Leukemia (2003) 17, 1605–1612

www.nature.com/leu

Proteomic analysis of the cell-surface membrane in chronic lymphocytic leukemia: identification of two novel proteins, BCNP1 and MIG2B

RS Boyd[1,5], PJ Adam[1,5], S Patel[1], JA Loader[1], J Berry[1], NT Redpath[1], HR Poyser[1], GC Fletcher[1], NA Burgess[1], AC Stamps[1], L Hudson[1], P Smith[1], M Griffiths[1], TG Willis[2], EL Karran[2,3], DG Oscier[4], D Catovsky[2], JA Terrett[1] and MJS Dyer[2,3]

[1]*Oxford Glycosciences (UK) Ltd, Abingdon, Oxford, UK;* [2]*Academic Haematology, Institute of Cancer Research, Sutton, Surrey, UK;* [3]*Department of Haematology, MRC Toxicology Unit/University of Leicester, UK; and* [4]*Department of Haematology, Royal Bournemouth Hospital, UK*

B-cell-specific plasma-membrane proteins are potential targets for either small molecule or antibody-based therapies. We have sought to annotate proteins expressed at the cell surface membrane in patients with chronic lymphocytic leukemia (CLL) using plasma-membrane-based proteomic analysis to identify previously uncharacterized and potentially B-cell-specific proteins. Proteins from plasma-membrane fractions were separated on one-dimensional gels and trypsinized fractions subjected to high-throughput MALDI-TOF mass spectrometry. Using this method, many known B-cell surface antigens were detected, but also known proteins not previously described in this disease or in this cellular compartment, including cell surface receptors, membrane-associated enzymes and secreted proteins, and completely unknown proteins. To validate the method, we show that BLK, a B-cell-specific kinase, is located in the CLL-plasma-membrane fraction. We also describe two novel proteins (MIG2B and B-cell novel protein #1, BCNP1), which are expressed preferentially in B cells. MIG2B is in a highly conserved and defined gene family containing two plasma-membrane-binding ezrin/radixin/moesin domains and a pleckstrin homology domain; the *Caenorhabditis elegans* homolog (UNC-112) is a membrane-associated protein that colocalizes with integrin at cell–matrix adhesion complexes. BCNP1 is a completely unknown protein with three predicted transmembrane domains, with three alternatively spliced final exons. Proteomic analysis may thus define new potential therapeutic targets.
Leukemia (2003) **17,** 1605–1612. doi:10.1038/sj.leu.2402993
Keywords: plasma-membrane; chronic lymphocytic leukemia; proteomics

Introduction

Chronic lymphocytic leukemia (CLL) is the commonest form of leukemia in North America and Europe accounting for 25% of all leukemias.[1] CLL remains incurable with the currently available chemotherapy, but the prognosis of these patients varies substantially. Some patients will require little or no therapy for many years, while others die rapidly with chemotherapy-resistant disease. The prognosis for fludarabine-resistant disease is particularly dismal with most patients dying within 1 year. Monoclonal antibodies (MAbs), such as the CD52 MAb CAMPATH-1H, may be useful in some patients with hematological disease, but there remains an urgent need for new therapeutic approaches.[2]

More recently, the application of genomic-scale gene expression profiling techniques has allowed the identification of a characteristic CLL gene profile, which is distinct from other subtypes of B-cell malignancy.[3–6] This signature is seen in patients both with and without mutations in the *IGHV* gene segments, indicating that both subsets might arise from the transformation of the same cell type via similar methods of transformation. The genes within the CLL signature included many genes not previously thought expressed in CLL. These genes included *WNT3*, *titin* and *ROR1* receptor tyrosine kinase. Proteomics provides information complementary to gene expression profiling, allowing the annotation of the protein content of cells, usually through two-dimensional (2D) gels and high-throughput analysis. Furthermore, individual subcellular fractions can be used in conjunction with this approach in order to study the protein content within a given cellular compartment. Hemopoietic malignancies provide an ideal situation to assess proteomic technology since pure tumor cells can be readily obtained. Studies on myeloid cell lines have shown that there may be significant differences between results obtained with gene expression profiling and proteomic analysis. This is not unexpected for the following reasons: the half-lives of proteins and their coding RNAs can be drastically different; and proteins exist in many different isoforms (caused by a range of post-translational mechanisms) than do alternatively spliced RNAs such that isoform specificity of a protein is controlled by modifying enzymes and not the coding RNA. Preliminary studies using 2D gels in CLL showed a correlation of large-scale protein expression profiles with clinical data; patients with shorter survival times exhibited increased levels of heat-shock protein 27 and decreased levels of thioredoxin peroxidase 2 or protein disulfide isomerase, which may be associated with altered drug resistance.[7]

