUMAN	17	1145	127	100
Actin, cytoplasmic 2	ACTG_HUMAN	17	1439	186	100
Myosin-9	MYH9_HUMAN	16	994	122	100
Ezrin	EZRI_HUMAN	16	964	88	100
Talin-1	TLN1_HUMAN	15	865	110	100
Heat shock cognate 71 kDa protein	HSP7C_HUMAN	15	945	109	100
Proteins identified in the SDS fraction					
Plectin	PLEC_HUMAN	35	2067	120	100
Vimentin	VIME_HUMAN	25	1898	123	100
Nuclear mitotic apparatus protein 1	NUMA1_HUMAN	24	1908	160	100
Lamin-B2	LMNB2_HUMAN	20	1359	124	100
Myosin-9	MYH9_HUMAN	20	1221	97	100
Lamin-B1	LMNB1_HUMAN	18	1573	160	100
Heterogeneous nuclear ribonucleoproteins A2/B1	ROA2_HUMAN	15	1296	229	100
Heterogeneous nuclear ribonucleoprotein M	HNRPM_HUMAN	14	1017	131	100
Actin, cytoplasmic 2	ACTG_HUMAN	13	1124	153	100
Poly [ADP-ribose] polymerase 1	PARP1_HUMAN	13	1019	117	100

Protein digest from the NP40 and SDS fractions were separately digested, separated on 2D nano-LC and analysed by MALDI mass spectrometry. Subsequently, MS/MS queries were searched against the Swiss- Prot database using the MASCOT database search engine embedded into GPS Explorer software. TIS: total ion score, BIS: best ion score, BIS C.I.%: best ion score confidence interval percentage.

an absolute call “present” or “marginal” were investigated. Of the 108 proteins, 93 (86%) had Affymetrix IDs and could be used for the analysis. Of these proteins, 92 (99%) had a cognate transcript expression in the CLL samples (range of Affymetrix signals: 356-47515, median: 2046). Figure 4.13 demonstrates the number of CLL samples that expressed the cognate transcript encoding these proteins. In addition, 10 examples of these proteins are shown in Table 4.6. This analysis increased the confidence of the protein identification based on a single peptide in a single experiment.

4.2.6 Analysis of the most and the least frequently identified proteins

Different factors such as poor solubility or low abundance are associated with the difficulty of identifying some proteins by mass spectrometry (Issaq, 2001, Brewis and Brennan, 2010). To explore why some of the 900 proteins were identified in only one MALDI mass spectrometry analysis with a single peptide ID (108 proteins), independent published transcriptomic data of six CLL samples (Huttmann et al., 2006) were used. The Affymetrix signals of the mRNA encoding the 900 proteins were used to potentially reflect the abundance of their protein products. The list of the 900 proteins were sorted in descending order according to the number of times the protein was detected in MALDI mass spectrometry analyses, followed by peptide count and best ion score (BIS) C.I.%. The Affymetrix signals (mean in six CLL samples) of the mRNA encoding the top proteins (n= 110: found in 10-20 MALDI MS runs with multiple peptides) were compared to that of the

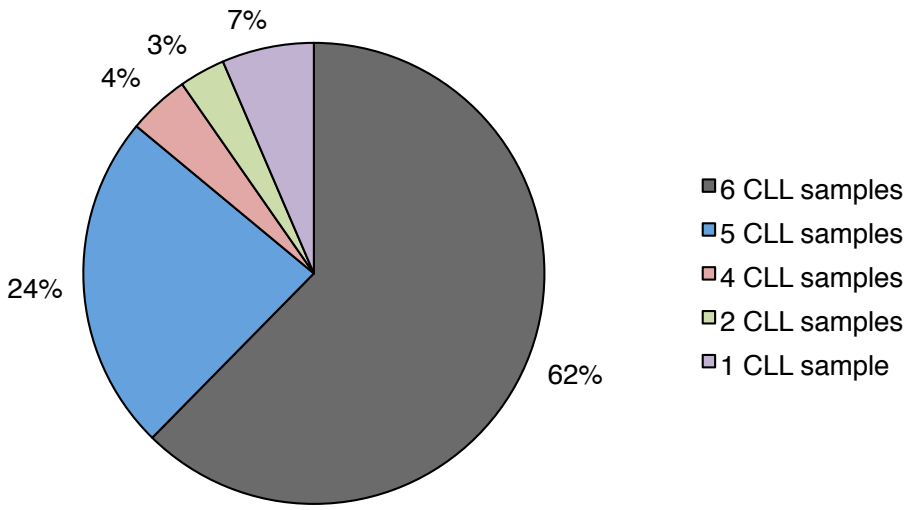


Figure 4.13: The number of CLL samples that express transcripts encoding the proteins that were identified with a single peptide in a single experiment. Transcriptomic data derived from 6 CLL samples were used to add more confidence to the protein identification based on one peptide in a single experiment (108 proteins). Of the 93 of these proteins that had match with Affymetrix IDs, 99% had a cognate transcript expressed in CLL samples. Transcriptomics data were taken from a previously published study (Huttmann et al., 2006).

Table 4.6: Transcriptomic data of CLL cells support protein identifications based on one peptide in a single MALDI mass spectrometry analysis.

Protein name	Accession Number	TIS	BIS	BIS C.I. %	Average of Affymetrix signal	Sample count
General vesicular transport factor p115	USO1_HUMAN	102	103	100	4091	6
Eukaryotic translation initiation factor 3 subunit I	EIF3I_HUMAN	94	94	100	3284	6
Chromatin target of PRMT1 protein	CHTOP_HUMAN	89	89	100	3677	6
Mps one binder kinase activator-like 1B	MOL1B_HUMAN	83	83	100	11781	6
26S protease regulatory subunit 4	PRS4_HUMAN	81	81	100	3434	6
E3 SUMO-protein ligase RanBP2	RBP2_HUMAN	47	48	99.92	3644	6
Kinectin	KTN1_HUMAN	48	48	99.92	6707	6
Microsomal glutathione S-transferase 3	MGST3_HUMAN	48	48	99.92	2141	6
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 6	NDUA6_HUMAN	48	48	99.91	3118	6
RNA-binding protein 4	RBM4_HUMAN	47	47	99.91	2676	6

Independently published transcriptomic data derived from six CLL samples (Huttmann et al,. 2006) were used to check whether CLL samples expressed the transcript encoding the proteins that were identified with a single peptide in a single experiment. The analysis showed that the vast majority of proteins that were identified with a single peptide in a single experiment had a transcript expressed in CLL samples. This table shows 10 examples of these proteins with the highest and lowest BIS C.I.%. TIS: total ion score, BIS: best ion score, BIS C.I.%: best ion score confidence interval percentage.

bottom proteins (n= 108: detected in a single MALDI MS run with a single peptide). The analysis demonstrated that the amount of mRNA of the most frequently detected proteins was 3.3 times more abundant than that of the least frequently detected proteins (p= 5.4 × 10^{-7} using unpaired Student's *t*-test; Figure 4.14). This analysis suggested that low abundance might have limited the identification of some proteins to only one MALDI mass spectrometry analysis based on a single peptide.

4.2.7 NP40 and SDS fractions: similar number of MS/MS spectra with different number of protein identifications

More proteins were consistently identified in the NP40 fraction than in the SDS fraction. To understand why this might occur, Protein Pilot software was used to analyse the number of spectra utilised and the number of distinct peptides identified in four NP40 fractions and four SDS fractions. Interestingly, a similar number of spectra were identified in the NP40 fractions (54%) and in the SDS fractions (46%). However, the number of distinct peptides sequenced from the NP40 fractions was almost 1.7-fold greater compared to the SDS fractions (Table 4.7). This implies that the difference in protein IDs reflects the internal complexity and relative protein abundance rather than any technical difference caused by the differential detergent extraction.

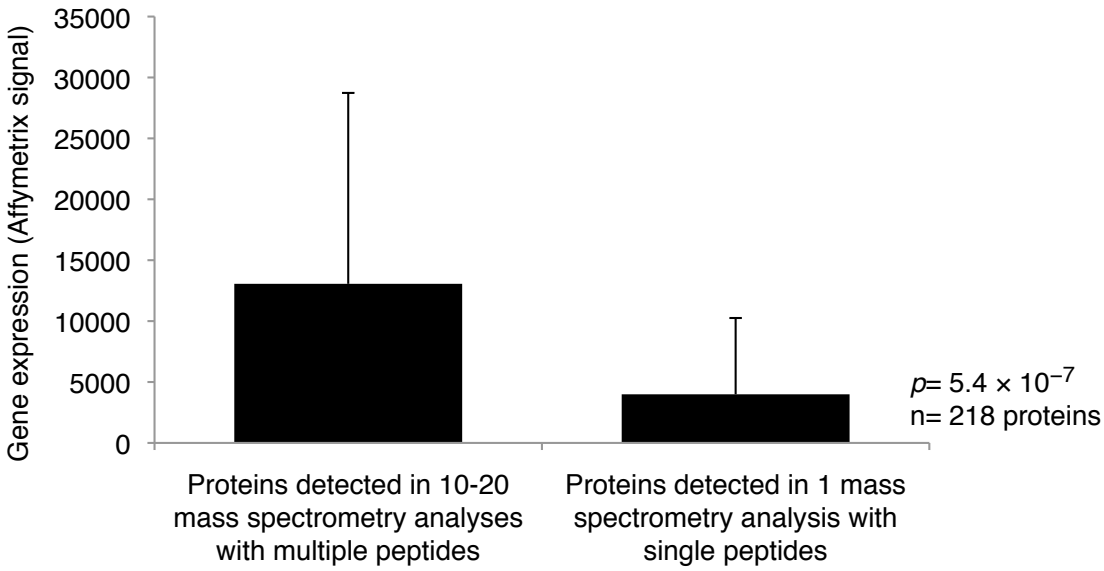


Figure 4.14: Abundance of the most frequently and least frequently identified proteins. Affymetrix signals of the mRNA encoding the most and least frequently detected proteins were used to potentially indicate the abundance of these two groups of proteins. The analysis was conducted on 218 proteins; 110 were found in 10-20 mass spectrometry analyses with multiple peptides, while 108 were detected in a single mass spectrometry analysis with a single peptide. This figure shows that the amount of mRNA of the most frequently detected proteins was 3.3 times more abundant than that of the least frequently detected proteins (mRNA average ± SD: 13062 ± 15670 versus 3991 ± 6246). The transcriptomics data were obtained from a previously published CLL study (Huttmann et al,. 2006)

Table 4.7: Summary of the total spectra and distinct peptides that were identified in the NP40 fractions and SDS fractions.

	Total numbers		% Total	
	NP40 Fractions	**SDS Fractions**	**NP40 Fractions**	**SDS Fractions**
Total spectra	25910	21236	55	45
Non-empty spectra	22777	18782	55	45
Spectra identified >95% confidence	11063	9387	54	46
Distinct peptides >95% confidence	3282	1959	63	37

The proteomic analyses of either 4 pooled SDS fractions or 4 NP40 fractions were separately analysed using Protein Pilot software coupled with the Paragon search algorithm to obtain the statistical summary of the spectra and distinct peptides. This table shows similar numbers of utilised spectra in the NP40 fractions and the SDS fractions, but a greater number of distinct peptides were identified in the NP40 fractions compared to the SDS fractions

4.2.8 Localisation of proteins identified in the NP40 fractions and SDS fractions

The rationale for developing the cellular fractionation method was that the NP40 fraction would be enriched with cytoplasmic proteins, while the SDS fraction would be enriched with nuclear proteins. This hypothesis was tested by using Gene Ontology data via the Quick GO-EBI tool (http://www.ebi.ac.uk/QuickGO/), to analyse the localisation of proteins that were uniquely found in each fraction (NP40 fractions = 541 proteins; SDS fractions = 138 proteins). Only proteins with one Gene Ontology location were analysed, as proteins with multiple locations would not be informative about the precision of the cellular fractionation procedure. Figure 4.15 shows that the NP40 fraction was predominantly comprised of proteins denoted cytoplasmic (48%) and also includes proteins denoted membrane (23%), nucleus (15%) and mitochondria (13%). In contrast, 82% of proteins from the SDS fraction were denoted as having a nuclear location.

4.2.9 Relationship between Affymerix signal and the feasibility of identifying a protein by mass spectrometry

In an attempt to establish a relationship between a transcript level of a gene and the possibility of identifying its protein product by mass spectrometry, gene expression profiles of six CLL samples derived from good and poor prognosis patients (Huttmann et al., 2006) and the CLL proteomics data (900 proteins) generated in this study were used. The analysis was only conducted for genes with an absolute call “present” that were identified in at

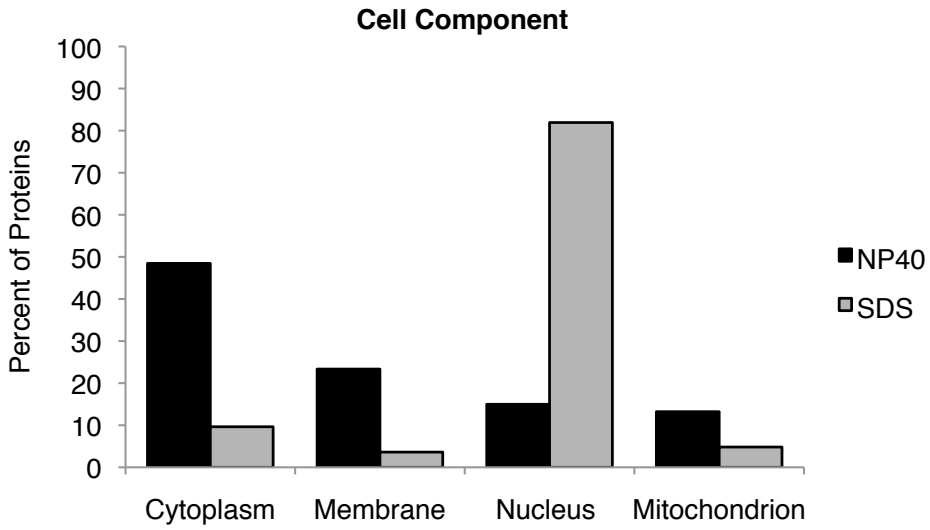


Figure 4.15: Cell component analysis of identified proteins. Localisation of proteins that were uniquely identified in the NP40 fractions and in the SDS fractions was determined from Gene Ontology data via the Quick GO-EBI tool. The analysis confirmed that the NP40 fraction was enriched with cytoplasmic proteins, while the SDS fraction was enriched with nuclear proteins.

least three different CLL samples. After converting Affymetrix IDs into UniProt IDs, transcripts were grouped according to their Affymetrix signal into six groups (100000-33000, 33000-11000, 11000-3666, 3666-1200, 1200-400 and 400-0). The CLL proteomics data were used to determine the percentage of identifications in each group (Figure 4.16). The analysis demonstrated that the higher the Affymetrix signal for a transcript, the greater the chance of identifying its cognate protein in the proteome list generated from my work. Table 4.8 shows the Affymetrix signal intensity of some genes known to be important in the pathology of CLL, and the probability of identifying their cognate protein products by MALDI mass spectrometry.

4.2.10 Qualitative proteomics and transcriptomic data highlighted proteins that may be relevant to the pathology of CLL

To identify proteins that may be relevant to the pathology of CLL, the CLL qualitative proteome generated in this study was compared to independently published Affymetrix gene array data of normal B-cells and CLL cells. First, previously published transcriptomics datat of CLL cells (Huttmann et al., 2006) were used to indicate whether or not the list of proteins reported in this study (900 proteins) had transcript expression in CLL cells. Of the 900 proteins, 125 did not have mach with the Affmetrix IDs and could not be used for the analysis. Of the rest (775 proteins), 759 proteins 98% had transcript expression in CLL cells (Appedix 2).

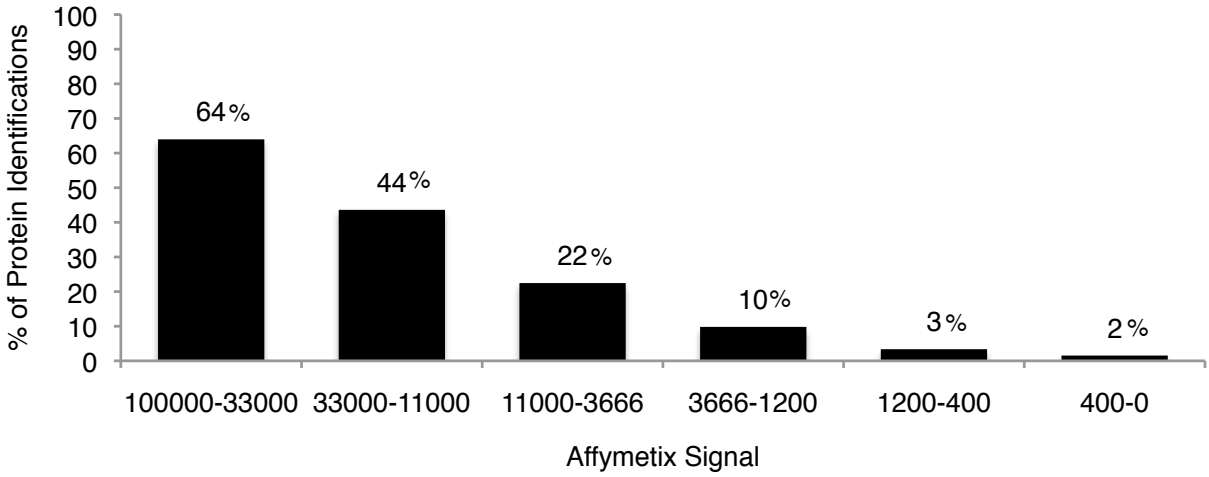


Figure 4.16: Relationship between Affymerix signal and the feasibility of identifying a protein by mass spectrometry. Transcriptomic data derived from six CLL samples and my CLL proteomic data were used to investigate whether the Affymetrix signal for a mRNA can be informative of the possibility of identifying its cognate protein by mass spectrometry. The analysis suggested that the higher the Affymetrix signal for a transcript the greater the chance of identifying its protein. The transcriptomics data were obtained from previously published CLL study (Huttmann et al,. 2006)

Table 4.8: Affymetrix signal of mRNA encoding some important proteins in CLL and the probability of identifying these proteins by MALDI mass spectrometry

Protein name	Accession number	Affymetrix signal Mean ± SD	Probability of identification by MALDI mass spectrometry
Induced myeloid leukemia cell differentiation protein Mcl-1	MCL1_HUMAN	7828 ± 2493	22%
Apoptosis regulator Bcl-2	BCL2_HUMAN	5038 ± 1585	22%
Transcription factor p65	TF65_HUMAN	1498 ± 332	10%
Nuclear factor NF-kappa-B p50 subunit	NFKB1_HUMAN	1334 ± 618	10%
CD38	CD38_HUMAN	1081 ± 305	3%
Tyrosine-protein kinase ZAP-70	ZAP70_HUMAN	788 ± 223	3%

Affymetrix signals of these genes were obtained from eight CLL samples (including poor and good prognosis CLL) (Huttmann et al., 2006). Protein name and accession number were obtained from Universal Protein Resource (UniProt) http://www.uniprot.org/ . SD: standard deviation.

Next, transcriptomic data of four CLL samples (Huttmann et al., 2006) and three normal B-cells samples (Hutcheson et al., 2008) were compared to identify genes that were preferentially expressed in CLL samples. The analysis was performed on genes with an absolute call of either “present” or “marginal”. The comparison showed that 6887 genes were expressed in both normal B-cells and CLL cells. In contrast, 503 genes appeared to be exclusively expressed in normal B-cells and 800 genes were specifically expressed in CLL cells (Figure 4.17). Encouragingly, *CD5* was among the 800 genes that were preferentially expressed in CLL cells. After determining genes that had the potential for being specific to CLL, my CLL proteomics data was utilised to investigate whether the protein products of these genes were evident in the CLL proteome. Of the 800 genes, only 20 (2.5%) encoded ptoteins were found in CLL proteome generated in this study (Figure 4.18). The absent expression of the 20 genes was confirmed in six normal B-cell samples that were examined by Hutcheson *et al.*, (2008). The protein products of these 20 genes are shown in Table 4.9.

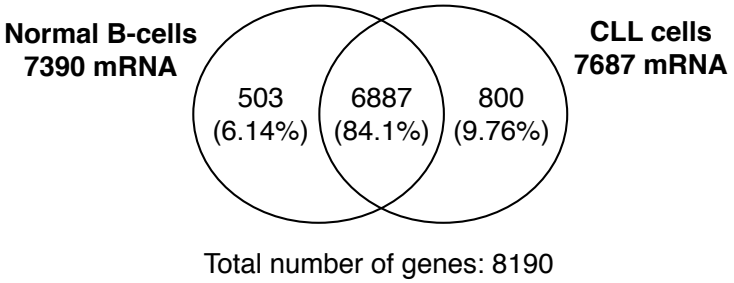


Figure 4.17: Venn diagram of genes expressed by normal B-cells and those expressed by CLL cells. Affymetrix gene array data derived from three normal B-cell samples were compared with that of four CLL samples. This figure demonstrates that 800 genes were preferentially expressed in CLL cells. The normal B-cells transcriptomics data were taken from Hutcheson et al., (2008) and the CLL cells transcriptomics data were obtained from Huttmann et al., (2006).

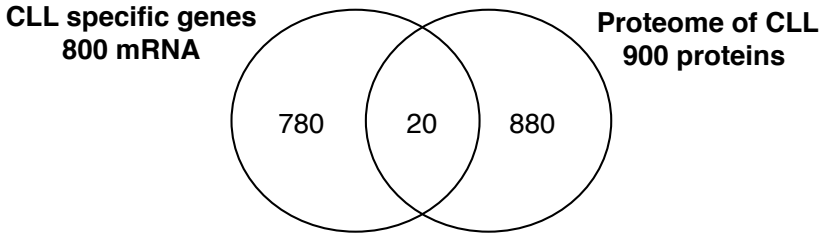


Figure 4.18: Venn diagram of CLL-specific mRNA and the CLL proteome generated in this study. In order to investigate whether the proteins encoded by the CLL-specific mRNA were identified in the CLL protein samples, the two data sets were compared with each other. The analysis showed that of the 800 mRNA, 20 were also detected in CLL samples at the protein level. The transcriptomics data were obtained from previously published CLL study (Huttmann et al,. 2006)

Table 4.9: Proteins whose transcripts expression is restricted to CLL cells but not normal B-cells

Protein name	Accession number	Biological process
Protein 4.1	41_HUMAN	Actin cytoskeleton organization
Calponin-3	CNN3_HUMAN	Actomyosin structure organization
ATP-dependent RNA helicase DDX3Y	DDX3Y_HUMAN	Helicase activity
2-oxoglutarate dehydrogenase, mitochondrial	ODO1_HUMAN	Metabolic process
CWF19-like protein 1	C19L1_HUMAN	Metabolic process
Tropomodulin-3	TMOD3_HUMAN	No information available
Heterochromatin protein 1-binding protein 3	HP1B3_HUMAN	Nucleosome assembly
Histone H1.4	H14_HUMAN	Nucleosome assembly
Histone H4-like protein type G	H4G_HUMAN	Nucleosome assembly
Pleckstrin homology domain-containing family A member 2	PKHA2_HUMAN	Positive regulation of cell-matrix adhesion
Exocyst complex component 5	EXOC5_HUMAN	Protein transport
Ras-related protein Rab-8B	RAB8B_HUMAN	Protein transport
Zinc finger protein Aiolos	IKZF3_HUMAN	Regulation of apoptotic process
3-ketoacyl-CoA thiolase, mitochondrial	THIM_HUMAN	Negative regulation of apoptotic process
Amyloid beta A4 precursor protein-binding family B member 1-interacting protein	AB1IP_HUMAN	Signal transduction
Guanine nucleotide-binding protein subunit alpha-13	GNA13_HUMAN	Signal transduction
Stathmin	STMN1_HUMAN	Signal transduction
T-cell surface glycoprotein CD5	CD5_HUMAN	Transmembrane signaling receptor activity
Thyroid hormone receptor-associated protein 3	TR150_HUMAN	Transcription cofactor activity
Eukaryotic translation initiation factor 1A, Y-chromosomal	IF1AY_HUMAN	Translation initiation factor activity

This table shows the protein products of the 20 genes that were specific to CLL cells but not normal B-cells. The biological process associated with these proteins was extracted from Gene Ontology using Quick GO-EBI tool and the list was sorted according to the biological processes. The transcriptomics data were obtained from previously published studies (Huttmann et al,. 2006, Hutcheson et al., 2008).

4.3 Discussion

After generating the NP40 fractions and the SDS fractions from good quality primary CLL samples, the next step was to introduce these proteins extracts into a proteomics workflow. In this project, a gel-free proteomics approach was chosen, in which the separation of digested proteins was performed using 2D nano-LC. This type of proteomics approach has the potential to identify a large number of proteins compared with gel-based proteomics using 2DE. This project returned 900 proteins, whereas other CLL proteomics studies that used 2DE as a protein separation method identified much lower numbers of proteins (e.g. 17 proteins, 31 proteins and 60 proteins) (Voss et al., 2001, Perrot et al., 2011, Cochran et al., 2003).

The application of 2D nano-LC using SCX and RP columns has gained popularity in the proteomics field as it separates peptides based on two different properties; charge and hydrophobicity (Yates et al., 2009). This in turn was shown to facilitate the identification of many proteins simultaneously (Yates et al., 2009, Washburn et al., 2001). In the context of CLL, only one previous CLL proteomics study exploited this technology to study crude membrane extracts leading to the identification of 695 proteins (Barnidge et al., 2005b). In this project, 2D nano-LC was an effective tool to sufficiently separate the peptide mixture to enable MALDI mass spectrometry to produce fragmentation data that facilitated the identification of 900 proteins.

Of the published CLL proteomics studies, four out of seven identified hundreds of proteins. Miguet et al. (2009) reported 371 proteins based on one or more peptides with FDR of <1% in membrane extracts. In addition, Boyd et al., (2003) identified 500 proteins with multiple or single peptides in plasma membrane fractions. Barnidge et al. (2005a) reported 326 proteins in cytosolic fractions and 212 proteins in membrane fractions. The total number of proteins was 538, but it was not specified whether this number excluded redundant proteins that were found in both fractions. None of these studies specified how many proteins were identified with multiple or single peptides. Furthermore, only the study by Miguet et al. (2009) reported the FDR of the detected proteins, which is important to indicate the percentage of proteins that may be falsely identified. In the absence of FDR calculations and statements about the number of proteins that were detected with either multiple or single peptides, it is difficult to evaluate the quality of the reported proteins by these studies.

The largest set of published CLL proteins was 695 that were detected with one or more peptides by Barnidge et al., (2005b) in crude membrane extracts. This study did not report the FDR of these proteins nor the number of proteins that were detected on the basis of multiple or single peptides. In contrast, in the current study, 900 proteins were reported, of which 625 were identified with multiple peptides and a minimum ion score of 30 corresponding to an ion score confidence interval of 95% or more. This group of proteins had a 0% FDR. Furthermore, for the 275 proteins that were

identified with single peptides a more stringent ion score of 47 was applied corresponding to an ion score confidence interval of more than 99.91%. This limited the FDR in these 275 proteins to 3.2%.

Different proteomics approaches are complementary. In this regard, a study used different proteomics approaches (LC-ESI/MS/MS and LC-MALDI/MS/MS) to study the proteome of a biological sample showed an overlap of 63% between the results generated by these two different approaches (Bodnar et al., 2003). In addition, another group studied membrane proteins using three different proteomics approaches (2D LC-ESI/MS/MS, LC-MALDI/MS/MS and 1DE-MS or 1DE-MS/MS) and found an overlap of 9% in the proteins identified by these approaches (Zhang et al., 2004). In the context of the published CLL proteome, 50% of the CLL proteins reported by Miguet et al. (2009) using 1DE-1D-LC/ESI/MS/MS were detected in the present study. In addition, 27% of CLL proteins reported by Barnidge et al. (2005b) using 2D-LC/ESI/MS/MS were identified in my study. From this one may conclude that different proteins can be identified using different proteomics approaches and thus exploiting a combination of different approaches can be an effective strategy to increase proteome coverage.

The increase in protein abundance was previously shown to be associated with an increase in the proteolytic peptides and thus MS/MS spectra (Liu et al., 2004). This current study showed that proteins that were identified with a large peptide count or TIS were those that were frequently

identified in MALDI mass spectrometry analyses and *vice versa*. In fact, there was a good correlation between the frequency at which proteins were identified by MALDI mass spectrometry and their peptide counts or TIS. This might indicate that the proteins that were commonly identified in MALDI mass spectrometry were those that were found in high abundance in CLL cells, whereas proteins that were rarely identified by MALDI mass spectrometry were those that exist at low concentrations in CLL cells. This was supported by the observation that the mRNA expression encoding the least frequently identified proteins were three times less than that encoding the most frequently identified proteins.

Given the nature of the proteome complexity, some proteins were identified with the best criteria and some were identified with the least, but acceptable, criteria. The latter were those protein identified with a single peptide in a single MALDI mass spectrometry analysis (108 proteins). CLL transcriptomic data (Huttmann et al., 2006) were found to be very useful to explore how accurate this type of protein identification was likely to be. The analysis demonstrated that CLL cells expressed the cognate transcript of the vast majority (99%) of these proteins. In addition, transcriptomic data suggested that the low abundance of these proteins restricted their identification by MALDI mass spectrometry to only one analysis based on a single peptide.

This chapter has described a method for calculating the probability of identifying a protein by mass spectrometry based on the Affymetrix signal, i.e. relative transcript level. Although mRNA quantity *per se* is not sufficient to predict protein abundance (Gygi et al., 1999b), the analysis showed that proteins with high levels of transcript were more likely to be identified by mass spectrometry. Almost 64% of proteins with high transcription signals (Affymetrix signal: 33,000-100,000) were identified by mass spectrometry, whereas only 2% of proteins with very low transcript signals (Affymetrix signal: ≤400) were identified using this approach. Some proteins known to be important in the pathology of CLL, for example, CD38 and ZAP-70, have low levels of Affymetrix signal (1081±305, 788±223 respectively) (Huttmann et al., 2006). Thus, the probability of finding these proteins by mass spectrometry was very low (3%). This analysis can be useful to indicate whether particularly target proteins are likely to be identified by mass spectrometry before starting a targeted proteomics project.

In this study, a strategy was described for combining transcriptomic and proteomics data to identify proteins with potential relevance to the pathology of CLL. By integrating my CLL proteomics data with published CLL and normal B-cells transcriptomic data (Huttmann et al., 2006, Hutcheson et al., 2008), a list of 20 proteins that were expressed in CLL but not detected in the normal B-cell transcriptome were generated.

Some of these 20 proteins were linked to an important biological role in cancer. For example, pleckstrin homology domain-containing family A member 2 (TAPP-2) was reported to participate in phosphatidylinositol 3-kinase (PI3K) signalling, an important pathway in B-lymphocyte activation and proliferation, following BCR stimulation (Marshall et al., 2002). Interestingly, TAPP-2 protein was reported to associate with poor prognosis CLL ($ZAP\text{-}70^{+}$ CLL and U-CLL) (Costantini et al., 2009). In addition, zinc finger protein Aiolos was linked to apoptosis, where it induces BCL2 expression and prevents cell death in T-cells (Romero et al., 1999). Furthermore, Stathmin was shown to be involved in cell differentiation and proliferation and is highly expressed in a number of different types of cancer such as ovary cancer and breast cancer (Sherbet and Cajone, 2005, Price et al., 2000, Curmi et al., 2000). This may indicate that this strategy is useful to highlight proteins with potential involvement in CLL using proteomics. Of these 20 proteins, 2 were selected on the basis of the quality of their mass spectrometry data for further study on normal B-cells and CLL cells (see chapter six).

A successful qualitative proteomics workflow that allowed identification of 900 proteins in CLL samples was developed in this chapter. The next step was to modify this workflow to allow relative quantification of proteins in the NP40 fractions and the SDS fractions from poor prognosis CLL versus good prognosis CLL. The next chapter explains the processes that allowed the identification of proteins with aberrant expression in the two forms of CLL.

Chapter Five

Quantitative Proteomics Analysis of Primary CLL Samples

5.1 Introduction

Although genetic predisposition contributes to the diverse clinical outcomes seen in CLL, differential protein expression has been shown to impact heavily on the clinical progression and response to treatment in this disease. In this regard, anti-apoptotic proteins and the NF-κB transcription factor subunits have been shown to be important (Kitada et al., 1998, Pepper et al., 2008, Hewamana et al., 2008). These studies all used conventional approaches based on detection of single proteins using specific antibodies. In contrast, quantitative proteomics approaches allow simultaneous global relative quantification of expressed proteins in a particular disease tissue; this has the potential to elucidate more of the key proteins driving the pathology of a disease such as CLL. Four CLL proteomics studies used different quantitative proteomics approaches to compare the proteome of poor prgnosis CLL with that of good prognosis CLL and reported number of proteins with possible involvment in CLL (discussed in detailes in 1.3)

The aim of the quantitative proteomics analysis in this chapter was to compare the proteomics expression profiles of the aggressive and indolent forms of CLL. CD38 expression was used to discriminate between poor prognosis CLL and good prognosis CLL. However, fuether analyses on the basis of other prognostic markers (ZAP-70, mutational status of *IGHV* genes and CLL satges) were also conducted

Given the success of the qualitative proteomics workflow (described in chapter four), it was modified by introducing iTRAQ labelling of the NP40 fractions and the SDS fractions to enable relative quantification of CLL proteins. Figure 5.1 illustrates the workflow that was used for the quantitative proteomics analysis.

5.2 Results

5.2.1 Quantification of CLL proteins in relation to a reference sample

To perform a relative quantification analysis on the proteome of poor prognosis CLL ($CD38^+$ CLL) and on the proteome of good prognosis CLL ($CD38^-$ CLL), the experiments were designed so that the proteome of both groups of patient samples would be compared to one reference sample. One reference sample was used for the analysis of the NP40 fractions and another reference sample was used for the SDS fractions analysis. The NP40 reference sample was prepared from the NP40 fraction of one CLL sample. However, in the course of optimising the quantitative proteomics workflow, the SDS reference sample was prepared from a mixture of the SDS fractions from the 12 CLL samples (poor prognosis and good prognosis CLL). After generating the relative quantification data of the CLL proteome, relative protein expression of poor prognosis CLL samples were compared to that of good prognosis CLL samples to identify proteins with altered expression in the two forms of CLL.

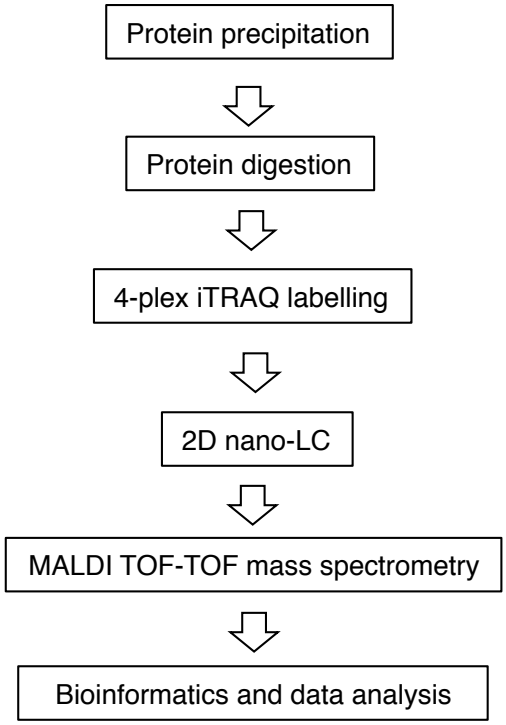


Figure 5.1: Quantitative proteomics workflow. Equivalent amounts of protein from $CD38^{+}$ CLL and $CD38^{-}$ CLL samples (NP40 fractions or SDS fractions) with a reference sample were separately precipitated using the 2D Clean-Up kit and digested with trypsin. The peptides were labelled with 4-plex iTRAQ reagents. Labelled samples were combined and separated on 2D nano-LC followed by analysis by MALDI mass spectrometry. Generated MS/MS spectra were searched against the Swiss-Prot database using ProteinPilot with the Paragon and ProGroup algorithms.

5.2.2 Sample labelling with iTRAQ reagents

Following protein precipitation and protein digestion by trypsin, peptides were labelled with iTRAQ reagents. In each iTRAQ LC-MALDI mass spectrometry experiment equivalent amount of peptide (20µg) from either the NP40 fraction or SDS fraction of three different CLL samples and the corresponding reference sample were labelled with iTRAQ reagents. Reference samples were consistently labelled with iTRAQ regents 114 or 117 (117 was used to label the NP40 reference sample and 114 was used to label the SDS reference sample). In contrast, an effort was made to avoid using one iTRAQ reagent to consistently label one group of the CLL samples (either $CD38^+$ or $CD38^-$). This was done to avoid the artifact that might result from constantly using a particular label for $CD38^+$ or $CD38^-$ CLL samples. Figure 5.2 demonstrates the different ways used for iTRAQ labelling.

5.2.3 Data analysis

Following iTRAQ LC-MALDI mass spectrometry experiments, MS/MS spectra from labelled peptides were searched against the Swiss-Prot database using ProteinPilot 2.0.1 software (Applied Biosystems) with the Paragon and ProGroup algorithms. The search was restricted to human taxonomy with trypsin specificity. Cysteine alkylation was performed with MMTS. Two different scores were reported for each protein: unused ProtScore and total ProtScore. The unused ProtScore specifically measures all of the peptide evidence for a protein that does not contribute to the

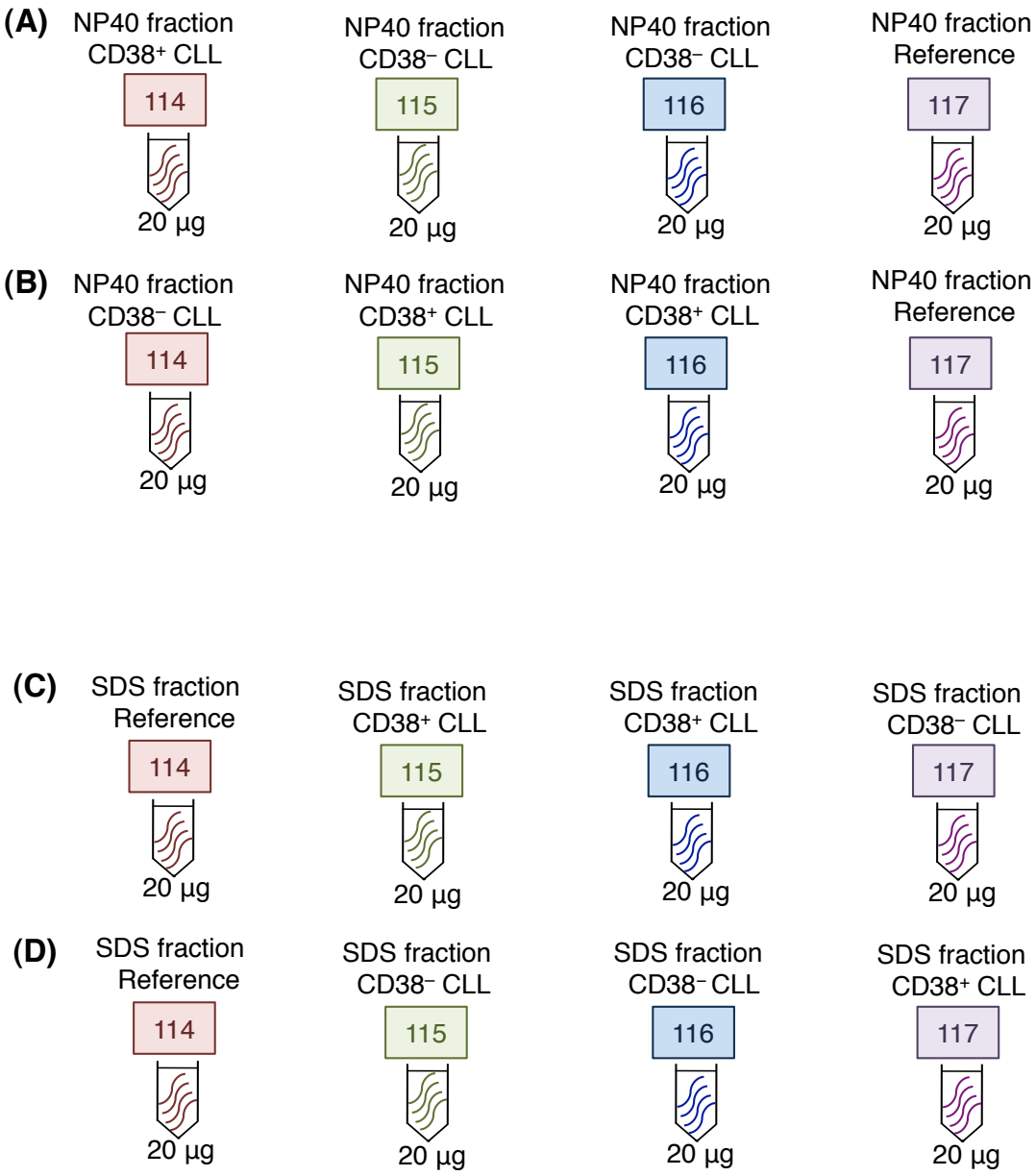


Figure 5.2: Different iTRAQ labelling of CLL samples. iTRAQ reagents were used in a different order to label the NP40 fractions of $CD38^+$ CLL samples and $CD38^-$ CLL samples (A and B). Similarly, iTRAQ reagents were used in different order to label the SDS fractions of $CD38^+$ CLL samples and $CD38^-$ CLL samples (C and D).

identification of a higher ranked protein. To consider a protein for further analysis it must be identified with a minimum unused ProtScore >1.3 and at least one peptide with a 95% confidence score.

5.2.4 Relative quantification of the CLL proteome

In total 12 iTRAQ LC-MALDI mass spectrometry runs were performed for 12 CLL samples. Six runs were carried out for the NP40 fractions and six runs were performed for the SDS fractions. The experiments were designed to study the relative protein expression of six $CD38^{+}$ CLL samples versus six $CD38^{-}$ CLL samples. In total, 655 proteins were quantified in at least three CLL samples with one peptide or more (confidence score ≥95%) and unused ProtScore >1.3 (Appendix 3). These proteins were identified with FDR 3%. From the NP40 fractions 488 proteins were quantified, whereas 288 proteins were quantified in the SDS fractions. 121 proteins were quantified in both fractions (Figure 5.3A). Of the 655 proteins, 386 proteins (59%) were identified with 2 peptides or more, while the rest (277 proteins, 41%) were identified with a single peptide (Figure 5.3B and C). The peptide count presented here was the mean peptide count in six iTRAQ LC-MALDI experiments with either NP40 fractions or SDS fractions.

5.2.5 Demonstration of the relative quantification of CLL proteins using iTRAQ reagents

Identification and relative quantification of labelled peptides was carried out using MALDI mass spectrometry and ProteinPilot software coupled

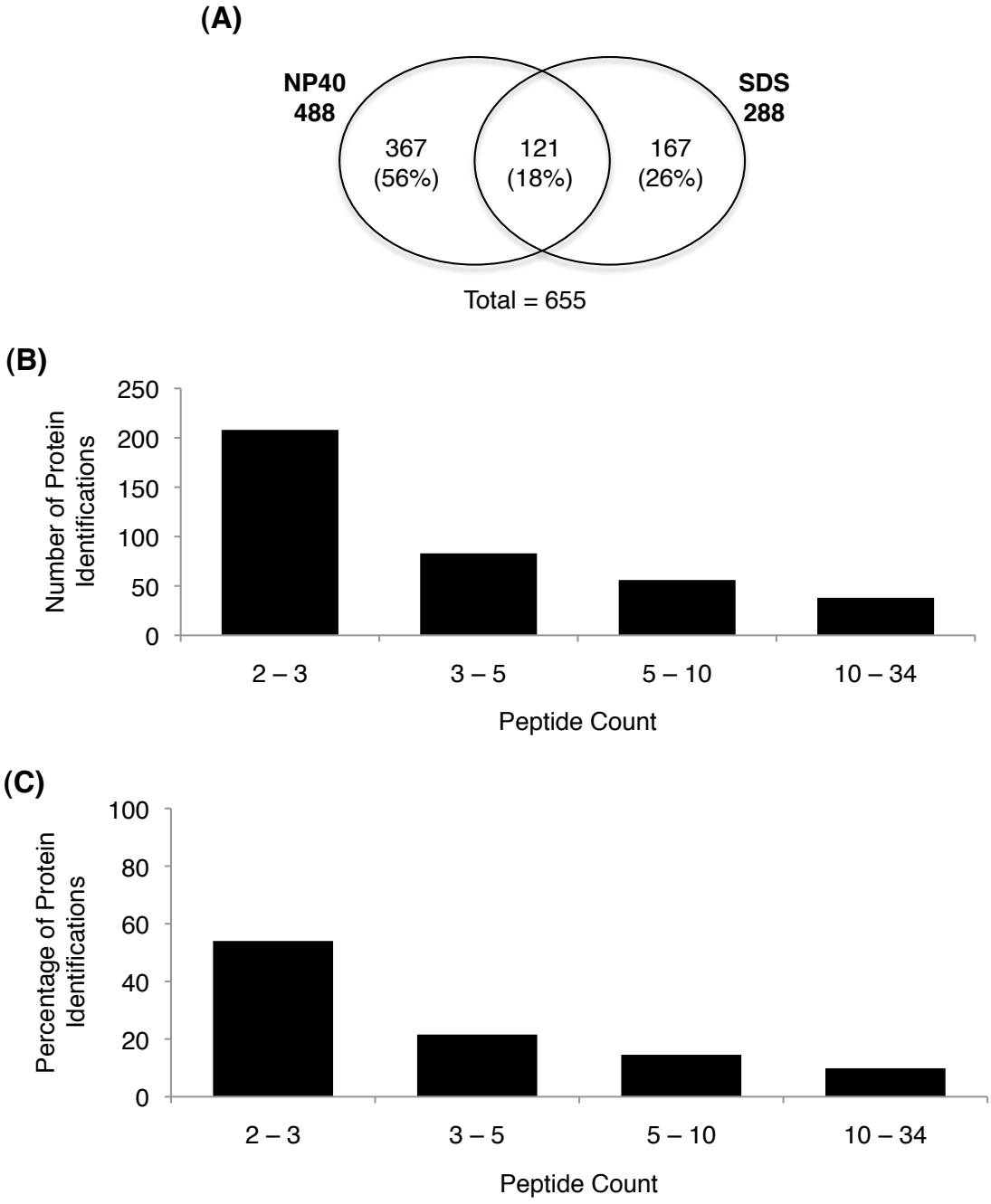


Figure 5.3: Venn diagram of proteins that were identified and quantified using iTRAQ reagents and LC-MALDI mass spectrometry in CLL samples. The total number of non redundant proteins that were identified and quantified in the NP40 fractions and the SDS fractions of three or more CLL samples was 655 proteins (A). Of these proteins, 386 (59%) were identified with 2 or more peptides (B and C).

with the Paragon and ProGroup algorithms. The assessment of protein expression in CLL samples was performed on the basis of the relative quantification of their corresponding peptides. Two examples have been chosen to illustrate the steps involved in quantification of peptides and proteins. From the NP40 fractions, Table 5.1 shows an example of the relative quantification of peptides that were used to measure the relative abundance of nucleophosmin in three CLL samples. The MS/MS spectrum of one of these peptides is shown in Figure 5.4. From the SDS fractions, Table 5.2 shows an example of the relative quantification of peptides that were used to evaluate the expression of splicing factor proline- and glutamine-rich in three CLL samples. The MS/MS spectrum of one of these peptides is shown in Figure 5.5.