We have sought to annotate the protein content of the cell-surface membrane fraction in CLL, since this may allow the rapid identification of new prognostic and therapeutic targets. Cell-surface molecules are generally of high molecular weight and hydrophobic, and are poorly resolved on 2D protein gels. We have therefore used one-dimensional (1D) gels to fractionate proteins from this compartment along with high-throughput MALDI-TOF mass spectrometry (ms) analysis of trypsinized fractions. This approach has not been reported before with leukemic cells and plasma-membrane fractions. Here, we report the preliminary data from this analysis, and show how proteomics in conjunction with genomic analysis was able to better define the open-reading frame of two novel proteins expressed preferentially in the B-cell lineage.

Materials and methods

Patient material

CLL samples were derived from patients attending the Royal Marsden and Bournemouth hospitals after Ethical Committee

Correspondence: Professor Dr MJS Dyer, Department of Haematology, University of Leicester, Robert Kilpatrick Clinical Sciences Building, Leicester Royal Infirmary PO Box 65, Leicester, LE2 7LX, UK; Fax: +44 116 252 3274
[5]These authors contributed equally to the work presented in this paper

review and after obtaining informed written consent. RNA samples from patients with lymphoma were derived from pathologically validated clinical lymphoma tissues and obtained with full Institutional Review Board approval from Ardais Corporation (Lexington, MA, USA). The samples comprised 18 cases of diffuse large B-cell lymphoma (DLBCL), five cases of follicular lymphoma (FCL) and six cases of nodular sclerosing Hodgkin's lymphoma (HL). These RNA samples were derived from unfractionated biopsies and thus contained RNA derived from residual normal and reactive cells. Other RNA samples were purchased from Clontech (Palo Alto, CA, USA). Cell line RNA samples used in this study were derived from Raji, Daudi, (A46, Namolwa, Ramos, Mutu111, kHM10B and BL58 (Burkitt's lymphoma cell lines), OZ, SUDHL6 (DLBCL cell lines), Jurkat (T-cell ALL cell line), THP-1 and HL60 (AML cell lines) and COLO 775 (CML cell line).

Isolation of purified CLL cells

Patients with lymphocyte counts greater than 50×10^6/ml were selected for this study in order to minimize possible contamination of tumor cells by normal hemopoietic cells. Blood samples were collected into lithium heparin, diluted in an equal volume of phosphate-buffered saline (PBS) and then separated on Ficoll–Isopaque medium. Mononuclear cells were collected from the interphase and washed in PBS and then incubated with magnetic microbeads coated with CD3 and CD14 (Miltenyi Biotec, UK); magnetically labeled cells were passed over a magnetic column to deplete residual T cells and monocytes specifically. The resulting CLL cells had a purity of $>95\%$ as determined by flow cytometry.

CLL-plasma-membrane isolation

Purified CLL cells (10^9) were washed in PBS three times. The cell pellet was resuspended in homogenization buffer containing 250 mM sucrose, 10 mM HEPES, 1 mM EDTA, 1 mM vanadate, 0.02% azide and protease inhibitors (Roche, Switzerland). Cells were fractionated using a ball bearing homogenizer (8.002 mm ball, HGM Lab equipment, EMBL, Germany) until approximately 95% of cells were disrupted. Membranes were fractionated using the method described by Pasquali *et al*.[8] The fractionated cells were centrifuged at 3000*g* for 10 min at 4°C and the post-nuclear supernatant was layered onto a 60% sucrose cushion and centrifuged at 100 000*g* for 45 min. The membranes were collected and layered on a preformed 15–60% sucrose gradient and spun at 100 000*g* for 17 h. Proteins from the fractionated sucrose gradient were run on a 4–20% 1D gel (Invitrogen, UK) and subject to Western blotting. Those fractions containing CD20 , flotillin or transferrin immunoreactivity but not oxido-reductase II or calnexin immunoreactivity were pooled and represented the plasma-membrane fraction. For cytosolic fractionation, CLL cells were resuspended in 5 ml of homogenization buffer and sonicated on ice. The cells were centrifuged at 1000*g* for 5 min and the supernatant was spun at 100 000*g* for 60 min at 4°C. The soluble fraction represented the cytosol fraction, and the pellet the membrane fraction.