Following the relative quantification of the proteins in all tested CLL samples, the protein expression of the poor prognosis CLL samples was compared with that of the good prognosis CLL samples. This was performed in order to identify proteins with altered expression in the two forms of CLL. The relative quantification of nucleophosmin and PSF in the 12 CLL samples that were analysed in this study is shown in Table 5.3.

It is worth mentioning here that not all peptides that were used for protein identification were used for protein relative quantification as some of these peptides contributed to the identification of multiple proteins (common peptides). This often happened when homologous proteins such as histone

Table 5.1: Examples of relative quantification of peptides that were used to determine the relative expression of nucleophosmin in CLL samples (NP40 fractions)

		iTRAQ relative quantification		
Peptide sequence	**Confidence %**	**$CD38^+$ CLL**	**$CD38^+$ CLL**	**$CD38^-$ CLL**
GPSSVEDIK	99	0.85	0.98	0.83
GPSSVEDIK	99	0.87	0.98	0.74
GPSSVEDIK	99	0.75	0.86	0.76
GPSSVEDIK	99	0.80	0.94	0.72
GPSSVEDIK	99	0.73	0.92	0.75
MQASIEK	99	0.93	0.98	0.81
MQASIEK	99	0.94	0.99	0.73
MQASIEK	99	0.88	0.85	0.72
MTDQEAIQDLWQWR	99	0.99	1.27	1.06
MTDQEAIQDLWQWR	99	0.97	1.10	0.90
VDNDENEHQLSLR	99	0.82	1.10	0.79
VDNDENEHQLSLR	99	0.83	0.89	0.87
FINYVK	98	0.78	0.97	0.71
FINYVK	98	0.71	0.85	0.66
FINYVK	97	0.97	1.04	0.86
FINYVK	97	1.01	1.13	0.81
GPSSVEDIK	87	1.11	0.99	1.27
Protein nucleophosmin	**Identified with 6 different peptides**	**0.82**	**0.94**	**0.76**

Identification and relative quantification of labelled peptides were performed by MALDI mass spectrometry and the ProteinPilot software using Paragon and ProGroup algorithms. This table shows relative quantification of peptides that were mapped to nucleophosmin. Relative quantification of these peptides were used to determine the relative quantification of nucleophosmin in the NP40 fractions of CLL samples. The examples shown in this table were extracted from one iTRAQ LC-MALDI mass spectrometry experiment that was carried out on the NP40 fractions of three CLL samples relative to the NP40 reference sample.

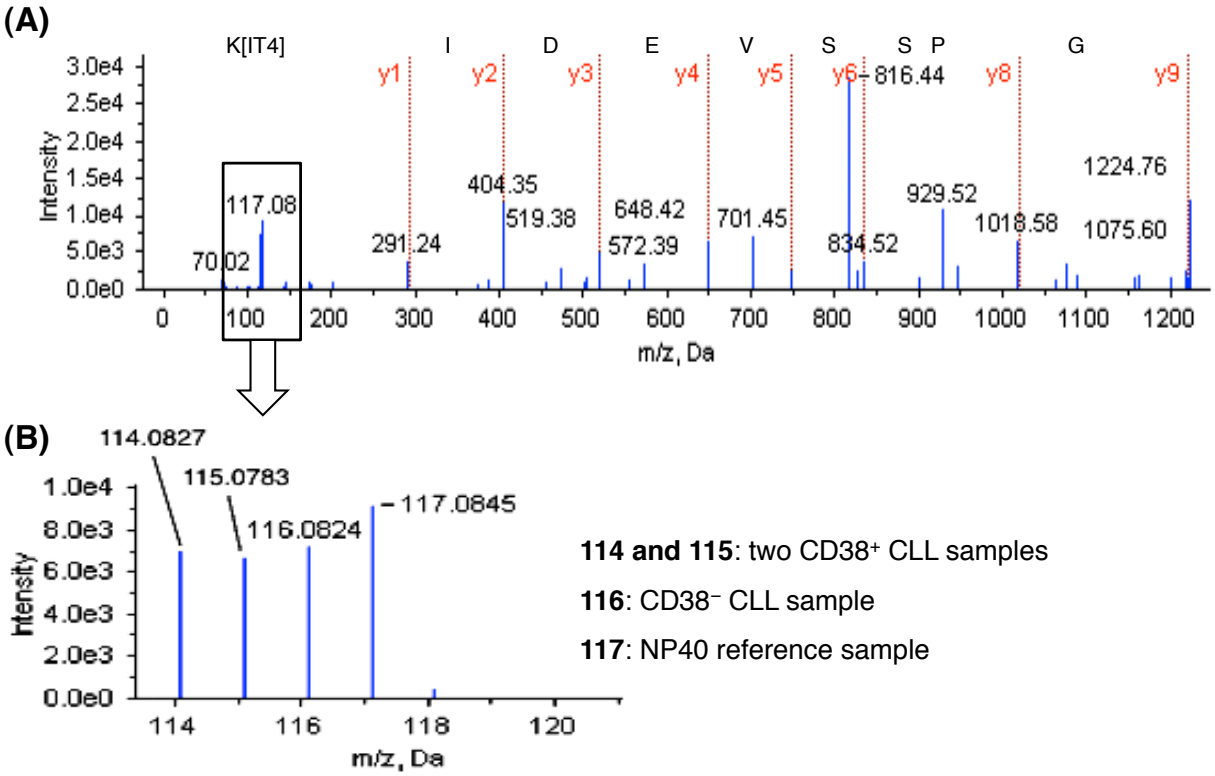


Figure 5.4: Identification and relative quantification of a specific peptide in nucleophosmin. Following separation of labelled peptides by 2D nano-LC, they were analysed by MALDI mass spectrometry and ProteinPilot software using the Paragon and ProGroup algorithms. This figure shows MS/MS spectra of a specific peptide derived from nucleophosmin (**GPSSVEDIK**) (A) as well as the relative quantification of this peptide in the NP40 fractions of three CLL samples in relation to the NP40 reference sample (B). For simplicity only y ions were shown in these MS/MS spectra. The sequence identified in the MS/MS spectra is read from the C-terminus.

Table 5.2: Examples of relative quantification of peptides that were used to measure the relative expression of splicing factor proline- and glutamine-rich (PSF) (SDS fractions of CLL samples).

		iTRAQ relative quantification		
Peptide sequence	**Confidence %**	**$CD38^+$ CLL**	**$CD38^+$ CLL**	**$CD38^-$ CLL**
AELDDTPMR	99	1.40	1.38	1.06
ANLSLLR	99	1.41	1.36	1.43
CSEGVFLLTTTPR	99	0.86	1.06	0.88
EEEMMIR	99	0.67	0.97	0.86
FAQHGTFEYEYSQR	99	0.78	1.49	1.64
FATHAAALSVR	99	1.12	1.33	1.20
FATHAAALSVR	99	0.97	1.51	1.38
FATHAAALSVR	99	1.04	1.23	1.19
FGQGGAGPVGGQGPR	99	1.11	1.29	1.40
FGQGGAGPVGGQGPR	99	1.23	1.24	1.12
FGQGGAGPVGGQGPR	99	0.98	1.16	1.20
FGQGGAGPVGGQGPR	99	1.15	1.28	1.22
GFGFIK	99	0.94	1.57	1.52
RMEELHNQEMQK	99	0.76	0.79	1.07
SPPPGMGLNQNR	99	0.74	1.03	1.30
YGEPGEVFINK	99	0.87	1.35	1.18
YGEPGEVFINK	99	0.97	1.50	1.25
ANLSLLR	98	0.79	1.27	0.98
SPPPGMGLNQNR	98	0.88	1.00	1.26
EMEEQMR	80	0.83	0.96	1.21
Protein PSF	**Identified with 13 different peptides**	**0.97**	**1.40**	**1.35**

Identification and relative quantification of labelled peptides were performed by MALDI mass spectrometry and ProteinPilot software using the Paragon and ProGroup algorithms. This table shows relative quantification of peptides that were mapped to PSF. Relative quantification of these peptides were used to determined the relative quantification of PSF in the SDS fractions of CLL samples. The examples shown in this figure were extracted from one iTRAQ LC-MALDI mass spectrometry experiment that was carried out on the SDS fractions of three CLL samples and the SDS reference sample.

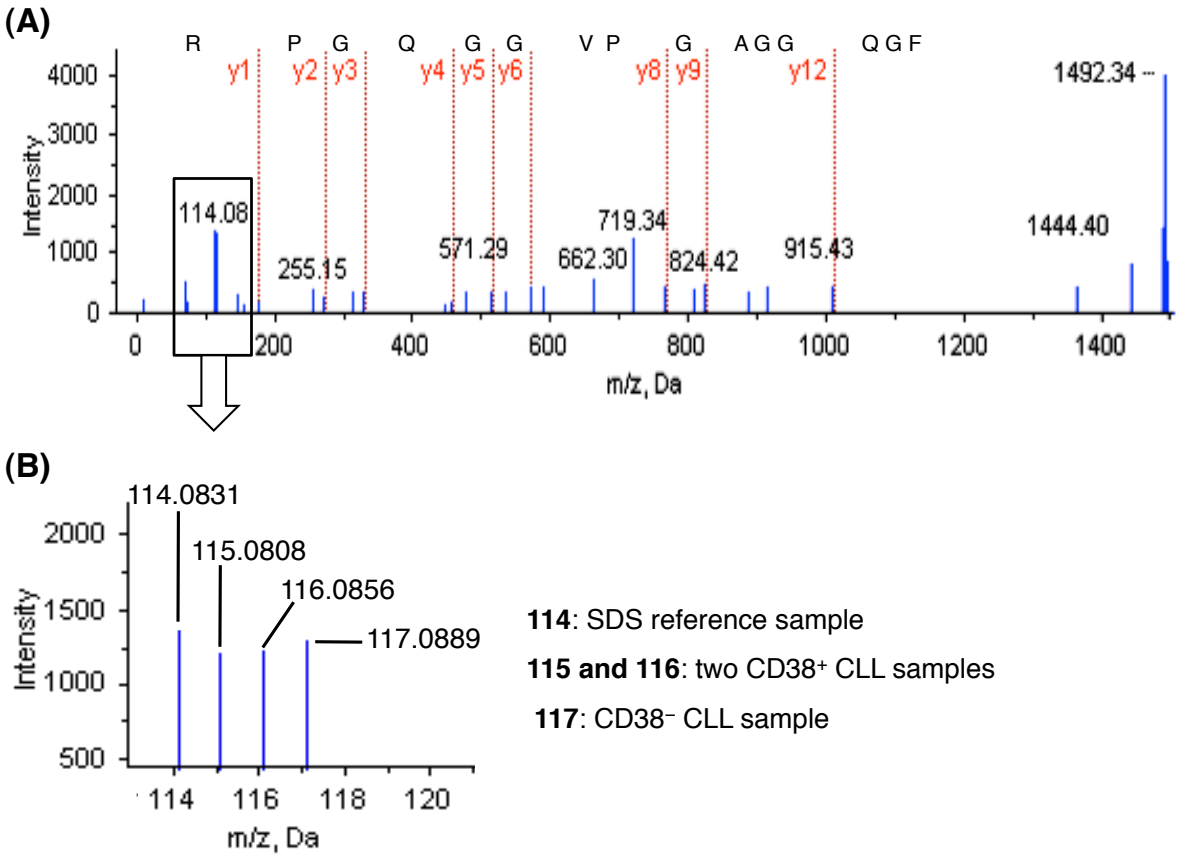


Figure 5.5: Identification and relative quantification of a specific peptide in splicing factor, proline- and glutamine-rich (PSF). Following separation of labelled peptides by 2D nano-LC, they were analysed by MALDI mass spectrometry and ProteinPilot software using the Paragon and ProGroup algorithms. This figure shows MS/MS spectra of a specific peptide in PFS (**FGQGGAGPVGGQGPR**) (A) as well as the relative quantification of this peptide in the SDS fractions of three CLL samples in relation to the SDS reference sample (B). For simplicity only y ions were shown in these MS/MS spectra. The sequence identified in the MS/MS spectra is read from the C terminus.

Table 5.3: Examples of the relative quantification of proteins in CD38$^+$ CLL samples compared with CD38$^-$ CLL samples.

		CD38$^+$ CLL samples								CD38$^-$ CLL samples								
Protein name	**Fraction**	**1**	**2**	**3**	**4**	**5**	**6**	**Mean**	**SD**	**1**	**2**	**3**	**4**	**5**	**6**	**Mean**	**SD**	***p* value**
Nucleophosmin	NP40	0.82	0.94	0.75	0.93	0.68	1.32	0.91	0.23	1.02	0.72	0.76	1.03	1.04	N/A	0.92	0.16	0.94
splicing factor, proline- and glutamine-rich	SDS	0.97	1.40	1.02	1.28	1.06	0.94	1.11	0.19	1.35	0.84	1.03	1.04	1.03	1.36	1.11	0.21	0.99

The relative quantification data of proteins in the tested CLL samples were copied from ProteinPilot software into an Excel spreadsheet for further analysis. Mean and standard deviation (SD) of iTRAQ data were calculated in the two groups of samples (CD38$^+$ CLL versus CD38$^-$ CLL). In addition, the *p* values were calculated using the unpaired *t*-test. This table shows examples of quantified proteins in CLL samples, which had similar expression in the two groups of samples.

H1.2, histone H1.4 and histone H1.5 were identified and quantified in CLL samples. In this case the relative quantification of these proteins was restricted to the relative quantification of their specific peptides. For example, while 13 peptides were used for identification of histone H1.2, only two peptides were shown to be specific to this protein and thus were used for the relative quantification of histone H1.2 (Table 5.4). This was an essential quality control step to avoid false positive identification of proteins with altered expression based on the relative quantification of common peptides.

5.2.6 Criteria for selecting altered proteins in the two forms of CLL

For a protein to be reported with altered expression in CLL samples it had to satisfy four criteria. First it must have been identified with at least one specific peptide (95% confidence score). In addition, a protein must have been quantified in at least five different CLL samples. A protein must have had an iTRAQ value ≤0.80 or ≥1.25, representing ≥20% change in protein expression, and a significant *p* value (≤0.05). Finally the protein must not be one of the contaminating proteins that are commonly identified in mass spectrometry. This study identified proteins that were likely to be detected in CLL samples dute to contamination during sample preparation or by other blood cells. These proteins included different types of keratin, hemoglobin subunits, neutrophil defensin 3, neutrophil elastase macrophage migration inhibitory factor.

Table 5.4: Examples of homologous proteins that were identified and quantified in CLL samples

Name	Accession Number	Peptides count (95% confidence)	Specific peptides
Histone H1.2	H12_HUMAN	13	2
Histone H1.4	H14_HUMAN	11	1
Histone H1.5	H15_HUMAN	4	2

Following identification of labelled peptides by MALDI mass spectrometry and ProteinPilot software using the Paragon and ProGroup algorithms, the software used these peptides to indicate their corresponding proteins. Some of these peptides were specific to their proteins, while some others were common to different proteins (usually proteins that belonged to the same family). While relative quantification was performed on both common and specific peptides, only relative quantification data derived from specific peptides were used to indicate the relative abundance of their corresponding proteins in CLL samples. The examples shown in this Table were extracted from one iTRAQ LC-MALDI mass spectrometry experiment (SDS fraction).

5.2.7 Analysis of protein expression in CLL samples based on CD38 expression

CD38 expression is one of the prognostic markers in CLL used to help predict the clinical outcome of the disease. In this regard, CLL patients with high expression of CD38 on their CLL cells are characterised by an aggressive clinical course and worse outcome compared to patients with low or no expression of CD38 on their CLL cells (Durig et al., 2002, Pepper et al., 2012). Therefore, CD38 expression was used to identify CLL patients with poor prognosis or good prognosis.

Protein expression of six $CD38^+$ CLL samples (CD38 expression >40%) was compared to that of six $CD38^-$ CLL samples (CD38 expression <5%). The analysis was separately done on the NP40 fractions and the SDS fractions. In total this study included 416 proteins as they were quantified in five or more samples. Figure 5.6 demonstrates the protein expression ratio of $CD38^+$ CLL samples to $CD38^-$ CLL samples together with statistical confidence scores that indicate whether the change in protein expression between the two groups of samples is significant. This score was calculated as follows: $(1 - p \text{ value}) \times 100$, where the p value was generated using an unpaired t-test. The majority of quantified proteins in $CD38^+$ CLL samples and $CD38^-$ CLL samples (407 proteins, 97.8%) did not show a significant change in their expression. However, nine proteins had iTRAQ values ≤0.80 or ≥1.25 with a significant p value. Five of these were found in the NP40 fraction (Figure 5.6A) and four were found in the SDS fraction (Figure 5.6B). These

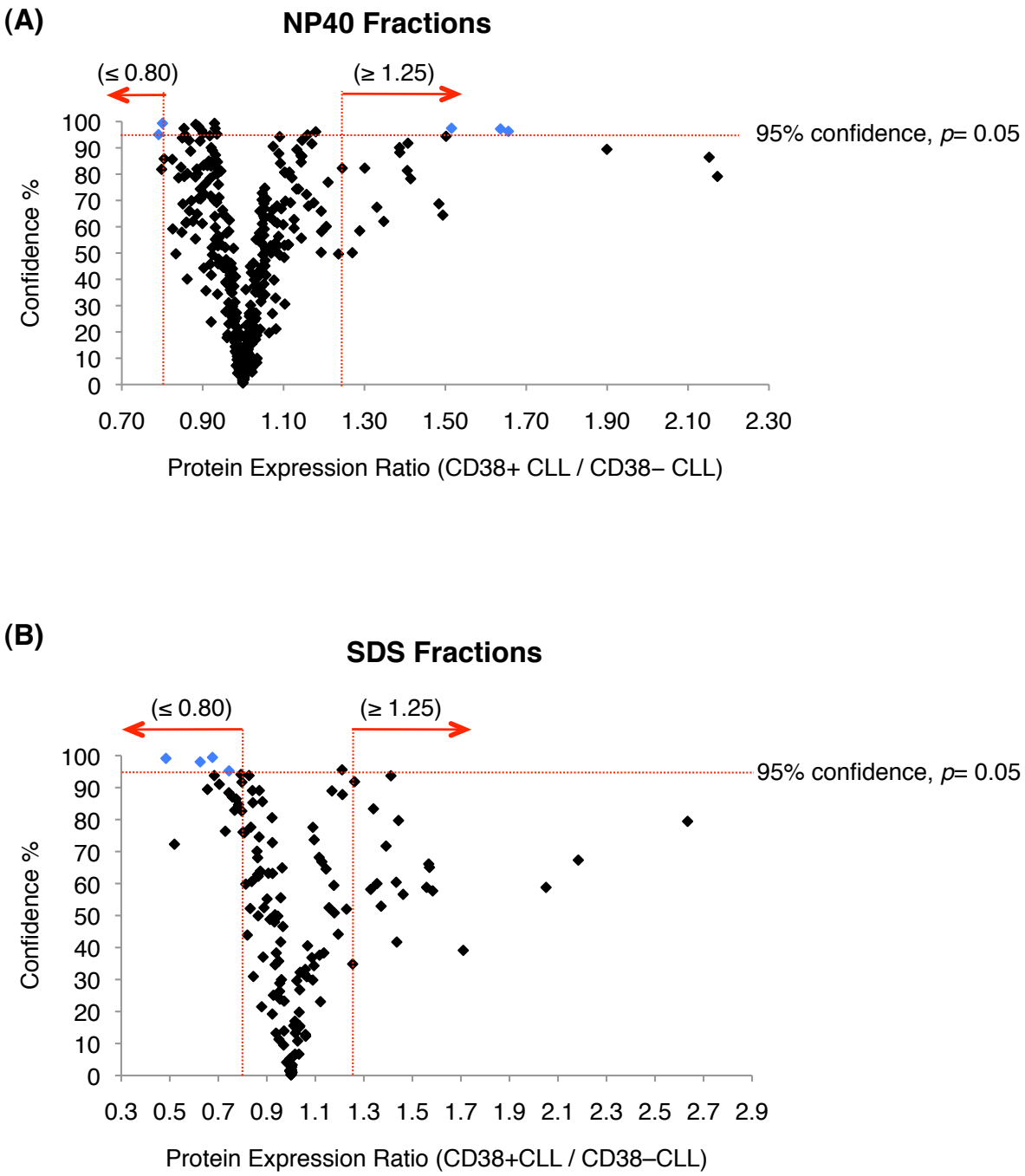


Figure 5.6: Protein expression of CD38⁺ CLL samples (poor prognosis) and CD38⁻ CLL samples (good prognosis). The analysis was performed on the NP40 fractions (A) and on the SDS fractions (B). The protein expression ratio was determined by dividing the proteomic relative quantification of CD38⁺ CLL samples by that of CD38⁻ CLL samples. Confidence % was generated using the calculated *p* value, where *p*=0.05 = 95% confidence. The blue dots represent proteins with iTRAQ values ≤0.80 or ≥1.25 and ≥95% confidence.

proteins are labelled in blue in Figure 5.6. Other proteins (19 proteins in the NP40 fractions and 31 proteins in the SDS fractions) were observed with iTRAQ values ≤0.80 or ≥1.25, yet their corresponding *p* value was >0.05 (confidence <95%) and thus were not included with the final list of the altered proteins in $CD38^+$ CLL samples compared to $CD38^-$ CLL samples. The reason why the altered expression of these proteins was not statistically significant (*p* >0.05) was because their expression was not consistently up-regulated or down-regulated in $CD38^+$ CLL samples compared with $CD38^-$ CLL samples. Table 5.5 and 5.6 show proteins with iTRAQ value ≤0.80 or ≥1.25 in the two groups of CLL samples both with and without a significant *p* value.

Of the nine proteins, six proteins satisfied the four criteria that were set to consider a protein with altered expression in CLL samples. From the NP40 fractions TCL1 showed higher expression in $CD38^+$ CLL samples (*p*= 0.05, n= 11, Figure 5.7A) while 14-3-3 protein (*p*= 0.05, n= 6, Figure 5.7B) and Septin-7 (*p*= 0.01, n= 6, Figure 5.7C) were found with lower expression in $CD38^+$ CLL samples. From the SDS fractions, lower expression was observed for Histone H2B type 2-E (*p*= 0.01, n= 6, Figure 5.7D), Splicing factor, arginine/serine-rich 2 (*p*= 0.005, n= 6, Figure 5.7E) and Histone H4 (*p*= 0.05, n= 12, Figure 5.7F) in $CD38^+$ CLL samples. These six proteins were identified with one or more specific peptides (Appendix 4). In addition, examples of the relative quantification of TCL-1 and histone H4 in CLL samples are demonstrated in Appendices 5-9.

Table 5.5: Proteins that had iTRAQ values ≤0.80 or ≥1.25, with or without significant *p* values, in the NP40 fractions of CD38⁺ CLL samples compared to CD38⁻ CLL samples

Protein Name	Accession Number	CD38⁺ Mean	CD38⁺ SD	CD38⁻ Mean	CD38⁻ SD	CD38⁺/ CD38⁻	Confidence %	*P* value	Sample count
Altered proteins with significant *p* value									
Septin-7	SEPT7_HUMAN	0.89	0.05	1.11	0.04	0.79	99.34	0.01	6.00
T-cell leukemia/ lymphoma protein 1A	TCL1A_HUMAN	1.99	0.70	1.20	0.17	1.66	96.28	0.04	11.00
14-3-3 protein theta	1433T_HUMAN	0.95	0.12	1.20	0.01	0.79	95.05	0.05	6.00
Altered proteins but with insignificant *p* value									
40S ribosomal protein S13	RS13_HUMAN	0.98	0.21	1.23	0.35	0.80	81.83	0.18	11.00
40S ribosomal protein S14	RS14_HUMAN	1.03	0.09	1.28	0.16	0.80	85.85	0.14	5.00
Heterogeneous nuclear ribonucleoproteins C1/C2	HNRPC_HUMAN	1.27	0.31	1.02	0.25	1.25	82.26	0.18	11.00
Protein S100-A8	S10A8_HUMAN	1.90	1.21	1.50	0.45	1.27	50.11	0.50	11.00
HLA class II histocompatibility antigen, DRB1-4 beta chain	2B14_HUMAN	1.49	0.57	1.15	0.25	1.29	58.44	0.42	5.00
60S ribosomal protein L7	RL7_HUMAN	1.07	0.13	0.82	0.16	1.30	82.29	0.18	5.00
Heterogeneous nuclear ribonucleoproteins A2/B1	ROA2_HUMAN	1.53	0.78	1.15	0.26	1.33	67.42	0.33	11.00

Relative protein expression in the CD38⁺ CLL samples and CD38⁻ CLL samples was quantified using the iTRAQ labelling technique followed by LC-MALDI mass spectrometry and ProteinPilot analysis. Data were transferred to an Excel spreadsheet for further analysis. Mean, standard deviation (SD) and *p* value (unpaired *t*-test) of iTRAQ data were calculated in the two groups of samples (CD38⁺ CLL versus CD38⁻ CLL). This table shows proteins with consistently altered expression in the CD38⁺ CLL samples compared to CD38⁻ CLL samples (Blue coloured). In addition, it shows proteins with inconsistently altered expression in the CD38⁺ CLL samples compared to CD38⁻ CLL samples (Black coloured). CD38⁺/CD38⁻: protein expression of CD38⁺ samples versus CD38⁻ CLL samples.

Table 5.5 continued: Proteins that had iTRAQ values ≤0.80 or ≥1.25, with or without significant *p* values, in the NP40 fractions of $CD38^+$ CLL samples compared to $CD38^-$ CLL samples

Protein Name	Accession Number	$CD38^+$ Mean	$CD38^+$ SD	$CD38^-$ Mean	$CD38^-$ SD	$CD38^+/CD38^-$	Confidence %	*P* value	Sample count
Altered proteins but with insignificant *p* value									
CapZ-interacting protein	CPZIP_HUMAN	1.20	0.04	0.87	0.38	1.39	88.23	0.12	6.00
Histone H4	H4_HUMAN	1.46	0.64	1.04	0.12	1.41	81.38	0.19	11.00
Histone H2A.J	H2AJ_HUMAN	1.39	0.67	0.98	0.13	1.41	78.25	0.22	11.00
Putative heterogeneous nuclear ribonucleoprotein A1-like 3	RA1L3_HUMAN	1.64	1.10	1.10	0.16	1.48	68.72	0.31	11.00
Protein S100-A9	S10A9_HUMAN	1.99	1.40	1.33	0.58	1.49	64.44	0.36	11.00

Table 5.6: Proteins that had iTRAQ values ≤0.80 or ≥1.25, with or without significant *p* values, in the SDS fractions of CD38⁺ CLL samples compared to CD38⁻ CLL samples.

Protein Name	Accession Number	CD38+ Mean	CD38+ SD	CD38⁻ Mean	CD38⁻ SD	CD38+/ CD38-	Confidence %	*p* value	Sample count
Altered proteins with significant *p* value									
Histone H4	H4_HUMAN	0.92	0.23	1.24	0.25	0.74	95.22	0.05	12
Histone H2B type 1-L	H2B1L_HUMAN	0.77	0.22	1.24	0.10	0.63	98.03	0.02	6
Histone H2B type 2-E	H2B2E_HUMAN	0.66	0.14	1.36	0.18	0.48	99.12	0.01	6
Splicing factor, arginine/serine-rich 2	SFRS2_HUMAN	1.14	0.12	1.69	0.10	0.68	99.42	0.01	6
Altered proteins but with insignificant *p* value									
Triosephosphate isomerase	TPIS_HUMAN	0.83	0.81	0.48	0.23	1.71	39.11	0.61	6
Tyrosine-protein phosphatase non-receptor type 6	PTN6_HUMAN	0.78	0.50	0.55	0.29	1.44	41.72	0.58	6
14-3-3 protein zeta/delta	1433Z_HUMAN	0.94	0.45	0.64	0.02	1.46	56.62	0.43	6
Hematopoietic lineage cell-specific protein	HCLS1_HUMAN	0.95	0.35	0.71	0.02	1.33	58.18	0.42	6
Lactotransferrin	TRFL_HUMAN	0.87	0.65	0.43	0.07	2.05	58.79	0.41	6
Cathepsin G	CATG_HUMAN	0.98	0.50	0.63	0.16	1.56	58.81	0.41	6
Myeloperoxidase	PERM_HUMAN	0.83	0.52	0.58	0.20	1.43	60.40	0.40	9
Heat shock protein HSP 90-alpha	HS90A_HUMAN	0.97	0.41	0.62	0.20	1.57	66.12	0.34	6
Azurocidin	CAP7_HUMAN	0.83	0.53	0.38	0.04	2.18	67.33	0.33	6
Histone H2B type 1-D	H2B1D_HUMAN	1.01	0.30	0.72	0.05	1.39	71.71	0.28	6
Histone H2A.J	H2AJ_HUMAN	0.54	0.33	1.03	0.49	0.52	72.29	0.28	6
Spectrin alpha chain, brain	SPTA2_HUMAN	0.85	0.03	1.06	0.26	0.80	76.08	0.24	6
RNA-binding protein 14	RBM14_HUMAN	0.83	0.23	1.14	0.33	0.73	76.36	0.24	6

This table shows proteins with consistently altered expression in the CD38⁺ CLL samples compared to CD38⁻ CLL samples (Blue coloured). In addition, it shows proteins with inconsistently altered expression in the CD38⁺ CLL samples compared to CD38⁻ CLL samples (Black coloured). SD: standard deviation, CD38⁺/CD38⁻: protein expression of CD38⁺ samples versus CD38⁻ CLL samples

Table 5.6 continued: Proteins that had iTRAQ values ≤0.80 or ≥1.25, with or without significant *p* values, in the SDS fractions of $CD38^+$ CLL samples compared to $CD38^-$ CLL samples.

Protein Name	Accession Number	CD38+ Mean	CD38+ SD	$CD38^-$ Mean	$CD38^-$ SD	CD38+/ CD38-	Confidence %	*p* value	Sample count
Altered proteins but with insignificant *p* value									
Heterogeneous nuclear ribonucleoproteins A2/B1	ROA2_HUMAN	1.09	0.43	1.37	0.18	0.80	82.66	0.17	12
Heterogeneous nuclear ribonucleoprotein K	HNRPK_HUMAN	0.98	0.37	1.28	0.12	0.77	82.90	0.17	9
60S acidic ribosomal protein P0-like	RLA0L_HUMAN	0.99	0.20	0.74	0.04	1.34	83.36	0.17	6
Heterogeneous nuclear ribonucleoprotein D0	HNRPD_HUMAN	1.01	0.37	1.29	0.25	0.78	84.64	0.15	12
Heterogeneous nuclear ribonucleoprotein A1	ROA1_HUMAN	0.99	0.38	1.28	0.21	0.77	86.45	0.14	12
High mobility group protein B2	HMGB2_HUMAN	0.75	0.17	1.00	0.25	0.76	87.05	0.13	9
Heterogeneous nuclear ribonucleoprotein G	HNRPG_HUMAN	0.97	0.42	1.31	0.22	0.74	88.38	0.12	12
Heterogeneous nuclear ribonucleoprotein A3	ROA3_HUMAN	1.13	0.54	1.72	0.38	0.66	89.42	0.11	9
Thyroid hormone receptor-associated protein 3	TR150_HUMAN	0.89	0.16	1.26	0.26	0.70	91.03	0.09	6
Splicing factor 3B subunit 2	SF3B2_HUMAN	1.06	0.04	1.32	0.26	0.80	91.68	0.08	6
Nucleolin	NUCL_HUMAN	1.29	0.29	1.02	0.18	1.26	91.83	0.08	12
ATPase family AAA domain-containing protein 3A	ATD3A_HUMAN	1.10	0.17	0.78	0.04	1.41	93.68	0.06	6
Polypyrimidine tract-binding protein 1	PTBP1_HUMAN	0.81	0.18	1.19	0.14	0.68	93.75	0.06	6
Far upstream element-binding protein 1	FUBP1_HUMAN	0.90	0.20	1.14	0.05	0.79	94.09	0.06	9

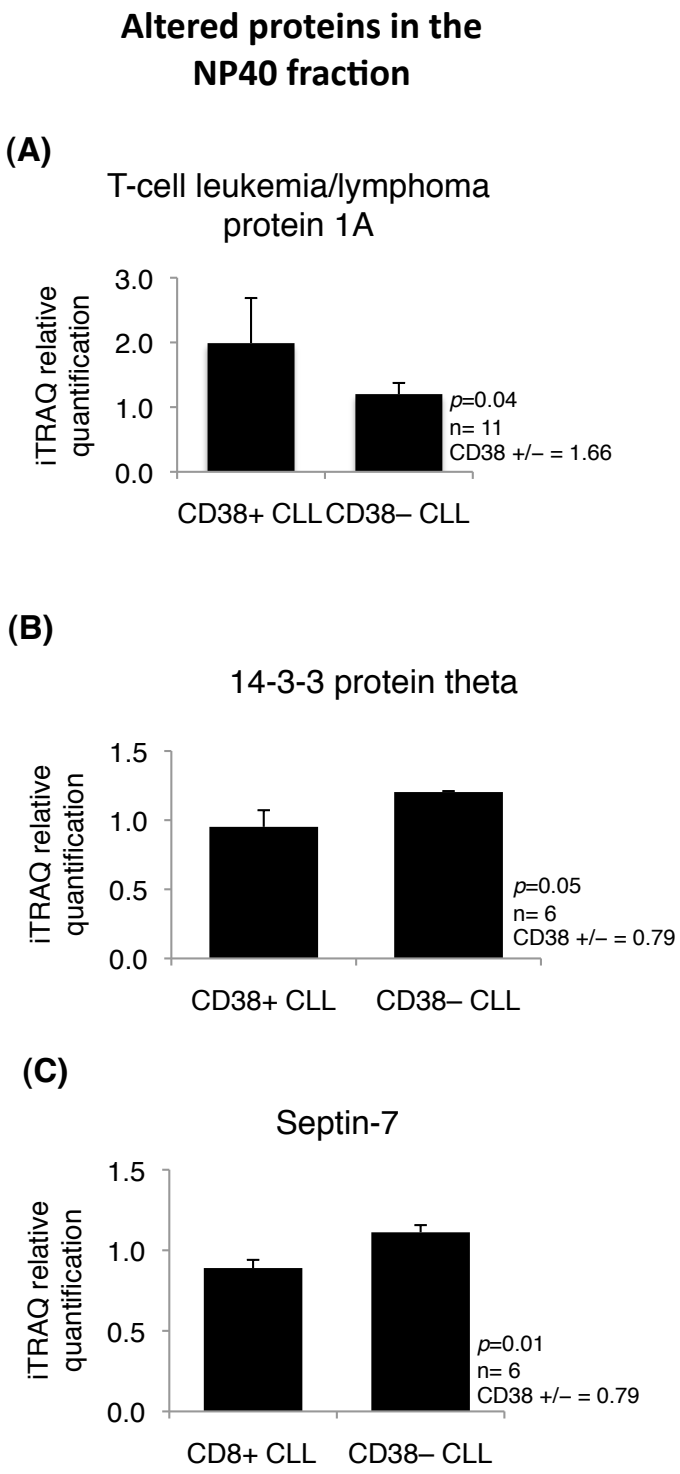


Figure 5.7: Proteins with different expression in CD38$^+$ CLL samples (poor prognosis) relative to CD38$^-$ CLL samples (good prognosis). Protein relative quantification using the iTRAQ technique coupled with LC-MS/MS was performed on the NP40 fractions and the SDS fractions of CLL samples. The expression of three proteins from the NP40 fractions of CLL samples was elevated in CD38$^+$ compared with CD38$^-$ samples (A-C).

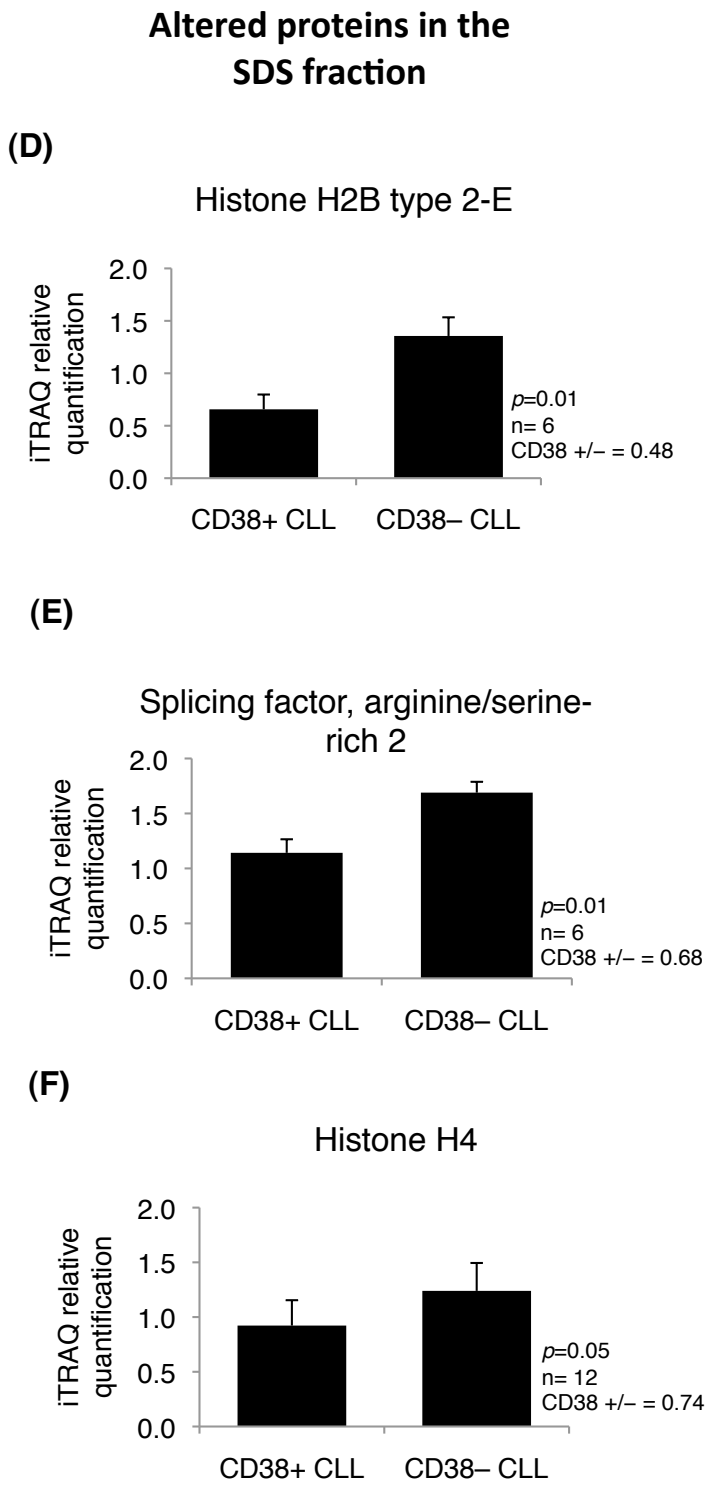


Figure 5.7 continued: Proteins with different expression in CD38^{+} CLL samples (poor prognosis) relative to CD38^{-} CLL samples (good prognosis). Protein relative quantification using iTRAQ technique coupled with LC-MS/MS was performed on the NP40 fractions and the SDS fractions of CLL samples. The expression of three proteins from the SDS fractions of CLL samples was down-regulated in CD38^{+} compared with CD38^{-} samples (D-F).

Of the proteins that had iTRAQ values ≤0.80 or ≥1.25 coupled with a significant *p* value, Keratin, type I cytoskeletal 10 (p= 0.03, n= 11) and Keratin, type II cytoskeletal 5 (p= 0.03, n= 6) were observed with higher expression in $CD38^+$ CLL samples. However, these two proteins were not included in the final annotated list of differentially expressed proteins in $CD38^+$ CLL samples as they are common contaminating proteins that are often identified by mass spectrometry (Ding et al., 2003, Lubec and Afjehi-Sadat, 2007).

The ninth protein that showed an iTRAQ value ≤0.80 or ≥1.25 with a significant *p* value between the $CD38^+$ CLL and $CD38^-$ CLL samples was Histone H2B type 1-L (p= 0.02, n= 6). This protein was observed with lower expression in $CD38^+$ CLL samples. However, all the peptides that were used for identification of this protein contributed to identification of other proteins, such as Histone H2B type 1-M and Histone H2B type 1-N. The reason why this occurred despite the use of high peptide count for protein identification (11 peptides) was because these three proteins are very similar in terms of their sequence (Figure 5.8). As a result, Histone H2B type 1-L was excluded from the final list and was not studied further.

As stated previously, the quantitative proteomics analysis was designed to study the CLL proteome associated with poor prognosis and good prognosis CLL as defined by CD38 expression. The clinical data that was available allowed a further subset analysis to be performed based on ZAP-70

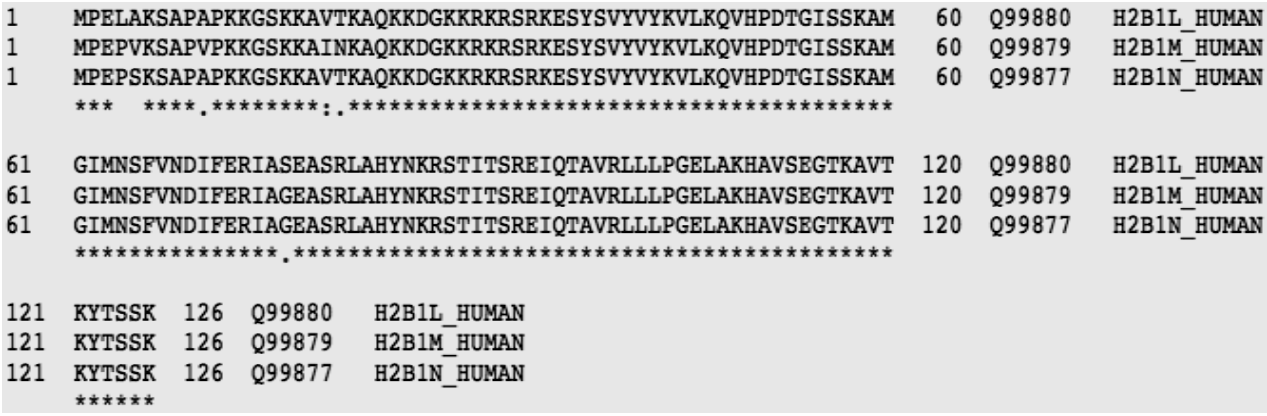


Figure 5.8: Alignment analysis of Histone H2B type 1-L (H2B1L_HUMAN), Histone H2B type 1-M (H2B1M_HUMAN) and Histone H2B type 1-N (H2B1N_HUMAN). This was performed using the alignment tool of the UniProt database (http://www.uniprot.org/). The asterisk symbols (*) represent identical positions, the colon symbol (:) indicates positions with strongly similar properties and the period symbol (.) demonstrates positions with weakly similar properties. The analysis demonstrated that of the full protein sequence (126 amino acids) these proteins had 120 identical positions.

expression, the presence or absence of *IGHV* gene mutations, and clinical staging. This was carried out separately on the NP40 fractions and the SDS fractions.

5.2.8 Analysis of protein expression in CLL samples based on ZAP-70 expression

The expression of ZAP-70 has prognostic value in CLL. High expression of ZAP-70 in CLL cells is associated with poor prognosis, whereas low expression of ZAP-70 in CLL cells is associated with good prognosis (Rassenti et al., 2004, Pepper et al., 2012).

In this analysis protein expression of three ZAP-70^{+} CLL samples (ZAP-70 expression ≥23%) was compared with that of seven ZAP-70^{-} CLL samples (ZAP70 expression <6%). The analysis was carried out on both the NP40 fractions and SDS fractions. In total 336 proteins were used for this study as they were relatively quantified in five or more samples. Figure 5.9 shows the protein expression ratio of ZAP-70^{+} CLL samples to ZAP-70^{-} CLL samples with statistical confidence score. The majority of the proteins (330 proteins, 98.2%) did not exhibit a significant change in their expression in ZAP-70^{+} CLL and ZAP-70^{-} CLL samples. However, six proteins were found to meet the four criteria used to report a protein with an altered expression in CLL samples. Four of these proteins were found in the NP40 fractions and two were found in the SDS fractions. These proteins are shown in Figure 5.9 and are highlighted in blue. From the NP40 fractions, four proteins were

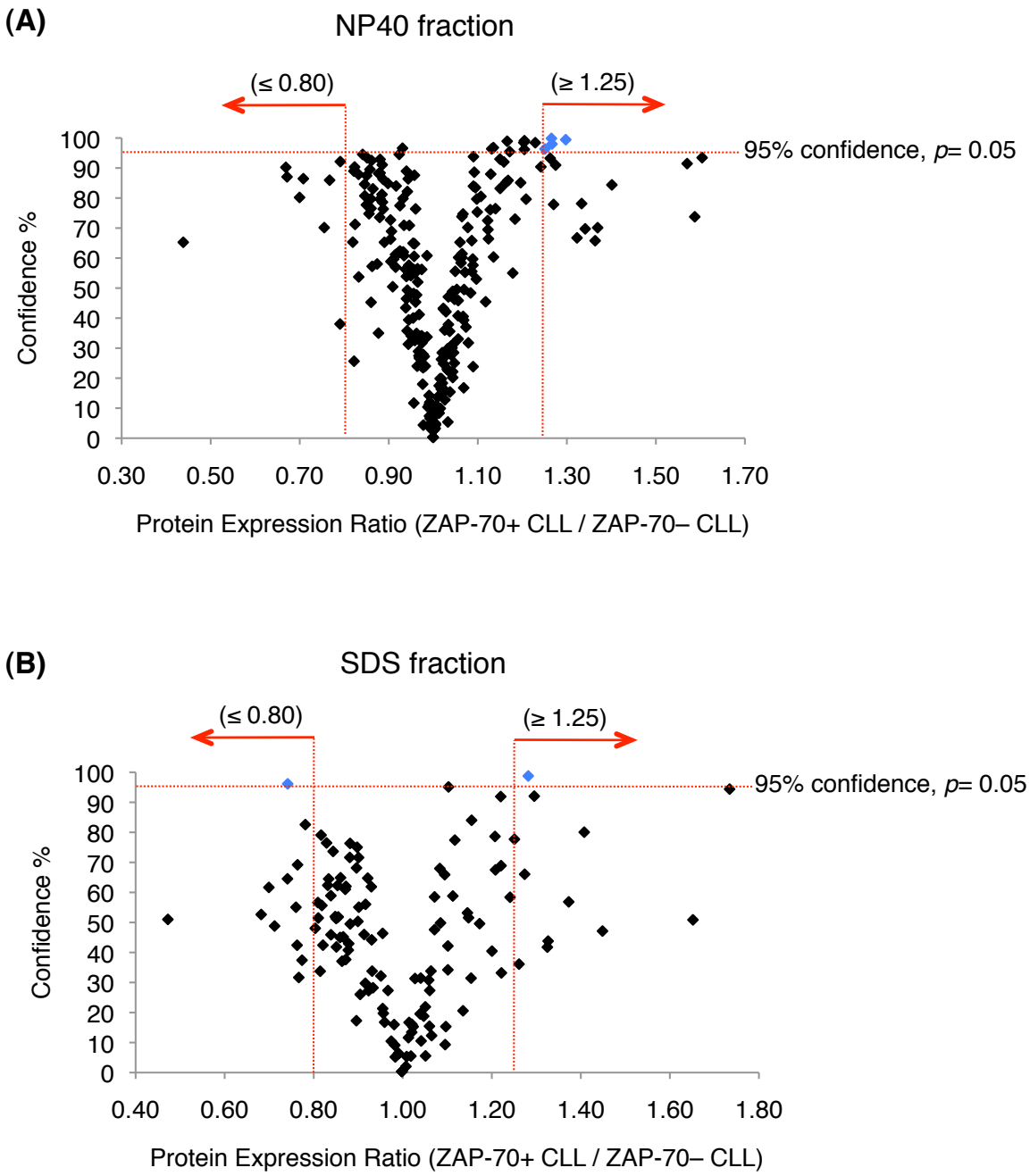


Figure 5.9: Protein expression of ZAP-70$^+$ CLL samples (poor prognosis) and ZAP-70$^-$ CLL samples (good prognosis). The analysis was performed on the NP40 fractions (A) and on the SDS fractions (B). The protein expression ratio was determined by dividing proteomic iTRAQ values of ZAP-70$^+$ CLL by that of ZAP-70$^-$ CLL. Percentage confidence was generated using *p* values, where *p*=0.05 is equal to 95% confidence. The blue dots represent proteins with an iTRAQ value ≤0.80 or ≥1.25 and ≥95% confidence.

detected with an up-regulated expression in the ZAP-70$^+$ CLL samples. These proteins included acidic leucine-rich nuclear phosphoprotein 32 family member A (p= 0.01, n= 7, Figure 5.10A), LIM and SH3 domain protein 1 (p= 0.001, n= 9, Figure 5.10B), U6 snRNA-associated Sm-like protein LSm3 (p= 0.02, n= 9, Figure 5.10C) and Peroxiredoxin-5, mitochondrial (p= 0.04, n= 6, Figure 5.10D). From the SDS fractions, RNA-binding protein FUS was shown with lower expression in ZAP-70$^+$ CLL samples (p= 0.04, n= 7, Figure 5.10E), while DNA-(apurinic or apyrimidinic site) lyase was found with higher expression in the ZAP-70$^+$ CLL samples (p= 0.01, n= 6, Figure 5.10F). All of these six proteins were identified with at least one specific peptide (Appendix 10).

5.2.9 Analysis of protein expression in CLL samples based on the mutational status of *IGHV*

The mutational status of the *IGHV* gene is regarded one of the most accurate and widely applicable prognostic markers in CLL. In this regard, patients with CLL cells bearing unmutated *IGHV* (U-CLL) were shown to have an aggressive form of CLL, while patients with CLL cells carrying mutated *IGHV* (M-CLL) were shown to have a more indolent form of CLL (Hamblin et al., 1999, Damle et al., 1999).

In this analysis protein expression of four U-CLL samples was compared to that of four M-CLL samples. The analysis was carried out on both the NP40 fractions and the SDS fractions using 322 protein IDs as they

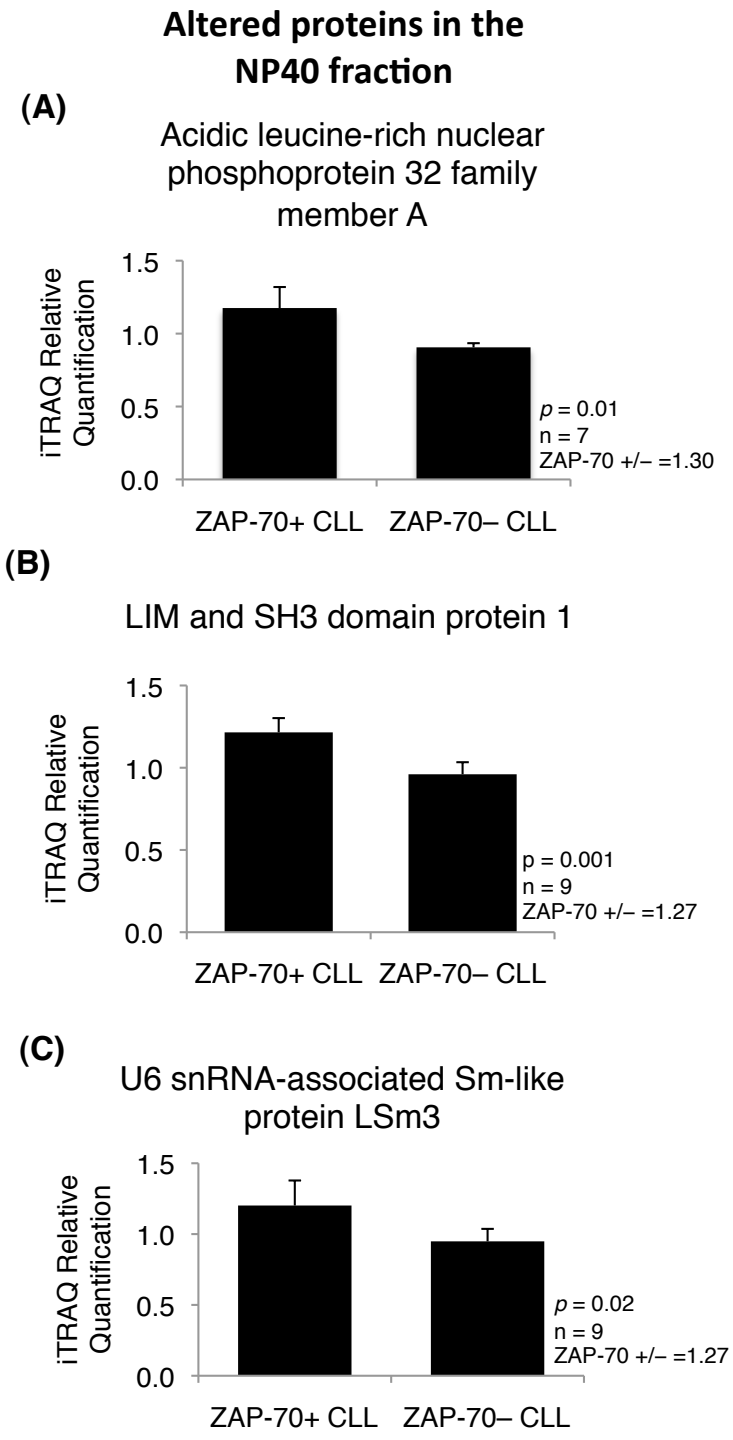


Figure 5.10: Proteins with an altered expression in ZAP-70^{+} CLL samples compared with ZAP-70^{-} CLL samples. Relative quantification of CLL proteins using the iTRAQ technique coupled with LC-MS/MS was performed on NP40 fractions and SDS fractions of CLL samples. Four proteins from the NP40 fractions of CLL samples with high or low ZAP-70 expression showed altered expression (A-C).

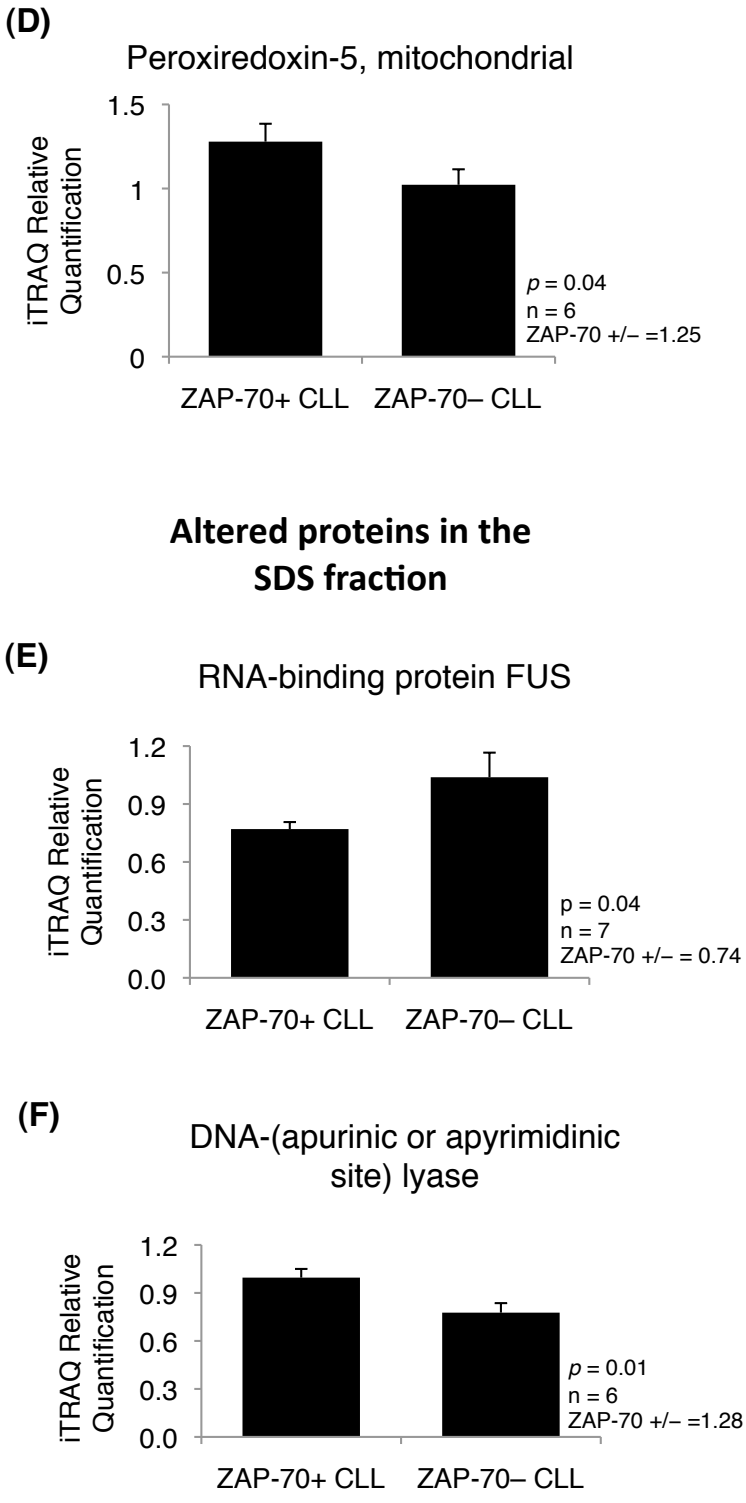


Figure 5.10 continued: Proteins with an altered expression in ZAP-70^{+} CLL samples compared with ZAP-70$^{–}$ CLL samples. Relative quantification of CLL proteins using the iTRAQ technique coupled with LC-MS/MS was performed on NP40 fractions and SDS fractions of CLL samples. The fourth protein that showed altered expression in the NP40 fractions of CLL samples with high or low ZAP-70 expression was peroxiredoxin-5, mitochondrial (D). From the SDS fractions of CLL samples with high or low ZAP-70 expression two proteins exhibited differential expression (E and F).

were quantified in five or more CLL samples. Figure 5.11 demonstrates the protein expression ratio of U-CLL samples to M-CLL samples with statistical confidence scores. The majority of the proteins (315 proteins, 97.8%) did not demonstrate a significant change in their expression in U-CLL versus M-CLL samples Nevertheless, seven proteins were found with iTRAQ values ≤0.80 or ≥1.25 and significant *p* values in U-CLL compared with M-CLL samples. Six of these proteins were indentified in the NP40 fractions and one was detected in the SDS fractions. These proteins are labelled in blue in Figure 5.11.

Of the seven proteins, three proteins satisfied the four criteria to be considered altered proteins in CLL samples. From the NP40 fractions, tropomyosin alpha-4 chain showed a decreased expression in U-CLL samples compared to M-CLL samples (p= 0.05, n= 6, Figure 5.12A), while LIM and SH3 domain protein 1 was shown with high expression in U-CLL samples compared to M-CLL samples (p= 0.02, n= 8, Figure 5.12B). From the SDS fractions, apoptotic chromatin condensation inducer in the nucleus was significantly lower in U-CLL samples relative to M-CLL samples (p= 0.003, n= 6, Figure 5.12C). All of these three proteins were identified with two or more specific peptides (Appendix 11).

Other proteins that had iTRAQ signal ≤ 0.80 or ≥ 1.25 and significant *p* value in U-CLL versus M-CLL samples were Keratin, type I cytoskeletal 9 (p= 0.004, n= 8), Keratin, type II cytoskeletal 2 epidermal (p= 0.01, n= 8), Keratin, type II cytoskeletal 1 (p= 0.01, n= 8) and Keratin, type I

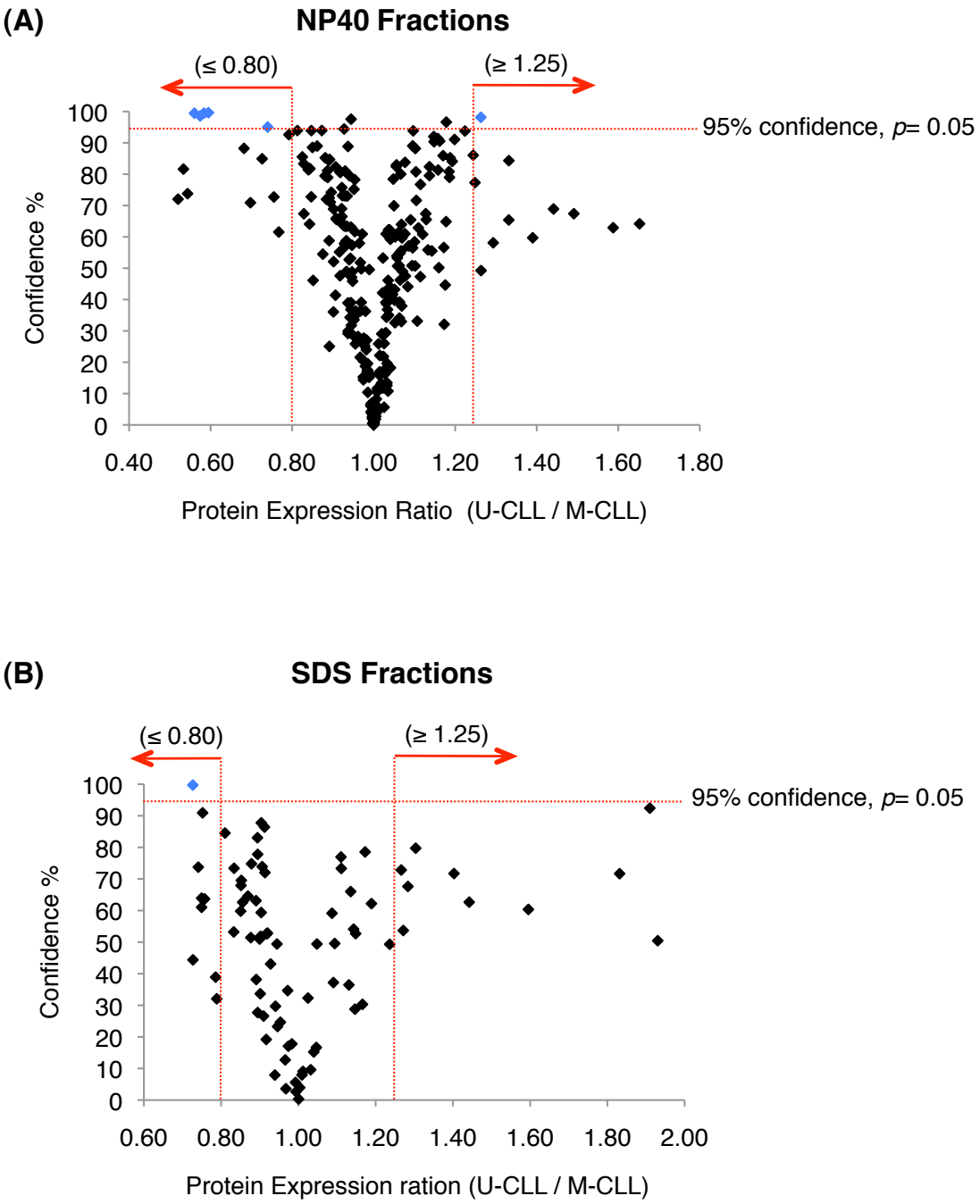


Figure 5.11: Protein expression of U-CLL samples (poor prognosis) and M-CLL samples (good prognosis). The analysis was performed on the NP40 fractions (A) and on the SDS fractions (B). Protein expression ratio was determined by dividing the proteomic relative quantification of U-CLL by that of M-CLL. Percentage confidence was generated using *p* values, where *p*=0.05 = 95% confidence. The blue dots represent proteins with an iTRAQ value ≤0.80 or ≥1.25 and ≥95% confidence.

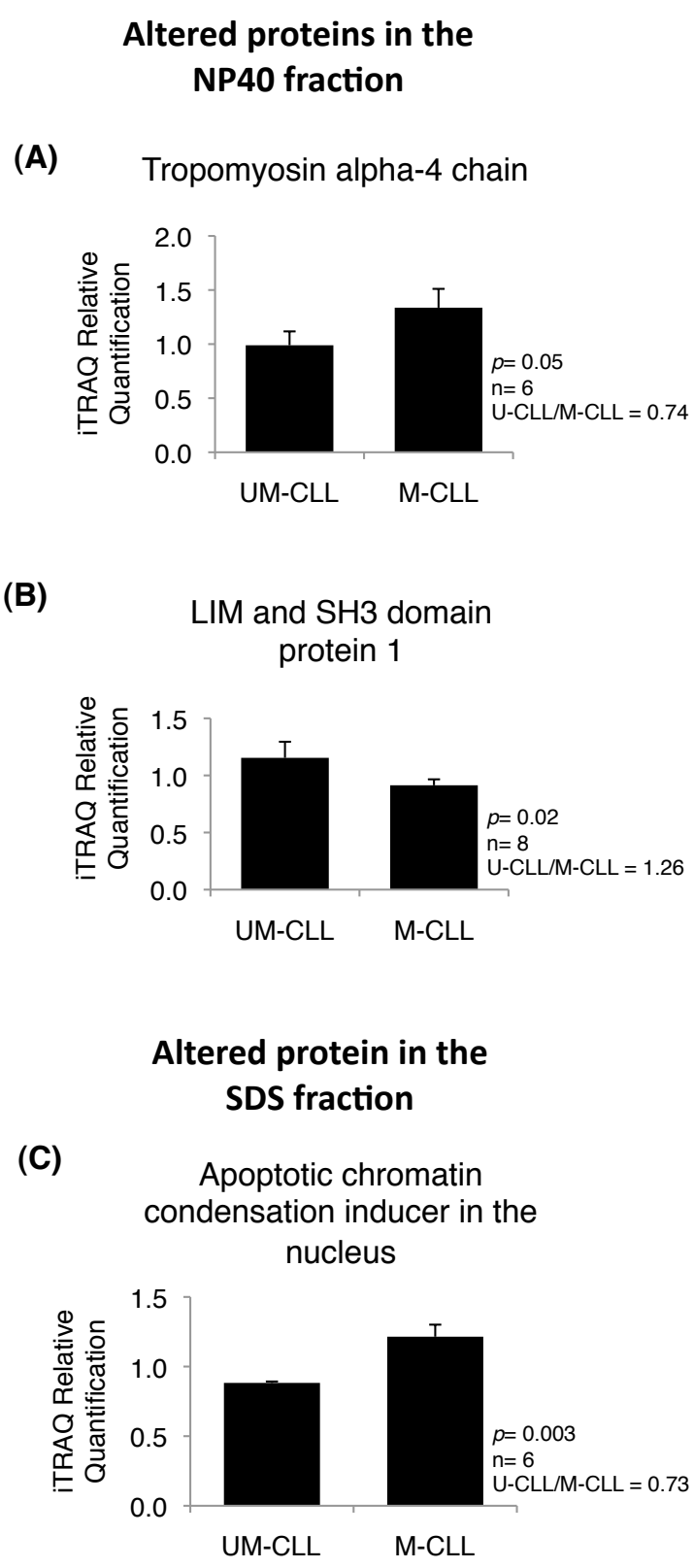


Figure 5.12: Proteins with an altered expression in U-CLL samples compared to M-CLL samples. Relative protein quantification using the iTRAQ technique coupled with LC/MS/MS was performed on the NP40 fractions and the SDS fractions of CLL samples. The expression level of two proteins from the NP40 fractions of CLL samples with or without *IGHV* mutations was different (A and B). Likewise, one protein from the SDS fractions of CLL samples with or without *IGHV* mutations demonstrated differential expression (C).

cytoskeletal 10 (p= 0.01, n= 8). All of these proteins were found with lower expression in U-CLL samples. For the same reason mentioned earlier, these three proteins were not included in the final list of differentially expressed proteins in U-CLL samples and M-CLL samples.

5.2.10 Analysis of protein expression in CLL samples based on disease stage

Historically there have been two different systems for CLL staging; the Rai system (Rai et al., 1975) and the Binet system (Binet et al., 1981). Both systems are based on observable clinical outcomes such as high white cell count, lymphadenopathy, hepatomegaly, splenomegaly, anaemia and thrombocytopenia. The majority of CLL patients (approximately 70%) are diagnosed at an early stage of the disease with a fraction of these developing a more advanced stage with time (Pepper et al., 2012). In the Binet staging system, the median survival of a patient with stage A is more than 10 years, five years for stage B CLL patients and three years for stage C CLL patients (Binet et al., 1981).

In this study, the proteomes of stage B/C CLL samples were compared to those of stage A CLL samples (three samples versus seven CLL samples, respectively). The analysis was performed on the NP40 fractions and the SDS fractions utilising 320 proteins that were quantified in five or more CLL samples. Figure 5.13 shows the protein expression ratio of stage B/C CLL samples to stage A CLL samples with statistical confidence. The

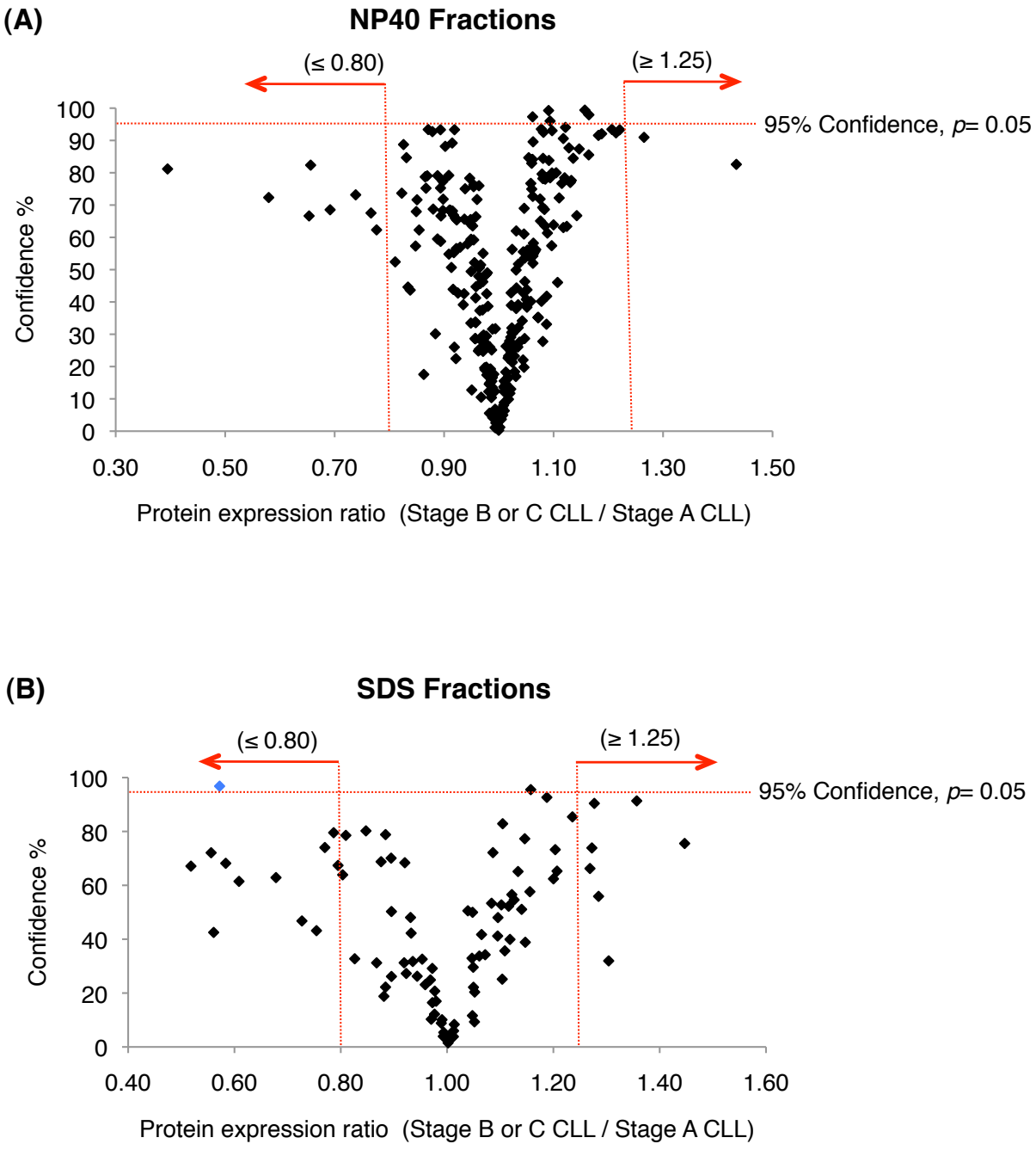


Figure 5.13: Protein expression of stage B/C CLL samples and stage A CLL samples. The analysis was performed on the NP40 fractions (A) and on the SDS fractions (B). The protein expression ratio was determined by dividing proteomic relative quantification of stages B/C CLL samples by that of stage A CLL samples. Confidence percent was generated using *p* value, where *p*=0.05 = 95% confidence. The blue dots represent proteins with an iTRAQ value ≤ 0.80 or ≥ 1.25 and ≥ 95% confidence.

majority of quantified proteins in the two groups of patient samples showed no significant change in their expression (319 proteins, 99.6%). However, one protein met the four criteria of altered proteins in CLL samples. This protein was myosin 9 that showed lower expression in the SDS fractions of stage B/C CLL samples compared to stage A CLL samples (p= 0.03, n =10, Figure 5.14). This protein was identified with two or more specific peptides (Appendix 12)

5.2.11 Summary of the altered proteins in poor prognosis CLL compared to good prognosis CLL

In this study the protein expression of poor prognosis CLL was compared to that of good prognosis CLL on the basis of four commonly used CLL prognostic markers (CD38, ZAP-70, mutational status of *IGHV* genes and Binet's CLL stages). These subset analyses indentified 15 proteins with differential expression in the two forms of CLL. These proteins with their biological functions that were extracted from their Gene Ontology data using Quick GO-EBI tool (http://www.ebi.ac.uk/QuickGO/) are shown in Table 5.7.

5.2.12 Proteins with the most heterogeneous expression in CLL samples

The standard deviation (SD) of relative expression of proteins in five or more CLL samples was used to indicate whether or not a protein was heterogeneously expressed in CLL. This analysis excluded proteins that were likely to be detected in CLL samples due to contamination. These proteins included different types of keratin, hemoglobin subunits, neutrophil defensin 3,

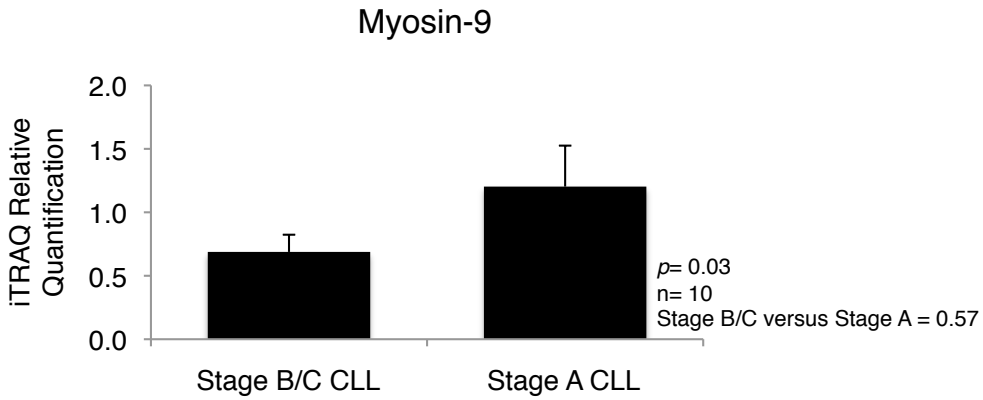


Figure 5.14: One protein (myosin-9) with different expression in CLL samples from patients in stages B and C and patients in stage A. Protein relative quantification using iTRAQ technique coupled with LC/MS/MS was performed on NP40 fractions and on the SDS fractions of CLL samples. Only myosin 9 was found with altered expression in stage B/C CLL samples versus stage A CLL samples.

Table 5.7: Summary of the altered proteins in the poor prognosis CLL versus the good prognosis CLL

Protein name	Accession number	Biological function	Comparison	Change ratio
T-cell leukemia/ lymphoma protein 1A	TCL1A_HUMAN	Stem cell maintenance	$CD38^+$ / $CD38^-$	1.66
Acidic leucine-rich nuclear phosphoprotein 32 family member A	AN32A_HUMAN	Intracellular signal transduction	ZAP-70^+ / ZAP-70^-	1.3
DNA-(apurinic or apyrimidinic site) lyase	APEX1_HUMAN	Positive regulation of DNA repair	ZAP-70^+ / ZAP-70^-	1.28
U6 snRNA-associated Sm-like protein LSm3	LSM3_HUMAN	Gene expression	ZAP-70^+ / ZAP-70^-	1.27
LIM and SH3 domain protein 1	LASP1_HUMAN	SH3/SH2 adaptor activity	ZAP-70^+ / ZAP-70^-	1.27
LIM and SH3 domain protein 1	LASP1_HUMAN	SH3/SH2 adaptor activity	UM / M	1.26
Peroxiredoxin-5, mitochondrial	PRDX5_HUMAN	Negative regulation of apoptotic process	ZAP-70^+ / ZAP-70^-	1.25
Septin-7	SEPT7_HUMAN	Protein heterooligomerization	$CD38^+$ / $CD38^-$	0.8
14-3-3 protein theta	1433T_HUMAN	Protein targeting	$CD38^+$ / $CD38^-$	0.79
Histone H4	H4_HUMAN	Nucleosome assembly	$CD38^+$ / $CD38^-$	0.74
RNA-binding protein FUS	FUS_HUMAN	Cell death	ZAP-70^+ / ZAP-70^-	0.74
Tropomyosin alpha-4 chain	TPM4_HUMAN	Cellular component movement	UM / M	0.74
Apoptotic chromatin condensation inducer in the nucleus	ACINU_HUMAN	Apoptotic process	UM / M	0.73
Splicing factor, arginine/ serine-rich 2	SFRS2_HUMAN	RNA splicing	$CD38^+$ / $CD38^-$	0.68
Myosin-9	MYH9_HUMAN	Cell-cell adhesion	B-C / A	0.57
Histone H2B type 2-E	H2B2E_HUMAN	Nucleosome assembly	$CD38^+$ / $CD38^-$	0.48

The 15 proteins that were identified with an altered expression when the proteome of poor prognosis CLL was compared with the proteome of good prognosis CLL on the basis of the following prognostic markers: CD38, ZAP-70, mutational status of *IGVH* and Binet's stags of CLL. The biological functions of these proteins were extracted from their Gene Ontology data using Quick Go – EPI (http://www.ebi.ac.uk/QuickGO/GAnnotation). This table is sorted according to the change in protein expression (change ratio).

neutrophil elastase and macrophage migration inhibitory factor. The SD of protein expression in CLL samples (≥5 samples) showed that 14 proteins exhibited the most variable expression in CLL samples (z-score > 2). Ten proteins were from the NP40 fractions representing 3.2% of the proteins that were quantified in five or more CLL samples (316 proteins). In addition, four proteins were from the SDS fractions representing 3.1% of the proteins that were quantified in at least five CLL samples (130 proteins). Table 5.8 shows these proteins with their biological functions that were extracted from their Gene Ontology data using Quick GO-EBI tool (http://www.ebi.ac.uk/QuickGO/).

5.2.13 Gene expression of proteins altered in CLL samples

To investigate the gene expression of the altered proteins in CLL samples, data from a published gene expression study that was conducted on $CD38^{+}$/ZAP-70^{+} CLL samples versus $CD38^{-}$/ZAP-70^{-} CLL samples were utilised (Huttmann et al., 2006). This analysis was limited to altered proteins that were found when the analysis was based on either CD38 expression or ZAP-70 expression (twelve proteins). The reason for this was because the gene expression study was based on CLL samples with negative or positive expression of CD38 and ZAP-70 (Huttmann et al., 2006). The accession numbers of these twelve proteins were converted to Affymetrix accession numbers (Affy_HG_U1331) using the g: Convert tool (http://biit.cs.ut.ee/gprofiler/gconvert.cgi).

Table 5.8: Proteins found with the most heterogeneous expression in CLL samples.

Protein name	Accession Number	Biological function	Average	SD	Sample count	Fraction	Z score
Protein S100-A9	S10A9_HUMAN	Positive regulation of NF-kappaB transcription factor activity	1.69	1.11	11	NP40	8.10
Protein S100-A8	S10A8_HUMAN	Positive regulation of NF-kappaB transcription factor activity	1.72	0.92	11	NP40	6.52
Putative heterogeneous nuclear ribonucleoprotein A1-like 3	RA1L3_HUMAN	Unknown	1.4	0.83	11	NP40	5.76
Triosephosphate isomerase	TPIS_HUMAN	Triose-phosphate isomerase activity	0.71	0.66	6	SDS	3.44
T-cell leukemia/ lymphoma protein 1A	TCL1A_HUMAN	Multicellular organismal development	1.63	0.65	11	NP40	4.26
Heterogeneous nuclear ribonucleoproteins A2/B1	ROA2_HUMAN	Gene expression	1.36	0.61	11	NP40	3.90
Lactotransferrin	TRFL_HUMAN	Cellular iron ion homeostasis	0.72	0.56	6	SDS	2.54
Galectin-1	LEG1_HUMAN	Positive regulation of I-kappaB kinase/ NF-kappaB cascade	1.35	0.54	11	NP40	3.32
Heterogeneous nuclear ribonucleoprotein A3	ROA3_HUMAN	Gene expression	1.39	0.54	9	SDS	2.45
Histone H2A.J	H2AJ_HUMAN	Nucleosome assembly	1.21	0.53	11	NP40	3.22
Histone H4	H4_HUMAN	Nucleosome assembly	1.27	0.51	11	NP40	3.09
Lysozyme C	LYSC_HUMAN	Lysozyme activity	0.97	0.50	9	SDS	2.10
Vimentin	VIME_HUMAN	Cellular component movement	0.99	0.41	11	NP40	2.26
Peroxiredoxin-2	PRDX2_HUMAN	Negative regulation of apoptotic process	1.07	0.40	11	NP40	2.21

Relative expression of proteins in CLL samples was determined using iTRAQ reagents followed by LC-MALDI mass spectrometry analysis. Data analysis was performed using PorteinPilot software coupled with the Paragon and ProGroup algorithms. This table shows proteins with the most heterogeneous expression that were quantified in at least five CLL samples. Average is the average of iTRAQ signals in five or more CLL samples, SD: standard deviation of iTRAQ signals. This table was sorted on the basis of SD followed by the sample count.

The Affymetrix accession numbers were then used to extract the cognate gene expression ratio of the altered proteins. Of the twelve proteins, two did not have a match with the Affymetrix IDs and could not be used in the analysis. The analysis demonstrated that some mRNA and the proteins they encod had a similar expression pattern, while some others showed an opposite expression patterns (Figure 5.15). For example, TCL-1 and LSAP-1 were up-regulated at both the transcripts and protein levels in the poor prognosis CLL. However, H4 and H2B2E had high transcripts yet low protein expression in the poor prognosis CLL. In addition, LSM3 showed low transcript but high protein expression in the poor prognosis CLL.

5.3 Discussion

Quantitative proteomics approaches have been previously used to study the protein expression of poor prognosis and good prognosis CLL. Three proteomics studies were conducted using gel-based quantitative approaches such as 2DE with silver staining or DIGE (Voss et al., 2001, Cochran et al., 2003, Perrot et al., 2011). In contrast, only one CLL proteomics study has been performed using gel-free quantitative approaches like cICAT (Barnidge et al., 2005a). Similarly, in the current project a gel-free quantitative approach using iTRAQ was utilised. In gel-based quantitative proteomics protein quantification occurs prior to protein identification. Therefore, it is usually only proteins with altered expression that are subjected to identification. This renders the number of quantified and identified proteins low e.g. 17 (Voss et al., 2001) or 31 proteins (Perrot et al., 2011) and leaves

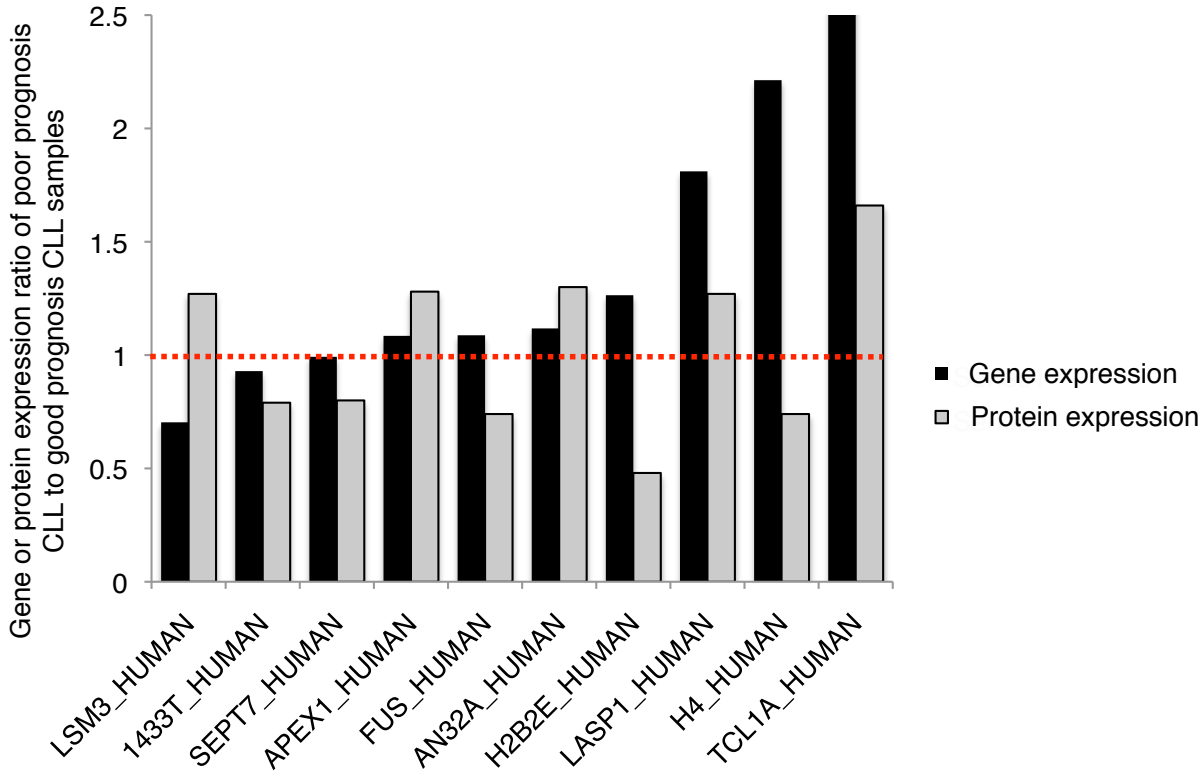


Figure 5.15: Gene expression of the altered proteins in CLL samples. Published CLL gene expression data (Huttmann et al., 2006) ($CD38^+$/$ZAP\text{-}70^+$ CLL samples versus $CD38^-$/$ZAP\text{-}70^-$ CLL samples) were used to study the gene expression ratio of the altered protein in CLL samples ($CD38^+$ CLL samples versus $CD38^-$ CLL samples and $ZAP\text{-}70^+$ CLL samples versus $ZAP\text{-}70^-$ CLL samples). The red dotted line represents a gene:protein expression ratio 1, where the poor prognosis and good prognosis CLL samples had an equal expression of a gene or a protein.

proteins with similar expression unidentified. However, in gel-free quantitative proteomics protein identification happens prior to protein quantification. This makes the number of identified and quantified proteins high; for example, 538 proteins (Barnidge et al., 2005a) and 655 proteins in the current project. From these proteins the main attention was paid to those with altered expression in CLL. However, proteins with similar expression can also be used to indicate biological functions that do not change in CLL.

Of the mass labels that have been introduced to facilitate quantitative proteomics, only one was previously used in CLL proteomics studies. Barnidge et al., (2005a) exploited the cICAT technique to study protein expression in one U-CLL sample versus one M-CLL sample. This mass labelling technique allowed simultaneous labelling of only two samples utilising only cysteine-containing peptides (Yi et al., 2005). In contrast, the mass labels used in the project described in this chapter were iTRAQ, which allowed simultaneous labelling of four samples (Ross et al., 2004).

The variability in the clinical outcome of CLL (Pepper et al., 2012) implies that the reference samples, which were used for the relative quantification of CLL proteome, should have been prepared from CLL samples of patients with different CLL prognosis. In principle, this would give an assurance that for every quantified peptide in the test samples there would be an analogous peptide in the reference sample to compare with in order to generate the relative quantification. This was the case with the relative

quantification of the proteins in the SDS fractions. However, the NP40 reference sample, which was used for the relative quantification of proteins in the NP40 fractions, was prepared from one CLL sample. While this had the possibility to create a situation where the relative quantification of peptides in the test samples are missed due to the absence of the analogous peptides in the reference sample, non of quantified proteins in the test samples (NP40 fractions) were absent in the NP40 reference sample.

Of the proteins that were found with iTRAQ signal ≤0.80 or ≥1.25 and significant *p* value in the poor prognosis CLL and the good prognosis CLL were different types of keratin. However, keratin has been shown in different reports to be one of the most contaminating proteins that are commonly identified by mass spectrometry (Ding et al., 2003, Lubec and Afjehi-Sadat, 2007). In fact, when a sample containing 20 highly purified recombinant human proteins was analysed by 27 different proteomics labs, all of these labs identified different types of keratin to be identified in the (Bell et al., 2009). This happened despite the fact that keratin was not included with the 20 recombinant proteins in the original sample.

Of the previous proteomics studies that have investigated protein expression in poor prognosis CLL and good prognosis CLL, the majority (3/4) used mutational status of *IGHV* to discriminate between the two forms of CLL (Cochran et al., 2003, Barnidge et al., 2005, Perrot et al., 2011). In fact, only one study was partially conducted on $CD38^+$ CLL and $CD38^-$ CLL

and showed no significant changes in the two groups of samples (Cochran et al., 2003). This project was initially designed to study protein expression in six $CD38^+$ CLL samples and six $CD38^-$ CLL samples. This led to the identification of six proteins with different expression in the two groups of samples such as TCL-1, which is linked to the tumourigenesis of CLL and poor prognosis of CLL (Bichi et al., 2002, Herling et al., 2006)

While the main goal of this chapter was to study protein expression in the two forms of CLL, patients with $CD38^+$ CLL do not always have an aggressive disease and patients with $CD38^-$ CLL do not always have an indolent disease (Durig et al., 2002, Pepper et al., 2012). As a result, subset analysis based on other common prognostic markers such as ZAP-70 expression, mutational status of *IGHV* and CLL stage were performed in an attempt to identity more proteins with altered expression in the two forms of CLL. These analyses increased the number of identified proteins with altered expression from six (CD38 expression based analysis) to 15 proteins. Interestingly, one of these proteins (LIM and SH3 domain protein 1) was observed with higher expression in ZAP-70^+ CLL samples as well as in U-CLL samples. This finding might be explained by the observation that there is a good, but not perfect, correlation between ZAP-70 positivity and unmutated *IGHV* genes (Crespo et al., 2003, Rassenti et al., 2004). In fact, in this study, 3 out of 4 of the U-CLL samples were ZAP-70^+ and all the four M-CLL were ZAP-70^-.

Of proteins that were demonstrated by previous CLL proteomics studies to be differentially expressed in the two forms of CLL, some were not identified in this project. For example, cytochrome c oxidase subunit 6B1 (CX6B1_HUMAN) was found with reduced expression in U-CLL compared to M-CLL (Barnidge et al., 2005a). The reasons for this might be due to the different proteomics approach used by Barnidge et al. (2005a) including different sample labelling (cICAT), peptide soft ionisation (electrospray ionisation) and the mass spectrometry instrumentation used (linear ion-trap mass spectrometer). Alternatively, the absence of this protein ID may simply reflect the biological differences observed between CLL patients.

This project yielded some contradictory results to those observed in previous proteomics studies. For instance, nucleophosmin was reported to be undetectable in U-CLL but present in M-CLL (Cochran et al., 2003). This protein was identified and quantified in all CLL samples in the current study, with at least 5 specific peptides (confidence score ≥ 98%). In addition, this protein did not show significantly altered expression in U-CLL samples versus M-CLL samples or in any other prognostic marker-based analyses. In the Cochran *et al.* (2003) study, nucleophosmin was identified with five peptides using peptide mass fingerprint (PMF) analysis. In contrast, in my study nucleophosmin was identified and quantified with at least 5 specific peptides based on tandem mass spectrometry analysis (peptide sequencing), which is reportedly more sensitive than the PMF method (Steen and Mann, 2004).

Consistent with previous CLL proteomics studies, I identified proteins that were previously shown to have different expression in the two forms of CLL. For example, my project and a previous CLL proteomics study reported an increased expression of acidic leucine-rich nuclear phosphoprotein 32 family member A in poor prognosis CLL (Barnidge et al., 2005a). Similarly, the current project and two other CLL proteomics studies showed that LIM and SH3 domain protein 1 was up-regulated in poor prognosis CLL (Voss et al., 2001, Barnidge et al., 2005). In addition, my project and a previous CLL proteomics study demonstrated that myosin-9 was found with low expression in poor prognosis CLL (Barnidge et al., 2005a).

While performing subset analysis based on different prognostic markers had the potential to identify more proteins with altered expression in the two forms of CLL, it is difficult to claim that it identified all relevant CLL proteins. As CLL is characterised by a heterogeneous clinical outcome, CLL relevant proteins might be those found amongst the proteins with the most heterogeneous expression. Some of the proteins identified with the most variable expression in the CLL samples examined, were linked to cancer and in some cases specifically to leukaemia. For example, Vimentin was previously linked to CLL prognosis, where vimentin expression inversely correlated with the percentage of smudge cells, a good prognostic marker (Nowakowski et al., 2007). The expression of this proteins was shown to increase in U-CLL following BCR stimulation (Perrot et al., 2011). In addition,

S100A8 and S100A9 were shown to be involved in the progression of breast cancer and prostate cancer (Cormier et al., 2009, Hermani et al., 2005) and were linked to NF-κB pathway (Benedyk et al., 2007), an important pathway in CLL (Hewamana et al., 2008). Furthermore, S100A8 was reported as a poor prognostic marker in acute myeloid leukaemia (AML) (Nicolas et al., 2011). More importantly, TCL-1 was linked to the tumourigenesis of CLL (Bichi et al., 2002) and was observed to associate with poor prognosis CLL (Herling et al., 2006). Consistently, in my project, TCL-1 was shown with an increased expression in $CD38^{+}$ CLL. This probably indicates that investigating heterogeneously expressed proteins in CLL is a good strategy to identify proteins with potential importance in this disease, whose expression might be independent of the prognostic markers that were used in this study.