1D gels, proteolysis and MALDI-TOF MS

Plasma-membrane fractions from the same CLL samples were run in adjacent lanes on the same 1D gels and sequential 0.5 mm gel slices were excised from the 1D gels between 250 and 20 kDa. The molecular weight of each gel slice was recorded and then each slice subjected to trypsin digestion. The resulting peptide fragments were analyzed by MALDI-TOF-MS (Voyager STR, Applied Biosystems, Framingham, MA, USA) using a 337 nm wavelength laser for desorption and the reflectron mode of analysis. Selected masses were further characterized by tandem MS using a QTOF-MS equipped with a nanospray ion source, (Micromass UK Ltd, Manchester, UK). Prior to MALDI analysis, the samples were desalted and concentrated using C18 Zip Tips™ (Millipore, Bedford, MA, USA). Samples for tandem MS were purified using a nano-LC system (LC Packings, Amsterdam, The Netherlands) incorporating C18 SPE material. Using the SEQUEST search program,[9] uninterpreted tandem mass spectra of tryptic digest peptides were searched against a database of public domain proteins constructed of protein entries in the nonredundant NCBI database (http://www.ncbi.nlm.nih.gov/) and Swissprot (http://www.expasy.com). Peptide matches identified by SEQUEST were filtered according to their crosscorrelation score (Xcorr), normalized difference correlation score (ΔC_n), compatibility with trypsin digestion, and number of observations of proteins and peptides. Peptides were only used for protein identification where Xcorr $\geqslant 1.2$ and $\Delta C_n \geqslant 0.2$. For conservative identification of abundant proteins two peptides with these scores were required, but for some of the novel lower abundance proteins identified we accepted a single tandem peptide with these scores so long as there were mass matches on the same protein sequence from the same gel slice. Thus, a virtual 1D molecular weight protein map was created, which also enabled comparisons of the molecular weight of proteins positively identified by tandem mass spectra in samples.

Western blotting

Western blots were performed as previously described.[10] Cell extract or membrane protein (10 μg) were run on a 4–20% 1D gel and transferred to PVDF membrane (Invitrogen). Membranes were blocked with 3% milk powder (Marvel), incubated with anti-BLK (Santa Cruz, 329), and then incubated with anti-rabbit HRP (Amersham). Immunoreactive bands were developed using the ECL detection system (Amersham).

Quantitative RT-PCR

Real-time RT-PCR was used to quantitatively measure *MIG2B* and *BCNP-1* expression in normal human tissue mRNAs (Clontech), cell lines (American Type Culture Collection), and CLL and B-NHL samples. The primer pairs used traversed introns and test products were sequenced to confirm specificity before use in these assays. PCR products from all samples were analyzed on agarose gels and positives shown to contain a single PCR product of the size predicted from cDNA. No fragments of the size predicted from genomic DNA were detected in any samples demonstrating the complete absence of genomic DNA contamination. All reactions were run twice and any samples showing a $>10\%$ variation in copy number excluded from analysis. The primers used for PCR were as follows:

B-cell novel protein #1 (*BCNP1*): sense 5′GGTGGTCGTGGGGAAGGGAAGA3′ (exon 1) and antisense: 5′ CCACGGTAGCAAGGCAGGAAGT3′ (exon 2, interexon distance = 3597 bp). *MIG2B*: sense 5′GCATCTCCTATGTCATGGTC-AG3′

Table 1 (a) Proteins characterized in this study and (b) proteins identified in this study remaining uncharacterized