Different studies utilised transcriptomic and proteomics approaches to explore whether a correlation can be found between data generated by both approaches. In this regard, (Nishizuka et al., 2003) demonstrated that protein expression and gene expression correlates well among structural proteins and poorly for non-structural proteins. Another study showed poor correlation between 1900 proteins and their cognate transcripts, but good correlation among tissue specific proteins and genes (Conrads et al., 2005). In the context of CLL, a CLL proteomics study showed that 3/25 proteins that exhibited different expression in U-CLL before and after BCR stimulation were altered at mRNA and protein levels in the same direction, whereas 8 were altered at mRNA and protein levels in opposite directions

(Perrot et al., 2011). In the present project, some of the highly expressed proteins in poor prognosis CLL samples were also up-regulated at the level of gene expression, albeit in a different set of samples. This may indicate that proteins such as TCL-1 and LIM and SH3 domain protein 1 are predominantly regulated at the level of gene transcription. In contrast, some proteins were reduced in poor prognosis CLL, yet their gene expression was over-expressed in poor prognosis CLL. This may suggest that proteins such as Histone H4 and Histone H2B type 2-E are regulated, at least to some extent, by post-translational modifications.

This chapter identified 15 proteins with altered expression in poor prognosis CLL compared with good prognosis CLL. Additionally, it highlighted 14 proteins with the most variable expression in CLL samples with either poor or good prognosis. The next step was to choose some of these proteins for validation and further investigation in additional CLL samples (see chapter six).

Chapter Six

Validation and Investigation of Proteins with Potential Relevance to CLL

6.1 Introduction

The CLL qualitative proteomics analysis (chapter four) and the CLL quantitative proteomics analysis (chapter five) identified 49 proteins with potential relevance to CLL. These proteins fell into three groups; firstly proteins (n= 20) whose gene expression was found to be specific to CLL cells but not B-cells from healthy donors. Secondly, proteins (n= 15) that were differentially expressed in poor prognosis CLL compared to good prognosis CLL. Finally, proteins (n= 14) that showed highly variable expression regardless of other CLL prognostic markers.

The aim of this chapter was to select six of the 49 proteins for validation and further investigation. A method independent from proteomics was used to measure the expression of the selected proteins in additional CLL samples. The expression of the selected proteins was always normalised to Actin expression as a loading control. This was performed using densitometric analysis and ImageJ software 1.44o. Expression of the selected proteins was subsequently analysed in the context of four prognostic markers (CD38 expression, ZAP-70 expression, mutational status of *IGHV* genes and Binet's staging) as well as time to first treatment (TTFT). Overall survival was not assessed as none of the patient samples used in this study were derived from patients who had died from to their disease at the time of analysis.

6.2. Results

6.2.1 Evaluation of protein expression by western blotting

Before using western blotting to study the expression of the proteins that showed possible importance in CLL (chapter four and five) in additional cohort of CLL samples, elevating doses of cell lysates were used to explore the sensitivity of protein quantification by western blotting. Increasing doses from whole cell lysate of CLL sample (3µg, 6µg and 12µg) were separated by SDS-PAGE followed by western blotting analysis using a specific antibody to actin. Detected bands were then subjected to densitometric analysis using ImageJ 1.44o software in order to measure their intensity. Figure 6.1 shows a very strong correlation between sample dose and band intensity. This analysis suggested that western blotting could be used to measure protein expression in patient samples.

6.2.2 Proteins whose transcripts are specific to CLL cells but not normal B-cells

Integrating my CLL proteome with published transcriptomes of CLL and normal B-cells (Huttmann et al., 2006, Hutcheson et al., 2008) highlighted proteins with possible relevance to CLL, as their gene expression was restricted to CLL cells compared to normal B-cells (described in chapter four). Two of these proteins were subsequently investigated using samples of normal B-cells and CLL cells. Three criteria were used for protein selection; a protein must have been detected in the majority of CLL samples (≥10 samples) that were analysed by proteomics, it must have been identified in

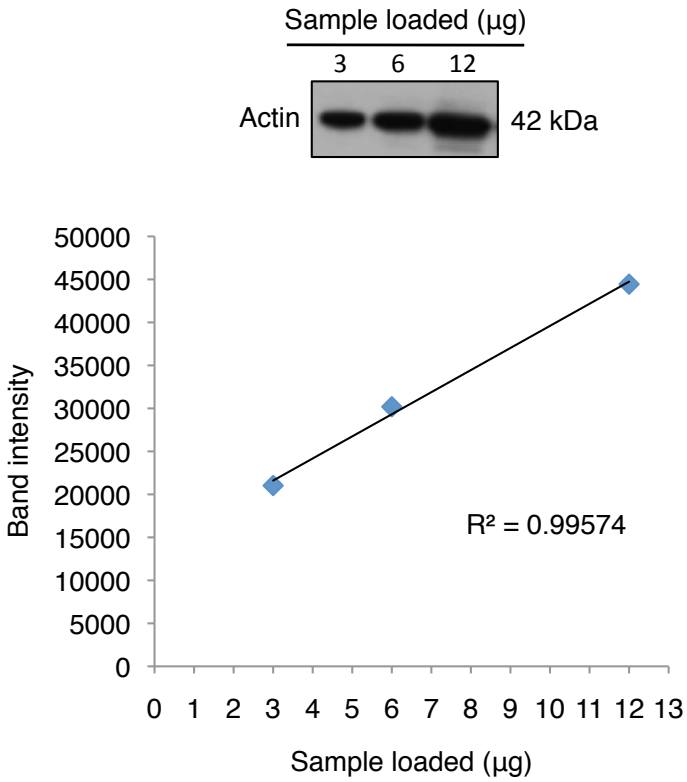


Figure 6.1: Western blotting analysis of actin in increasing doses of cell lysate from CLL sample. Elevating doses from whole cell lysate of CLL sample (3µg, 6µg and 12µg) were resolved by SDS-PAGE followed by transfer onto PVDF membrane and antibody detection using a specific antibody to actin (A). The intensity of the detected bands was measured using ImageJ 1.44o software. The analysis showed a very strong correlation between sample dose and band intensity (B).

CLL samples with multiple peptide IDs (>4), and it must have been called 'absent' in all 11 normal B-cell samples tested by Hutcheson *et al.*, (2008). Thyroid hormone receptor-associated protein 3 (TR150_HUMAN) and Heterochromatin protein 1-binding protein 3 (HP1B3_HUMAN) were found to satisfy these criteria; they were identified by proteomics in ≥11 out of 12 CLL samples with at least six different peptides (Ion Score for each peptide ≥95% C.I.) (Figures 6.2 and 6.3).

6.2.2.1 Isolation of B-cells from peripheral blood of healthy donors

To isolate B-cells from buffy coat samples of healthy donors, which were received 24 hours following isolation, the low-density layer of mononuclear cells containing B-cells were firstly separated using Ficoll. An aliquot of the separated cells were then stained with an anti-CD19 labelled with allophycocyanin (APC) and subjected to analysis by flow cytomerty in order to assess the percentage of CD19-expressing cells (B-cells). The analysis showed that further purification of B-cells was still required to increase the percentage of $CD19^+$ cells in the samples (Figure 6.4A). This was achieved by positive isolation of B-cells using magnetic beads labelled with a CD19 antibody. Following cell isolation, the bead-bound cells were released from the beads using DETACHaBEAD CD19 reagent. Positively isolated and detached cells were stained with an anti-CD19 antibody-APC followed by flow cytometric analysis to measure the percentage of CD19-expressing cells. The analysis demonstrated that the populations other than

(A)

Full sequence of Thyroid hormone receptor-associated protein 3

MSKTNKSKSGSRSSRSRSASRSRSRSFSKSRSRSRSLSRSRKRRLSSRSRSRSYSPAHNRERNHP
RVYQNRDFRGHNRGYRRPYYFRGRNRGFYPWGQYNRGGYGNYRSNWQNYRQAYSPRRGRSRS
RSPKRRSPSPRSRSHSRNSDKSSSDRSRRSSSSRSSSNHSRVESSKRKSAKEKKSSSKDSRPSQA
AGDNQGDEAKEQTFSGGTSQDTK**ASESSKPWPDATYGTGSASR**<u>**ASAVSELSPR**</u>ERSPALK**SPLQS**
VVVRRRSPRPSPVPKPSPPLSSTSQMGSTLPSGAGYQSGTHQGQFDHGSGSLSPSKKSPVGKSPP
STGSTYGSSQKEESAASGGAAYTKRYLEEQKTENGKDKEQKQTNTDKEKIKEKGSFSDTGLGDGKM
KSDSFAPKTDSEKPFRGSQSPKRYKLRDDFEKKMADFHKEEMDDQDKDKAKGRKESEFDDEPKFM
SKVIGANKNQEEEKSGKWEGLVYAPPGKEKQR**KTEELEEESFPER**SKKEDRGKRSEGGHRGFVPE
KNFRVTAYKAVQEKSSSPPPRKTSESRDKLGAKGDFPTGKSSFSITREAQVNVRMDSFDEDLARPSG
LLAQERKLCRDLVHSNKKEQEFR**SIFQHIQSAQSQR**SPSELFAQHIVTIVHHVKEHHFGSSGMTLHER
FTKYLKRGTEQEAAKNKKSPEIHRRIDISPSTFRKHGLAHDEMKSPREPGYKAEGKYKDDPVDLRLDI
ERRKKHKERDLKRGKSRESVDSRDSSHSRERSAEKTEKTHKGSKKQKKHRRARDRSRSSSSSSQS
SHSYKAEEYTEETEEREESTTGFDKSRLGTKDFVGPSERGGGRARGTFQFRARGRGWGRGNYSG
NNNNNSNNDFQKR**NREEEWDPEYTPK**SKKYYLHDDREGEGSDKWVSRGRGRGAFPRGRGRFMF
RKSSTSPKWAHDKFSGEEGEIEDDESGTENREEKDNIQPTTE

(B)

MS/MS spectra of peptide of mass 1529.7

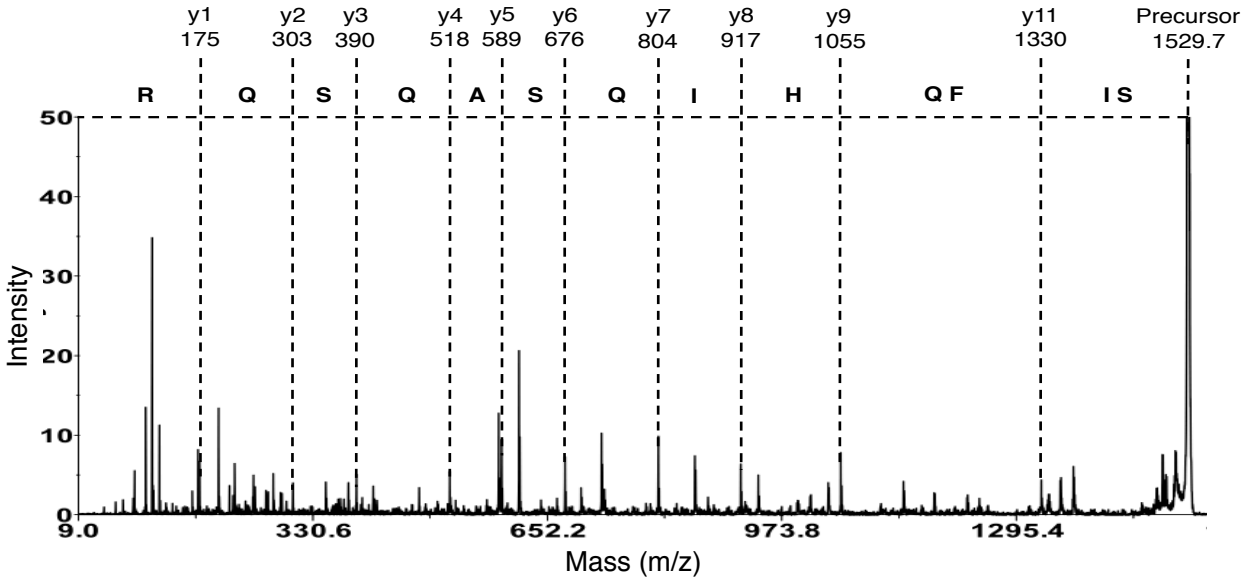


Figure 6.2: Identification of thyroid hormone receptor-associated protein 3 (TR150_HUMAN or TRAP150) by proteomics. TR150 was identified with six good quality peptides (ion score confidence interval ≥ 95%). These peptides are shown in (**Red Bold**) in the full sequence of TR150 (A). The MS/MS spectra of one peptide (**SIFQHIQSAQSQR**), which is marked with a rectangle, is shown to illustrate the assignment of sequence from the MS/MS data (B). The sequence identified in the MS/MS spectra is read from the C terminus as only y ions were shown in the MS/MS spectra.

(A) **Full sequence of Heterochromatin protein 1-binding protein 3**

MATDTSQGELVHPKALPLIVGAQLIHADKLGEKVEDSTMPIRRTVNSTRETPPKSKLAEGEEEKPEPDISSEESV
STVEEQENETPPATSSEAEQPKGEPENEEKEENKSSEETKKDEKDQSKEKEKKVKKTIPSWATLSASQLARA
QKQTPMASSPRPKMDAILTEAIKACFQKSGASVVAIRKYIIHKYPSLELERRGYLLKQALKRELNRGVIKQVKG
KGASGSFVVVQKSRKTPQKSRNRKNRSSAVDPEPQVKLEDVLPLAFTRLCEPKEASYSLIRKYVSQYYPKLR
VDIRPQLLKNALQRAVERGQLEQITGKGASGTFQLKKSGEKPLLGGSLMEYAILSAIAAMNEPKTCSTTALKKYV
LENHPGTNSNYQMHLLKKTLQKCEKNGWMEQISGKGFSGTFQLCFPYYPSPGVLFPKKEPDDSRDEDEDED
ESSEEDSEDEEPPPKRRLQKKTPAKSPGKAASVKQRGSKPAPKVSAAQRGKARPLPKKAPPKAKTPAKKTRP
SSTVIKKPSGGSSKKPATSARKEVKLPGKGKSTMKKSFRVKK

(B) **MS/MS spectra of peptide of mass 1601.8**

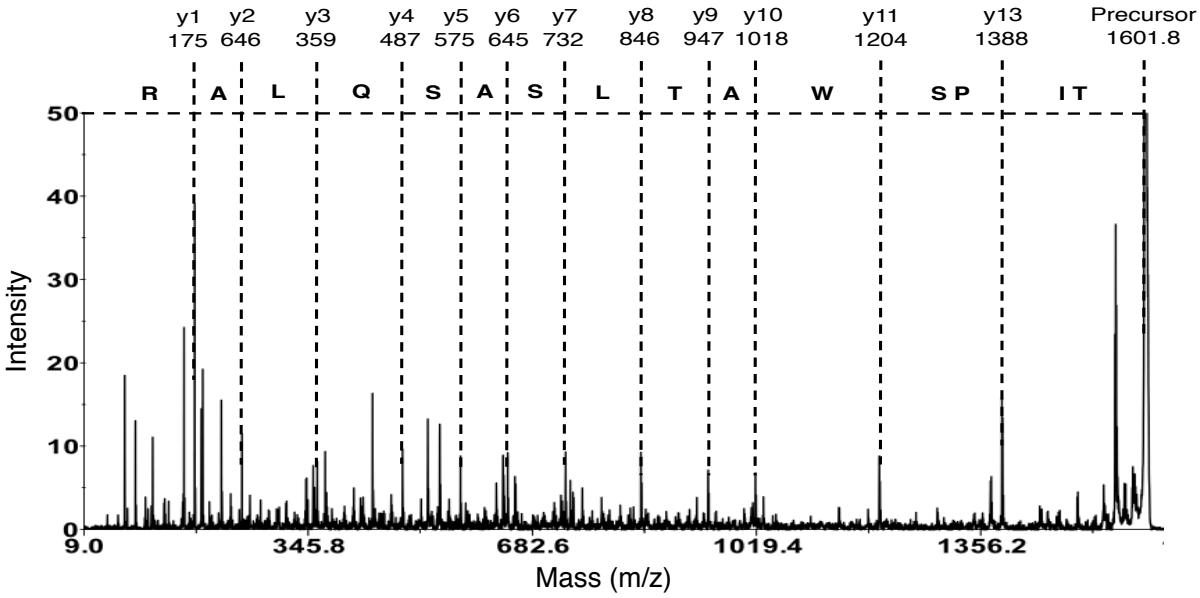


Figure 6.3: Identification of heterochromatin protein 1-binding protein 3 (HP1BP3_HUMAN) by proteomics. HP1BP3 was identified with eight good quality peptides (ion score confidence interval ≥ 95%). These peptides are shown in (**Red Bold**) in the full sequence of HP1BP3 (A). The MS/MS spectra of one peptide (**TIPSWATLSASQLAR**) which is marked with a rectangle, is shown to illustrate the assignment of sequence from the MS/MS data (B). The sequence identified in the MS/MS spectra is read from the C terminus as only y ions were shown in the MS/MS spectra.

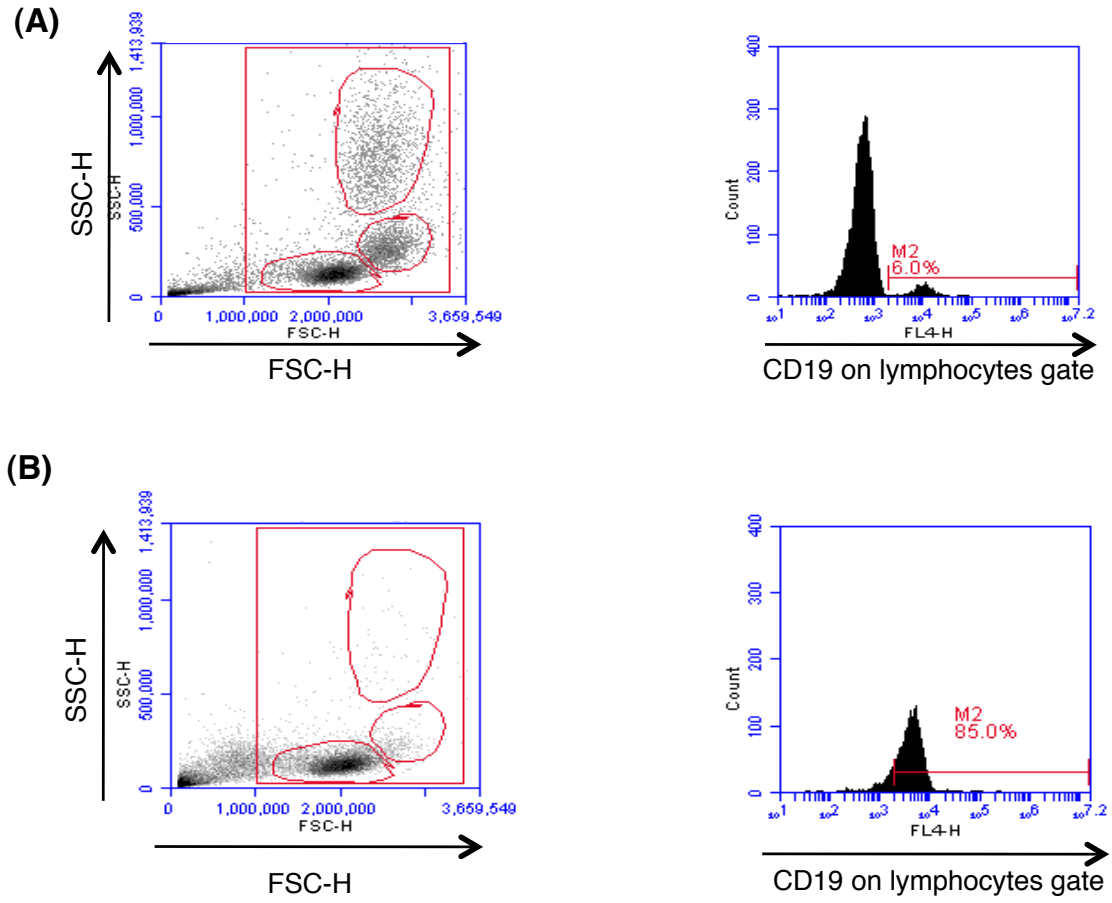


Figure 6.4: Isolation of B cells from buffy coat samples of healthy donor. Blood mononuclear cells were separated from buffy coat samples and stained with an anti CD19 antibody followed by flow cytometric analysis (A). More purification of B-cells was needed. This was achieved by performing positive isolation of B cells using CD19 Dynabeads followed by detaching the isolated cells from the Dynabeads. Flow cytometric analysis following staining with an anti CD19 antibody of the positively isolated and detached cells showed that 85% of the cells were $CD19^{+}$ (B cells) (B).

lymphocytes were diminished (Figure 6.4B), and that the percentage of $CD19^+$ cells increased dramatically (average = 85%, SD = 3% B-cells).

6.2.2.2 Expression of TR150 and HP1BP3 in normal B-cells samples

The analysis showed that TR150 and HP1BP3 were absent in the B-cells transcriptome (Hutcheson et al., 2008), but present in the CLL transcriptome (Huttmann et al., 2006) as well as the proteome of CLL samples generated in this project. To investigate whether TR150 and HP1BP3 proteins are expressed by normal B-cells, whole cell lysates from four normal B-cell samples, as well as two CLL samples, were resolved by SDS-PAGE. Separated proteins were transferred onto a PVDF membrane for antibody detection using specific antibodies against TR150 and HP1BP3. The analysis demonstrated that these two proteins were detected in the two CLL samples and in two out of four normal B-cells samples (Figure 6.5A). To investigate whether the amount of TR150 and HP1BP3 that was detected in 2/4 of the normal B-cells samples was derived from 'contaminating' non-B-cells, these two proteins were measured in normal B-cells and in matched (unseparated) peripheral blood mononuclear cells using western blotting. The analysis demonstrated that the amount of TR150 and HP2BP3 that was detected in the B-cell samples was greater than that found in the matched peripheral blood mononuclear cells (Figure 6.5B). This suggested that the amount of TR150 and BP1BP3 detected in these normal B-cell samples was not due to contamination from other peripheral blood mononuclear cells.

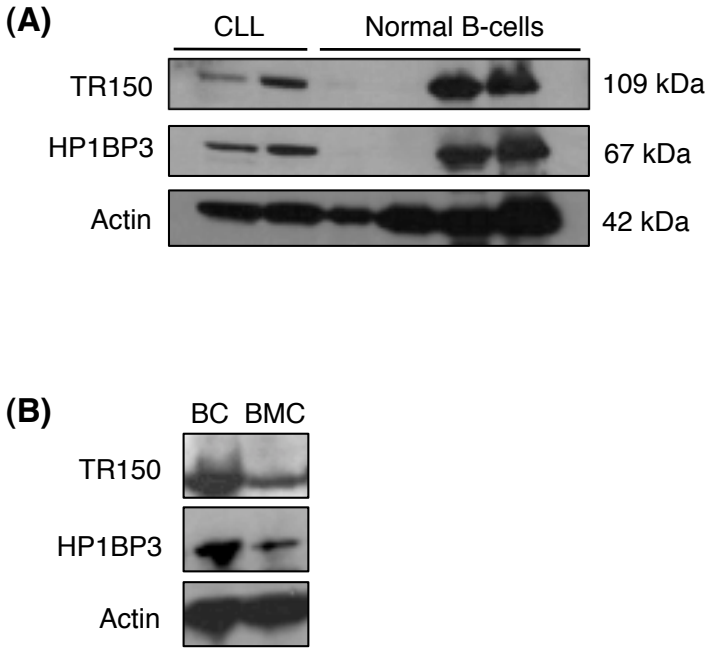


Figure 6.5: Expression of TR150 and HP1BP3 in CLL cells, normal B-cells and peripheral blood mononuclear cells. Proteins from whole cell lysates of CLL and normal B-cell samples were resolved in SDS-PAGE followed by transfer onto PVDF membrane and antibody detection using a specific antibody to TR150 and HP1BP3. This figure shows that HP1BP3 and TR150 are present in all CLL samples and in 50% of normal B-cell samples (A). In addition, the amount of TR150 and HP1BP3 in the normal B-cells (BC) is larger than that in peripheral blood mononuclear cells (BMC) (B).

TR150 and HP1BP3 were detected by proteomics in ≥11/12 CLL samples. However, these two proteins were also detected by western blotting in 2/4 of normal B-cell samples. These two normal B-cells samples showed higher expression of both TR150 and HP1BP3 compared to the two CLL samples. This indicated that these proteins were not specific to the pathological transformation to CLL but did not rule out a prognostic role for one or both of these proteins during the course of the disease. In addition, the strategy described in Chapter four of integrating the identified proteome of CLL cells with published transcriptome of CLL cells and normal B-cells identified proteins with potential link to CLL such as TAPP-2 (Costantini et al., 2009), although it was detected in both normal and malignant B-cells. Consequently, analyses were performed to determine if these two proteins have a potential impact on the prognosis of CLL.

6.2.2.3 Expression of TR150 in CLL samples

To explore whether the expression of TR150 (also known as TRAP150) had a potential influence on the clinical outcome of CLL, this protein was measured in 16 CLL samples exhibiting different prognostic markers. A specific antibody was used to assess TR150 expression in CLL samples following SDS-PAGE and western blotting. Figure 6.6A demonstrated that the expression of TR150 in CLL samples is relatively variable. In addition, TR150 expression showed no significant change in poor prognosis CLL compared to good prognosis CLL on the basis of CD38 (p= 0.23, n= 16, Figure 6.6B), ZAP-70 (p= 0.35, n= 16, Figure 6.6C) and *IGHV*

mutational status (p= 0.47, n= 12, Figure 6.6D). In contrast, CLL samples from patients in advanced stage had a significantly higher expression of TR150 compared to those from patients in the early stage of CLL (p= 0.05, n= 16, Figure 6.6E). This data suggest that TR150 may play a role in CLL progression, yet this remains to be confirmed in a largercohort of CLL patients.

Given the possible association of TR150 high expression with more advanced stage disease, I next investigated whether the expression of TR150 was predictive of the requirement for treatment in CLL patients. The median TR150 expression (TR150/Actin ratio = 0.55) was used to divide CLL patients into two groups; a high TR150 group with median follow-up 5.12 years and a low TR150 group with median follow-up of 7.32 years. The TTFT of the patients in each group was analysed using the Log-rank test and graphically represented using Kaplan-Meier curves. The analysis showed that the median TTFT in the high TR150 group was 0.78 years, while the median of TTFT in the low TR150 group was not reached (p= 0.02, n= 16, Figure 6.7). Although high expression of TR150 was significantly associated with an early need for treatment, not all patients with high TR150 required treatment. This preliminary analysis suggests the potential importance of TR150 in disease progression in at least some CLL patients. The biological reasons for this are yet to be elucidated.

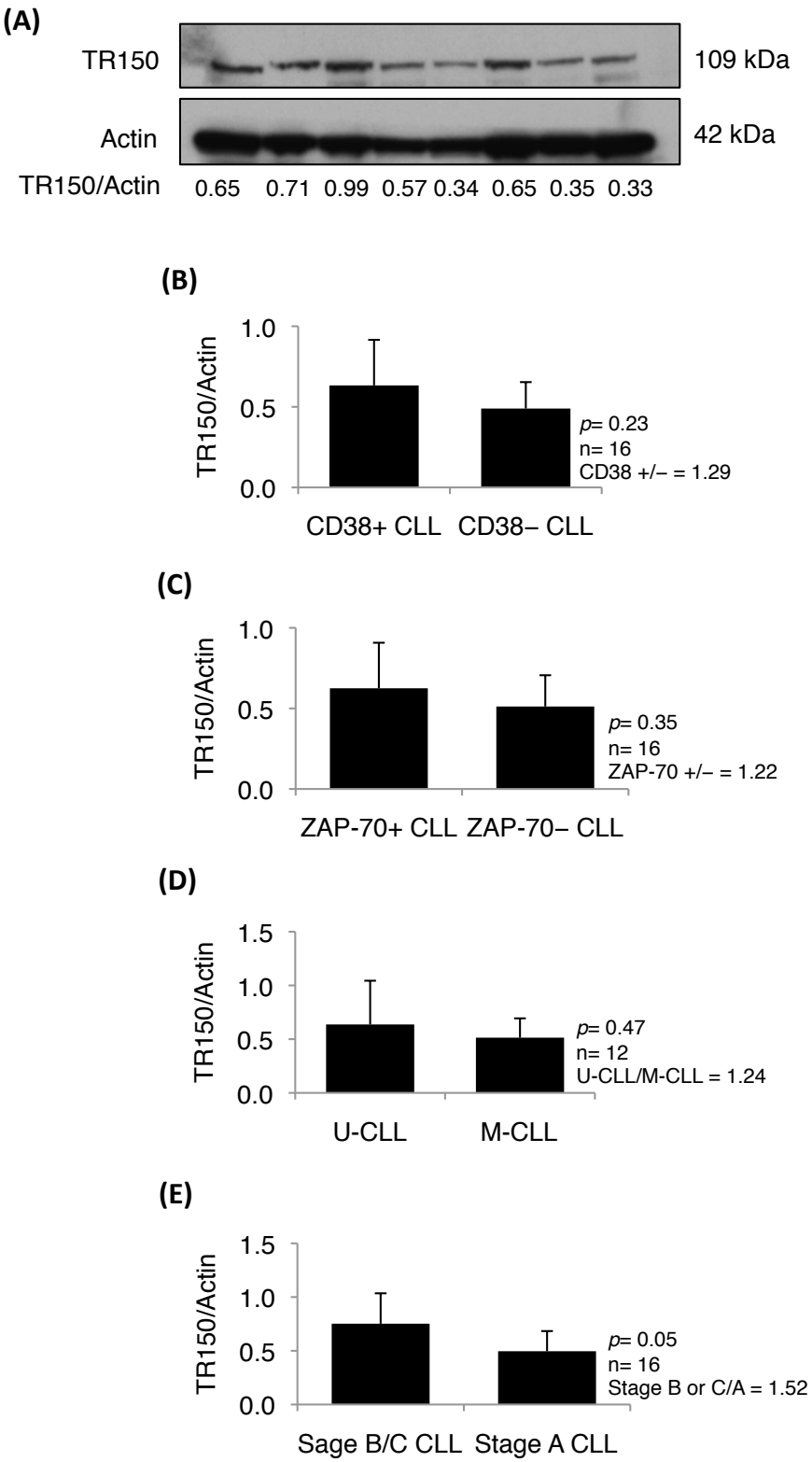


Figure 6.6: Expression of TR150 in CLL samples with different prognostic markers. Whole cell lysate from 16 CLL samples was separated by SDS-PAGE followed by transfer onto PVDF membrane and antibody detection using a specific antibody to TR150 (A). CLL samples with poor prognostic markers appear to express a higher amount of TR150 compared to those with good prognostic markers, yet the difference in the TR150 expression was statistically significant only in late versus early stages of CLL.

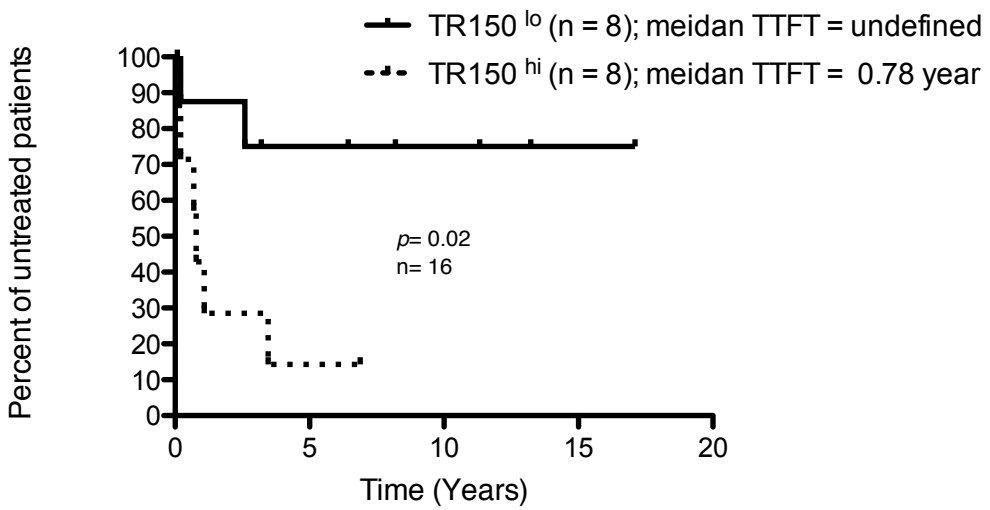


Figure 6.7: Kaplan-Meier curve comparing time period from diagnosis to the first treatment in CLL patients with high or low expression of TR150. Patients were divided into two groups using the median of TR150 expression; high TR150 group (TR150/Actin ratio > 0.55) and low TR150 group (TR150/Actin ratio < 0.55). TTFT was 0.78 years in the high TR150 group and undefined in the low TR150 group.

6.2.2.4 Expression of HP1BP3 in CLL samples

Measuring the abundance of HP1BP3 by western and antibody detection in 16 CLL samples from low and high risk CLL patients made it possible to investigate whether HP1BP3 had impact on CLL outcome. The analysis revealed that the expression of HP1BP3 in CLL samples was variable (Figure 6.8A). However, no significant association of HP1BP3 expression with other poor prognostic markers was observed. This included CD38 (p= 0.14, n= 16, Figure 6.8B), ZAP-70 (p= 0.69, n= 16, Figure 6.8C), *IGHV* mutational status (p= 0.57, n= 12, Figure 6.8D) and Binet stage (p= 0.43, n= 16, Figure 6.8E).

Despite the lack of association with known prognostic markers, I next investigated whether a particular pattern of HP1BP3 expression was associated with early need for treatment in CLL patients. The TTFT of patients with high or low HP1BP3, as defined by the median expression of HP1BP3 (HP1BP3/Actin ratio = 0.94), was analysed using the Log-rank test and graphically represented using Kaplan-Meier curve. The median follow-up was 7.54 years in the high HP1BP3 group and was 3.68 years in the low HP1BP3 group. The analysis revealed no association between increased HP1BP3 expression and an early need for treatment in CLL patients. The median TTFT was 9.1 years in the high HP1BP3 group versus 3.5 years in the low HP1BP3 group (p= 0.79, n= 16, Figure 6.9). Although the median TTFT in the HP1BP3 group (n= 8) was 9.1 years, 4/8 patients required early treatment (TTFT≤1.08 years). This highlights the dangers of over interpreting small datasets in terms

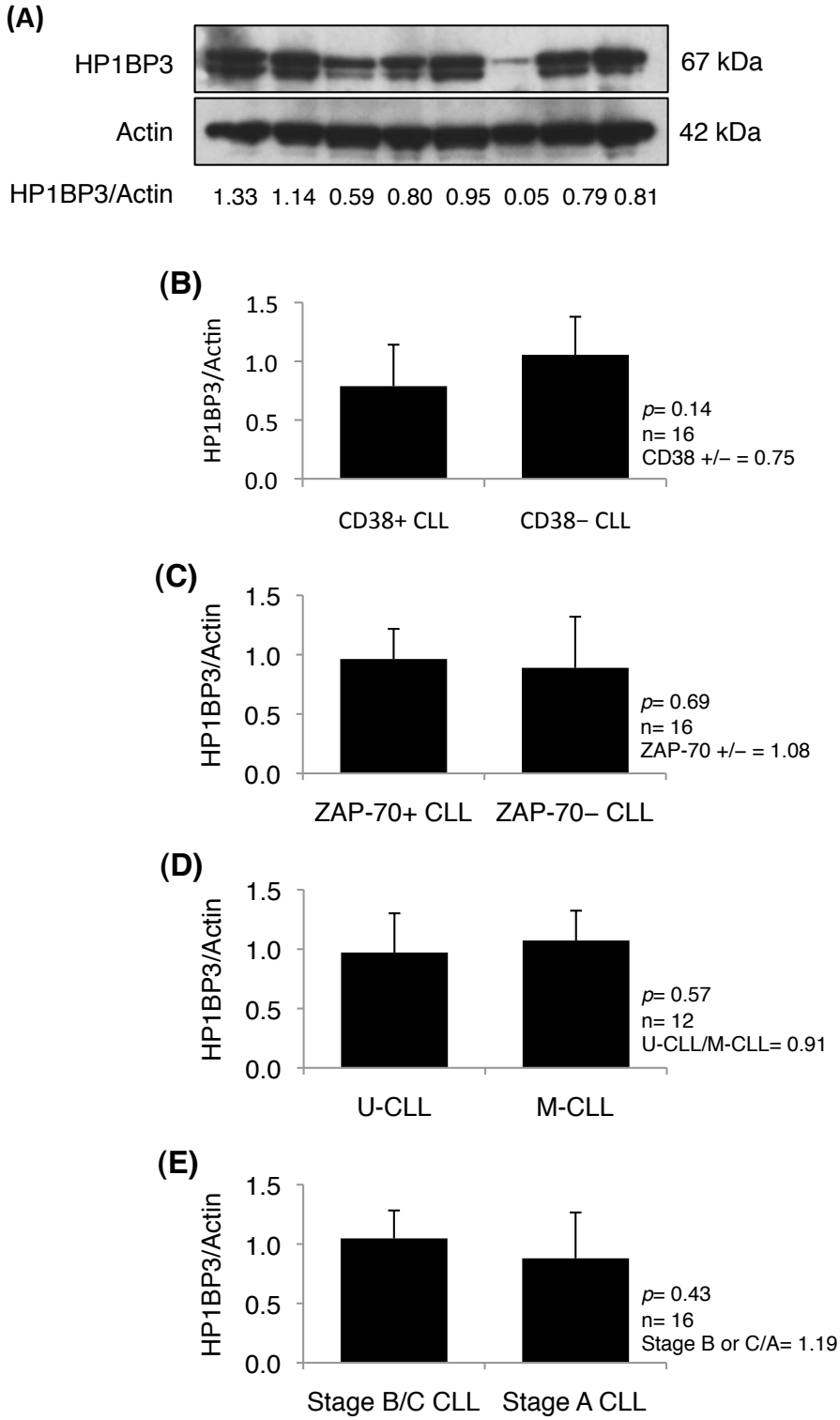


Figure 6.8: Expression of HP1BP3 in CLL samples with different prognostic markers. Whole cell lysate from 16 CLL samples was separated by SDS-PAGE followed by transfer onto PVDF membrane and antibody detection using a specific antibody to HP1BP3 (A). HP1BP3 expression is reduced in CLL samples with some poor prognostic markers (B and D) and up-regulated in CLL samples with some other prognostic markers (C and E). However, the differences in the expression of HP1BP3 were not statically significant.

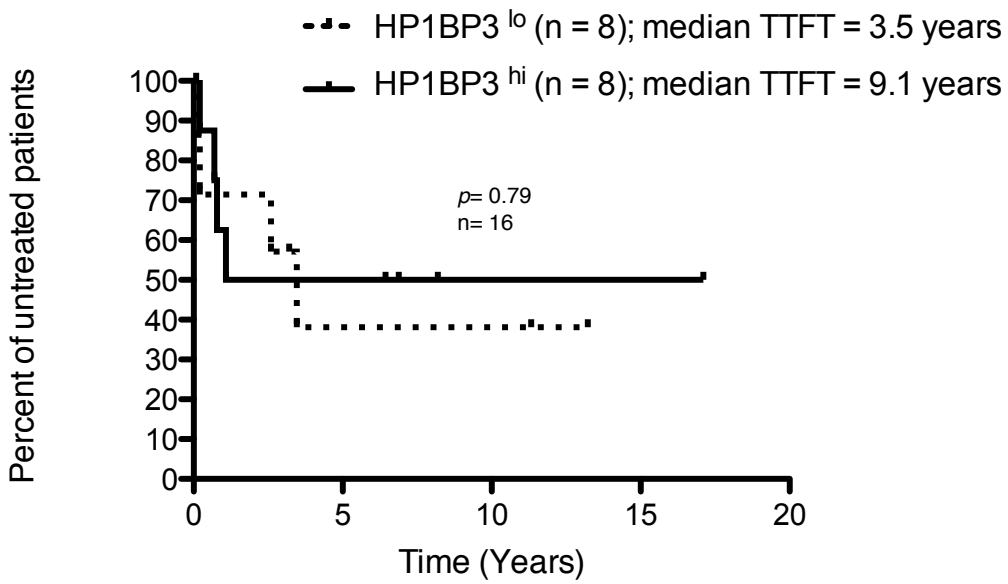


Figure 6.9: Kaplan-Meier curve comparing time period from diagnosis to the first treatment in CLL patients with high or low expression of HP1BP3. Patients were divided into two groups using the median of HP1BP3 expression; high HP1BP3 group (HP1BP3/Actin ratio > 0.94) and low HP1BP3 group (HP1BP3/Actin ratio < 0.94). Median TTFT was 9.1 years in the high HP1BP3 group and 3.5 in the low HP1BP3 group with insignificant *p* value.

of clinical prognosis. This was reinforced by a closer evaluation of the four untreated patients with high HP1BP3 expression; two had poor prognostic markers ($CD38^+$ and $ZAP\text{-}70^+$) and the other two had good prognostic markers ($CD38^-$, $ZAP\text{-}70^-$, M-CLL, Binet stage A). This indicates that in some cases even high expression of HP1BP3 combined with some other known poor prognostic markers such $CD38^+$ and $ZAP\text{-}70^+$ does not predispose patients to early treatment.

6.2.3 Proteins with altered expression in CLL samples

The quantitative proteomics analysis on CLL samples identified 15 proteins with an altered expression in poor prognosis CLL compared to good prognosis CLL (described in chapter five). Three of these proteins were selected for further analysis in additional CLL samples to validate their altered expression. Two criteria were used for protein selection; a protein must have been identified with the most altered expression and must have been quantified in the majority of CLL samples (≥10 samples) used for quantitative proteomics analysis. TCL-1 and myosin-9 were found to meet these criteria; they were the most altered proteins that were quantified in ≥10/12 CLL samples. Histone H4 was also validated in more CLL samples although it did not show the most altered expression in CLL samples. This was done because this project was initially designed to study the proteome associated with poor prognosis CLL based on CD38 expression. Histone H4 was among the proteins that showed altered expression in $CD38^+$ CLL compared to $CD38^-$ CLL.

6.2.3.1 Validation of TCL-1 in CLL samples

As demonstrated in the previous chapter, TCL-1 was over-expressed in the NP40 fractions of $CD38^+$ CLL samples compared to $CD38^-$ CLL samples (p=0.04, n= 11, Figure 6.10A). The NP40 fractions of 24 CLL samples (12 $CD38^+$ CLL samples versus 12 $CD38^-$ CLL samples) were used to analyse the expression of TCL-1. A specific antibody was used to determine TCL-1 expression in CLL samples following SDS-PAGE and western blotting (Figure 6.10B). The data confirmed that TCL-1 was over-expressed in $CD38^+$ CLL samples compared to $CD38^-$ CLL samples (p= 0.03, n= 24, Figure 6.10C).

Based on the high expression of TCL-1 in $CD38^+$ CLL samples, the association of TCL-1 with other poor prognostic markers was evaluated in the same CLL samples (n= 24). Example of the expression of TCL-1 in some of these samples is shown in Figure 6.11A. The normalised TCL-1/Actin data showed that TCL-1 was increased in CLL samples with other poor prognostic markers. This included ZAP-70^+ CLL (p= 0.03, n= 20, Figure 6.11B), U-CLL (p=0.05, n=17, Figure 6.11C) and Binet stage B/C CLL (p= 0.02, n= 22, Figure 6.11D).

Given the association of TCL-1 with poor prognosis CLL, the next question to address was whether TCL-1 expression could predict for early treatment in CLL patients. Based on the median of TCL-1 expression (TCL-1/Actin ratio = 0.56) patients were considered to have either high or low

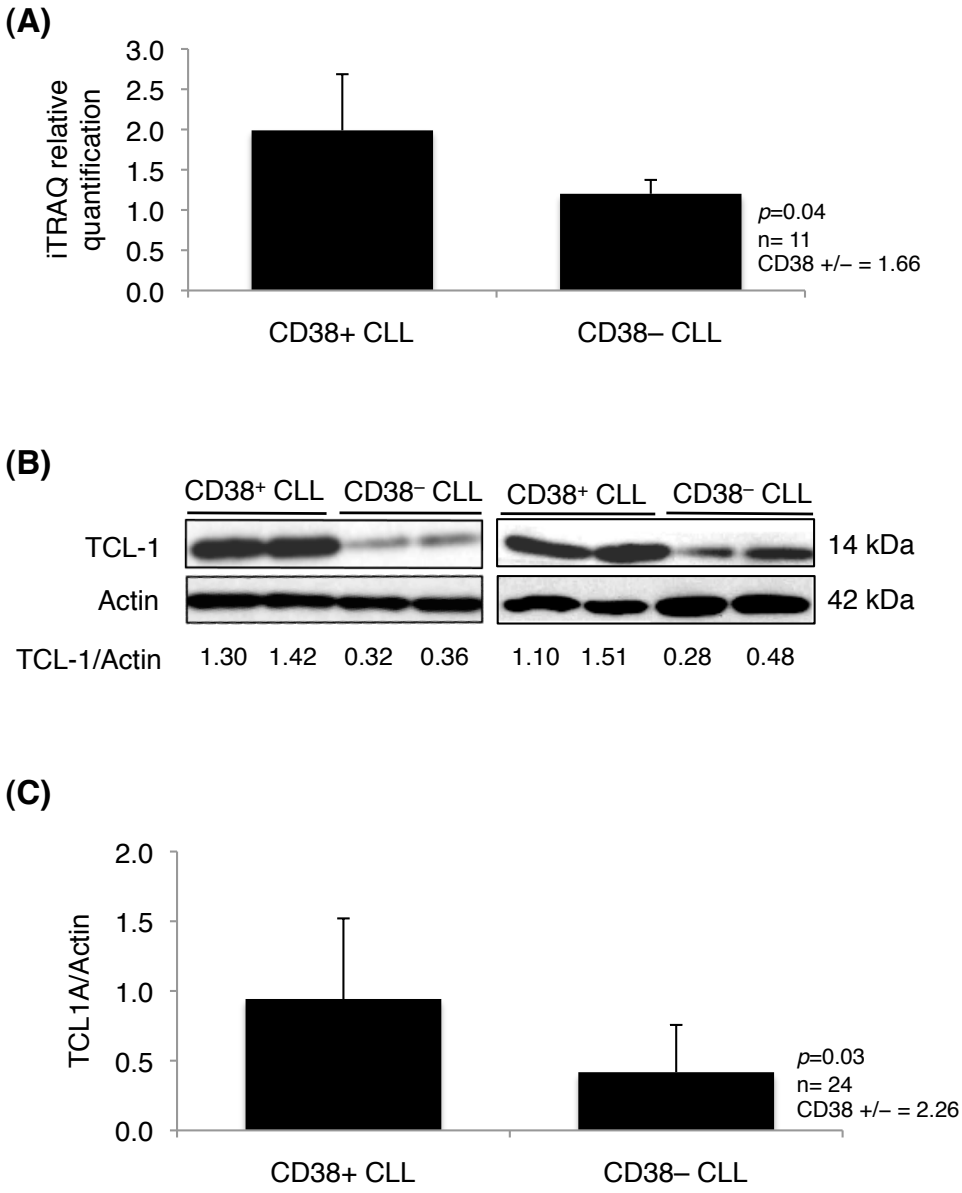


Figure 6.10: Validation of the altered expression of T-cell leukemia/ lymphoma protein 1A (TCL1) in CLL samples with high or low CD38 expression. iTRAQ quantification showed that TCL-1 was over-expressed in the NP40 fractions of CD38+ CLL (A). Proteins from the NP40 fractions of CLL samples were resolved by SDS-PAGE followed by proteins transfer onto PVDF membrane and antibody detection using a specific antibody to TCL-1 (B). The analysis demonstrated that TCL-1 was highly expressed in CD38+ CLL samples (C).

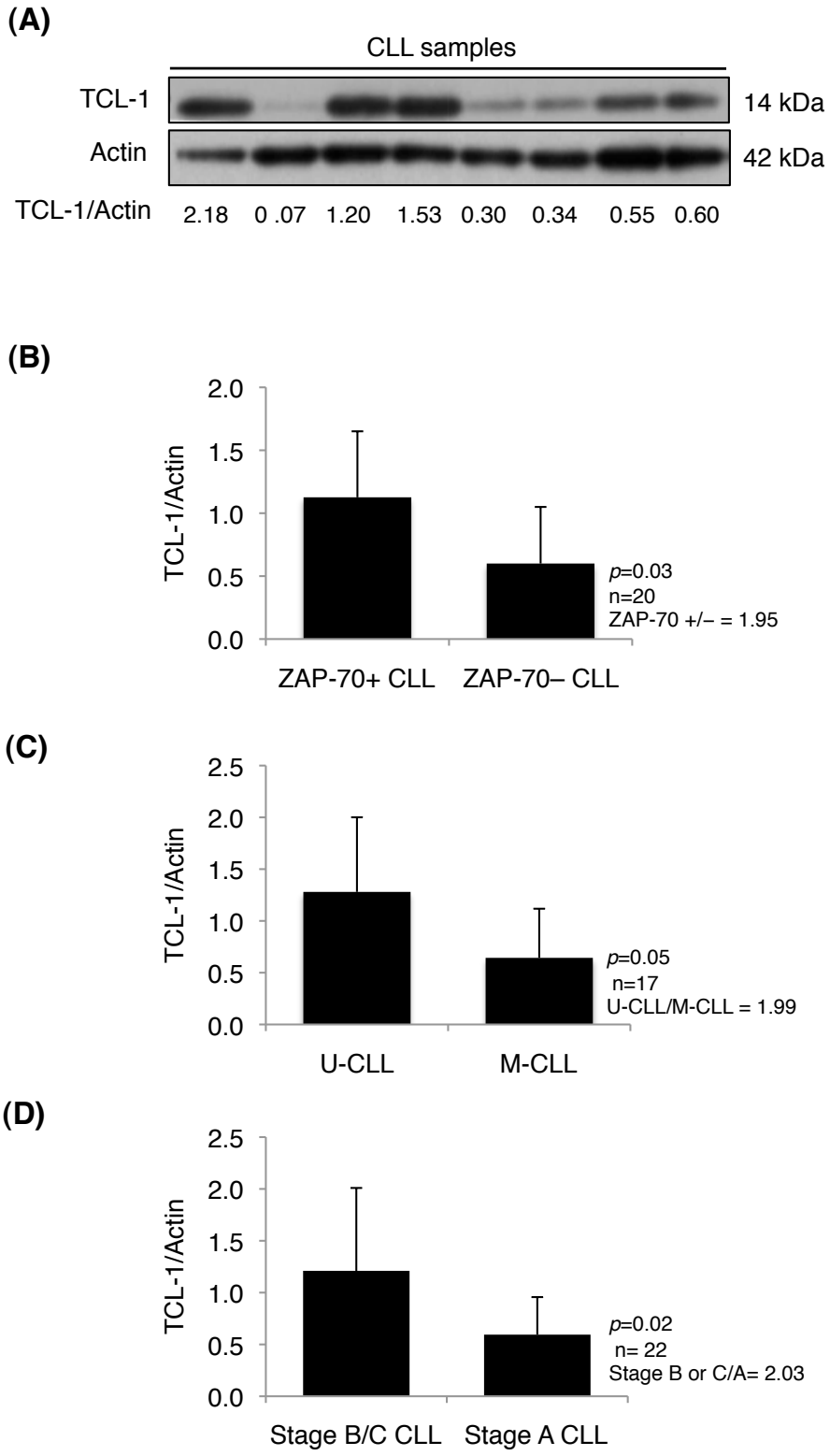


Figure 6.11: TCL-1 expression in CLL samples with other prognostic markers. Example of TCL-1 expression in CLL samples (A). TCL-1 was up-regulated in ZAP-70^{+} CLL (B), U-CLL (C) and stage B or C CLL (C).

expression of TCL-1. The median follow-up in the high TCL-1 group was 6.88 years and 4.23 years in the low TCL-1 group. The TTFT of patients in each group was analysed using the Log-rank test and graphically represented using Kaplan-Meier curves. As shown in Figure 6.12 the median TTFT in the high TCL-1 group was 5.4 years and was not reached in the low TCL-1 group (p= 0.01, n= 22). Nevertheless, some patients with high TCL-1 did not require treatment even 13.2 years after diagnosis suggesting that patients with elevated TCL-1 expression do not uniformly require early treatment. Despite the limited number of samples used this study and the larger follow-up in the high TCL-1 group compared to the low TCL-1 group, this analysis provided preliminary evidence of the potential prediction of TTFT on the basis of TCL-1 expression.

6.2.3.2 Validation of Histone H4 in CLL samples

One of the proteins that exhibited altered expression in the SDS fractions of $CD38^+$ CLL samples compared with $CD38^-$ CLL samples was Histone H4 (p= 0.05, n= 12, Figure 6.13A). To further study the expression of Histone H4 in additional CLL samples, proteins in the SDS fraction of 17 CLL samples (8 $CD38^+$ CLL samples versus 9 $CD38^-$ CLL samples) were resolved by SDS-PAGE. Proteins were transferred onto PVDF membranes followed by antibody detection using a specific antibody against Histone H4 (Figure 6.13B). The analyses showed no significant change in the expression of Histone H4 in $CD38^+$ CLL samples compared to $CD38^-$ CLL samples (p= 0.8,

n= 17, Figure 6.13C), which is not consistent with thr quantitative proteomics data (chapter five).

Following the validation analyses of Histone H4 in 17 CLL samples, normalized Histone H4/Actin data were utilised to explore the expression of Histone H4 in the context of other poor prognostic markers in CLL samples. Example of histone H4 expression in some of these samples is shown in Figure 6.14A. The analyses indicated no significant change in the expression of Histone H4 in CLL samples with other prognostic markers. This included ZAP-70 (p= 0.11, n= 16, Figure 6.14B), mutational status of *IGHV* genes (p= 0.98, n= 10, Figure 6.14C) and Binet stage (p= 0.90, n= 16, Figure 6.14D).

Although there was no association found between Histone H4 expression and the poor prognostic markers, I explored whether patients with an early TTFT exhibited altered expression of Histone H4. The median of Histone H4 expression (Histone H4/Actin = 0.74) was used to segregate CLL patients with high or low expression of Histone H4 in CLL cells. Patients with high Histone H4 had a median follow-up 7.13 years and patients with low Histone H4 had a median follow-up of 4.20 years. The TTFT of patients in each group were compared using Kaplan-Meier analysis. This revealed that

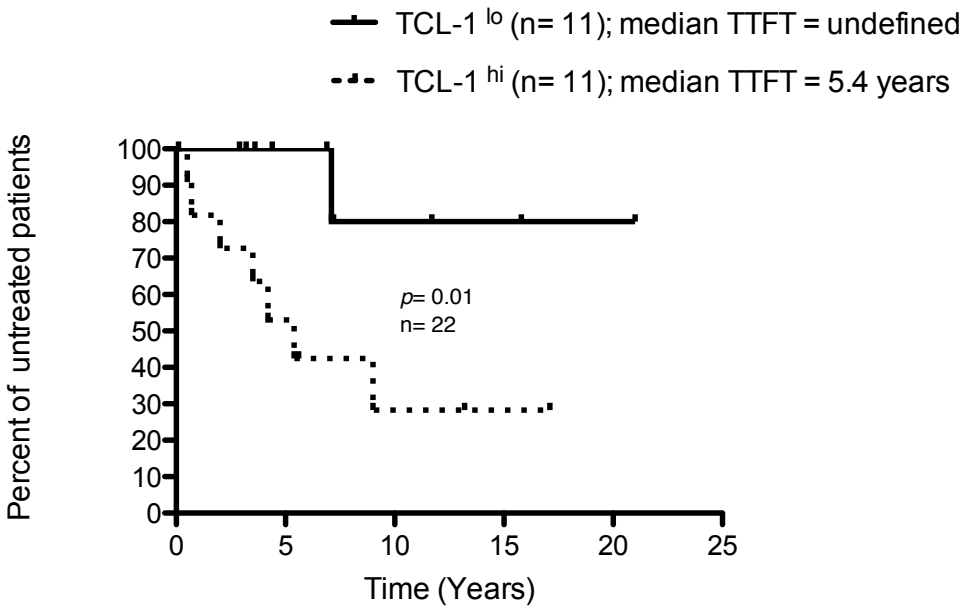


Figure 6.12: Kaplan-Meier curve comparing time period from diagnosis to the first treatment in CLL patients with high or low expression of TCL-1. Patients were divided into two groups using the median of TCL-1 expression; high TCL-1 group (TCL-1/Actin ratio > 0.56) and low TCL-1 group (TCL-1/Actin ratio < 0.56). Median TTFT was 5.4 years in the high TCL-1 group and undefined in the low TCL-1 group.

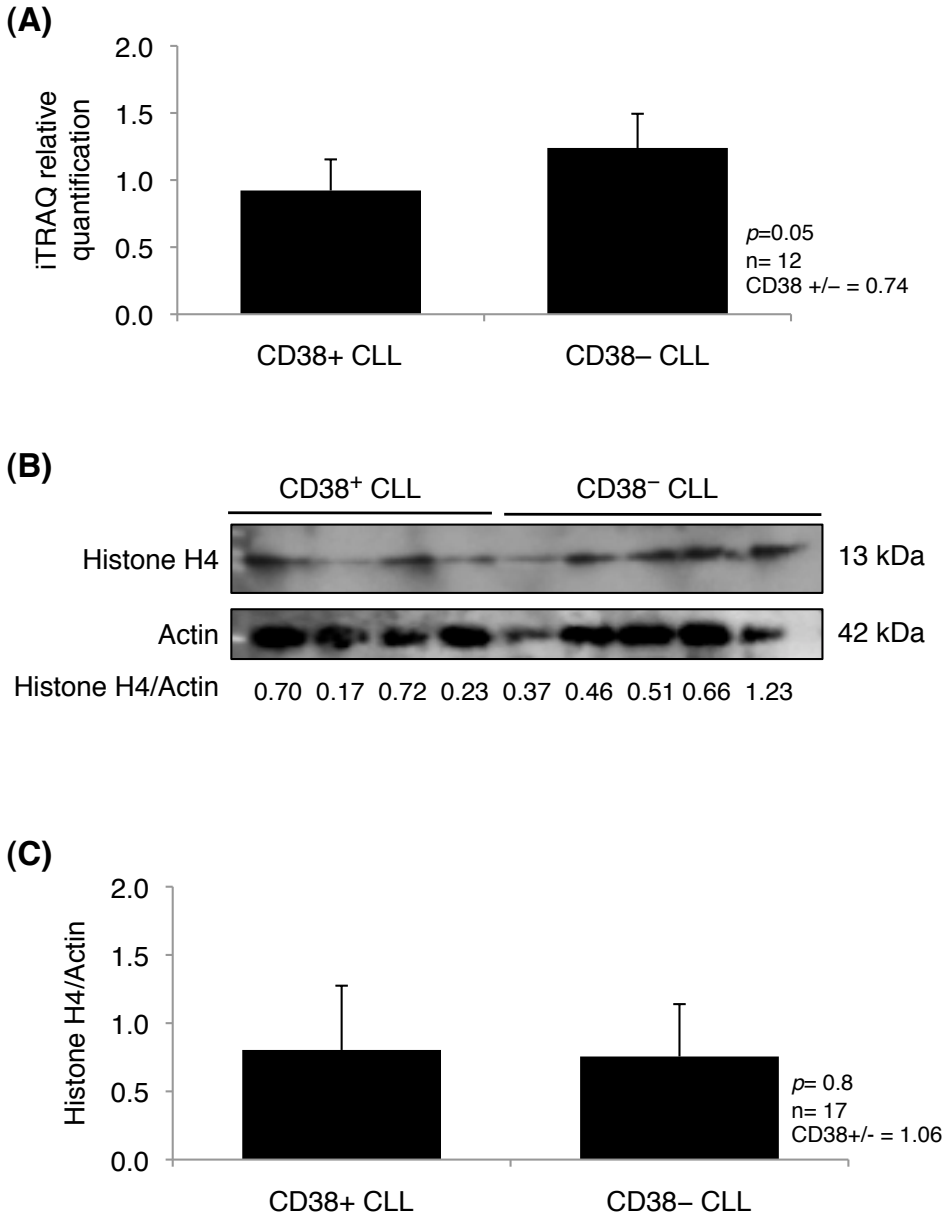


Figure 6.13: Evaluation of the relative expression of Histone H4 in CLL samples with high or low CD38 expression. iTRAQ quantification of proteins in the SDS fractions showed a decreased expression of Histone H4 in CD38^{+} CLL compared with CD38^{-} CLL samples (A). Proteins from the SDS fractions of CLL samples were resolved by SDS-PAGE followed by proteins transfer onto PVDF membrane and antibody detection using a specific antibody to Histone H4 (B). The analysis showed no significant change in Histone H4 expression in CD38^{+} CLL compared to CD38^{-} CLL (C).

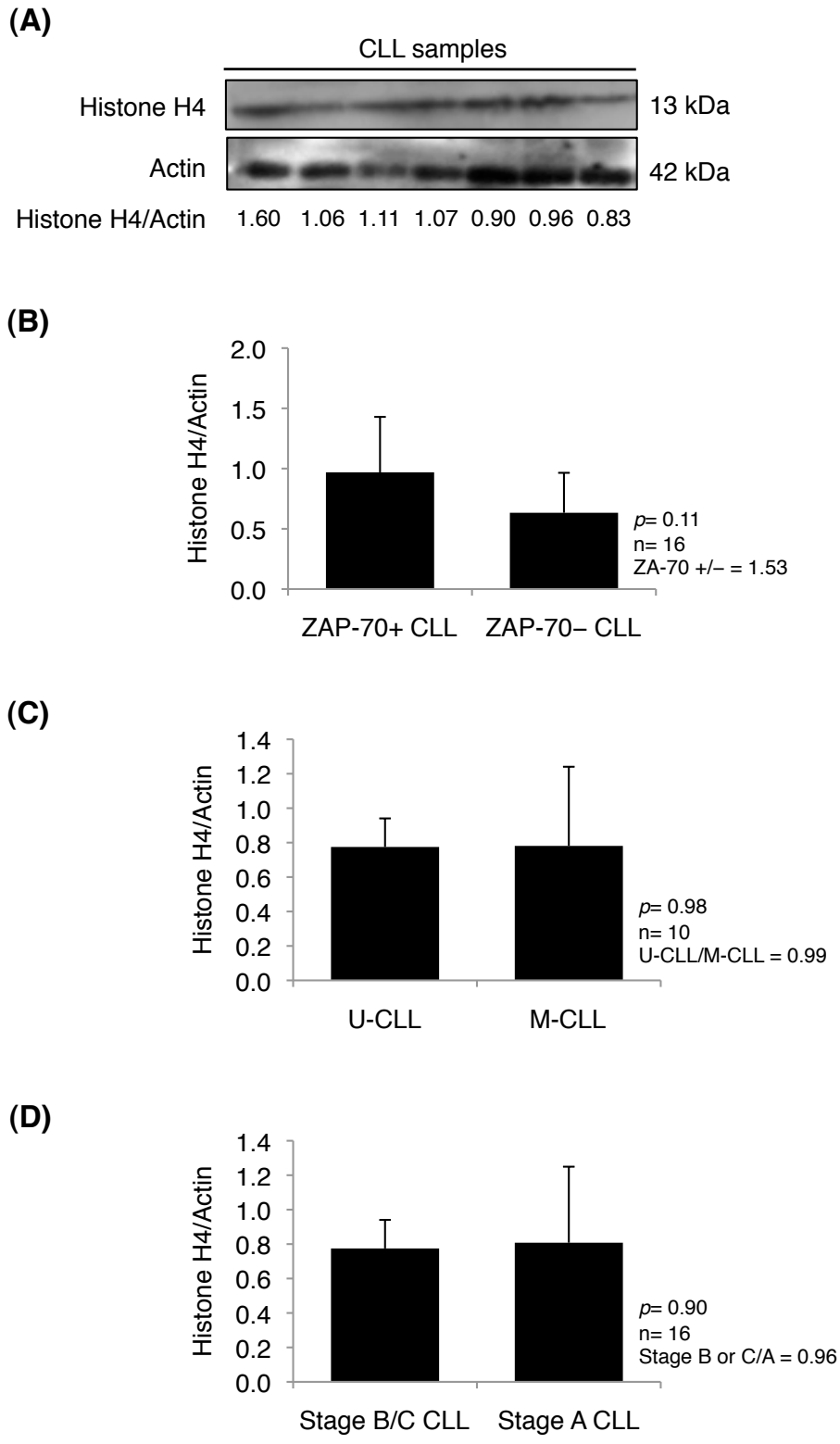


Figure 6.14: Expression of Histone H4 in CLL cells with other prognostic markers. Example of histone H4 expression in CLL samples (A). The analysis of Histone H4 in CLL samples showed that it did not exhibit significantly altered expression in patients with other poor prognostic (B-D).

median TTFT was 7.1 years in the high Histone H4 group and 3.9 years in the low Histone H4 group. However, both groups included patients who did not require treatment 11.7 years following diagnosis. This undoubtedly contributed to the lack of significant difference in TTFT between the two groups (p= 0.31, n= 15, Figure 6.15).

6.2.3.3 Validation of myosin-9 in CLL samples

Quantitative proteomics analysis demonstrated that mysoin-9 was reduced in the SDS fraction of CLL samples from patients in more advanced stages of CLL compared to patients with early stage disease (p= 0.03, n= 10, Figure 6.16A). The altered expression of myosin-9 in CLL samples was validated in the SDS fractions of 16 CLL samples (8 Binet stage B/C samples versus 8 Binet stage A CLL samples). Following SDS-PAGE and western blotting myosin-9 was measured using a specific antibody (Figure 6.16B). The normalised myosin-9/Actin data from 16 patients confirmed that myosin-9 expression was decreased in CLL samples from patients in Binet stage B/C (p= 0.0001, n= 16, Figure 6.16C).

The next step was to investigate if myosin-9 expression also changed in CLL samples on the basis of other prognostic markers. Example of the expression of myosin-9 in CLL samples is shown in Figure 6.17A. The analysis showed no significant change in myosin-9 expression in $CD38^+$ CLL compared to $CD38^-$ CLL (p= 0.47, n= 15, Figure 6.17B). In contrast, myosin-9

was significantly lower-expressed in ZAP-70^{+} CLL (p= 0.05, n= 16, Figure 1.17C) and U-CLL (p= 0.01, n= 12, Figure 6.17D).

The potential association of myosin-9 with poor prognostic markers (ZAP-70^{+}, unmutated *IGHV* genes and Binet stage B/C) prompted an evaluation of the influence of low myosin-9 expression on the need for treatment in CLL patients. The TTFT of patients with high and low of myosin-9, as defined by the median expression of myosin-9, was analysed using the Log-rank test and graphically represented using Kaplan-Meier curves. The median follow up was 7.41 years in the high myosin-9 group and was 4.84 years in the low myosin-9 group. The two groups exhibited different TTFT; median TTFT was 1.64 years in the low myosin-9 group and was undefined in high myosin-9 group but this was not statistically significant (p= 0.1, n= 16, Figure 6.18).

6.2.4 Proteins with the most heterogeneous expression in CLL samples

The quantitative proteomics analysis (chapter five) showed that 14 proteins were found with a heterogeneous expression in CLL. However, the expression of these proteins was not significantly associated with commonly used prognostic markers in CLL such as CD38, ZAP-70, *IGHV* mutational status and Binet stage. The largest dynamic range of expression in CLL samples as defined by standard deviation was observed for S100A9 and

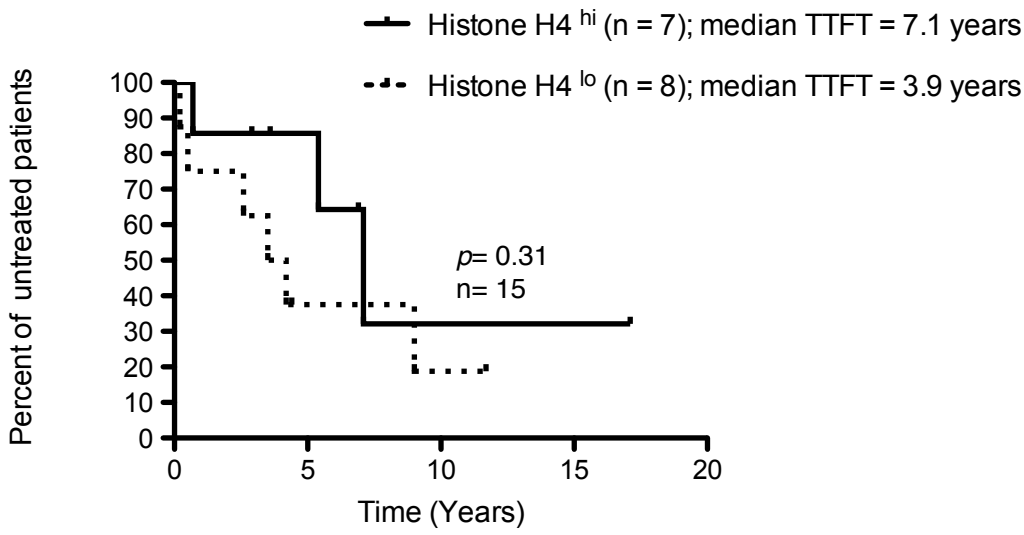


Figure 6.15: Kaplan-Meier curve comparing time period from diagnosis to the first treatment in CLL patients with high or low expression of histone H4. Patients were divided into two groups using the median of Histone H4 expression; high Histone H4 group (Histone H4/Actin ratio > 0.74) and low Histone H4 group (Histone H4/Actin ratio < 0.74). Median TFT was 7.1 years in the high Histone H4 CLL patients group and 3.9 years in low Histone H4 CLL patients group. The analysis demonstrated that the difference in TTFT of the two groups was not significant (p= 0.31).

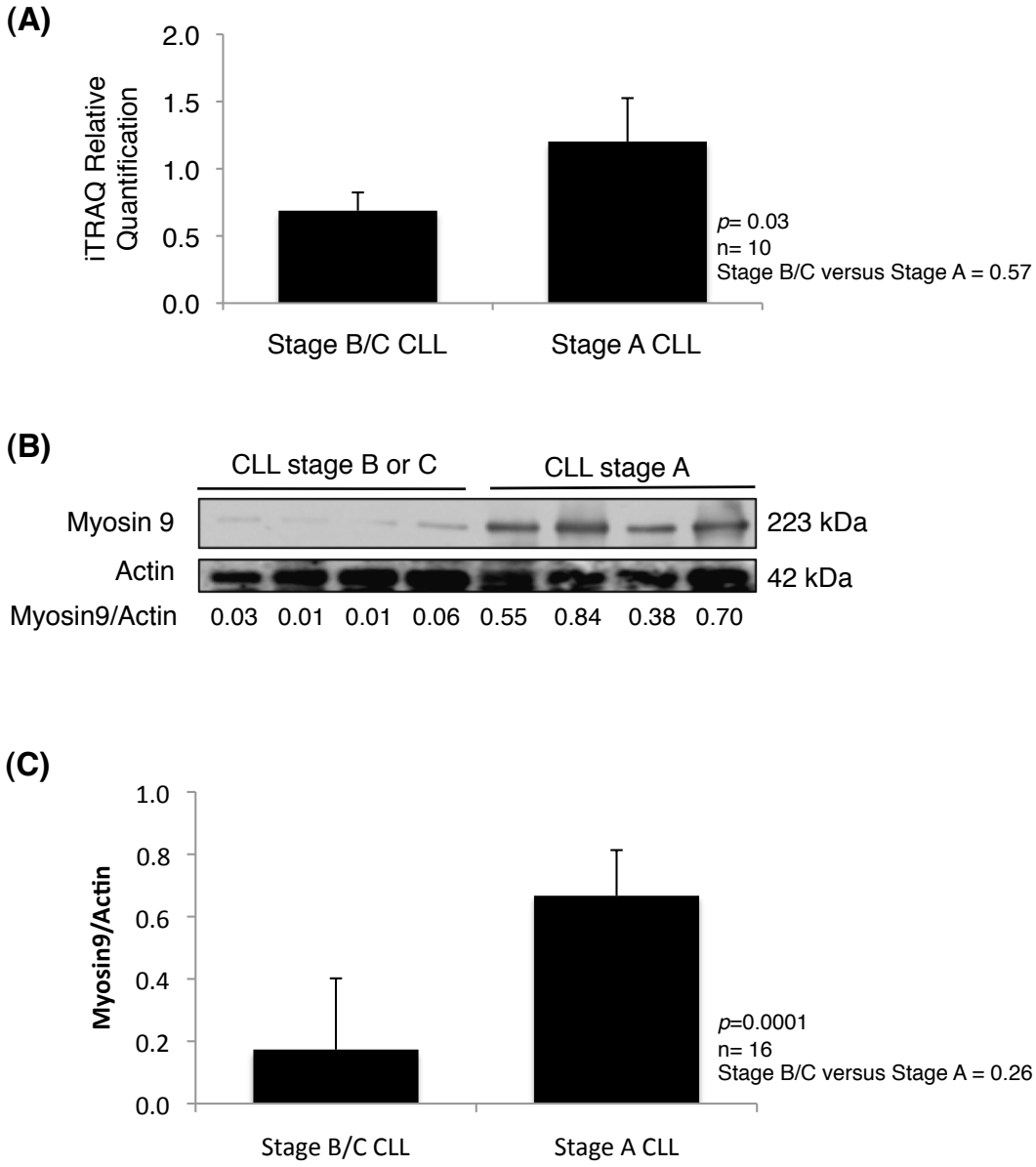


Figure 6.16: Validation of the altered expression of myosin-9 in CLL samples. iTRAQ quantification demonstrated that myosin-9 expression was reduced in the SDS fractions of CLL samples from patients in stage B or C (A). Proteins from the SDS fractions of CLL samples were resolved by SDS-PAGE followed by proteins transfer onto PVDF membrane and antibody detection using a specific antibody to Myosin-9 (B). The analysis demonstrated that myosin-9 was down-regulated in CLL samples from patients in late stages of CLL (C).

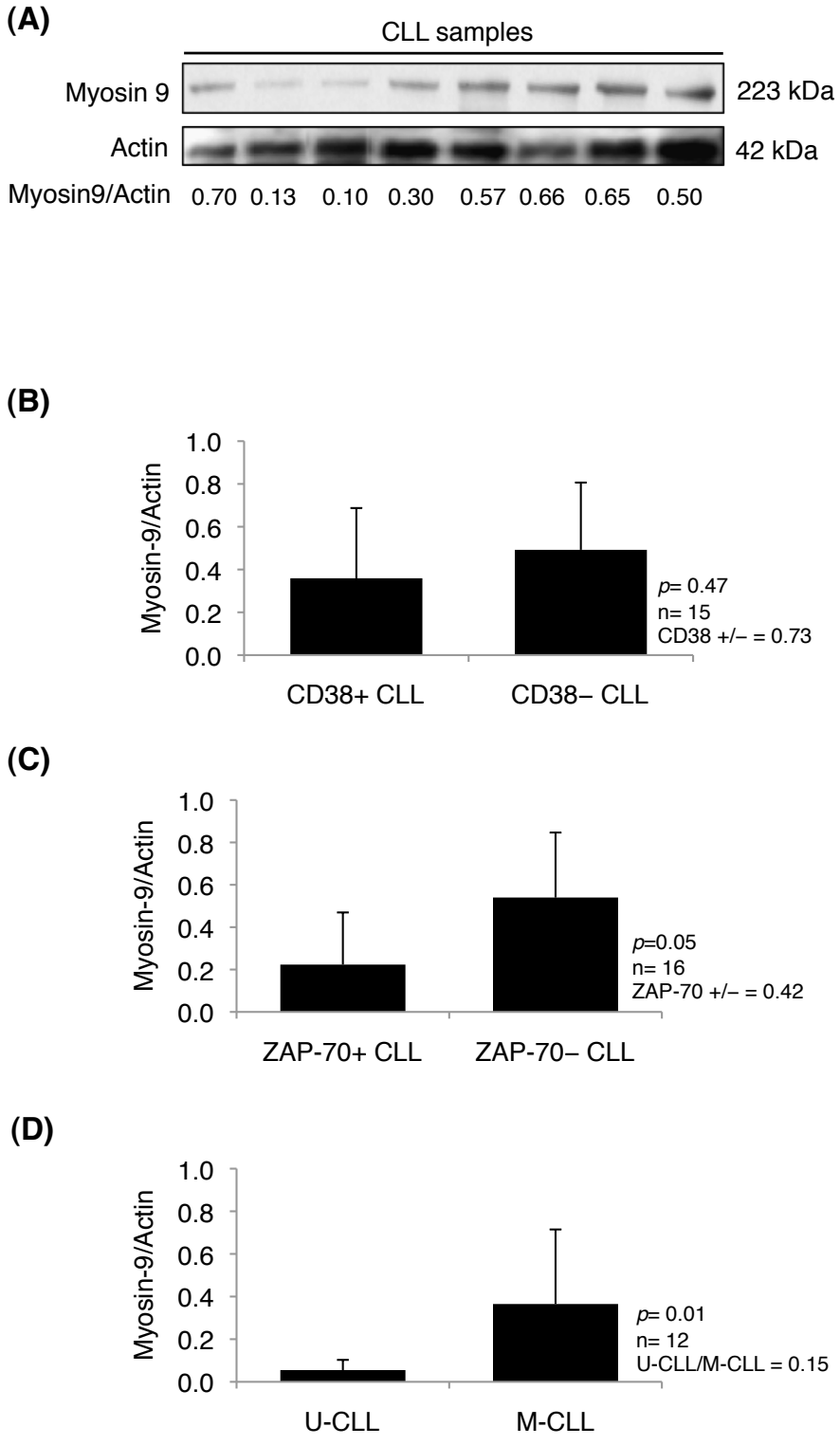


Figure 6.17: Expression of myosin-9 in CLL samples with other prognostic markers. Example of myosin-9 expression in CLL samples (A). The analysis of myosin-9 showed no significant change in its expression in CLL samples with high or low CD38 expression (B). In contrast, myosin-9 was down-regulated in ZAP70^{+} CLL samples (C) as well as in U-CLL samples (D).

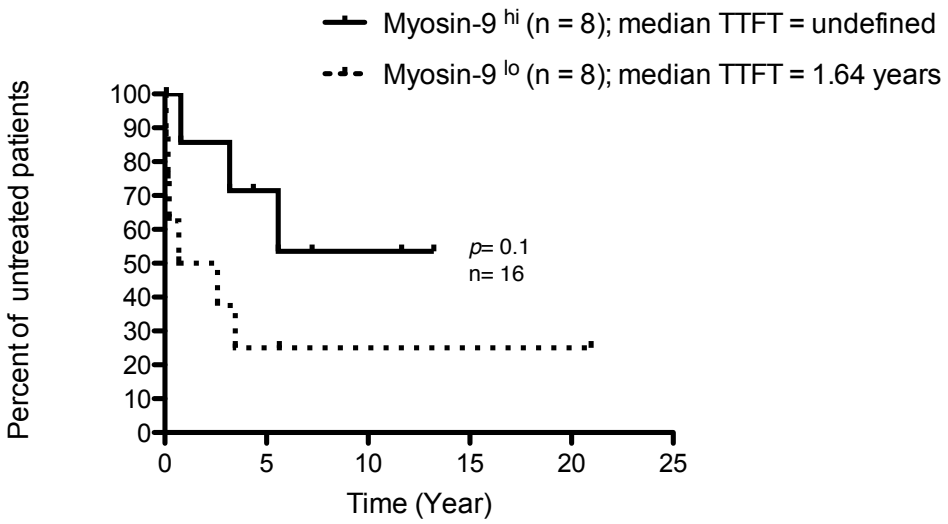


Figure 6.18: Kaplan-Meier curve comparing time period from diagnosis to the first treatment in CLL patients with high or low expression of myosin-9. Patients were divided into two groups using the median of myosin-9 expression; high myosin-9 group (myosin-9/ Actin ratio > 0.57) and low myosin-9 group (myosin-9/Actin ratio < 0.57). Median TTFT was 1.64 years in the low myosin-9 group and undefined in the high myosin-9 group. The analysis demonstrated that the difference between the two groups in terms of TTFT was not significant (p= 0.1).

S100A8. Both proteins were linked to different types of cancer such as breast cancer and prostate cancer (Cormier et al., 2009, Hermani et al., 2005). S100A8 was further linked to poor prognosis of acute myeloid leukaemia (Nicolas et al., 2011). As a result, S100A8 was chosen for further analysis in additional CLL samples.

6.2.4.1 Investigation of S100A8 expression in CLL samples

The iTRAQ relative quantification showed that S100A8 was variably expressed in the NP40 fraction of CLL samples (Figure 6.19A). The expression of S100A8 was assessed in the NP40 fraction of 18 CLL samples using western blotting (Figure 6.19B). The normalised S100A8/Actin data confirmed the heterogeneous expression of S100A8 in CLL samples and demonstrated that its high expression was associated with more rapid progression of CLL (p= 0.03, n= 18, Figure 6.19C).

Given the association of high S100A8 expression with rapid progression of CLL, altered expression of S100A8 in CLL samples exhibiting poor prognostic markers was explored. Example of S100A8 expression in CLL samples is shown in Figure 6.20A. The analysis showed no significant change in S100A8 expression in $CD38^+$ CLL compared to $CD378^-$ CLL (p= 0.87, n= 18, Figure 6.20B) or ZAP-70^+ compared to ZAP-70^- CLL (p= 0.14, n= 17, Figure 6.20C). Of the CLL samples that were analysed for S100A8 expression only one was U-CLL and two were in stage B/C. Therefore, no conclusions

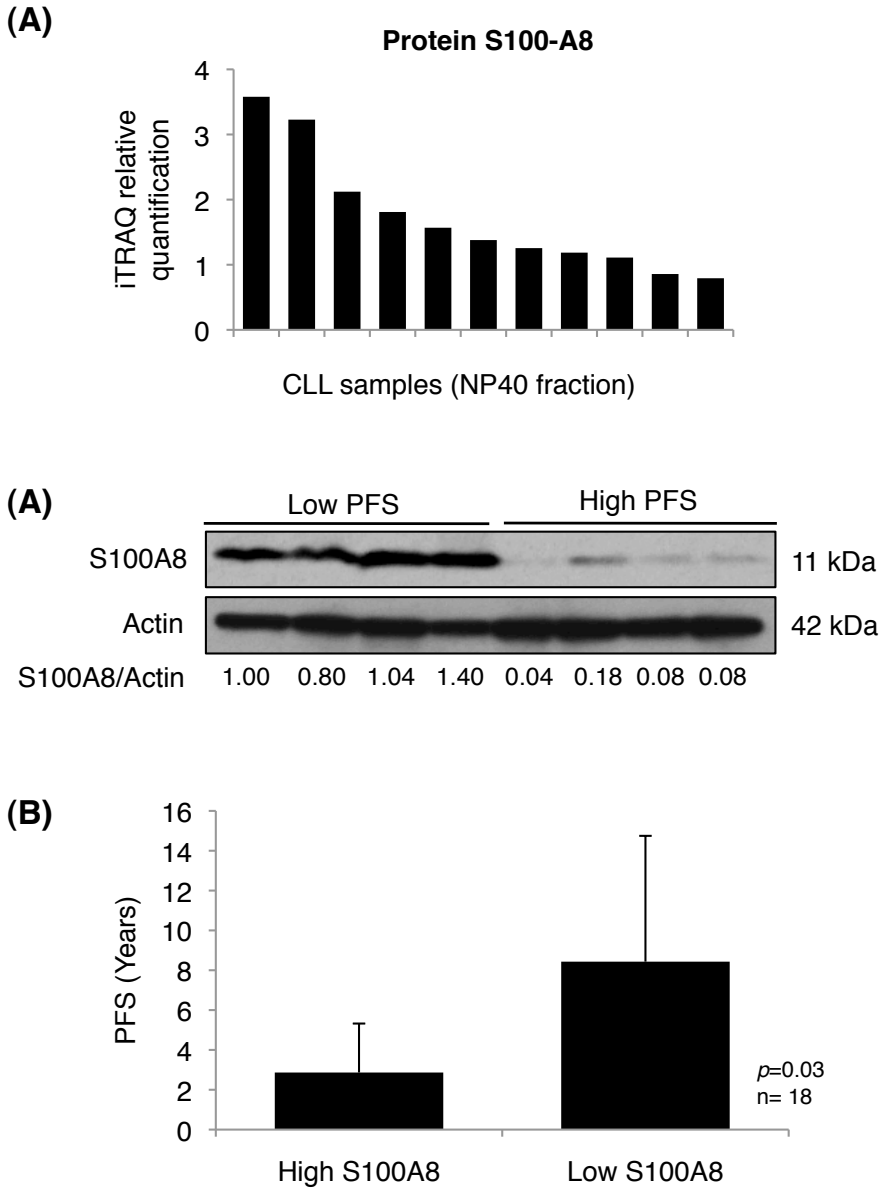


Figure 6.19: Expression of protein S100A8 in CLL samples. iTRAQ quantification showed heterogeneous expression of S100A8 in the NP40 fractions of CLL samples (A) Proteins from the NP40 fractions of CLL samples were resolved by SDS-PAGE followed by proteins transfer onto PVDF membrane and antibody detection using a specific antibody to S100A8 (B). The analysis confirmed the heterogeneous expression of S100A8 in CLL samples and showed that the high expression of S100A8 in CLL cells was associated with more rapid progression of the disease (C).

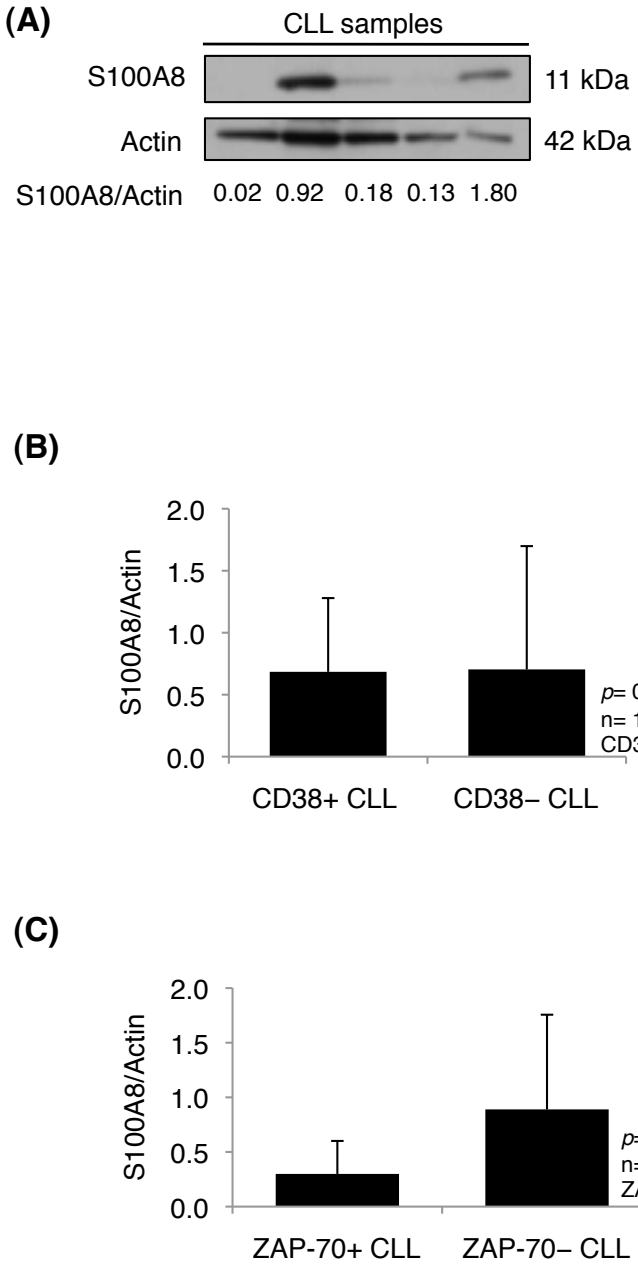


Figure 6.20: Expression of S100A8 in CLL samples with other prognostic markers. Example of S100A8 expression in CLL samples (A). The analysis of S100A8 showed no significant change in its expression in CLL samples with high or low CD38 expression (B) or ZAP-70 expression (C).

can be drawn about the expression of S100A8 on the basis of these two prognostic markers.

Based on the potential link between elevated expression of S100A8 and the progression of CLL, I investigated whether S100A8 expression was associated with TTFT in CLL patients. As defined by median S100A8 expression (S100A8/Actin ratio = 0.39) patients were divided into two groups: a high S100A8 group and a low S100A8 group. The median follow-up was 4.30 years in the high S100A8 group and 10.35 years in the low S100A8 group. TTFT data of the patients in each group were analysed by using Kaplan-Meier curves. The analysis revealed a significantly different TTFT in the two groups of patients; the median TTFT was 3.2 years in the high S100A8 group and not reached in the low S100A8 group (p= 0.01, n= 19, Figure 6.21). In fact, some patients with low S100A8 expression did not required treatment even >15 years following diagnosis. In contrast, all of the high S100A8 group received treatment despite the shorter follow-up in this subset. This analysis indicates potential importance of S100A8 in predicting TTFT in CLL patients but this remains to be confirmed in a much larger CLL cohort.

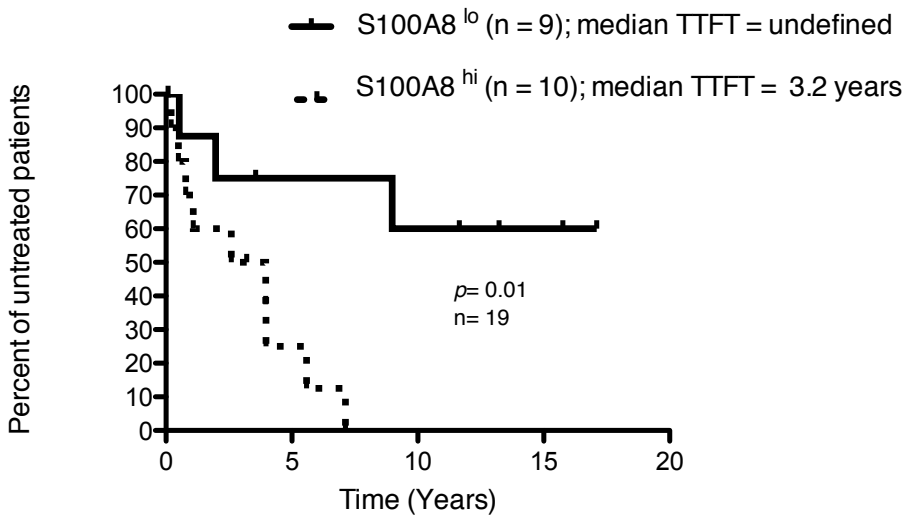


Figure 6.21: Kaplan-Meier curve comparing time period from diagnosis to the first treatment in CLL patients with high or low expression of S100A8. Patients were divided into two groups using the median of S100A8 expression; high S100A8 group (S100A8/Actin ratio > 0.39) and low S100A8 group (S100A8/Actin ratio < 0.39). Median time to first treatment (TTFT) was 3.2 years in the high S100A8 group and undefined in the low S100A8 group.

6.3 Discussion

In this chapter, an independent method from proteomics was used to validate and further invistigate six of the proteins that were identified as having possible involvement in CLL in chapters four and five. Historically, western blotting followed by specific antibody detection was the method often used to validate altered proteins identified by proteomics (Voss et al., 2001, Boyd et al., 2003, Barnidge et a., 2005, Perrot et al., 2011). In this study the number of proteins that were subjected to further investigations was six. This compares favourably with previous CLL proteomics studies in which the highest number of proteins chosen for validation, following proteomics analysis, was two (Barnidge et al., 2005, Perrot et al., 2011).

Of the six proteins that were subjected to further analyses in additionl cohort of CLL samples, four (TR150, TCL-1, myosin-9 and S100A8) showed a significant change in their expression in poor prognosis CLL compared with good prognosis CLL. These proteins (TCL-1, myosin-9 and S100A8) have been previously linked to the pathology of CLL, as will be discussed later in this section. This supports their identification in this project as having a potential role in CLL. Nevertheless, given the heterogeneity of CLL, investigating the expression of these four proteins in a larger cohort of CLL samples may be required in order to draw a more definitive conclusion about their prognostic impact on CLL.

This chapter showed a potential association of TR150 with advanced stage of CLL as well as with an early need for treatment in CLL patients. No extensive work has been done on TR150 in the context of CLL, but it was reported to play a role in transcriptional co-activation and in mRNA splicing (Fondell et al., 1996, Lee et al., 2010). More specifically, TR150 is coordinately recruited to the *Cyclin D1* gene and cyclin D1 mRNA to control its expression (Bracken et al., 2008). In the same study, knock down of TR150 using siRNA was shown to reduce the production of mature cyclin D1 transcripts. Cyclins, including cyclin D1, are well known regulators of the cell cycle (Israels and Israels, 2001).

Elevated expression of cyclin D1 has been reported in different types of cancer including breast cancer and prostate cancer (Kenny et al., 1999, Drobnjak et al., 2000). In the context of leukaemia, over-expression of cyclin D1 is a hallmark of mantle cell lymphoma (MCL) and was also found in a percentage (21%) of CLL cases (Jain et al., 2002). Interestingly, cyclin D1-positive CLL cells were found to localise to the proliferation centres of lymph nodes suggesting that cyclin D1 is an important factor for CLL proliferation (Abboudi et al., 2009). As a result, the high expression of TR150 observed in CLL samples from patients with advanced stage disease and those who required an earlier treatment is consistent with the increased propensity of these cells to undergo proliferation.

The present study demonstrated an association of TCL-1 expression with poor prognosis CLL and an early requirement for treatment. *In vitro* and *in vivo* studies have shown that over-expression of TCL-1 enhanced the activity of AKT kinase, which phosphorylates a wide range of proteins that are involved in variety of cellular processes including apoptosis and proliferation (Laine et al., 2000). More specifically, the same study showed that up-regulation of TCL-1 expression enhanced cell proliferation and protected cells from apoptosis. Interestingly, over-expression of TCL-1 in murine B-cells led to a CLL-like illness seen in a TCL-1 transgenic mouse model, indicating a primary role for TCL-1 in tumourigenesis (Bichi et al., 2002).