Novel proteins	*Observed approx mol wt (kDa)*	*Peptides identified*	*Accession*	*Homolog accession*	*Homology (expect value)*	*Locus link*	*Chromosomal localization*
(a)							
MIG2B	60	LLVPSPEGMSEIYLR VFVGEEDPEAESVTLR EILGIANNR	AAH13366.1| AAH15584.1|	CAA80852 (MIG2)	1×10^{-75}	83706	11q13.1
BCNP1	80	ELGPQEPTGSQLLR	AK097121.1| AK074069	BAB15951 (Niban) (R. norvegicus)	4×10^{-14}	199786	19p13.1
(b)							
KIAA1273	70	TAGTLFGEGFR LAQFDYGR	Q9NVI7	–	–	55210	1p36.3
KIAA0068	152	FINMFAVLDELK DFVSEAYLITLGK	Q9ULQ2	AAK81821 (Cyfip) (*Mus. musculus*)	0.0 (85%)	23191	15q11
KIAA0906	200	LPSQYNFAMNVLGR	O94980	AAF21969 (nuclear pore membrane glycoprotein POM210) (*Mus. musculus*)	0.0 (80%)	23225	3p25.1
KIAA0171	89	ATNVVMNYSEIESK ELVEFAQDDDR	Q14677	AAF43421 (epsin-like protein) (*D. melanogaster*)	1×10^{-61}	9685	5q23–33
KIAA1430	31	VLHDTMDLNHLLK	Q9P2B7	–	–	57587	4q35.1
flj10702	17	GVNAIVYMIDAADR MNLSAIQDR	Q9NVJ2	AAM11162 (LD29185p) (*D. melanogaster*)	4×10^{-94}	55207	3p26.1
dj889n15.1	38	EQLSIQWSFFHK	O95532	CAA74390 (ChT1 thymocyte antigen) (*Gallus gallus*)	5×10^{-64}	-	Xq23
flj11301	25	QIENIVDK	Q9H089	AAF45628 CG14788-PA (*D. melanogaster*)	$\times 10^{-151}$	55341	3q29
Hypothetical 60.5 kDa protein	61	YQQELEEEIK	Q9BSE0	–	–	25923	11q13.1

Figure 1 BLK immunoreactivity in matched B-CLL membrane and cytosol samples. Membrane (M1–M5) and cytosol fractions (C1–C5) from five matched patients were analyzed by immunoblotting with anti-BLK antibody.

(exon 4) and antisense 5′GATGTCCCAGTTGACA-TTCCAC3′ (exon 5, interexon distance = 2061 bp).

PCR reactions containing 5 ng cDNA, SYBR green sequence detection reagents (PE Biosystems), *Taq* polymerase, sense and antisense primers were assayed on an ABI7700 sequence detection system (PE Biosystems). The PCR conditions were one cycle at 50°C for 2 min, one cycle at 95°C for 10 min, and 40 cycles of 95°C for 15 s, 65°C for 1 min. The accumulation of PCR product was measured in real time as the increase in SYBR green fluorescence, and the data were analyzed using the Sequence Detector Program v1.6.3 (PE Biosystems). Standard curves relating initial template copy number to fluorescence and amplification cycle were generated using the amplified PCR product as a template, and were used to calculate copy number in each sample.

Results

Protein identification

CLL-plasma-membrane protein fractions were resolved on a 1D SDS-polyacrylamide gel and sequential 0.5 mm gel slices containing proteins subjected to exhaustive trypsinolysis. The resulting tryptic peptides were analyzed by MALDI-TOF MS. Sets of peptide masses thus obtained from each gel slice were compared to theoretical tryptic masses of known primate protein sequences (GenBank PRI division http://www.ncbi.nlm.nih.gov) using the MOWSE algorithm.[11] Three successive rounds of mass matching were carried out for the peptides from each gel slice and peptide masses were arranged into three clusters representing the protein sequence with the highest number of matches for each round. This process enabled selection of peptides for subsequent QTOF tandem mass spectrometric analysis to be biased away from those found in proteins most frequently identified by MS,[12] such as structural, heat-shock and MHC proteins. In all, 500 proteins were identified by mass- and fragment-based searching of the public databases. A total of 365 (74%) of these were previously characterized proteins of known subcellular localization; the other 26% of proteins comprised both hypothetical proteins (15%) and proteins whose subcellular localization has not been determined (11%). Of the 365 previously characterized proteins, 238 (65%) were known membrane-associated proteins, while the remaining 35% were known non-membrane proteins, reflecting

contamination of the membrane preparation with soluble cytosolic proteins.