In the context of CLL prognosis, high gene expression of *TCL1A* was reported to associate with poor prognosis CLL including U-CLL, *IGHV*3-21 CLL regardless of the mutational status of their *IGHV* genes, $CD38^+$ CLL and chromosomal aberrations of 11q and 17p (Mansouri et al., 2010). The same study also showed that high and low transcripts of TCL-1 segregated between patients with shorter survival (median = 7.0 years) and patients with longer survival (median = 12.2 years). In line with these findings, high protein expression of TCL-1 was also observed in poor prognosis CLL such as U-CLL, $ZAP70^+$ CLL and a chromosomal abnormality of 11q (Herling et al., 2006). TCL-1 was shown to interact with ATM and activate the NF-κB pathway in primary CLL cells (Gaudio et al., 2012). The same study also reported that knock down of TCL-1 inhibited cell proliferation in lymphoma cells.

TCL-1 expression was also linked to the response to treatment in CLL patients. The reduction of peripheral blood $CD5^+/CD19^+$ leukaemic B-cells to less than 1% following treatment was significantly associated with patients who had low TCL-1 and those patients were more likely to achieve a complete response (Browning et al., 2007).

Although cell proliferation was reported to be inducible by over-expression of TCL-1, Herling *et al*., (2006) showed that CLL differentiation and proliferation induced by IL-4 *in vitro* was associated with reduction in TCL-1 expression. A similar expression pattern of TCL-1 is perhaps seen in lymph nodes, where IL4 is secreted from tumour-associated T-cells. This may suggest that TCL-1 high expression is not essential for CLL proliferation and survival in lymph nodes where other pro-survival and proliferation signals exist. Nevertheless, high expression of TCL-1 may have a key role in protecting CLL cells in the peripheral blood from apoptosis, at least through the enhanced phosphorylation (inactivation) of the pro-death protein Bad (Laine et al., 2000).

The current study showed primary evidence that myosin-9 (also known as myosin IIA) was associated with poor prognosis of CLL. Myosin-9 belongs to the class II myosin family, which is involved in variety of cellular processes such as protrusion, migration and locomotion (Sellers, 2000). Myosin-9 plays a role in mediating uropodal detachment from highly adhesive molecules (Morin et al., 2008). Importantly, loss of myosin-9 in T-cells led to a prolonged contact with high-endothelial venules (HEVs) (Jacobelli et al., 2010).

Interestingly these HEVs express large amounts of CD31 (Clark et al., 1998). Therefore, the low expression of myosin-9 in a subgroup of CLL samples may facilitate a long-lasting interaction between CD38 on CLL cells and its ligand CD31 on the endothelial cells. This interaction was demonstrated to promote CLL survival and proliferation and (Deaglio et al., 2005, Deaglio et al., 2010). Moreover, *in vivo* analysis reported that Mysoin-9 deficient T-cells accumulated in lymph nodes for a longer time period compared to control T-cells (Jacobelli et al., 2010). Subsequently, CLL cells expressing low amounts of myosin-9 may be retained in lymph nodes for a longer time period exploiting the pro-survival and proliferation signals that exist in lymph nodes leading to CLL survival and proliferation (Mainou-Fowler et al., 2001, Burger, 2012).

Autoreactivity of BCR to myosin-9 exposed on dead cells was linked to poor survival in CLL patients. This autoreactivity was found in a subgroup of CLL patients (Chu et al., 2008, Chu et al., 2010). BCRs with binding activity to myosin-9 on apoptotic cells were commonly found amongst U-CLL and this autoreactivity was observed to inversely correlate with CLL patient survival (Chu et al., 2010). This indicates that myosin-9 may play a role in providing pro-survival signals to at least some CLL patients.

My study also indicated an association of S100A8 with the progression of CLL and the early need for treatment. S100 proteins are involved in a variety of cellular processes, such as cell cycle regulation, cell

growth, cell differentiation and motility, which are important biological functions in cancer (Heizmann et al., 2002). S100A8 was reported to be up-regulated in different types of cancer including cancer and prostate cancer (Hermani et al., 2005, Cormier et al., 2009). In addition S100A8 was shown to support cell survival; suppression of S100A8 expression in HeLa contamination cells (Hep2 cells) was co-incident with increased apoptosis and down-regulation of *BCL2* gene expression (Huang et al., 2008). In the context of leukaemia S100A8 mRNA was six fold over-expressed in $CD38^+$ sub-clones (Pepper et al., 2007). In addition, high expression of S100A8 is associated with short survival of AML patients (Nicolas et al., 2011).

Importantly, a relationship between the critical transcription factor NF-κB (Hewamana et al., 2008), and S100A8 was found. *S100A8* is an NF-κB target gene (Nemeth et al., 2009) and over-expression of S100A8/S100A9 leads to greater activity of NF-κB (Benedyk et al., 2007). Thus, high expression of S100A8 may be indicative of increased NF-κB activation, which is a poor prognosis marker in CLL (Benedyk et al., 2007). Therefore, the potential impact of S100A8 on the pathology of CLL may explain the association of S100A8 with rapid progression of CLL and early requirement for treatment in CLL patients.

This chapter showed a potential association of TR150, TCL-1 and S100A8 with high risk form CLL and early need for teatment. In addotion it demonstrated reduced expression of myosin-9 in the aggressive form of the

disease. These proteins were studied in cohort of CLL samples using western blotting. While this technique is often used to validate different expression of proteins following proteomics analysis (Voss et al., 2001, Boyd et al., 2003, Barnidge et a., 2005, Perrot et al., 2011), other methods, such as ELISA, would serve as more robust and reliable means of quantifying differences in protein expression across patient samples.

Chapter Seven

Summary and General Discussion

7.1 Summary of the key achievements of this thesis

- A two-step cellular fractionation method that dissolves all the CLL cell components and reduces sample complexity was developed.
- A qualitative proteomics workflow (2D nano-LC-MS/MS) was performed to study the proteome of CLL
    - This workflow identified the most conservitive and largest number of proteins (n= 900) to date in CLL cells by combining the identifications from the NP40 and SDS fractions. These protein identifications were achieved with 0% FDR for the proteins identified with ≥2 peptides (625 proteins) and a 3.2% FDR for the proteins identified with single peptides (275 proteins).

    - A method of combining transcriptomic data of CLL cells and normal B-cells (Huttmann et al., 2006, Hutcheson et al., 2008) with the proteomics data of CLL cells (generated in this project) was described and was used to highlight proteins (n= 20) with possible involvement in CLL. To the best of my knowledge this method has not been described before for CLL cells or for other cell systems.

- A quantitative proteomics workflow (iTRAQ 2D nano-LC-MS/MS) was performed to explore protein expression in poor prognosis versus good prognosis CLL samples.

- This workflow provided relative quantification of 655 proteins in primary CLL samples and identified preferentially altered proteins (n= 15) and those with the most heterogeneous expression that were not associated with existing prognostic markers (n= 14).

- Of the proteins found with possible relevance to CLL (n= 49), six were further studied in additional cohorts of CLL samples using specific antibodies. The analyses indicated that TCL-1, myosin-9 and S100A8 (section 6.3) are important in the pathology of CLL and identified TR150 as a novel protein with potential to be important in the prognosis and pathology of this disease.

7.2 General discussion

As proteins are the main functional molecules in a biological system (Twyla, 2004), great effort has been made to explore differences between poor prognosis CLL and good prognosis CLL on the basis of protein expression. This was performed using conventional methods such antibody detection of a single protein (Marschitz et al., 2000, Pepper et al., 2008, Damle et al., 2007) and proteomics approaches, which allow global identification and quantification of proteins simultaneously (Voss et al., 2001, Cochran et al., 2003, Barnidge et al., 2005a, Perrot et al., 2011). This project employed a cellular fractionation method with qualitative and quantitative proteomics workflows as well as western blotting/antibody detection to extend this type of analysis to identify more proteins with a potential role in the prognosis or/and the pathology of CLL.

In this project, three different strategies were applied to seek proteins with possible involvement in CLL. Firstly, the identified proteome of primary CLL samples was integrated with published transcriptomes of CLL cells and normal B-cells to identify proteins with preferential expression in CLL cells but not in normal B-cells. Secondly, using quantitative proteomics data, the proteome of poor prognosis CLL was compared with the proteome of good prognosis CLL to identify proteins with potential association with either form of the disease. Thirdly, the standard deviation of protein expression in five or more CLL samples (regardless of their known prognostic markers) was used to indicate proteins with the most variable expression in CLL, which had the

possibility of driving the heterogeneous outcome of CLL. Together these methods identified 49 proteins that fulfilled one of these criteria.

Among the 49 proteins, some, such as TCL-1, were already known to be important in CLL and high expression of TCL-1 mRNA and protein has been reported to be associated with poor prognosis (Bichi et al., 2002, Mansouri et al., 2010, Herling et al., 2006). Therefore, the identification of TCL-1 with an altered expression in poor prognosis CLL compared to good prognosis CLL provided some independent confirmation of the proteomics workflow that was used.

Progression of CLL is thought to be largely driven by the pro-survival and pro-proliferation signals that exist in the cellular microenvironment, such as lymph nodes (section 1.1.4.2). Signal transduction following BCR stimulation, CD40L engagement with its receptor CD40, or CD38/CD31 interaction has been reported to promote survival and proliferation of CLL cells (Guarini et al., 2008, Quiroga et al., 2009, Granziero et al., 2001, Deaglio et al., 2005, Deaglio et al., 2010). This may suggest that active signal transduction is important for the progression of CLL. In agreement with this, the Gene Ontology data of the 20 proteins whose gene expression is preferentially expressed in CLL cells but not normal B-cells showed that four of them (AB1IP, GNA13, STMN1 and CD5) were involved in signal transduction. In particular, STMN1 was reported to mediate the transduction of differentiation and proliferation signals upon the activation of

different protein kinases including protein kinase C (PKC) and mitogen activated protein kinase (MAPK) (Di Paolo et al., 1996, Drouva et al., 1998, Sherbet and Cajone, 2005). Consistently, high gene expression of *STMN1* was demonstrated to associate with highly proliferating breast and ovarian cancer cells (Price et al., 2000, Curmi et al., 2000). This supports the concept of aberrant signal transduction in CLL cells compared to normal B-cells.

One of the pathways reported to influence CLL cells is activation of the transcription factor, NF-κB, which is involved in cellular survival, differentiation, and proliferation (section: 1.1.4.3.4). Previously, CLL cells have been shown to exhibit high activity of NF-κB compared to normal B-cells, but importantly NF-κB activity was variable among CLL patients, possibly reflecting the heterogeneous clinical course of CLL (Hewamana et al., 2008). Interestingly, patients with high activity of NF-κB, as defined by Rel A DNA binding, were those with short TTFT and short survival time (Hewamana et al., 2009). In line with these studies, 3/14 proteins (S100A8, S100A9 and Galectin-1), which were found to be the most variablly expressed proteins in CLL samples, positively regulate the activity of NF-κB as implied by their Gene Ontology data. For example, over-expression of S100A8 and S100A9 was shown to increase the activation of NF-κB in a human keratinocyte cell line (Benedyk et al., 2007). This suggests that the variable activity of NF-κB in CLL patients is perhaps partly driven by the heterogeneous expression of these three proteins. Furthermore, the association of S100A8 with short TTFT in CLL patients as shown by western blotting/antibody detection supports the

previous finding that high activity of NF-κB was associated with short TTFT and short survival of CLL patients (Hewamana et al., 2009).

Deregulation of proteins involved in apoptosis was reported in CLL (sections: 1.1.4.3.5 and 1.1.4.3.6). In fact, CLL was initially considered an accumulative disease that results from defective programmed cell death in the malignant B-cells (Dighiero, 2003, Lanasa, 2010). High expression of the anti-apoptotic protein BCL2 is a hallmark of CLL (Hanada et al., 1993, Del Gaizo Moore et al., 2007, McCarthy et al., 2008). In addition, increased expression of another anti-apoptotic protein MCL1 was shown to be associated with early need for treatment as well as resistance to treatment of CLL patients (Kitada et al., 1998, Pepper et al., 2008). Consistently, of the 49 proteins that were identified with potential importance in CLL, six (IKZF3, THIM, PRDX5, PRDX2, FUS and ACINU) were involved in apoptosis, as indicated by their Gene Ontology data. For example, over-expression of THIM, which was absent from the transcriptomes of normal B-cells but present in the transcriptomes and proteomes of CLL cells, was shown to protect from apoptosis in human hepatocellular carcinoma cell line and human osteosarcoma cell line (Cao et al., 2008). In addition, depletion of PRDX5, which was up-regulated in ZAP-70^{+} CLL compared to ZAP-70^{-} CLL, was reported to make a human neuroblastoma cell line more prone to apoptosis and oxidative damage (De Simoni et al., 2008). Similarly, over-expression of PRDX5 in primary human tendon cells was shown to protect these cells from apoptosis (Yuan et al., 2004). In addition, high expression of PRDX2, which was observed with

variable expression in CLL samples, was reported to protect from apoptosis in murine pancreatic beta cell line (Zhao and Wang, 2012). This may re-emphasize the relevance of the aberrant apoptosis in the pathology of CLL.

Although the clinical course of poor prognosis CLL and good prognosis CLL is remarkably different, these two forms share common gene expression profiles. Two independent studies compared the gene expression profile of U-CLL with that of M-CLL and found that only a small proportion of genes exhibited altered expression in these subtypes of CLL samples (Klein et al., 2001); Rosenwald et al., 2001). Of approximately 12,000 transcripts that were studied by Klein et al. (2001), 23 mRNA were found with altered expression in U-CLL versus M-CLL. In addition, by using more comprehensive microarray chips containing 17,856 human cDNAs, Rosenwald et al. (2001) showed that approximately 175 mRNA were differentially expressed in U-CLL compared to M-CLL. In the context of protein expression profiling, three independent CLL proteomics studies compared the proteome of U-CLL with that of M-CLL using different proteomics approaches and identified a small number of differentially expressed proteins in the two subtypes of CLL: 4 proteins (Cochran et al., 2003), 13 proteins (Barnidge et al., 2005) and 5 proteins (Perrot et al., 2011). In the present study, multiple comparisons on the basis of different prognostic markers (CD38, ZAP-70, mutational status of *IGHV* and Binet staging system) were conducted on the proteome of poor prognosis CLL versus the proteome of good prognosis CLL. Of the 655 proteins that were relatively quantified in CLL samples, 15 proteins were

identified with an altered expression in the two forms of CLL. The small number of mRNA and proteins with altered expression in the poor prognosis CLL compared to good prognosis CLL may indicate that CLL cells, irrespective of their prognostic markers, follow the same mechanism of malignant transformation or they originate from a common precursor (Rosenwald et al., 2001, Seifert et al., 2012).

Reflecting on some of the limitations of this study, three different issues have arisen. Firstly, only viable CLL cells were utilized for proteomics analysis. This avoided attributing changes in protein expression to variation in the viability of CLL samples. However, patient samples vary in their viability, as part of CLL heterogeneity, and thus it is possible that I selected a more homogeneous population that did not reflect the full range of CLL pathology. The inclusion of samples with more heterogeneous levels of viability might have identified more altered proteins in poor prognosis CLL versus good prognosis CLL. However, this may have biased my analysis towards pro-apoptotic protein discovery.

A second limitation of the workflow is the larger number of LC-MALDI runs that must be completed because of the two-step cellular fractionation. As discussed previously, reducing sample complexity is a valuable step in proteomics analyses (Ahmed, 2009, Huber et al., 2003, Dredger, 2003). While this fractionation method had the potential to extract all

cellular proteins and reduce the sample complexity, it doubled the number of samples required for proteomics analyses.

Finally, quantitative proteomics analysis (chapter five) showed that Histone H4 was down-regulated in the SDS fractions of $CD38^+$ CLL compared to $CD38^-$ CLL (iTRAQ signal: $CD38^+$ CLL versus $CD38^-$ CLL= 0.75, p= 0.05, n= 12, Figure 5.7F). However, the validation of Histone H4 expression in additional $CD38^+$ CLL versus $CD38^-$ CLL samples demonstrated no significant change in the expression of Histone H4 (p= 0.8, n= 17, Figure 6.13C). In an attempt to understand why this happened, the proteomics quantitative data of Histone H4 was re-analysed. In particular, I looked at whether post-translational modifications may have led to a cleavage in Histone H4 in $CD38^+$ CLL but not in $CD38^-$ CLL samples. This might cause the number of distinct peptides that were used for Histone H4 quantification to be less in $CD38^+$ CLL compared to $CD38^-$ CLL samples. Consequently, low iTRAQ signals in $CD38^+$ CLL compared to $CD38^-$ CLL samples would be reported despite the similar expression of histone H4 in the two types of CLL samples. The analysis showed that the same peptides (that belong to Histone H4) were, in fact, identified and quantified in both sets of CLL samples. Given the heterogeneity of CLL, the different findings of the expression of Histone H4 reported by quantitative proteomics and western blotting in a separate validation cohort might simply reflect the heterogeneity between CLL samples.

In summary, this project utilised a cellular fractionation method combined with gel-free proteomics approaches to characterize the proteome

of poor prognosis and good prognosis CLL. In total 900 proteins were identified, of which 49 had the potential to be important in CLL (as defined by the criteria described in this thesis). Of these proteins, six were studied in additional patient samples and 4/6 (TR150, TCL-1, myosin-9 and S100A8) showed a potential link to the prognosis and the pathology of CLL. Given the variability in CLL, these proteins now need to be evaluated in a larger cohort in order to confirm the importance of these proteins in the prognosis and the pathology of CLL.

7.3 Future work

Following the development of the cellular fractionation method and the proteomics workflows, it is now possible to use these methods to explore how the proteome of CLL probably changes following the exposure to various pro-survival and proliferation signals including:

- BCR stimulation
- CD38/CD31 interaction
- Accessory cells such as T-cells, nurse-like cells and bone marrow stroma in order to mimic the impact of the microenvironment

This could improve our understanding about CLL and probably help to identify therapeutic targets for CLL treatment

Based on the potential importance of the 49 proteins in CLL, reported in this study, it would be interesting to explore these proteins in a larger CLL sample cohort in order to investigate more fully their potential impact on CLL prognosis and pathology, in particular:

- The link between TR150 expression and cellular proliferation through the enhanced expression of Cyclin D1 transcripts (Bracken et al., 2008)
- Given the positive influence of S100A8 and S100A9 on the activation of NF-κB (Benedyk et al., 2007), it would be interesting to explore the potential association of these two proteins with activation of the NF-κB pathway in CLL. This may allow the use of S100A8 as a surrogate marker of NF-κB activity in CLL. In addition,

S100A8 and S100A9 may serve as therapeutic targets to inhibit the NF-κB pathway for CLL treatment.

- Based on the preliminary evidence shown in this study of the potential importance of TR150 and S100A8 in predicting TTFT in CLL patients, assessing their expression in a larger cohort of CLL samples may be a valid option to confirm their value in predicting TTFT in CLL patients.

References

ABBOTT, B. L. 2006. Chronic lymphocytic leukemia: recent advances in diagnosis and treatment. *Oncologist,* 11**,** 21-30.

ABBOUDI, Z., PATEL, K. & NARESH, K. N. 2009. Cyclin D1 expression in typical chronic lymphocytic leukaemia. *Eur J Haematol,* 83**,** 203-7.

ADAMS, J. M. & CORY, S. 1998. The Bcl-2 protein family: arbiters of cell survival. *Science,* 281**,** 1322-6.

AHMED, F. E. 2009. Sample preparation and fractionation for proteome analysis and cancer biomarker discovery by mass spectrometry. *J Sep Sci,* 32**,** 771-98.

ANDERSON, L. & SEILHAMER, J. 1997. A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis,* 18**,** 533-7.

ANDERSON, N. L. & ANDERSON, N. G. 2002. The Human Plasma Proteome. *Molecular & Cellular Proteomics,* 1**,** 845-867.

BANTSCHEFF, M., SCHIRLE, M., SWEETMAN, G., RICK, J. & KUSTER, B. 2007. Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem,* 389**,** 1017-31.

BARNIDGE, D. R., JELINEK, D. F., MUDDIMAN, D. C. & KAY, N. E. 2005a. Quantitative protein expression analysis of CLL B cells from mutated and unmutated IgV(H) subgroups using acid-cleavable isotope-coded affinity tag reagents. *J Proteome Res,* 4**,** 1310-7.

BARNIDGE, D. R., TSCHUMPER, R. C., JELINEK, D. F., MUDDIMAN, D. C. & KAY, N. E. 2005b. Protein expression profiling of CLL B cells using replicate off-line strong cation exchange chromatography and LC-MS/MS. *J Chromatogr B Analyt Technol Biomed Life Sci,* 819**,** 33-9.

BARTAL, A., BENTWICH, Z., MANNY, N. & IZAK, G. 1978. Ethnical and clinical aspects of chronic lymphocytic leukemia in Israel: a survey on 288 patients. *Acta Haematol,* 60**,** 161-71.

BELL, A. W., DEUTSCH, E. W., AU, C. E., KEARNEY, R. E., BEAVIS, R., SECHI, S., NILSSON, T., BERGERON, J. J. & GROUP, H. T. S. W. 2009. A HUPO test sample study reveals common problems in mass spectrometry-based proteomics. *Nat Methods,* 6**,** 423-30.

BENEDYK, M., SOPALLA, C., NACKEN, W., BODE, G., MELKONYAN, H., BANFI, B. & KERKHOFF, C. 2007. HaCaT keratinocytes overexpressing the S100 proteins S100A8 and S100A9 show increased NADPH oxidase and NF-kappaB activities. *J Invest Dermatol,* 127**,** 2001-11.

BICHI, R., SHINTON, S. A., MARTIN, E. S., KOVAL, A., CALIN, G. A., CESARI, R., RUSSO, G., HARDY, R. R. & CROCE, C. M. 2002. Human chronic lymphocytic leukemia modeled in mouse by targeted TCL1 expression. *Proc Natl Acad Sci U S A,* 99**,** 6955-60.

BINET, J. L., AUQUIER, A., DIGHIERO, G., CHASTANG, C., PIGUET, H., GOASGUEN, J., VAUGIER, G., POTRON, G., COLONA, P., OBERLING, F., THOMAS, M., TCHERNIA, G., JACQUILLAT, C., BOIVIN, P., LESTY, C., DUAULT, M. T., MONCONDUIT, M., BELABBES, S. & GREMY, F. 1981. A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis. *Cancer,* 48**,** 198-206.

BINSKY, I., LANTNER, F., GRABOVSKY, V., HARPAZ, N., SHVIDEL, L., BERREBI, A., GOLDENBERG, D. M., LENG, L., BUCALA, R., ALON, R., HARAN, M. & SHACHAR, I. 2010. TAp63 regulates VLA-4 expression and chronic lymphocytic leukemia cell migration to the bone marrow in a CD74-dependent manner. *J Immunol,* 184**,** 4761-9.

BODNAR, W. M., BLACKBURN, R. K., KRISE, J. M. & MOSELEY, M. A. 2003. Exploiting the complementary nature of LC/MALDI/MS/MS and LC/ESI/MS/MS for increased proteome coverage. *J Am Soc Mass Spectrom,* 14**,** 971-9.

BOICE, J. D., JR., MORIN, M. M., GLASS, A. G., FRIEDMAN, G. D., STOVALL, M., HOOVER, R. N. & FRAUMENI, J. F., JR. 1991. Diagnostic x-ray procedures and risk of leukemia, lymphoma, and multiple myeloma. *JAMA,* 265**,** 1290-4.

BOYD, R. S., ADAM, P. J., PATEL, S., LOADER, J. A., BERRY, J., REDPATH, N. T., POYSER, H. R., FLETCHER, G. C., BURGESS, N. A., STAMPS, A. C., HUDSON, L., SMITH, P., GRIFFITHS, M., WILLIS, T. G., KARRAN, E. L., OSCIER, D. G., CATOVSKY, D., TERRETT, J. A. & DYER, M. J. 2003. Proteomic analysis of the cell-surface membrane in chronic lymphocytic leukemia: identification of two novel proteins, BCNP1 and MIG2B. *Leukemia,* 17**,** 1605-12.

BRACKEN, C. P., WALL, S. J., BARRE, B., PANOV, K. I., AJUH, P. M. & PERKINS, N. D. 2008. Regulation of cyclin D1 RNA stability by SNIP1. *Cancer Res,* 68**,** 7621-8.

BRENNAN, P., SHORE, A. M., CLEMENT, M., HEWAMANA, S., JONES, C. M., GILES, P., FEGAN, C., PEPPER, C. & BREWIS, I. A. 2009. Quantitative nuclear proteomics reveals new phenotypes altered in lymphoblastoid cells. *PROTEOMICS – Clinical Applications,* 3**,** 359-369.

BREWIS, I. A. & BRENNAN, P. 2010. Proteomics technologies for the global identification and quantification of proteins. *Adv Protein Chem Struct Biol,* 80**,** 1-44.

BROWNING, R. L., GEYER, S. M., JOHNSON, A. J., JELINEK, D. F., TSCHUMPER, R. C., CALL, T. G., SHANAFELT, T. D., ZENT, C. S., BONE, N. D., DEWALD, G. W., LIN, T. S., HEEREMA, N. A., GREVER, M. R., KAY, N. E., BYRD, J. C. & LUCAS, D. M. 2007. Expression of TCL-1 as a potential prognostic factor for treatment outcome in B-cell chronic lymphocytic leukemia. *Leukemia research,* 31**,** 1737-1740.

BURGER, J. A. 2012. Targeting the microenvironment in chronic lymphocytic leukemia is changing the therapeutic landscape. *Curr Opin Oncol*.

BURGER, J. A., TSUKADA, N., BURGER, M., ZVAIFLER, N. J., DELL'AQUILA, M. & KIPPS, T. J. 2000. Blood-derived nurse-like cells protect chronic lymphocytic leukemia B cells from spontaneous apoptosis through stromal cell,Äìderived factor-1. *Blood,* 96**,** 2655-2663.

BURGER, M., HARTMANN, T., KROME, M., RAWLUK, J., TAMAMURA, H., FUJII, N., KIPPS, T. J. & BURGER, J. A. 2005. Small peptide inhibitors of the CXCR4 chemokine receptor (CD184) antagonize the activation, migration, and antiapoptotic responses of CXCL12 in chronic lymphocytic leukemia B cells. *Blood,* 106**,** 1824-30.

BYRD, J. C., STILGENBAUER, S. & FLINN, I. W. 2004. Chronic lymphocytic leukemia. *Hematology Am Soc Hematol Educ Program***,** 163-83.

CALIGARIS-CAPPIO, F. 2011. Inflammation, the microenvironment and chronic lymphocytic leukemia. *Haematologica,* 96**,** 353-5.

CALIGARIS-CAPPIO, F. & HAMBLIN, T. J. 1999. B-cell chronic lymphocytic leukemia: a bird of a different feather. *J Clin Oncol,* 17**,** 399-408.

CALIN, G. A., DUMITRU, C. D., SHIMIZU, M., BICHI, R., ZUPO, S., NOCH, E., ALDLER, H., RATTAN, S., KEATING, M., RAI, K., RASSENTI, L., KIPPS, T., NEGRINI, M., BULLRICH, F. & CROCE, C. M. 2002. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A,* 99**,** 15524-9.

CAMPAS, C., COSIALLS, A. M., BARRAGAN, M., IGLESIAS-SERRET, D., SANTIDRIAN, A. F., COLL-MULET, L., DE FRIAS, M., DOMINGO, A., PONS, G. & GIL, J. 2006. Bcl-2 inhibitors induce apoptosis in chronic lymphocytic leukemia cells. *Exp Hematol,* 34**,** 1663-9.

CANNON-ALBRIGHT, L. A., THOMAS, A., GOLDGAR, D. E., GHOLAMI, K., ROWE, K., JACOBSEN, M., MCWHORTER, W. P. & SKOLNICK, M. H. 1994. Familiality of cancer in Utah. *Cancer Res,* 54**,** 2378-85.

CAO, W., LIU, N., TANG, S., BAO, L., SHEN, L., YUAN, H., ZHAO, X. & LU, H. 2008. Acetyl-Coenzyme A acyltransferase 2 attenuates the apoptotic effects of BNIP3 in two human cell lines. *Biochim Biophys Acta,* 1780**,** 873-80.

CASTRO, J. E., PRADA, C. E., LORIA, O., KAMAL, A., CHEN, L., BURROWS, F. J. & KIPPS, T. J. 2005. ZAP-70 is a novel conditional heat shock protein 90 (Hsp90) client: inhibition of Hsp90 leads to ZAP-70 degradation, apoptosis, and impaired signaling in chronic lymphocytic leukemia. *Blood,* 106**,** 2506-12.

CATOVSKY, D., RICHARDS, S., MATUTES, E., OSCIER, D., DYER, M. J., BEZARES, R. F., PETTITT, A. R., HAMBLIN, T., MILLIGAN, D. W., CHILD, J. A., HAMILTON, M. S., DEARDEN, C. E., SMITH, A. G., BOSANQUET, A. G., DAVIS, Z., BRITO-BABAPULLE, V., ELSE, M., WADE, R., HILLMEN, P., GROUP, U. K. N. C. R. I. H. O. C. S. & GROUP, N. C. L. L. W. 2007. Assessment of fludarabine plus cyclophosphamide for patients with chronic lymphocytic leukaemia (the LRF CLL4 Trial): a randomised controlled trial. *Lancet,* 370**,** 230-9.

CHELIUS, D. & BONDARENKO, P. V. 2002. Quantitative Profiling of Proteins in Complex Mixtures Using Liquid Chromatography and Mass Spectrometry. *Journal of Proteome Research,* 1**,** 317-323.

CHEN, F., CASTRANOVA, V., SHI, X. & DEMERS, L. M. 1999. New insights into the role of nuclear factor-kappaB, a ubiquitous transcription factor in the initiation of diseases. *Clin Chem,* 45**,** 7-17.

CHEN, L., APGAR, J., HUYNH, L., DICKER, F., GIAGO-MCGAHAN, T., RASSENTI, L., WEISS, A. & KIPPS, T. J. 2005. ZAP-70 directly enhances IgM signaling in chronic lymphocytic leukemia. *Blood,* 105**,** 2036-41.

CHEN, L., WIDHOPF, G., HUYNH, L., RASSENTI, L., RAI, K. R., WEISS, A. & KIPPS, T. J. 2002. Expression of ZAP-70 is associated with increased B-cell receptor signaling in chronic lymphocytic leukemia. *Blood,* 100**,** 4609-14.

CHEN, R., GUO, L., CHEN, Y., JIANG, Y., WIERDA, W. G. & PLUNKETT, W. 2011. Homoharringtonine reduced Mcl-1 expression and induced apoptosis in chronic lymphocytic leukemia. *Blood,* 117**,** 156-64.

CHU, C. C., CATERA, R., HATZI, K., YAN, X. J., ZHANG, L., WANG, X. B., FALES, H. M., ALLEN, S. L., KOLITZ, J. E., RAI, K. R. & CHIORAZZI, N. 2008. Chronic lymphocytic leukemia antibodies with a common stereotypic rearrangement recognize nonmuscle myosin heavy chain IIA. *Blood,* 112**,** 5122-9.

CHU, C. C., CATERA, R., ZHANG, L., DIDIER, S., AGAGNINA, B. M., DAMLE, R. N., KAUFMAN, M. S., KOLITZ, J. E., ALLEN, S. L., RAI, K. R. & CHIORAZZI, N. 2010. Many chronic lymphocytic leukemia antibodies recognize apoptotic cells with exposed nonmuscle myosin heavy chain IIA: implications for patient outcome and cell of origin. *Blood,* 115**,** 3907-15.

CIMMINO, A., CALIN, G. A., FABBRI, M., IORIO, M. V., FERRACIN, M., SHIMIZU, M., WOJCIK, S. E., AQEILAN, R. I., ZUPO, S., DONO, M., RASSENTI, L., ALDER, H., VOLINIA, S., LIU, C. G., KIPPS, T. J., NEGRINI, M. & CROCE, C. M. 2005. miR-

15 and miR-16 induce apoptosis by targeting BCL2. *Proc Natl Acad Sci U S A,* 102**,** 13944-9.

CLAMP, M., FRY, B., KAMAL, M., XIE, X., CUFF, J., LIN, M. F., KELLIS, M., LINDBLAD-TOH, K. & LANDER, E. S. 2007. Distinguishing protein-coding and noncoding genes in the human genome. *Proc Natl Acad Sci U S A,* 104**,** 19428-33.

CLARK, R. A., FUHLBRIGGE, R. C. & SPRINGER, T. A. 1998. L-Selectin ligands that are O-glycoprotease resistant and distinct from MECA-79 antigen are sufficient for tethering and rolling of lymphocytes on human high endothelial venules. *J Cell Biol,* 140**,** 721-31.

COCHRAN, D. A., EVANS, C. A., BLINCO, D., BURTHEM, J., STEVENSON, F. K., GASKELL, S. J. & WHETTON, A. D. 2003. Proteomic analysis of chronic lymphocytic leukemia subtypes with mutated or unmutated Ig V(H) genes. *Mol Cell Proteomics,* 2**,** 1331-41.

COLLINS, R. J., VERSCHUER, L. A., HARMON, B. V., PRENTICE, R. L., POPE, J. H. & KERR, J. F. R. 1989. Spontaneous programmed death (apoptosis) of B-chronic lymphocytic leukaemia cells following their culture in vitro. *British Journal of Haematology,* 71**,** 343-350.

CONRADS, K. A., YI, M., SIMPSON, K. A., LUCAS, D. A., CAMALIER, C. E., YU, L. R., VEENSTRA, T. D., STEPHENS, R. M., CONRADS, T. P. & BECK, G. R., JR. 2005. A combined proteome and microarray investigation of inorganic phosphate-induced pre-osteoblast cells. *Mol Cell Proteomics,* 4**,** 1284-96.

COON, J. J., SYKA, J. E., SHABANOWITZ, J. & HUNT, D. F. 2005. Tandem mass spectrometry for peptide and protein sequence analysis. *Biotechniques,* 38**,** 519, 521, 523.

CORMIER, K. A., FOURNIER, S., O'BRIEN, J., TESSIER, P. & ROBICHAUD, G. A. 2009. Role of S100A8 and S100A9 proteins on breast cancer aggressivity. *The FASEB Journal,* 23**,** LB276.

CORY, S. & ADAMS, J. M. 2002. The Bcl2 family: regulators of the cellular life-or-death switch. *Nat Rev Cancer,* 2**,** 647-56.

COSTANTINI, J. L., CHEUNG, S. M., HOU, S., LI, H., KUNG, S. K., JOHNSTON, J. B., WILKINS, J. A., GIBSON, S. B. & MARSHALL, A. J. 2009. TAPP2 links phosphoinositide 3-kinase signaling to B-cell adhesion through interaction with the cytoskeletal protein utrophin: expression of a novel cell adhesion-promoting complex in B-cell leukemia. *Blood,* 114**,** 4703-12.

COX, J. R. & MANN, M. 2007. Is Proteomics the New Genomics? *Cell,* 130**,** 395-398.

CRESPO, M., BOSCH, F., VILLAMOR, N., BELLOSILLO, B., COLOMER, D., ROZMAN, M., MARCE, S., LOPEZ-GUILLERMO, A., CAMPO, E. & MONTSERRAT, E.

2003. ZAP-70 expression as a surrogate for immunoglobulin-variable-region mutations in chronic lymphocytic leukemia. *N Engl J Med,* 348**,** 1764-75.

CURMI, P. A., NOGUES, C., LACHKAR, S., CARELLE, N., GONTHIER, M. P., SOBEL, A., LIDEREAU, R. & BIECHE, I. 2000. Overexpression of stathmin in breast carcinomas points out to highly proliferative tumours. *Br J Cancer,* 82**,** 142-50.

DAMLE, R. N., TEMBURNI, S., CALISSANO, C., YANCOPOULOS, S., BANAPOUR, T., SISON, C., ALLEN, S. L., RAI, K. R. & CHIORAZZI, N. 2007. CD38 expression labels an activated subset within chronic lymphocytic leukemia clones enriched in proliferating B cells. *Blood,* 110**,** 3352-3359.

DAMLE, R. N., WASIL, T., FAIS, F., GHIOTTO, F., VALETTO, A., ALLEN, S. L., BUCHBINDER, A., BUDMAN, D., DITTMAR, K., KOLITZ, J., LICHTMAN, S. M., SCHULMAN, P., VINCIGUERRA, V. P., RAI, K. R., FERRARINI, M. & CHIORAZZI, N. 1999. Ig V Gene Mutation Status and CD38 Expression As Novel Prognostic Indicators in Chronic Lymphocytic Leukemia. *Blood,* 94**,** 1840-1847.

DAVIDS, M. S. & BURGER, J. A. 2012. Cell Trafficking in Chronic Lymphocytic Leukemia. *Open J Hematol,* 3.

DE SIMONI, S., GOEMAERE, J. & KNOOPS, B. 2008. Silencing of peroxiredoxin 3 and peroxiredoxin 5 reveals the role of mitochondrial peroxiredoxins in the protection of human neuroblastoma SH-SY5Y cells toward MPP+. *Neurosci Lett,* 433**,** 219-24.

DEAGLIO, S., AYDIN, S., GRAND, M. M., VAISITTI, T., BERGUI, L., D'ARENA, G., CHIORINO, G. & MALAVASI, F. 2010. CD38/CD31 interactions activate genetic pathways leading to proliferation and migration in chronic lymphocytic leukemia cells. *Mol Med,* 16**,** 87-91.

DEAGLIO, S., AYDIN, S., VAISITTI, T., BERGUI, L. & MALAVASI, F. 2008. CD38 at the junction between prognostic marker and therapeutic target. *Trends in molecular medicine,* 14**,** 210-218.

DEAGLIO, S., CAPOBIANCO, A., BERGUI, L., DURIG, J., MORABITO, F., DUHRSEN, U. & MALAVASI, F. 2003. CD38 is a signaling molecule in B-cell chronic lymphocytic leukemia cells. *Blood,* 102**,** 2146-55.

DEAGLIO, S. & MALAVASI, F. 2009. Chronic lymphocytic leukemia microenvironment: shifting the balance from apoptosis to proliferation. *Haematologica,* 94**,** 752-6.

DEAGLIO, S., VAISITTI, T., AYDIN, S., BERGUI, L., D'ARENA, G., BONELLO, L., OMED, P., SCATOLINI, M., JAKSIC, O., CHIORINO, G., EFREMOV, D. &

MALAVASI, F. 2007. CD38 and ZAP-70 are functionally linked and mark CLL cells with high migratory potential. *Blood,* 110**,** 4012-4021.

DEAGLIO, S., VAISITTI, T., AYDIN, S., FERRERO, E. & MALAVASI, F. 2006. In-tandem insight from basic science combined with clinical research: CD38 as both marker and key component of the pathogenetic network underlying chronic lymphocytic leukemia. *Blood,* 108**,** 1135-44.

DEAGLIO, S., VAISITTI, T., BERGUI, L., BONELLO, L., HORENSTEIN, A. L., TAMAGNONE, L., BOUMSELL, L. & MALAVASI, F. 2005. CD38 and CD100 lead a network of surface receptors relaying positive signals for B-CLL growth and survival. *Blood,* 105**,** 3042-50.

DEL GAIZO MOORE, V., BROWN, J. R., CERTO, M., LOVE, T. M., NOVINA, C. D. & LETAI, A. 2007. Chronic lymphocytic leukemia requires BCL2 to sequester prodeath BIM, explaining sensitivity to BCL2 antagonist ABT-737. *J Clin Invest,* 117**,** 112-21.

DI BERNARDO, M. C., CROWTHER-SWANEPOEL, D., BRODERICK, P., WEBB, E., SELLICK, G., WILD, R., SULLIVAN, K., VIJAYAKRISHNAN, J., WANG, Y., PITTMAN, A. M., SUNTER, N. J., HALL, A. G., DYER, M. J., MATUTES, E., DEARDEN, C., MAINOU-FOWLER, T., JACKSON, G. H., SUMMERFIELD, G., HARRIS, R. J., PETTITT, A. R., HILLMEN, P., ALLSUP, D. J., BAILEY, J. R., PRATT, G., PEPPER, C., FEGAN, C., ALLAN, J. M., CATOVSKY, D. & HOULSTON, R. S. 2008. A genome-wide association study identifies six susceptibility loci for chronic lymphocytic leukemia. *Nat Genet,* 40**,** 1204-10.

DI PAOLO, G., PELLIER, V., CATSICAS, M., ANTONSSON, B., CATSICAS, S. & GRENNINGLOH, G. 1996. The phosphoprotein stathmin is essential for nerve growth factor-stimulated differentiation. *J Cell Biol,* 133**,** 1383-90.

DIGHIERO, G. 2003. Unsolved issues in CLL biology and management. *Leukemia,* 17**,** 2385-91.

DIGHIERO, G. 2005. CLL biology and prognosis. *Hematology Am Soc Hematol Educ Program***,** 278-84.

DIGHIERO, G. & HAMBLIN, T. J. 2008. Chronic lymphocytic leukaemia. *Lancet,* 371**,** 1017-29.

DING, Q., XIAO, L., XIONG, S., JIA, Y., QUE, H., GUO, Y. & LIU, S. 2003. Unmatched masses in peptide mass fingerprints caused by cross-contamination: an updated statistical result. *Proteomics,* 3**,** 1313-7.

DOHNER, H., STILGENBAUER, S., BENNER, A., LEUPOLT, E., KROBER, A., BULLINGER, L., DOHNER, K., BENTZ, M. & LICHTER, P. 2000. Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med,* 343**,** 1910-6.

DOHNER, H., STILGENBAUER, S., DOHNER, K., BENTZ, M. & LICHTER, P. 1999. Chromosome aberrations in B-cell chronic lymphocytic leukemia: reassessment based on molecular cytogenetic analysis. *J Mol Med (Berl),* 77**,** 266-81.

DREGER, M. 2003. Subcellular proteomics. *Mass Spectrom Rev,* 22**,** 27-56.

DROBNJAK, M., OSMAN, I., SCHER, H. I., FAZZARI, M. & CORDON-CARDO, C. 2000. Overexpression of Cyclin D1 Is Associated with Metastatic Prostate Cancer to Bone. *Clinical Cancer Research,* 6**,** 1891-1895.

DROUVA, S. V., POULIN, B., MANCEAU, V. & SOBEL, A. 1998. Luteinizing hormone-releasing hormone-signal transduction and stathmin phosphorylation in the gonadotrope alphaT3-1 cell line. *Endocrinology,* 139**,** 2235-9.

DURIG, J., NASCHAR, M., SCHMUCKER, U., RENZING-KOHLER, K., HOLTER, T., HUTTMANN, A. & DUHRSEN, U. 2002. CD38 expression is an important prognostic marker in chronic lymphocytic leukaemia. *Leukemia,* 16**,** 30-5.

DYER, M. J., ZANI, V. J., LU, W. Z., O'BYRNE, A., MOULD, S., CHAPMAN, R., HEWARD, J. M., KAYANO, H., JADAYEL, D., MATUTES, E. & ET AL. 1994. BCL2 translocations in leukemias of mature B cells. *Blood,* 83**,** 3682-8.

EL ROUBY, S., THOMAS, A., COSTIN, D., ROSENBERG, C., POTMESIL, M., SILBER, R. & NEWCOMB, E. 1993. p53 gene mutation in B-cell chronic lymphocytic leukemia is associated with drug resistance and is independent of MDR1/MDR3 gene expression. *Blood,* 82**,** 3452-3459.

ENJUANES, A., BENAVENTE, Y., BOSCH, F., MARTIN-GUERRERO, I., COLOMER, D., PEREZ-ALVAREZ, S., REINA, O., ARDANAZ, M. T., JARES, P., GARCIA-ORAD, A., PUJANA, M. A., MONTSERRAT, E., DE SANJOSE, S. & CAMPO, E. 2008. Genetic variants in apoptosis and immunoregulation-related genes are associated with risk of chronic lymphocytic leukemia. *Cancer Res,* 68**,** 10178-86.

FERRI, C., LA CIVITA, L., CARACCIOLO, F., BELLESI, G. & ZIGNEGO, A. L. 1996. Hepatitis C virus and lymphoproliferative disorders. *Blood,* 88**,** 4730-1.

FEYCHTING, M., FORSSEN, U. & FLODERUS, B. 1997. Occupational and residential magnetic field exposure and leukemia and central nervous system tumors. *Epidemiology,* 8**,** 384-9.

FIGEYS, D. 2005. Proteomics: the basic overview. *Methods Biochem Anal,* 45**,** 1-62.

FLODERUS, B., PERSSON, T., STENLUND, C., WENNBERG, A., OST, A. & KNAVE, B. 1993. Occupational exposure to electromagnetic fields in relation to leukemia and brain tumors: a case-control study in Sweden. *Cancer Causes Control,* 4**,** 465-76.

FONDELL, J. D., GE, H. & ROEDER, R. G. 1996. Ligand induction of a transcriptionally active thyroid hormone receptor coactivator complex. *Proc Natl Acad Sci U S A,* 93**,** 8329-33.

FURMAN, R. R., ASGARY, Z., MASCARENHAS, J. O., LIOU, H. C. & SCHATTNER, E. J. 2000. Modulation of NF-kappa B activity and apoptosis in chronic lymphocytic leukemia B cells. *J Immunol,* 164**,** 2200-6.

GAUDIO, E., SPIZZO, R., PADUANO, F., LUO, Z., EFANOV, A., PALAMARCHUK, A., LEBER, A. S., KAOU, M., ZANESI, N., BOTTONI, A., COSTINEAN, S., RASSENTI, L. Z., NAKAMURA, T., KIPPS, T. J., AQEILAN, R. I., PEKARSKY, Y., TRAPASSO, F. & CROCE, C. M. 2012. Tcl1 interacts with Atm and enhances NF-kappaB activation in hematologic malignancies. *Blood,* 119**,** 180-7.

GOTTARDI, D., ALFARANO, A., DE LEO, A. M., STACCHINI, A., ARAGNO, M., RIGO, A., VENERI, D., ZANOTTI, R., PIZZOLO, G. & CALIGARIS-CAPPIO, F. 1996. In leukaemic CD5+ B cells the expression of BCL-2 gene family is shifted toward protection from apoptosis. *British Journal of Haematology,* 94**,** 612-618.

GRANZIERO, L., CIRCOSTA, P., SCIELZO, C., FRISALDI, E., STELLA, S., GEUNA, M., GIORDANO, S., GHIA, P. & CALIGARIS-CAPPIO, F. 2003. CD100/Plexin-B1 interactions sustain proliferation and survival of normal and leukemic CD5+ B lymphocytes. *Blood,* 101**,** 1962-9.

GRANZIERO, L., GHIA, P., CIRCOSTA, P., GOTTARDI, D., STROLA, G., GEUNA, M., MONTAGNA, L., PICCOLI, P., CHILOSI, M. & CALIGARIS-CAPPIO, F. 2001. Survivin is expressed on CD40 stimulation and interfaces proliferation and apoptosis in B-cell chronic lymphocytic leukemia. *Blood,* 97**,** 2777-2783.

GRIBBEN, J. G. 2010. How I treat CLL up front. *Blood,* 115**,** 187-197.

GUARINI, A., CHIARETTI, S., TAVOLARO, S., MAGGIO, R., PERAGINE, N., CITARELLA, F., RICCIARDI, M. R., SANTANGELO, S., MARINELLI, M., DE PROPRIS, M. S., MESSINA, M., MAURO, F. R., DEL GIUDICE, I. & FOA, R. 2008. BCR ligation induced by IgM stimulation results in gene expression and functional changes only in IgV H unmutated chronic lymphocytic leukemia (CLL) cells. *Blood,* 112**,** 782-92.

GUNZ, F. W., GUNZ, J. P., VEALE, A. M., CHAPMAN, C. J. & HOUSTON, I. B. 1975. Familial leukaemia: a study of 909 families. *Scand J Haematol,* 15**,** 117-31.

GUNZ, F. W. & VEALE, A. M. 1969. Leukemia in close relatives--accident or predisposition? *J Natl Cancer Inst,* 42**,** 517-24.

GYGI, S. P., RIST, B., GERBER, S. A., TURECEK, F., GELB, M. H. & AEBERSOLD, R. 1999a. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol,* 17**,** 994-9.

GYGI, S. P., ROCHON, Y., FRANZA, B. R. & AEBERSOLD, R. 1999b. Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol,* 19**,** 1720-30.

HAJRA, K. M. & LIU, J. R. 2004. Apoptosome dysfunction in human cancer. *Apoptosis,* 9**,** 691-704.

HALLEK, M., CHESON, B. D., CATOVSKY, D., CALIGARIS-CAPPIO, F., DIGHIERO, G., DOHNER, H., HILLMEN, P., KEATING, M. J., MONTSERRAT, E., RAI, K. R., KIPPS, T. J. & INTERNATIONAL WORKSHOP ON CHRONIC LYMPHOCYTIC, L. 2008. Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines. *Blood,* 111**,** 5446-56.

HALLEK, M., FISCHER, K., FINGERLE-ROWSON, G., FINK, A. M., BUSCH, R., MAYER, J., HENSEL, M., HOPFINGER, G., HESS, G., VON GRUNHAGEN, U., BERGMANN, M., CATALANO, J., ZINZANI, P. L., CALIGARIS-CAPPIO, F., SEYMOUR, J. F., BERREBI, A., JAGER, U., CAZIN, B., TRNENY, M., WESTERMANN, A., WENDTNER, C. M., EICHHORST, B. F., STAIB, P., BUHLER, A., WINKLER, D., ZENZ, T., BOTTCHER, S., RITGEN, M., MENDILA, M., KNEBA, M., DOHNER, H., STILGENBAUER, S., INTERNATIONAL GROUP OF, I. & GERMAN CHRONIC LYMPHOCYTIC LEUKAEMIA STUDY, G. 2010. Addition of rituximab to fludarabine and cyclophosphamide in patients with chronic lymphocytic leukaemia: a randomised, open-label, phase 3 trial. *Lancet,* 376**,** 1164-74.

HAMBLIN, T. J., DAVIS, Z., GARDINER, A., OSCIER, D. G. & STEVENSON, F. K. 1999. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood,* 94**,** 1848-54.

HAMBLIN, T. J., ORCHARD, J. A., IBBOTSON, R. E., DAVIS, Z., THOMAS, P. W., STEVENSON, F. K. & OSCIER, D. G. 2002. CD38 expression and immunoglobulin variable region mutations are independent prognostic variables in chronic lymphocytic leukemia, but CD38 expression may vary during the course of the disease. *Blood,* 99**,** 1023-9.

HANADA, M., DELIA, D., AIELLO, A., STADTMAUER, E. & REED, J. C. 1993. bcl-2 gene hypomethylation and high-level expression in B-cell chronic lymphocytic leukemia. *Blood,* 82**,** 1820-8.

HAWLEY, R. S. & FRIEND, S. H. 1996. Strange bedfellows in even stranger places: the role of ATM in meiotic cells, lymphocytes, tumors, and its functional links to p53. *Genes Dev,* 10**,** 2383-8.

HAYDEN, M. S. & GHOSH, S. 2012. NF-kappaB, the first quarter-century: remarkable progress and outstanding questions. *Genes Dev,* 26**,** 203-34.

HE, Z., YANG, C. & YU, W. 2008. Peak bagging for peptide mass fingerprinting. *Bioinformatics,* 24**,** 1293-9.

HEGDE, P. S., WHITE, I. R. & DEBOUCK, C. 2003. Interplay of transcriptomics and proteomics. *Curr Opin Biotechnol,* 14**,** 647-51.

HEIZMANN, C. W., FRITZ, G. & SCHAFER, B. W. 2002. S100 proteins: structure, functions and pathology. *Front Biosci,* 7**,** d1356-68.

HERLING, M., PATEL, K. A., KHALILI, J., SCHLETTE, E., KOBAYASHI, R., MEDEIROS, L. J. & JONES, D. 2006. TCL1 shows a regulated expression pattern in chronic lymphocytic leukemia that correlates with molecular subtypes and proliferative state. *Leukemia,* 20**,** 280-5.

HERMANI, A., HESS, J., DE SERVI, B., MEDUNJANIN, S., GROBHOLZ, R., TROJAN, L., ANGEL, P. & MAYER, D. 2005. Calcium-Binding Proteins S100A8 and S100A9 as Novel Diagnostic Markers in Human Prostate Cancer. *Clinical Cancer Research,* 11**,** 5146-5152.

HEWAMANA, S., ALGHAZAL, S., LIN, T. T., CLEMENT, M., JENKINS, C., GUZMAN, M. L., JORDAN, C. T., NEELAKANTAN, S., CROOKS, P. A., BURNETT, A. K., PRATT, G., FEGAN, C., ROWNTREE, C., BRENNAN, P. & PEPPER, C. 2008. The NF-kappaB subunit Rel A is associated with in vitro survival and clinical disease progression in chronic lymphocytic leukemia and represents a promising therapeutic target. *Blood,* 111**,** 4681-9.

HEWAMANA, S., LIN, T. T., ROWNTREE, C., KARUNANITHI, K., PRATT, G., HILLS, R., FEGAN, C., BRENNAN, P. & PEPPER, C. 2009. Rel a is an independent biomarker of clinical outcome in chronic lymphocytic leukemia. *J Clin Oncol,* 27**,** 763-9.

HORIE, R., WATANABE, M., OKAMURA, T., TAIRA, M., SHODA, M., MOTOJI, T., UTSUNOMIYA, A., WATANABE, T., HIGASHIHARA, M. & UMEZAWA, K. 2006. DHMEQ, a new NF-kappaB inhibitor, induces apoptosis and enhances fludarabine effects on chronic lymphocytic leukemia cells. *Leukemia,* 20**,** 800-6.

HUANG, D., FU, W., GUO, Y., XU, Z., SUN, X. & SUN, K. 2008. <i>S100A8</i> mediates the activation of <i>P65/HLA-B/S100A8/BCL-2/Caspase-9</i> (<i>−3</i>) pathway in laryngeal carcinogenesis. *Chinese Science Bulletin,* 53**,** 2017-2024.

HUBER, L. A., PFALLER, K. & VIETOR, I. 2003. Organelle proteomics: implications for subcellular fractionation in proteomics. *Circ Res,* 92**,** 962-8.

HUTCHESON, J., SCATIZZI, J. C., SIDDIQUI, A. M., HAINES, G. K., 3RD, WU, T., LI, Q. Z., DAVIS, L. S., MOHAN, C. & PERLMAN, H. 2008. Combined deficiency of proapoptotic regulators Bim and Fas results in the early onset of systemic autoimmunity. *Immunity,* 28**,** 206-17.

HUTTMANN, A., KLEIN-HITPASS, L., THOMALE, J., DEENEN, R., CARPINTEIRO, A., NUCKEL, H., EBELING, P., FUHRER, A., EDELMANN, J., SELLMANN, L.,

DUHRSEN, U. & DURIG, J. 2006. Gene expression signatures separate B-cell chronic lymphocytic leukaemia prognostic subgroups defined by ZAP-70 and CD38 expression status. *Leukemia,* 20**,** 1774-82.

ISRAELS, E. D. & ISRAELS, L. G. 2001. The Cell Cycle. *Stem Cells,* 19**,** 88-91.

ISSAQ, H. J. 2001. The role of separation science in proteomics research. *Electrophoresis,* 22**,** 3629-38.

ISSAQ, H. J., CONRADS, T. P., JANINI, G. M. & VEENSTRA, T. D. 2002. Methods for fractionation, separation and profiling of proteins and peptides. *Electrophoresis,* 23**,** 3048-61.

JACOBELLI, J., FRIEDMAN, R. S., CONTI, M. A., LENNON-DUMENIL, A.-M., PIEL, M., SORENSEN, C. M., ADELSTEIN, R. S. & KRUMMEL, M. F. 2010. Confinement-optimized three-dimensional T cell amoeboid motility is modulated via myosin IIA-regulated adhesions. *Nature immunology,* 11**,** 953-961.

JAIN, P., GIUSTOLISI, G. M., ATKINSON, S., ELNENAEI, M. O., MORILLA, R., OWUSU-ANKOMAH, K., RAFIQ-MOHAMMED, F., MATUTES, E., WOTHERSPOON, A. & CATOVSKY, D. 2002. Detection of cyclin D1 in B cell lymphoproliferative disorders by flow cytometry. *J Clin Pathol,* 55**,** 940-5.

JAMROZIAK, K., SZEMRAJ, Z., GRZYBOWSKA-IZYDORCZYK, O., SZEMRAJ, J., BIENIASZ, M., CEBULA, B., GIANNOPOULOS, K., BALCERCZAK, E., JESIONEK-KUPNICKA, D., KOWAL, M., KOSTYRA, A., CALBECKA, M., WAWRZYNIAK, E., MIROWSKI, M., KORDEK, R. & ROBAK, T. 2009. CD38 gene polymorphisms contribute to genetic susceptibility to B-cell chronic lymphocytic leukemia: evidence from two case-control studies in Polish Caucasians. *Cancer Epidemiol Biomarkers Prev,* 18**,** 945-53.

JEVTOVIC-STOIMENOV, T., KOCIC, G., PAVLOVIC, D., MACUKANOVIC-GOLUBOVIC, L., MARJANOVIC, G., DJORDJEVIC, V., TOSIC, N. & PAVLOVIC, S. 2008. Polymorphisms of tumor-necrosis factor-alpha - 308 and lymphotoxin-alpha + 250: possible modulation of susceptibility to apoptosis in chronic lymphocytic leukemia and non-Hodgkin lymphoma mononuclear cells. *Leuk Lymphoma,* 49**,** 2163-9.

KARAS, M. & HILLENKAMP, F. 1988. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem,* 60**,** 2299-301.

KENNY, F. S., HUI, R., MUSGROVE, E. A., GEE, J. M. W., BLAMEY, R. W., NICHOLSON, R. I., SUTHERLAND, R. L. & ROBERTSON, J. F. R. 1999. Overexpression of Cyclin D1 Messenger RNA Predicts for Poor Prognosis in Estrogen Receptor-positive Breast Cancer. *Clinical Cancer Research,* 5**,** 2069-2076.

KIPPS, T. J. 2007. The B-cell receptor and ZAP-70 in chronic lymphocytic leukemia. *Best Pract Res Clin Haematol,* 20**,** 415-24.

KITADA, S., ANDERSEN, J., AKAR, S., ZAPATA, J. M., TAKAYAMA, S., KRAJEWSKI, S., WANG, H. G., ZHANG, X., BULLRICH, F., CROCE, C. M., RAI, K., HINES, J. & REED, J. C. 1998. Expression of apoptosis-regulating proteins in chronic lymphocytic leukemia: correlations with In vitro and In vivo chemoresponses. *Blood,* 91**,** 3379-89.

KLEIN, U., TU, Y., STOLOVITZKY, G. A., MATTIOLI, M., CATTORETTI, G., HUSSON, H., FREEDMAN, A., INGHIRAMI, G., CRO, L. & BALDINI, L. 2001. Gene expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory B cells. *The Journal of experimental medicine,* 194**,** 1625-1638.

KONG, G. H., BU, J. Y., KUROSAKI, T., SHAW, A. S. & CHAN, A. C. 1995. Reconstitution of Syk function by the ZAP-70 protein tyrosine kinase. *Immunity,* 2**,** 485-92.

KROBER, A., SEILER, T., BENNER, A., BULLINGER, L., BRUCKLE, E., LICHTER, P., DOHNER, H. & STILGENBAUER, S. 2002. V(H) mutation status, CD38 expression level, genomic aberrations, and survival in chronic lymphocytic leukemia. *Blood,* 100**,** 1410-6.

KUMAGAI, M., COUSTAN-SMITH, E., MURRAY, D. J., SILVENNOINEN, O., MURTI, K. G., EVANS, W. E., MALAVASI, F. & CAMPANA, D. 1995. Ligation of CD38 suppresses human B lymphopoiesis. *J Exp Med,* 181**,** 1101-10.

LAINE, J., K¸NSTLE, G., OBATA, T., SHA, M. & NOGUCHI, M. 2000. The Protooncogene TCL1 Is an Akt Kinase Coactivator. *Molecular cell,* 6**,** 395-407.

LANASA, M. C. 2010. Novel insights into the biology of CLL. *Hematology Am Soc Hematol Educ Program,* 2010**,** 70-6.

LEE, K.-M., HSU, I.-W. & TARN, W.-Y. 2010. TRAP150 activates pre-mRNA splicing and promotes nuclear mRNA degradation. *Nucleic Acids Research,* 38**,** 3340-3350.

LETAI, A., SORCINELLI, M. D., BEARD, C. & KORSMEYER, S. J. 2004. Antiapoptotic BCL-2 is required for maintenance of a model leukemia. *Cancer cell,* 6**,** 241-249.

Liebler, D. (2002) Introduction to Proteomics: Tools for the New Biology, New Jersey: Humana Press Inc.

LIM, M. S. & ELENITOBA-JOHNSON, K. S. 2004. Proteomics in pathology research. *Lab Invest,* 84**,** 1227-44.

LIU, H., SADYGOV, R. G. & YATES, J. R., 3RD 2004. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem,* 76**,** 4193-201.

LONGO, P. G., LAURENTI, L., GOBESSI, S., SICA, S., LEONE, G. & EFREMOV, D. G. 2008. The Akt/Mcl-1 pathway plays a prominent role in mediating antiapoptotic signals downstream of the B-cell receptor in chronic lymphocytic leukemia B cells. *Blood,* 111**,** 846-55.

LUBEC, G. & AFJEHI-SADAT, L. 2007. Limitations and pitfalls in protein identification by mass spectrometry. *Chem Rev,* 107**,** 3568-84.

LUPPI, M., GRAZIA FERRARI, M., BONACCORSI, G., LONGO, G., NARNI, F., BAROZZI, P., MARASCA, R., MUSSINI, C. & TORELLI, G. 1996. Hepatitis C virus infection in subsets of neoplastic lymphoproliferations not associated with cryoglobulinemia. *Leukemia,* 10**,** 351-5.

MAINOU-FOWLER, T., MILLER, S., PROCTOR, S. J. & DICKINSON, A. M. 2001. The levels of TNF alpha, IL4 and IL10 production by T-cells in B-cell chronic lymphocytic leukaemia (B-CLL). *Leuk Res,* 25**,** 157-63.

MANN, M., HENDRICKSON, R. C. & PANDEY, A. 2001. Analysis of proteins and proteomes by mass spectrometry. *Annu Rev Biochem,* 70**,** 437-73.

MANSOURI, M. R., SEVOV, M., ALESKOG, A., JONDAL, M., MERUP, M., SUNDSTROM, C., OSORIO, L. & ROSENQUIST, R. 2010. IGHV3-21 gene usage is associated with high TCL1 expression in chronic lymphocytic leukemia. *Eur J Haematol,* 84**,** 109-16.

MARSCHITZ, I., TINHOFER, I., HITTMAIR, A., EGLE, A., KOS, M. & GREIL, R. 2000. Analysis of Bcl-2 protein expression in chronic lymphocytic leukemia. A comparison of three semiquantitation techniques. *Am J Clin Pathol,* 113**,** 219-29.

MARSHALL, A. J., KRAHN, A. K., MA, K., DURONIO, V. & HOU, S. 2002. TAPP1 and TAPP2 are targets of phosphatidylinositol 3-kinase signaling in B cells: sustained plasma membrane recruitment triggered by the B-cell antigen receptor. *Mol Cell Biol,* 22**,** 5479-91.

MARTIN, S. J., REUTELINGSPERGER, C. P., MCGAHON, A. J., RADER, J. A., VAN SCHIE, R. C., LAFACE, D. M. & GREEN, D. R. 1995. Early redistribution of plasma membrane phosphatidylserine is a general feature of apoptosis regardless of the initiating stimulus: inhibition by overexpression of Bcl-2 and Abl. *J Exp Med,* 182**,** 1545-56.

MCCARTHY, B. A., BOYLE, E., WANG, X. P., GUZOWSKI, D., PAUL, S., CATERA, R., TROTT, J., YAN, X. J., CROCE, C. M., DAMLE, R., YANCOPOULOS, S., MESSMER, B. T., LESSER, M., ALLEN, S. L., RAI, K. R. & CHIORAZZI, N. 2008. Surface expression of Bcl-2 in chronic lymphocytic leukemia and other B-cell leukemias and lymphomas without a breakpoint t(14;18). *Mol Med,* 14**,** 618-27.

MESSMER, B. T., MESSMER, D., ALLEN, S. L., KOLITZ, J. E., KUDALKAR, P., CESAR, D., MURPHY, E. J., KODURU, P., FERRARINI, M., ZUPO, S., CUTRONA, G., DAMLE, R. N., WASIL, T., RAI, K. R., HELLERSTEIN, M. K. & CHIORAZZI, N. 2005. In vivo measurements document the dynamic cellular kinetics of chronic lymphocytic leukemia B cells. *J Clin Invest,* 115**,** 755-64.

MIGUET, L., BECHADE, G., FORNECKER, L., ZINK, E., FELDEN, C., GERVAIS, C., HERBRECHT, R., VAN DORSSELAER, A., MAUVIEUX, L. & SANGLIER-CIANFERANI, S. 2009. Proteomic analysis of malignant B-cell derived microparticles reveals CD148 as a potentially useful antigenic biomarker for mantle cell lymphoma diagnosis. *J Proteome Res,* 8**,** 3346-54.

MITTAL, A. K., HEGDE, G. V., AOUN, P., BOCIEK, R. G., DAVE, B. J., JOSHI, A. D., SANGER, W. G., WEISENBURGER, D. D. & JOSHI, S. S. 2007. Molecular basis of aggressive disease in chronic lymphocytic leukemia patients with 11q deletion and trisomy 12 chromosomal abnormalities. *Int J Mol Med,* 20**,** 461-9.

MOLICA, S. 2006. Sex differences in incidence and outcome of chronic lymphocytic leukemia patients. *Leuk Lymphoma,* 47**,** 1477-80.

MONTEOLIVA, L. & ALBAR, J. P. 2004. Differential proteomics: an overview of gel and non-gel based approaches. *Brief Funct Genomic Proteomic,* 3**,** 220-39.

MONTES-ARES, O., MOYA-QUILES, M. R., MONTES-CASADO, M., GUERRA-PEREZ, N., CAMPILLO, J. A., GONZALEZ, C., LOPEZ-BERMEJO, A., TAMAYO, M., MAJADO, M. J., PARRADO, A., MURO, M., MARIN, L. & ALVAREZ-LOPEZ, M. R. 2006. Human leucocyte antigen-C in B chronic lymphocytic leukaemia. *Br J Haematol,* 135**,** 517-9.

MONTSERRAT, E. 2006. New prognostic markers in CLL. *Hematology Am Soc Hematol Educ Program***,** 279-84.

MORIN, N. A., OAKES, P. W., HYUN, Y. M., LEE, D., CHIN, Y. E., KING, M. R., SPRINGER, T. A., SHIMAOKA, M., TANG, J. X., REICHNER, J. S. & KIM, M. 2008. Nonmuscle myosin heavy chain IIA mediates integrin LFA-1 de-adhesion during T lymphocyte migration. *J Exp Med,* 205**,** 195-205.

NADLER, L. M., ANDERSON, K. C., MARTI, G., BATES, M., PARK, E., DALEY, J. F. & SCHLOSSMAN, S. F. 1983. B4, a human B lymphocyte-associated antigen expressed on normal, mitogen-activated, and malignant B lymphocytes. *J Immunol,* 131**,** 244-50.

NAKAMURA, T. & ODA, Y. 2007. Mass spectrometry-based quantitative proteomics. *Biotechnol Genet Eng Rev,* 24**,** 147-63.

NEMETH, J., STEIN, I., HAAG, D., RIEHL, A., LONGERICH, T., HORWITZ, E., BREUHAHN, K., GEBHARDT, C., SCHIRMACHER, P., HAHN, M., BEN-NERIAH, Y., PIKARSKY, E., ANGEL, P. & HESS, J. 2009. S100A8 and S100A9

are novel nuclear factor kappa B target genes during malignant progression of murine and human liver carcinogenesis. *Hepatology,* 50**,** 1251-62.

NICOLAS, E., RAMUS, C., BERTHIER, S., ARLOTTO, M., BOUAMRANI, A., LEFEBVRE, C., MOREL, F., GARIN, J., IFRAH, N., BERGER, F., CAHN, J. Y. & MOSSUZ, P. 2011. Expression of S100A8 in leukemic cells predicts poor survival in de novo AML patients. *Leukemia,* 25**,** 57-65.

NILSEN, T. W. & GRAVELEY, B. R. 2010. Expansion of the eukaryotic proteome by alternative splicing. *Nature,* 463**,** 457-63.

NISHIZUKA, S., CHARBONEAU, L., YOUNG, L., MAJOR, S., REINHOLD, W. C., WALTHAM, M., KOUROS-MEHR, H., BUSSEY, K. J., LEE, J. K., ESPINA, V., MUNSON, P. J., PETRICOIN, E., 3RD, LIOTTA, L. A. & WEINSTEIN, J. N. 2003. Proteomic profiling of the NCI-60 cancer cell lines using new high-density reverse-phase lysate microarrays. *Proc Natl Acad Sci U S A,* 100**,** 14229-34.

ONG, S.-E., BLAGOEV, B., KRATCHMAROVA, I., KRISTENSEN, D. B., STEEN, H., PANDEY, A. & MANN, M. 2002. Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Molecular & Cellular Proteomics,* 1**,** 376-386.

ORMEROD, M. G., SUN, X. M., SNOWDEN, R. T., DAVIES, R., FEARNHEAD, H. & COHEN, G. M. 1993. Increased membrane permeability of apoptotic thymocytes: a flow cytometric study. *Cytometry,* 14**,** 595-602.

OSCIER, D., FEGAN, C., HILLMEN, P., ILLIDGE, T., JOHNSON, S., MAGUIRE, P., MATUTES, E. & MILLIGAN, D. 2004. Guidelines on the diagnosis and management of chronic lymphocytic leukaemia. *British Journal of Haematology,* 125**,** 294-317.

OTAKE, Y., SOUNDARARAJAN, S., SENGUPTA, T. K., KIO, E. A., SMITH, J. C., PINEDA-ROMAN, M., STUART, R. K., SPICER, E. K. & FERNANDES, D. J. 2007. Overexpression of nucleolin in chronic lymphocytic leukemia cells induces stabilization of bcl2 mRNA. *Blood,* 109**,** 3069-75.

PAN, J. W., COOK, L. S., SCHWARTZ, S. M. & WEIS, N. S. 2002. Incidence of leukemia in Asian migrants to the United States and their descendants. *Cancer Causes Control,* 13**,** 791-5.

PANAYIOTIDIS, P., JONES, D., GANESHAGURU, K., FORONI, L. & HOFFBRAND, A. V. 1996. Human bone marrow stromal cells prevent apoptosis and support the survival of chronic lymphocytic leukaemia cells in vitro. *Br J Haematol,* 92**,** 97-103.

PARKER, T. L. & STROUT, M. P. 2011. Chronic lymphocytic leukemia: prognostic factors and impact on treatment. *Discov Med,* 11**,** 115-23.

PATTEN, P. E. M., BUGGINS, A. G. S., RICHARDS, J., WOTHERSPOON, A., SALISBURY, J., MUFTI, G. J., HAMBLIN, T. J. & DEVEREUX, S. 2008. CD38 expression in chronic lymphocytic leukemia is regulated by the tumor microenvironment. *Blood,* 111**,** 5173-5181.

PEKARSKY, Y., SANTANAM, U., CIMMINO, A., PALAMARCHUK, A., EFANOV, A., MAXIMOV, V., VOLINIA, S., ALDER, H., LIU, C. G., RASSENTI, L., CALIN, G. A., HAGAN, J. P., KIPPS, T. & CROCE, C. M. 2006. Tcl1 expression in chronic lymphocytic leukemia is regulated by miR-29 and miR-181. *Cancer Res,* 66**,** 11590-3.

PEPPER, C., LIN, T. T., PRATT, G., HEWAMANA, S., BRENNAN, P., HILLER, L., HILLS, R., WARD, R., STARCZYNSKI, J., AUSTEN, B., HOOPER, L., STANKOVIC, T. & FEGAN, C. 2008. Mcl-1 expression has in vitro and in vivo significance in chronic lymphocytic leukemia and is associated with other poor prognostic markers. *Blood,* 112**,** 3807-3817.

PEPPER, C., MAJID, A., LIN, T. T., HEWAMANA, S., PRATT, G., WALEWSKA, R., GESK, S., SIEBERT, R., WAGNER, S., KENNEDY, B., MIALL, F., DAVIS, Z. A., TRACY, I., GARDINER, A. C., BRENNAN, P., HILLS, R. K., DYER, M. J., OSCIER, D. & FEGAN, C. 2012. Defining the prognosis of early stage chronic lymphocytic leukaemia patients. *Br J Haematol,* 156**,** 499-507.

PEPPER, C., THOMAS, A., HOY, T., COTTER, F. & BENTLEY, P. 1999. Antisense-mediated suppression of Bcl-2 highlights its pivotal role in failed apoptosis in B-cell chronic lymphocytic leukaemia. *Br J Haematol,* 107**,** 611-5.

PEPPER, C., WARD, R., LIN, T. T., BRENNAN, P., STARCZYNSKI, J., MUSSON, M., ROWNTREE, C., BENTLEY, P., MILLS, K., PRATT, G. & FEGAN, C. 2007. Highly purified CD38+ and CD38- sub-clones derived from the same chronic lymphocytic leukemia patient have distinct gene expression signatures despite their monoclonal origin. *Leukemia,* 21**,** 687-96.

PEREZ-CHACON, G., CONTRERAS-MARTIN, B., CUNI, S., ROSADO, S., MARTIN-DONAIRE, T., LOSADA-FERNANDEZ, I., VARGAS, J. A., JORDA, J., ALVAREZ, N., GARCIA-MARCO, J. & PEREZ-ACIEGO, P. 2005. Polymorphism in the CD5 gene promoter in B-cell chronic lymphocytic leukemia and mantle cell lymphoma. *Am J Clin Pathol,* 123**,** 646-50.

PERROT, A., PIONNEAU, C., NADAUD, S., DAVI, F., LEBLOND, V., JACOB, F., MERLE-BERAL, H., HERBRECHT, R., BENE, M. C., GRIBBEN, J. G., BAHRAM, S. & VALLAT, L. 2011. A unique proteomic profile on surface IgM ligation in unmutated chronic lymphocytic leukemia. *Blood,* 118**,** e1-15.

PERSENGIEV, S. P., DEVIREDDY, L. R. & GREEN, M. R. 2002. Inhibition of apoptosis by ATFx: a novel role for a member of the ATF/CREB family of mammalian bZIP transcription factors. *Genes Dev,* 16**,** 1806-14.

PINILLA-IBARZ, J. & MCQUARY, A. 2010. Chronic lymphocytic leukemia: putting new treatment options into perspective. *Cancer Control,* 17**,** 4-15; quiz 16.

PRICE, D. K., BALL, J. R., BAHRANI-MOSTAFAVI, Z., VACHRIS, J. C., KAUFMAN, J. S., NAUMANN, R. W., HIGGINS, R. V. & HALL, J. B. 2000. The phosphoprotein Op18/stathmin is differentially expressed in ovarian cancer. *Cancer Invest,* 18**,** 722-30.

QUIROGA, M. P., BALAKRISHNAN, K., KURTOVA, A. V., SIVINA, M., KEATING, M. J., WIERDA, W. G., GANDHI, V. & BURGER, J. A. 2009. B-cell antigen receptor signaling enhances chronic lymphocytic leukemia cell migration and survival: specific targeting with a novel spleen tyrosine kinase inhibitor, R406. *Blood,* 114**,** 1029-37.

RAI, K., SAWITSKY, A., CRONKITE, E., CHANANA, A., LEVY, R. & PASTERNACK, B. 1975. Clinical staging of chronic lymphocytic leukemia. *Blood,* 46**,** 219-234.

RASSENTI, L. Z., HUYNH, L., TOY, T. L., CHEN, L., KEATING, M. J., GRIBBEN, J. G., NEUBERG, D. S., FLINN, I. W., RAI, K. R., BYRD, J. C., KAY, N. E., GREAVES, A., WEISS, A. & KIPPS, T. J. 2004. ZAP-70 Compared with Immunoglobulin Heavy-Chain Gene Mutation Status as a Predictor of Disease Progression in Chronic Lymphocytic Leukemia. *New England Journal of Medicine,* 351**,** 893-901.

REDAELLI, A., LASKIN, B. L., STEPHENS, J. M., BOTTEMAN, M. F. & PASHOS, C. L. 2004. The clinical and epidemiological burden of chronic lymphocytic leukaemia. *Eur J Cancer Care (Engl),* 13**,** 279-87.

RICHARDSON, S. J., MATTHEWS, C., CATHERWOOD, M. A., ALEXANDER, H. D., CAREY, B. S., FARRUGIA, J., GARDINER, A., MOULD, S., OSCIER, D., COPPLESTONE, J. A. & PRENTICE, A. G. 2006. ZAP-70 expression is associated with enhanced ability to respond to migratory and survival signals in B-cell chronic lymphocytic leukemia (B-CLL). *Blood,* 107**,** 3584-3592.

ROMERO, F., MARTINEZ, A. C., CAMONIS, J. & REBOLLO, A. 1999. Aiolos transcription factor controls cell death in T cells by regulating Bcl-2 expression and its cellular localization. *EMBO J,* 18**,** 3419-30.

ROSENWALD, A., ALIZADEH, A. A., WIDHOPF, G., SIMON, R., DAVIS, R. E., YU, X., YANG, L., PICKERAL, O. K., RASSENTI, L. Z., POWELL, J., BOTSTEIN, D., BYRD, J. C., GREVER, M. R., CHESON, B. D., CHIORAZZI, N., WILSON, W. H., KIPPS, T. J., BROWN, P. O. & STAUDT, L. M. 2001. Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia. *J Exp Med,* 194**,** 1639-47.

ROSS, P. L., HUANG, Y. N., MARCHESE, J. N., WILLIAMSON, B., PARKER, K., HATTAN, S., KHAINOVSKI, N., PILLAI, S., DEY, S., DANIELS, S.,

PURKAYASTHA, S., JUHASZ, P., MARTIN, S., BARTLET-JONES, M., HE, F., JACOBSON, A. & PAPPIN, D. J. 2004. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics,* 3**,** 1154-69.

SEIFERT, M., SELLMANN, L., BLOEHDORN, J., WEIN, F., STILGENBAUER, S., DURIG, J. & KUPPERS, R. 2012. Cellular origin and pathophysiology of chronic lymphocytic leukemia. *J Exp Med,* 209**,** 2183-98.

SELLERS, J. R. 2000. Myosins: a diverse superfamily. *Biochim Biophys Acta,* 1496**,** 3-22.

SHAW, G. R. & KRONBERGER, D. L. 2000. TP53 Deletions but not Trisomy 12 Are Adverse in B-Cell Lymphoproliferative Disorders. *Cancer Genetics and Cytogenetics,* 119**,** 146-154.

SHERBET, G. V. & CAJONE, F. 2005. Stathmin in Cell Proliferation and Cancer Progression. *Cancer Genomics - Proteomics,* 2**,** 227-237.

SRIYAM, S., SINCHAIKUL, S., TANTIPAIBOONWONG, P., TZAO, C., PHUTRAKUL, S. & CHEN, S. T. 2007. Enhanced detectability in proteome studies. *J Chromatogr B Analyt Technol Biomed Life Sci,* 849**,** 91-104.

STAROSTIK, P., MANSHOURI, T., O'BRIEN, S., FREIREICH, E., KANTARJIAN, H., HAIDAR, M., LERNER, S., KEATING, M. & ALBITAR, M. 1998. Deficiency of the ATM protein expression defines an aggressive subgroup of B-cell chronic lymphocytic leukemia. *Cancer Res,* 58**,** 4552-7.

STEEN, H. & MANN, M. 2004. The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol,* 5**,** 699-711.

STEVENSON, F. K. & CALIGARIS-CAPPIO, F. 2004. Chronic lymphocytic leukemia: revelations from the B-cell receptor. *Blood,* 103**,** 4389-4395.

STEVENSON, F. K., KRYSOV, S., DAVIES, A. J., STEELE, A. J. & PACKHAM, G. 2011. B-cell receptor signaling in chronic lymphocytic leukemia. *Blood,* 118**,** 4313-20.

STILGENBAUER, S., LIEBISCH, P., JAMES, M. R., SCHRODER, M., SCHLEGELBERGER, B., FISCHER, K., BENTZ, M., LICHTER, P. & DOHNER, H. 1996. Molecular cytogenetic delineation of a novel critical genomic region in chromosome bands 11q22.3-923.1 in lymphoproliferative disorders. *Proc Natl Acad Sci U S A,* 93**,** 11837-41.

THIEDE, B., HOHENWARTER, W., KRAH, A., MATTOW, J., SCHMID, M., SCHMIDT, F. & JUNGBLUT, P. R. 2005. Peptide mass fingerprinting. *Methods,* 35**,** 237-47.

THOMAS, A., EL ROUBY, S., REED, J. C., KRAJEWSKI, S., SILBER, R., POTMESIL, M. & NEWCOMB, E. W. 1996. Drug-induced apoptosis in B-cell chronic lymphocytic leukemia: relationship between p53 gene mutation and bcl-2/bax proteins in drug resistance. *Oncogene,* 12**,** 1055-62.

THOMPSON, A. A., TALLEY, J. A., DO, H. N., KAGAN, H. L., KUNKEL, L., BERENSON, J., COOPER, M. D., SAXON, A. & WALL, R. 1997. Aberrations of the B-cell receptor B29 (CD79b) gene in chronic lymphocytic leukemia. *Blood,* 90**,** 1387-94.

TSUKADA, N., BURGER, J. A., ZVAIFLER, N. J. & KIPPS, T. J. 2002. Distinctive features of ‚Äúnurselike‚Äù cells that differentiate in the context of chronic lymphocytic leukemia. *Blood,* 99**,** 1030-1037.

Twyman, R. (2004) Principles of Proteomics, Abingdon,: Bios Scientific.

VAISITTI, T., AYDIN, S., ROSSI, D., COTTINO, F., BERGUI, L., D'ARENA, G., BONELLO, L., HORENSTEIN, A. L., BRENNAN, P., PEPPER, C., GAIDANO, G., MALAVASI, F. & DEAGLIO, S. 2010. CD38 increases CXCL12-mediated signals and homing of chronic lymphocytic leukemia cells. *Leukemia,* 24**,** 958-69.

VAN BOCKSTAELE, F., VERHASSELT, B. & PHILIPPE, J. 2009. Prognostic markers in chronic lymphocytic leukemia: a comprehensive review. *Blood Rev,* 23**,** 25-47.

VOSS, T., AHORN, H., HABERL, P., DOHNER, H. & WILGENBUS, K. 2001. Correlation of clinical data with proteomics profiles in 24 patients with B-cell chronic lymphocytic leukemia. *Int J Cancer,* 91**,** 180-6.

VUILLIER, F., DUMAS, G., MAGNAC, C., PREVOST, M. C., LALANNE, A. I., OPPEZZO, P., MELANITOU, E., DIGHIERO, G. & PAYELLE-BROGARD, B. 2005. Lower levels of surface B-cell-receptor expression in chronic lymphocytic leukemia are associated with glycosylation and folding defects of the mu and CD79a chains. *Blood,* 105**,** 2933-40.

WATERHOUSE, D., CARMAN, W. J., SCHOTTENFELD, D., GRIDLEY, G. & MCLEAN, S. 1996. Cancer incidence in the rural community of Tecumseh, Michigan: a pattern of increased lymphopoietic neoplasms. *Cancer,* 77**,** 763-70.

WEISS, N. S. 1979. Geographical variation in the incidence of the leukemias and lymphomas. *Natl Cancer Inst Monogr***,** 139-42.

WELTON, J. L., KHANNA, S., GILES, P. J., BRENNAN, P., BREWIS, I. A., STAFFURTH, J., MASON, M. D. & CLAYTON, A. 2010. Proteomics analysis of bladder cancer exosomes. *Mol Cell Proteomics,* 9**,** 1324-38.

Westermeier R, Naven T, Hopker H (2008) Proteomics in Practice: A Guide to Successful Experiment Design, 2nd edition, Weinheim: Wiley-VCH.

WU, L. & HAN, D. K. 2006. Overcoming the dynamic range problem in mass spectrometry-based shotgun proteomics. *Expert Rev Proteomics,* 3**,** 611-9.

WYSOCKI, V. H., RESING, K. A., ZHANG, Q. & CHENG, G. 2005. Mass spectrometry of peptides and proteins. *Methods,* 35**,** 211-22.

YAN, W. & CHEN, S. S. 2005. Mass spectrometry-based quantitative proteomic profiling. *Brief Funct Genomic Proteomic,* 4**,** 27-38.

YATES, J. R., 3RD, GILCHRIST, A., HOWELL, K. E. & BERGERON, J. J. 2005. Proteomics of organelles and large cellular structures. *Nat Rev Mol Cell Biol,* 6**,** 702-14.

YATES, J. R., RUSE, C. I. & NAKORCHEVSKY, A. 2009. Proteomics by mass spectrometry: approaches, advances, and applications. *Annu Rev Biomed Eng,* 11**,** 49-79.

YUAN, J., MURRELL, G. A., TRICKETT, A., LANDTMETERS, M., KNOOPS, B. & WANG, M. X. 2004. Overexpression of antioxidant enzyme peroxiredoxin 5 protects human tendon cells against apoptosis and loss of cellular function during oxidative stress. *Biochim Biophys Acta,* 1693**,** 37-45.

YUILLE, M. R., MATUTES, E., MAROSSY, A., HILDITCH, B., CATOVSKY, D. & HOULSTON, R. S. 2000. Familial chronic lymphocytic leukaemia: a survey and review of published studies. *Br J Haematol,* 109**,** 794-9.

ZANINONI, A., IMPERIALI, F. G., PASQUINI, C., ZANELLA, A. & BARCELLINI, W. 2003. Cytokine modulation of nuclear factor-kappaB activity in B-chronic lymphocytic leukemia. *Exp Hematol,* 31**,** 185-90.

ZHANG, N., LI, N. & LI, L. 2004. Liquid chromatography MALDI MS/MS for membrane proteome analysis. *J Proteome Res,* 3**,** 719-27.

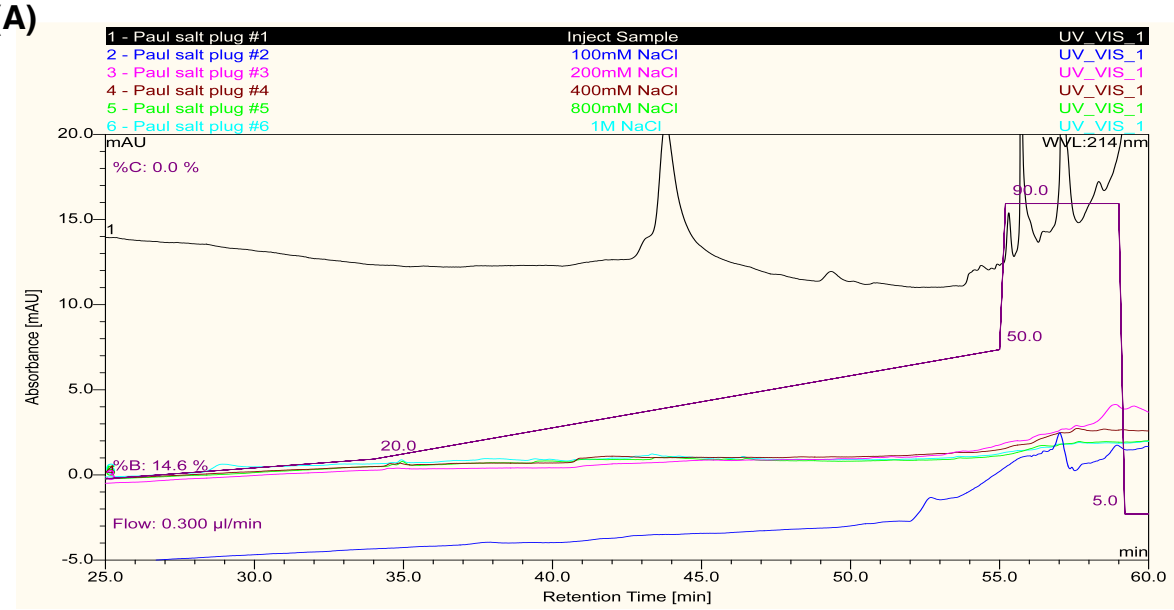
ZHAO, F. & WANG, Q. 2012. The protective effect of peroxiredoxin II on oxidative stress induced apoptosis in pancreatic beta-cells. *Cell Biosci,* 2**,** 22.

ZHU, W., SMITH, J. W. & HUANG, C. M. 2010. Mass spectrometry-based label-free quantitative proteomics. *J Biomed Biotechnol,* 2010**,** 840518.

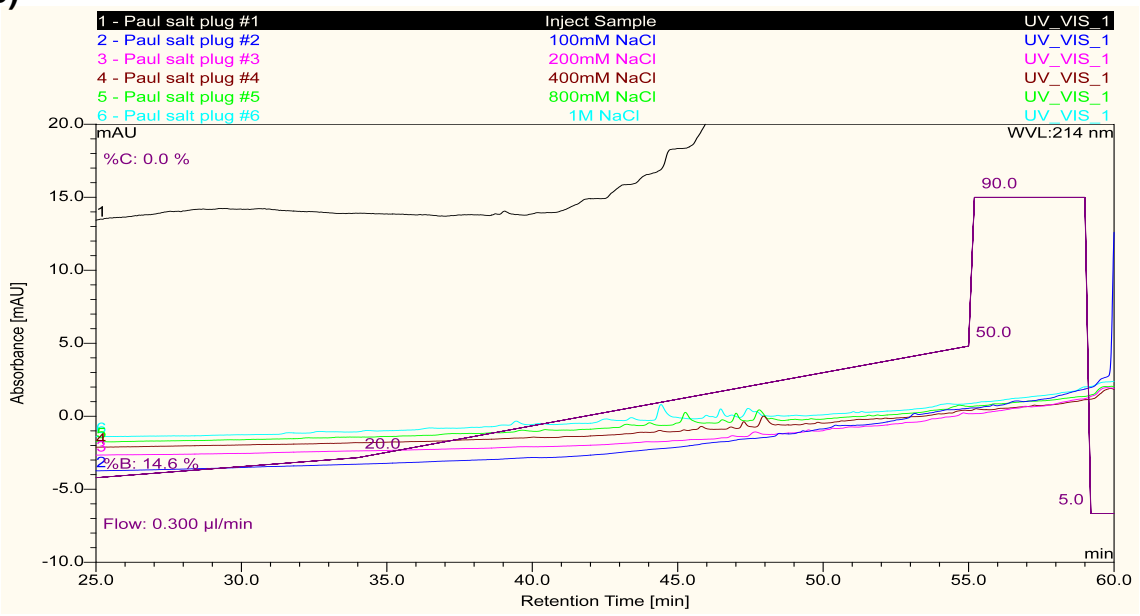
ZUPO, S., RUGARI, E., DONO, M., TABORELLI, G., MALAVASI, F. & FERRARINI, M. 1994. CD38 signaling by agonistic monoclonal antibody prevents apoptosis of human germinal center B cells. *Eur J Immunol,* 24**,** 1218-22.

Appendices

(A)



(B)



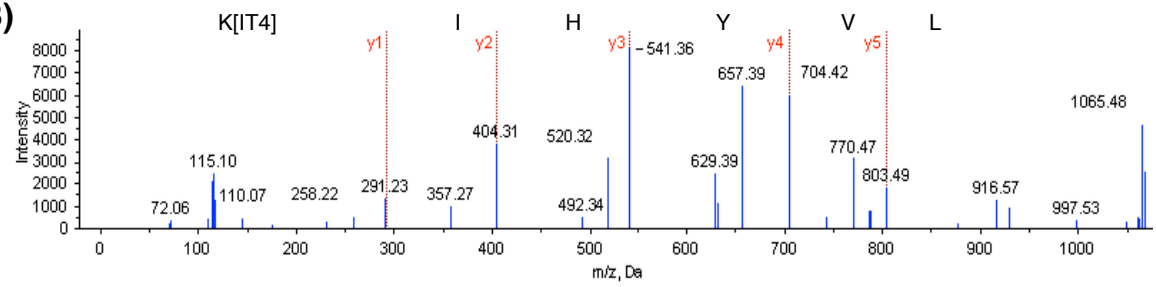
Appendix 1: Examples of peptides separation on the 2D nano-LC with poorly functioning columns. Peptides were separated on an SCX column using six salt fractions with an increasing salt concentration. Each salt fraction was then further separated on an RP column. This figure shows that the majority of the peptides did not bind to the SCX column as the breakthrough fraction (black line) was observed with very high absorbance signal. In addition, very few peaks were seen in the rest of the salt fractions especially in Appendix A.

(A)

T-cell leukemia/lymphoma protein 1A (TCL1A_HUMAN)

MAECPTLGEAVTDHPDR**LWAWEKFVYLDEK**QHAWLPLTIEIKDRLQLRVLLRREDVVLGRPMTPTQIGPSLL
PIMWQLYPDGRYRSSDSSFWR**LVYHIK**IDGVEDMLLELLPDD

(B)

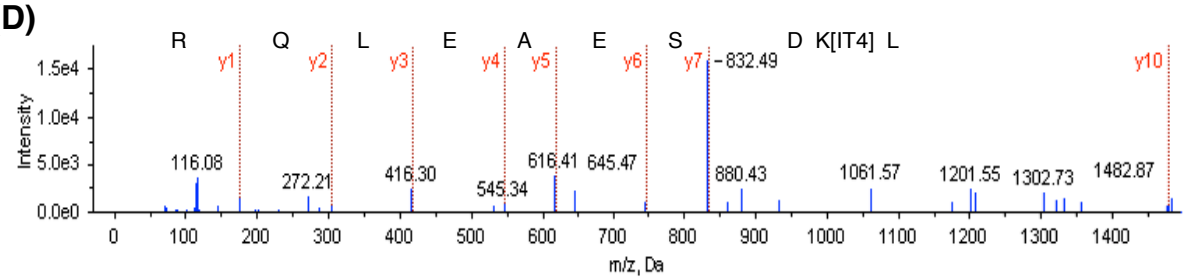


(C)

Septin-7 (SEPT7_HUMAN)

MSVSARSAAAEERSVNSSTMVAQQKNLEGYVGFANLPNQVYRKSVKRGFEFTLMVVGESGLGKSTLINSLFLT
DLYSPEYPGPSHRIKKTVQVEQSKVLIKEGGVQLLLTIVDTPGFGDAVDNSNCWQPVIDYIDSKFEDYLNAESRV
NRRQMPDNRVQCCLYFIAPSGHGLKPLDIEFMKRLHEKVNIIPLIAKADTLTPEECQQFKKQIMKEIQEHKIKIYEF
PETDDEEENKLVKKIKDRLPLAVVGSNTIIEVNGKRVRGRQYPWGVAEVENGEHCDFTILRNMLIRTHMQDLKDV
TNNVHYENYRSRKLAAVTYNGVDNNKNKGQLTKSPLAQMEEERREHVAKMKKMEMEMEQVFEMKVKEKVQK
LKDSEAELQRRHEQMKKNLEAQHKELEEKRRQFEDEKANWEAQQRILEQQNSSRTLEKNKKKGKIF

(D)



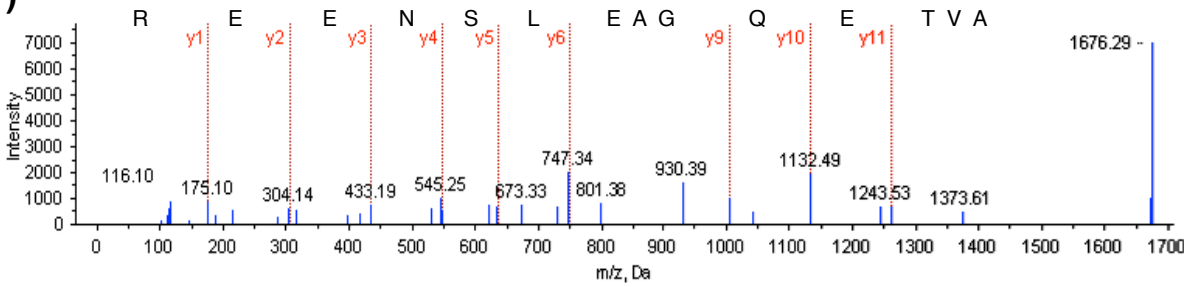
Appendix 4: Examples of peptides that were used to identify proteins with altered expression in CD38^{+} CLL samples compared to CD38^{-} CLL samples. This figure shows in "**Red Bold**" specific peptides and in "**Blue Bold**" common peptides that were used to identify these proteins (A, C, E, G, I and L). One MS/MS spectra is also shown for one specific peptide in each protein (B, D, F, H, J and L). For simplicity only y ions are shown in these MS/MS spectra. Peptides, of which MS/MS spectra are shown, are marked with a rectangle **P E P T I D E**. The sequence identified in each MS/MS spectra is read from the C-terminus as only y ions were shown in the MS/MS spectra. Underlining was used to discriminate between adjacent peptides that were identified.

(E)

14-3-3 protein theta (1433T_HUMAN)

MEKTELIQKAKLAEQAERYDDMATCMKAVTEQGAELSNEERNLLSVAYKNVVGGRRSAWRVISSIEQKTDTSDKK
LQLIKDYREKVESELRSICTTVLELLDKYLIANATNPESKVFYLKMKGDYFRYLAEVACGDDRKQTIDNSQGAYQEA
FDISKKEMQPTHPIRLGLALNFSVFYYEILNNPELACTLAKTAFDEAIAELDTLNEDSYKDSTLIMQLLRDNLTLWTSD
SAGEECDAAEGAEN

(F)

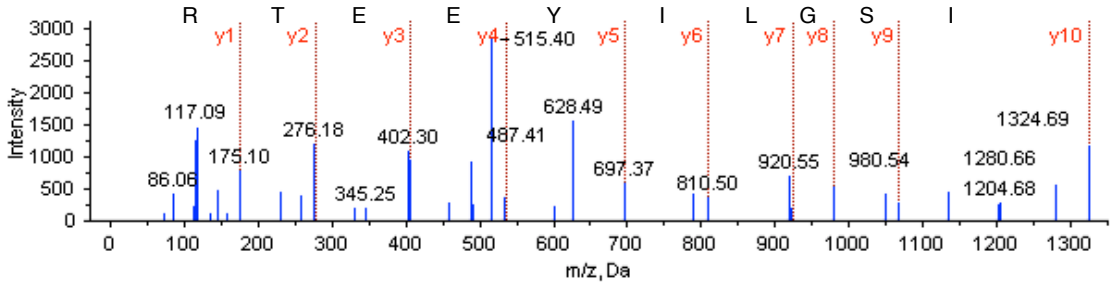


(G)

Histone H4 (H4_HUMAN)

MSGRGKGGKGLGKGGAKRHRKVLRDNIQGITKPAIRRLARRGGVKRISGLIYEETRGVLKVFLENVIRDAVTY
TEHAKRKTVTAMDVVYALKRQGRTLYGFGG

(H)

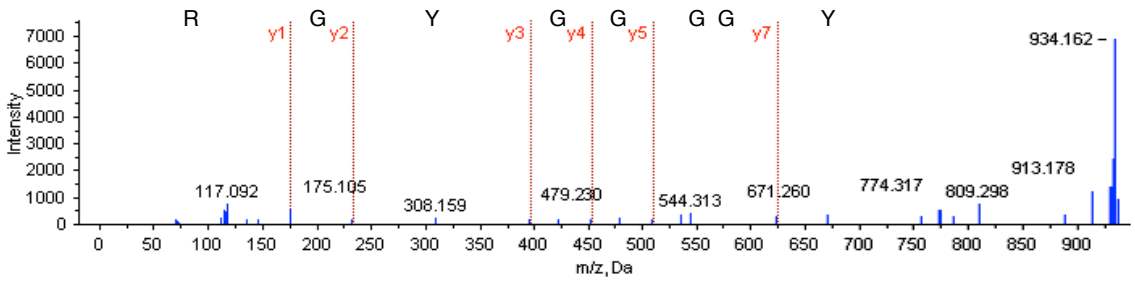


(I)

Splicing factor, arginine/serine-rich 2 (SFRS2_HUMAN)

MSYGRPPPDVEGMTSLKVDNLTYRTSPDTLRRVFEKYGRVGDVYIPRDRYTKESRGFAFVRFHDKRDAEDAM
DAMDGAVLDGRELRVQMARYGRPPDSHHSRRGPPPRRYGGGGYGRRSRSPRRRRRSRSRSRSRSRSRSRS
RYSRSKSRSRTRSRSRSTSKSRSARRSKSKSSSVSRSRSRSRSRSRSRSPPPVSKRESKSRSRSKSPPKSPEE
EGAVSS

(J)



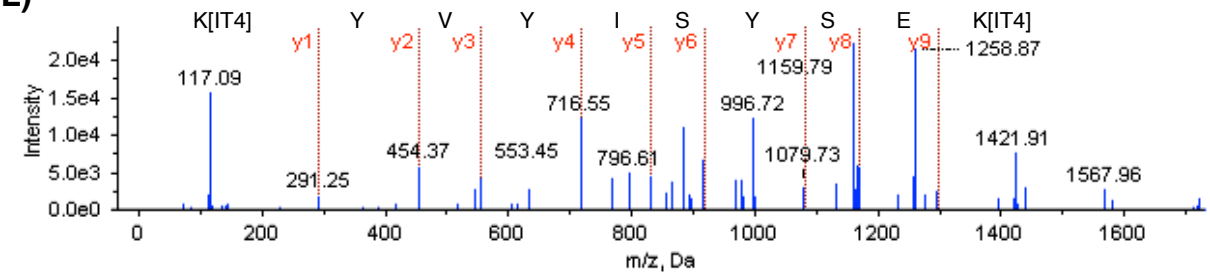
Appendix 4 continued: Examples of peptides that were used to identify proteins with altered expression in CD38⁺ CLL samples compared to CD38⁻ CLL samples.

(K)

Histone H2B type 2-E (H2B2E_HUMAN)

MPEPAKSAPAPKKGSKKAVTKAQKKDGKKRKRSR**KESYSIYVYK**VLKQVHPDTGISSKAMGIMNSFVNDIFERIAGEAS
RLAHYNKRSTITSREIQTAVRLLLPGELAKHAVSEGTKAVT KYTSSK

(L)

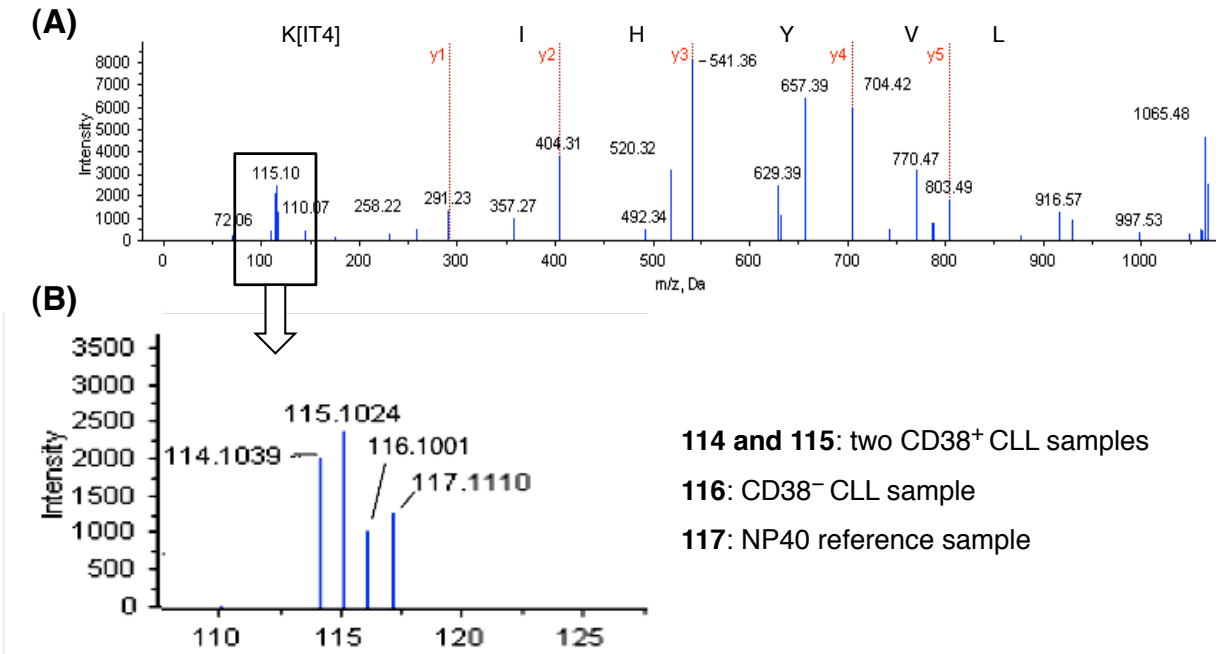


Appendix 4 continued: Examples of peptides that were used to identify proteins with altered expression in CD38^{+} CLL samples compared to CD38^{-} CLL samples.

Appendix 5 : Examples of the relative quantification of peptides that were used to measure the relative expression of TCL-1 in the NP40 fractions of CLL samples

Peptide sequence	Confidence %	$CD38^+$ CLL	$CD38^+$ CLL	$CD38^-$ CLL
FVYLDEK	99	1.55	1.74	1.13
FVYLDEK	99	1.72	2.27	0.84
LWAWEK	99	2.45	2.70	1.13
LVYHIK	99	2.17	2.66	0.89
LVYHIK	98	1.93	2.64	1.04
LWAWEK	97	3.04	3.77	0.93
Protein TCL-1	**Identified with 3 different peptides**	**2.26**	**2.80**	**0.97**

Identification and relative quantification of labelled peptides were performed by MALDI mass spectrometry and ProteinPilot using the Paragon and ProGroup algorithms. This table shows the relative quantification of peptides that were mapped to TCL-1. The relative quantification of these peptides was used to determine the relative expression of TCL-1 in CLL samples. The examples shown in this table were extracted from one iTRAQ LC-MALDI mass spectrometry experiment (NP40 fraction) containing three CLL samples with a reference sample.

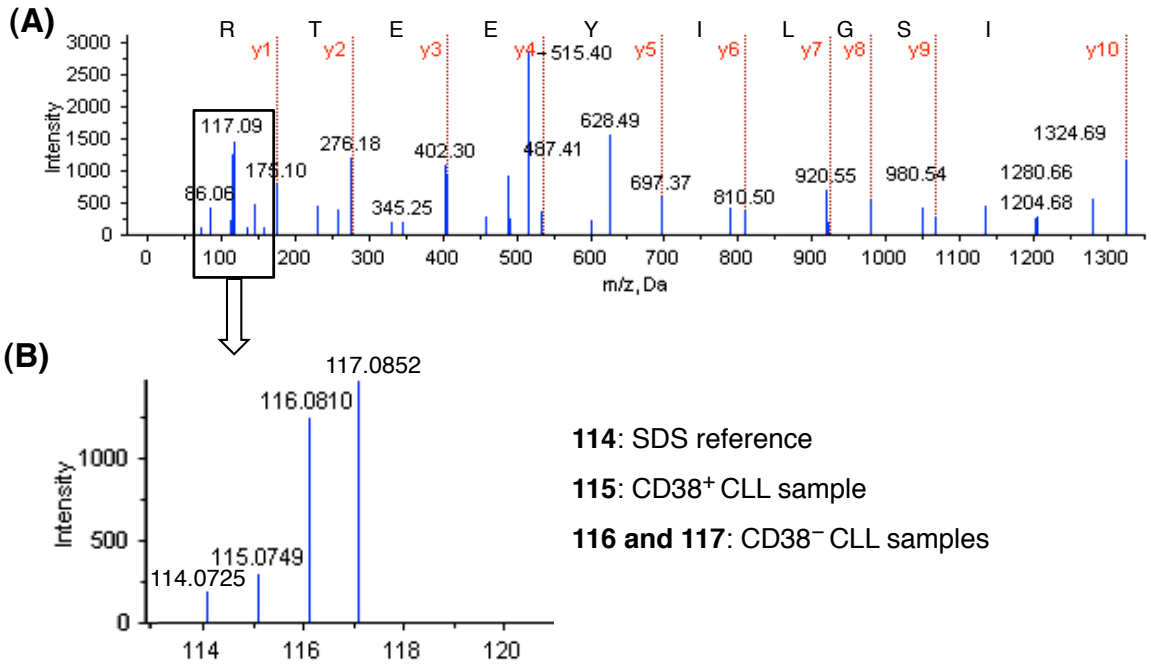


Appendix 6: Identification and relative quantification of a specific peptide in TCL-1. Following separation of labelled peptides by 2D nano-LC, they were analysed by MALDI mass spectrometry and ProteinPilot software using the Paragon and ProGroup algorithms. This figure shows MS/MS spectra of a specific peptide in TCL-1 (**LVYHIK**) (A) as well as the relative quantification of this peptide in the NP40 fractions of three CLL samples in relation to the NP40 reference sample (B). For simplicity only y ions were shown in these MS/MS spectra. The sequence identified in the MS/MS spectra is read from the C-terminus.

Appendix 7: Examples of the relative quantification of peptides that were used to measure the relative expression of Histone H4 in the SDS fractions of CLL samples

Peptide sequence	Confidence %	$CD38^+$ CLL	$CD38^-$ CLL	$CD38^-$ CLL
DAVTYTEHAK	99	0.53	1.21	1.24
DAVTYTEHAK	99	0.87	1.64	1.64
DAVTYTEHAK	99	0.83	1.66	1.58
DAVTYTEHAK	99	1.05	1.62	1.60
DNIQGITKPAIR	99	0.78	1.41	1.29
DNIQGITKPAIR	99	0.78	1.56	1.43
DNIQGITKPAIR	99	0.85	1.43	1.35
DNIQGITKPAIR	99	0.83	1.58	1.47
ISGLIYEETR	99	0.49	1.52	1.47
ISGLIYEETR	99	0.49	1.33	1.41
ISGLIYEETR	99	0.58	1.49	1.51
ISGLIYEETR	99	0.52	1.32	1.14
ISGLIYEETR	99	0.55	1.36	1.40
ISGLIYEETR	99	0.97	1.75	1.81
ISGLIYEETR	99	0.79	0.82	0.77
TVTAMDVVYALK	99	1.11	1.23	1.46
TVTAMDVVYALK	99	0.94	1.20	1.28
TVTAMDVVYALK	99	0.89	1.14	1.29
VFLENVIR	99	0.54	1.09	1.10
VFLENVIR	99	0.70	1.29	1.25
VFLENVIR	99	0.71	1.42	1.35
VFLENVIR	99	0.92	1.72	1.76
VFLENVIR	99	0.76	1.44	1.40
VFLENVIR	99	0.70	1.50	1.22
VFLENVIR	99	0.74	1.50	1.44
RISGLIYEETR	95	0.71	1.16	0.86
Protein Histone H4	**Identified with 17 different peptides**	**0.72**	**1.43**	**1.39**

Identification and relative quantification of labelled peptides were performed by MALDI mass spectrometry and the ProteinPilot using Paragon and ProGroup algorithms. This table shows realtive quantification of peptides that were mapped to Histone H4 .The relative quantification of these peptides were used to determine the relative expression of Histone H4 in CLL samples. The examples shown in this figure were extracted from one iTRAQ LC-MALDI mass spectrometry experiment (SDS fraction) containing three CLL samples with a reference sample.



Appendix 8: Identification and the relative quantification of a specific peptide in Histone H4. Following separation of labelled peptides by 2D nano-LC, they were analysed by MALDI mass spectrometry and the ProteinPilot software using Paragon and ProGroup algorithms. This figure shows MS/MS spectra of a specific peptide in TCL-1 (**ISGLIYEETR**) (A) as well as the relative quantification of this peptide in the SDS fractions of three CLL samples in relation to the SDS reference sample (B). For simplicity only y ions were shown in these MS/MS spectra. The sequence identified in the MS/MS spectra is read from the C-terminus.

Appendix 9: Relative quantification of some altered proteins in CD38$^+$ CLL samples compared to CD38$^-$ CLL samples.

		CD38$^+$ CLL samples								CD38$^-$ CLL samples								
Protein name	**Fraction**	**1**	**2**	**3**	**4**	**5**	**6**	**Mean**	**SD**	**1**	**2**	**3**	**4**	**5**	**6**	**Mean**	**SD**	***p* value**
TCL-1	NP40	2.58	0.90	1.79	1.60	2.26	2.80	1.99	0.70	0.97	1.19	1.10	1.38	1.36	N/A	1.20	0.17	0.04
Histone H4	SDS	0.92	0.67	0.99	0.72	1.32	0.92	0.92	0.23	0.80	1.27	1.45	1.43	1.39	1.09	1.24	0.25	0.05

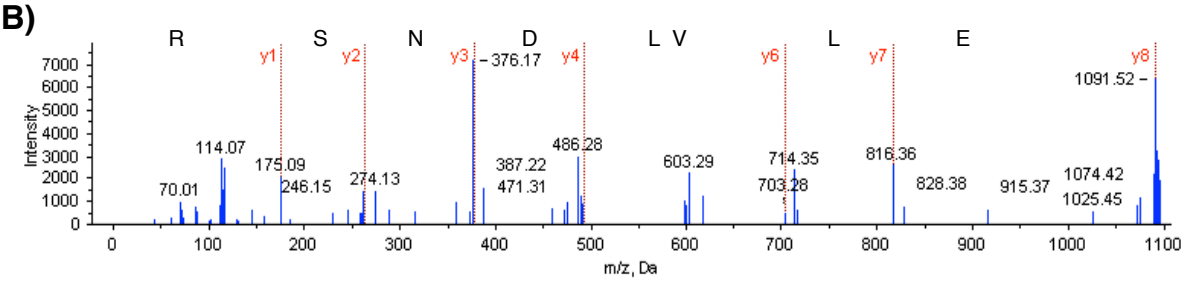
Following relative quantification of proteins in the NP40 fractions and SDS fractions of 12 CLL samples, protein relative quantifications were used to indicate proteins with altered expression in CLL samples (CD38$^+$ CLL versus CD38$^-$ CLL). This table shows examples of proteins that were observed with significantly altered expression in CLL samples with high or low expression of CD38.

(A)

Acidic leucine-rich nuclear phosphoprotein 32 family member A (AN32A_HUMAN)

MEMGRR**IHLELR**NRTPSDVK**ELVLDNSR**SNEGKLEGLTDEFEELEFLSTINVGLTSIANLPKLNKLKKLELSDNR
VSGGLEVLAEKCPNLTHLNLSGNKIKDLSTIEPLKKLENLKSLDLFNCEVTNLNDYRENVFKLLPQLTYLDGYDR
DDKEAPDSDAEGYVEGLDDEEEDEDEEEYDEDAQVVEDEEDEDEEEEGEEEDVSGEEEEDEEGYNDGEVDD
EEDEEELGEEERGQKRKREPEDEGEDDD

(B)

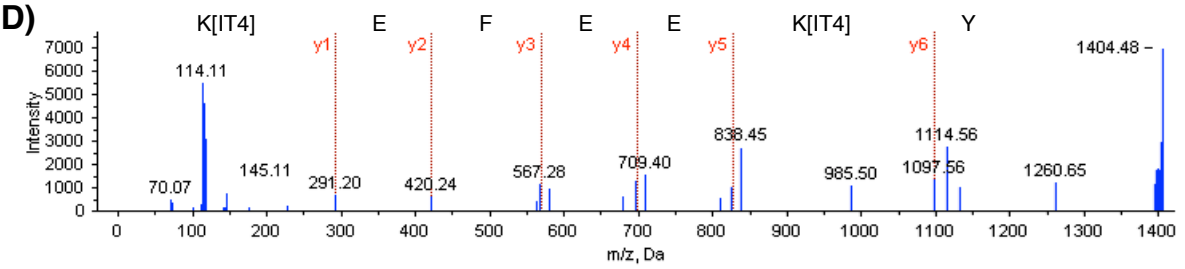


(C)

LIM and SH3 domain protein 1 (LASP1_HUMAN)

MNPNCARCGKIVYPTEKVNCLDKFWHKACFHCETCKMTLNMKNYKGYEKKPYCNAHYPKQSFTMVADTPEN
LRLKQQSELQSQVR**YKEEFEK**NKGK**GFSVVADTPELQR**IKKTQDQISNIK**YHEEFEK**SRMGPSGGEGMEPER
RDSQDGSSYRRPLEQQQPHHIPTSAPVYQQPQQQPVAQSYGGYKEPAAPVSIQRSAPGGGGKRYRAVYDYS
AADEDEVSFQDGDTIVNVQQIDDGWMYGTVERTGDTGMLPANYVEAI

(D)

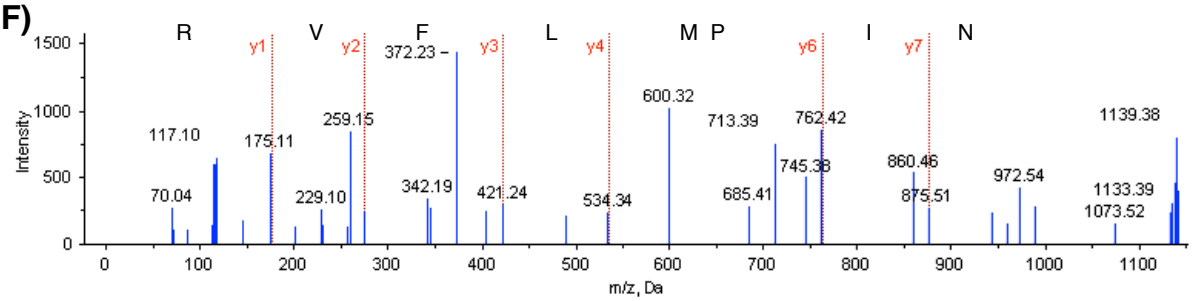


(E)

U6 snRNA-associated Sm-like protein LSm3 (LSM3_HUMAN)

MADDVDQQQTTNTVEEPLDLIRLSLDERIYVKMRNDRELRGRLHAYDQHLNMILGDVEETVTTIEIDEETYE
EIYKSTKR**NIPMLFVR**GDGVVLVAPPLRVG

(F)



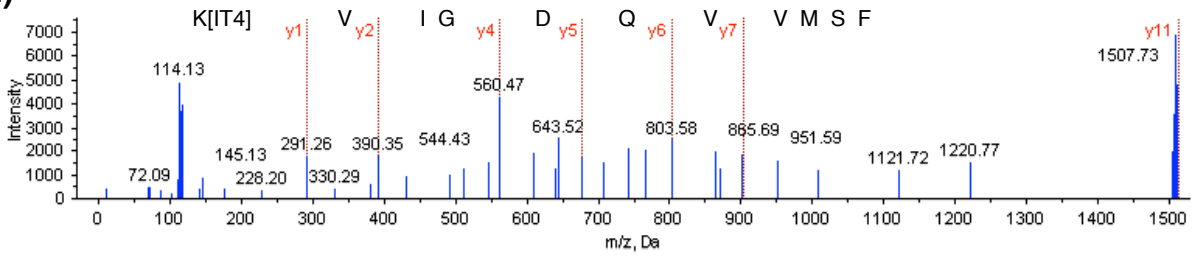
Appendix 10: Examples of peptides that were used to identify proteins with altered expression in ZAP-70^{+} CLL samples compared to ZAP-70^{-} CLL samples. This figure shows in "**Red Bold**" specific peptides and in "**Blue Bold**" common peptides that were used to identify these proteins (A, C, E, G, I and K). One MS/MS spectra is shown for one specific peptide in each protein (B, D, F, H, J and L). For simplicity only y ions were shown in these MS/MS spectra. Peptides, of which MS/MS spectra are shown, were marked with a rectangle **P E P T I D E**. The sequence identified in each MS/MS spectra is read from the C-terminus as only y ions were shown in the MS/MS spectra. Underlining was used to discriminate between adjacent peptides that were identified.

(G)

Peroxiredoxin-5, mitochondrial (PRDX5_HUMAN)

MGLAGVCALRRSAGYILVGGAGGQSAAAAARRYSEGEWASGGVRSFSRAAAAMAPIKVGDAIPAVEVFEGEPGNKVN
LAELFKGKKGVLFGVPGAFTPGCSKTHLPGFVEQAEALKAKGVQVVACLSVNDAFVTGEWGRAHKAEGKVRLLADPT
GAFGKETDLLLDDSLVSIFGNRRLKRFSMVVQDGIVKALNVEPDGTGLTCSLAPNIISQL

(H)

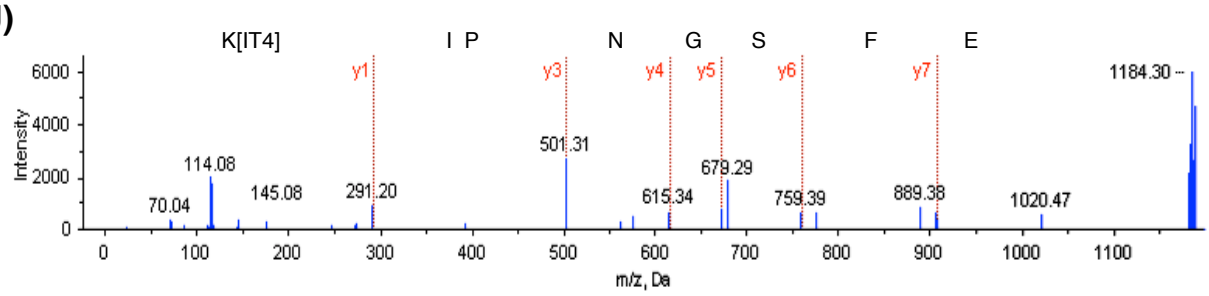


(I)

RNA-binding protein FUS (FUS_HUMAN)

MASNDYTQQATQSYGAYPTQPGQGYSQQSSQPYGQQSYSGYSQSTDTSGYGQSSYSSYGQSQNTGYGTQSTPQGYG
STGGYGSSQSSQSSYGQQSSYPGYGQQPAPSSTSGSYGSSSQSSSYGQPQSGSYSQQPSYGGQQQSYGQQQSYNP
PQGYGQQNQYNSSSGGGGGGGGGGNYGQDQSSMSSGGGSGGGYGNQDQSGGGGSGGYGQQDRGGRGRGGSGG
GGGGGGGGYNRSSGGYEPRGRGGGRGGRGGMGGSDRGGFNKFGGPRDQGSRHDSEQDNSDNNTIFVQGLGENVTI
ESVADYFKQIGIIKTNKKTGQPMINLYTDRETGKLKGEATVSFDDPPSAKAAIDWFDGKEFSGNPIKVSFATRRADFNRGG
GNGRGGRGRGGPMGRGGYGGGGSGGGGRGGFPSGGGGGGGQQRAGDWKCPNPTCENMNFSWRNECNQCKAPK
PDGPGGGPGGSHMGGNYGDDRRGGRGGYDRGGYRGRGGDRGGFRGGRGGGDRGGFGPGKMDSRGEHRQDRRE
RPY

(J)

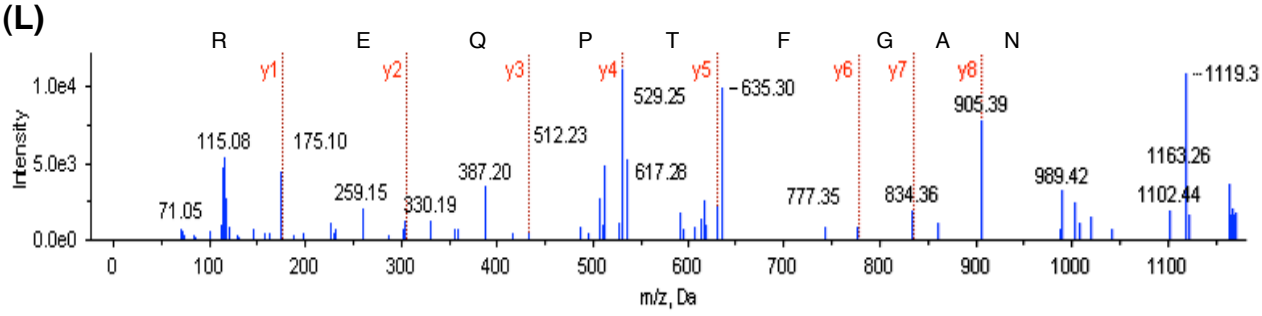


(K)

DNA-(apurinic or apyrimidinic site) lyase (APEX1_HUMAN)

MPKRGKKGAVAEDGDELRTEPEAKKSKTAAKKNDKEAAGEGPALYEDPPDQKTSPSGKPATLKICSWNVDGLRAW
IKKKGLDWVKEEAPDILCLQETKCSENKLPAELQELPGLSHQYWSAPSDKEGYSGVGLLSRQCPLKVSYGIGDEEH
DQEGRVIVAEFDSFVLVTAYVPNAGRGLVRLEYRQRWDEAFRKFLKGLASRKPLVLCGDLNVAHEEIDLRNPKGNKK
NAGFTPQERQGFGELLQAVPLADSFRHLYPNTPYAYTFWTYMMNARSKNVGWRLDYFLLSHSLLPALCDSKIRSK
ALGSDHCPITLYLAL

(L)



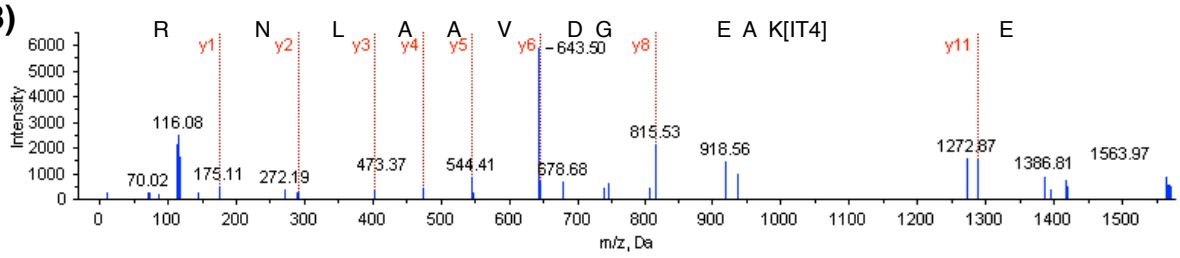
Appendix 10 continued: Examples of peptides that were used to identify proteins with altered expression in ZAP-70^{+} CLL samples compared to ZAP-70^{-} CLL samples.

(A)

Tropomyosin alpha-4 chain (TPM4_HUMAN)

MAGLNSLEAVKR**KIQALQQQADEAEDR**AQGLQRELDGERERR**EKAEGDVAALNR**R**IQLVEEELDR**AQER**LATALQ**
KLEEAEKAADESERGMKVIENRAMKDEEKMEIQMQLKEAKHIAEEADRKYEEVARKLVILEGELERAEERAEVSELKC
GDLEEELKNVTNNLKSLEAASEKYSEK**EDKYEEEIK**LLSDKLKEAETRAEFAERTVAKLEKTIDDLEEKLAQAKEENVG
LHQTLDQTLNELNCI

(B)

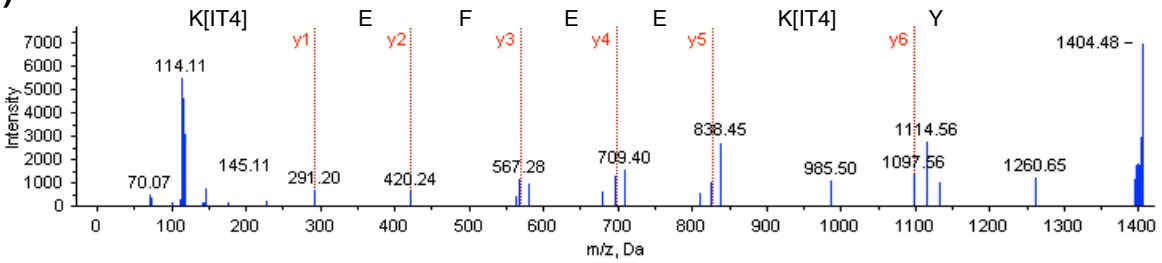


(C)

LIM and SH3 domain protein 1 (LASP1_HUMAN)

MNPNCARCGKIVYPTEKVNCLDKFWHKACFHCETCKMTLNMKNYKGYEKKPYCNAHYPKQSFTMVADTPENLRLK
QQSELQSQVR**YKEEFEK**NKGK**GFSVVADTPELQR**IKKTQDQISNIK**YHEEFEK**SRMGPSGGEGMEPERRDSQDGS
SYRRPLEQQQPHHIPTSAPVYQQPQQQPVAQSYGGYKEPAAPVSIQRSAPGGGGKRYRAVYDYSAADEDEVSFQD
GDTIVNVQQIDDGWMYGTVERTGDTGMLPANYVEAI

(D)



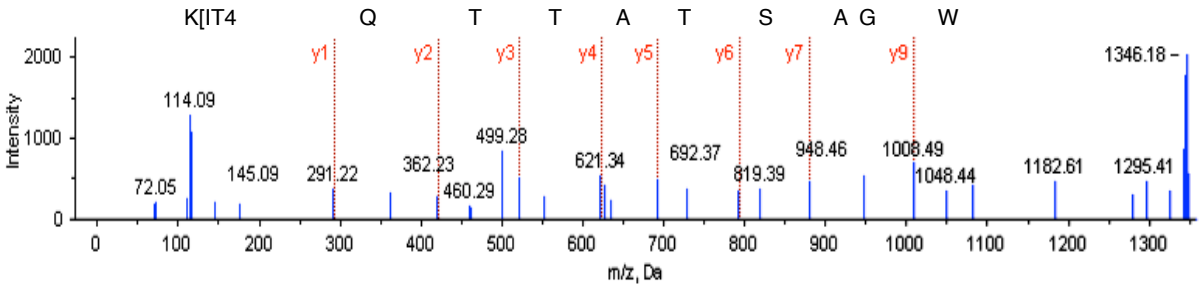
Appendix 11: Examples of peptides that were used to identify proteins with altered expression in UM-CLL samples compared to M-CLL samples. This figure shows in "**Red Bold**" specific peptides and in "**Blue Bold**" common peptides that were used to identify these proteins (A, C and E). One MS/MS spectra is shown for one specific peptide in each protein (B, D and F). For simplicity only y ions were shown in these MS/MS spectra. Peptides, of which MS/MS spectra are shown, are marked with a rectangle **P E P T I D E**. The sequence identified in each MS/MS spectra is read from the C-terminus as only y ions were shown in the MS/MS spectra.

(E)

Apoptotic chromatin condensation inducer in the nucleus (ACINU_HUMAN)

MWRRKHPRTSGGTRGVLSGNRGVEYGSGRGHLGTFEGRWRKLPKMPEAVGTDPSTSRKMAELEEVTLDGKPLQALRV
TDLKAALEQRGLAKSGQKSALVKRLKGALMLENLQKHSTPHAAFQPNSQIGEEMSQNSFIKQYLEKQQELLRQRLEREAR
EAAELEEASAESEDEMIHPEGVASLLPPDFQSSLERPELELSRHSPRKSSSISEEKGDSDDEKPRKGERRSSRVRQARAA
KLSEGSQPAEEEEDQETPSRNLRVRADRNLKTEEEEEEEEEEEEDDEEEEGDDEGQKSREAPILK**EFKEEGEEIPR**VKPE
EMMDERPKTRSQEQEVLERGGRFTRSQEEARKSHLARQQQEKEMKTTSPLEEEEREIKSSQGLKEKSKSPSPPRLTEDR
KKASLVALPEQTASEEETPPPLLTKEASSPPPHPQLHSEEEIEPMEGPAPAVLIQLSPPNTDADTRELLVSQHTVQLVGGLSP
LSSPSDTKAESPAEKVPEESVLPLVQKSTLADYSAQKDLEPESDRSAQPLPLKIEELALAKGITEECLKQPSLEQKEGRRAS
HTLLPSHRLKQSADSSSSRSSSSSSSSSRSRSRSPDSSGSRSHSPLRSKQRDVAQARTHANPRGRPKMGSRSTSESRS
RSRSRSRSASSNSRKSLSPGVSRDSSTSYTETKDPSSGQEVATPPVPQLQVCEPKERTSTSSSSVQARRLSQPESAEKH
VTQRLQPERGSPKKCEAEEAEPPAATQPQTSETQTSHLPESERIHHTVEEKEEVTMDTSENRPENDVPEPPMPIADQVSN
DDRPEGSVEDEEKKESSLPKSFKRKISVVSATKGVPAGNSDTGGQPGRKRR**WGASTATTQK**KPSISITTESLKSLIPDIKPL
AGQEAVVDLHADDSRISEDETERNGDDGTHDKGLKICRTVTQVVPAEGQENGQREEEEEEKEPEAEPPVPPQVSVEVAL
PPPAEHEVKKVTLGDTLTRRSISQQK**SGVSITIDDPVR**TAQVPSPPRGKISNIVHISNLVRPFTLGQLKELLGRTGTLVEEAF
WIDKIKSHCFVTYSTVEEAVATRTALHGVKWPQSNPKFLCADYAEQDELDYHRGLLVDRPSETKTEEQGIPRPLHPPPPPP
VQPPQHPRAEQREQERAVREQWAEREREMERRERTRSEREWDRDKVREGPRSRSRSRDRRRKERAKSKEKKSEKKE
KAQEEPPAKLLDDLFRKTKAAPCIYWLPLTDSQIVQKEAERAERAKEREKRRKEQEEEEQKEREKEAERERNRQLEREKR
REHSRERDRERERERERDRGDRDRDRERDRERGRERDRRDTKRHSRSRSRSTPVRDRGGRR

(F)



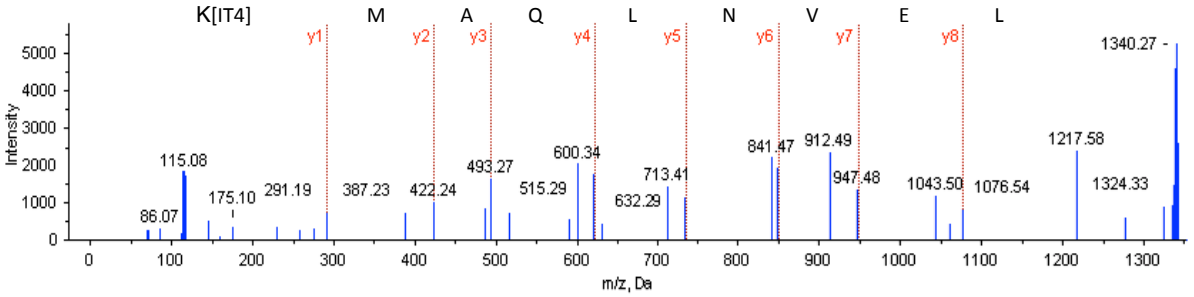
Appendix 11 continued: Examples of peptides that were used to identify proteins with altered expression in UM-CLL samples compared to M-CLL samples.

(A)

Myosin-9 (MYH9_HUMAN)

MAQQAADKYLYVDKNFINNPLAQADWAAKKLVWVPSDKSGFEPASLKEEVGEEAIVELVENGKKVK**VNKDDIQK**MNPP
KFSKVEDMAELTCLNEASVLHNLKERYYSGLIYTYSGLFCVVINPYKNLPIYSEEIVEMYKGKKRHEMPPHIYAITDTAYR
SMMQDREDQSILCTGESGAGKTENTKKVIQYLAYVASSHKSKKDQGELERQLLQANPILEAFGNAKTVKNDNSSRFGK
FIRINFDVNGYIVGANIETYLLEKSRAIRQAKEERTFHIFYYLLSGAGEHLKTDLLLEPYNKYRFLSNGHVTIPGQQDKDM
FQETMEAMRIMGIPEEEQMGLLRVISGVLQLGNIVFKKER**NTDQASMPDNTAAQK**VSHLLGINVTDFTRGILTPRIKVG
RDYVQKAQTKEQADFAIEALAKATYERMFRWLVLRINKALDKTKRQGASFIGILDIAGFEIFDLNSFEQLCINYTNEKLQQ
LFNHTMFILEQEEYQREGIEWNFIDFGLDLQPCIDLIEKPAGPPGILALLDEECWFPKATDKSFVEK**VMQEQGTHPK**FQK
PKQLKDKADFCIIHYAGKVDYKADEWLMKNMDPLNDNIATLLHQSSDK**FVSELWK**DVDRIIGLDQVAGMSETALPGAFK
TRKGMFR**TVGQLYK**EQLAKLMATLRNTNPNFVRCIIPNHEKKAGKLDPHLVLDQLRCNGVLEGIRICRQGFPNR**VVFQE**
FRQRYEILTPNSIPKGFMDGKQACVLMIKALELDSNLYRIGQSKVFFRAGVLAHLEEERDLKITDVIIGFQACCRGYLARK
AFAKR**QQQLTAMK**VLQRNCAAYLKLRNWQWWRLFTKVKPLLQVSRQEEEMMAKEEELVKVREKQLAAENRLTEMET
LQSQLMAEKLQLQEQLQAETELCAEAEELRARLTAKKQELEEICHDLEARVEEEEER**CQHLQAEK**KKMQQNIQELEEQ
LEEEESARQKLQLEKVTTEAKLKKLEEEQIILEDQNCKLAKEKKLLEDRIAEFTTNLTEEEEKSKSLAKLKNKHEAMITDL
EERLRREEKQRQELEKTRRKLEGDSTDLSDQIAELQAQIAELKMQLAK**KEEELQAALAR**VEEEAAQKNMALKKIRELE
SQISELQEDLESERASRNKAEKQKRDLGEELEALKTELEDTLDSTAAQQELRSKREQEVNILKKTLEEEAKTHEAQIQE
MRQKHSQAVEELAEQLEQTKRVKANLEKAKQTLENERGELANEVKVLLQGKGDSEHKRKKVEAQLQELQVKFNEGE
RVRTELADKVTKLQVELDNVTGLLSQSDSKSSKLTKDFSALESQLQDTQELLQEENRQKLSLSTKLKQVEDEKNSFRE
QLEEEEEAKHNLEKQIATLHAQVADMKKKMEDSVGCLETAEEVKRKLQKDLEGLSQRHEEK**VAAYDKLEK**TKTRLQQE
LDDLLVDLDHQRQSACNLEKKQKKFDQLLAEEKTISAKYAEERDRAEAEAREKETKALSLARALEEAMEQKAELERLNK
QFRTEMEDLMSSKDDVGKSVHELEKSKR**ALEQQVEEMK**TQLEELEDELQATEDAKLR**LEVNLQAMK**AQFERDLQGR
DEQSEEKKKQLVRQVREMEAELEDERKQRSMAVAARKKLEMDLKDLEAHIDSANKNRDEAIKQLRKLQAQMKDCMRE
LDDTR**ASREEILAQAK**ENEKKLKSMEAEMIQLQEELAAAERAKRQAQQERDELADEIANSSGK**GALALEEK**RRLEARI
AQLEEELEEEQGNTELINDRLKKANLQIDQINTDLNLERSHAQKNENARQQLERQNKELKVK**LQEMEGTVK**SKYKASIT
ALEAKIAQLEEQLDNETKERQAACKQVRRTEKKLKDVLLQVDDERR**NAEQYKDQADK**ASTRLKQLKR**QLEEAEEEAQ**
RANASRRKLQRELEDATETADAMNREVSSLKNKLRRGDLPFVVPRRMARKGAGDGSDEEVDGKADGAEAKPAE

(B)



Appendix 12: Examples of peptides that were used to identify myosin 9 that was altered in stages B/C CLL samples compared to stage A CLL samples. This figure shows in "**Red Bold**" specific peptides and in "**Blue Bold**" common peptides that were used to identify myosin-9 (A). One MSMS spectra is shown for one specific peptide in myosin-9 (B). For simplicity only y ions were shown in the MS/MS spectra. Peptide, of which MS/MS spectra is shown, is marked with a rectangle **P E P T I D E**. The sequence identified in the MS/MS spectra is read from the C-terminus as only y ions were shown in the MS/MS spectra.

Appendix 13: Clinical details of the CLL patients cohort.

IVGH <98% mutated CLL (Good prognosis); IGVH ≥98% unmutated CLL (poor prognosis); CD38 <20% (good prognosis); CD38 ≥20% (poor prognosis); ZAP-70 <20% (good prognosis); ZAP-70 ≥20%(poor prognosis); B and C are advanced stages of CLL; A0 and A are early stage of the disease; TTFT: time of the first treatment; TSD: time since diagnosis with CLL. The first 12 samples were the samples used for quantitative proteomic analysis.

Samples	Age	% CD38	% ZAP-70	*%IGVH* status	Stage	TTFT (days)	TSD (days)
1		90	23	100	C	1245	1459
2	62	75	2		A0	932	1212
3	87	3	5	90	A0		4194
4	53	39	2	96	B		2028
5	75	85	0	96	A	281	2150
6	80	1	4	98	A0		1155
7	73	67	0	89	A		4764
8	66	75	35	100			1001
9		4	3		A		30
10		2	39	99	C	77	212
11		2					
12	60	2	19		A0		1568
13	76	3	5	92	A		2478
14	63	4	1	90	A0		1722
15	62	5	10	94	A	1064	4050
16		5			A		
17	68	5	10		A0		1030
18	75	6	0		A0		5674
19		6	35	91	B	1150	2733
20	62	6	6	93	A	1427	1484
21	87	9	54		C	58	1321
22	65	10	8	92	A	1430	1563
23	76	34	1	100	B	21	3688
24	85	39	3	91	A0	2010	2740
25	60	41	51	99	B	247	3438
26	78	52	2	91	A0	190	2737
27		70	18		A	1944	2577
28	74	80	36	84	A		6160
29	57	80	44		A	71	1534
30		83	29		A	176	330
31	75	90	20	95	C		7547
32	80	92	2	94	A0		2605
33	67	93	12	93	A	2568	2568
34	85	97	4	95	A	1524	1734
35	77	1	24	96	A0	3232	3303
36	55	3	28		A0		1282
37	70	3	5	95	A	708	3725
38	70	3	2	96	A	389	3287