Previously described CLL-associated cell surface proteins were identified in the derived plasma-membrane fractions.[13,14] These included molecules utilized as diagnostic markers in CLL including CD5, CD19, CD20, CD22 and CD23, but also other B-cell-restricted molecules such as CD72. Many integrin-related molecules including CD11a/b/c, CD29, CD41, CD49d/e and ICAMs CD51 and CD53 were also detected in this fraction, as were CD73 and CD166 (data not shown). Although highly abundant, CD52 was not detected, presumably since this is a small peptide of 12 amino acids that generates on trypsinolysis few peptides of very low molecular weight that may not be visible with these methods.[15] Table 1 lists some of the uncharacterized proteins identified in this study that formed the basis of our efforts to describe novel B-cell-specific membrane proteins. The selection of *BCNP1* and *MIG2B* for further investigation was based primarily on the paucity of expressed sequence tags (EST) data (the peptide identified for *BCNP-1* only matched to genomic sequences at the time of discovery) or a hint of specificity in B cells from publicly available ESTs (*MIG2B*).

Novel findings from proteomic analysis of CLL membrane fractions

The combination of subcellular proteomics and genomics can also add significantly to the definition of genes currently defined only by ESTs, cDNA clones or theoretical transcripts predicted from raw genomic sequences by algorithms such as Genscan.[16] In all, 115 of the proteins we identified by MS analysis of CLL membranes had only previously been identified by the above methods. Given that 238 (65%) of the 365 previously characterized proteins we identified were known membrane proteins, and that 133 (56%) of these were in some way localized to the plasma membrane, it might be expected that up to 42 of the 115 uncharacterized proteins would therefore also be plasma membrane proteins. The known tissue distribution of encoding mRNAs (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db = unigene), and/or homology to known proteins (http://www.ncbi.nlm.nih.gov/BLAST), or the presence of predicted transmembrane domains (http://www.ch.embnet.org/software/TMPRED_form.html), were assessed where possible for this set of proteins. Proteins with either restricted expression, or homology to protein families known to be involved in cancer or with predicted transmembrane domains were then subjected to mRNA profiling in a panel of primary CLL and lymphoma samples.

We discuss here the membrane specificity of the BLK kinase and the preliminary characterization of two of the novel proteins, MIG2B and BCNP1 (Table 1).

Identification of BLK at CLL membrane BLK is a 55 kDa non-receptor tyrosine kinase of the SRC family of kinases that is expressed preferentially in B cells. The human gene is overall 87% identical to the murine, but has an NH2-terminal insertion of six amino acids, which was thought to indicate possible insertion into the plasma membrane. The function of BLK is poorly understood, but it may play an important role in the signaling pathways controlling cell proliferation and differentiation.[17] Its cellular localization in leukemic cells had not been determined. To confirm the results of the proteomic analysis, which suggested an association with the plasma membrane, we assessed the level of BLK immunoreactivity in purified CLL

Full length *BCNP1* (Isoform 1)

```
   1 GCTGAGCAGGAGATGGGAATTGAAACCTGCGCAGAGGCGGCTGTGCAGGGTGAGAGTGGA 60
  61 GCCGAAACCACAGAAAGTGAAGTTTGCTTCAACGTCTTGTCCCGGCAGGCCACTCAGATG 120
 121 TGAGAGTGAGGAAGTGGGATGGGGCCTGACCGGAAGGAGGTGCCCCTGAGCCGAGGAACG 180
                     M  G  P  D  R  K  E  V  P  L  S  R  G  T
 181 CAGGCGGTGGTCGTGGGGAAGGGAAGAGGAGCCCCGGGAGACGACAGCAGCATGGGTGGG 240
     Q  A  V  V  V  G  K  G  R  G  A  P  G  D  D  S  S  M  G  G
 241 CGGCCTTCGAGCCCTCTGGACAAGCAGCAGCGGCAGCACCTAAGGGGTCAGGTGGACACC 300
     R  P  S  S  P  L  D  K  Q  Q  R  Q  H  L  R  G  Q  V  D  T
 301 CTGCTGAGGAACTTCCTGCCTTGCTACCGTGGGCAGCTGGCAGCGTCTGTCCTGCGGCAG 360
     L  L  R  N  F  L  P  C  Y  R  G  Q  L  A  A  S  V  L  R  Q
 361 ATCTCTCGAGAGCTGGGCCCTCAGGAGCCGACCGGAAGCCAGTTGCTACGCAGCAAAAAG 420
     I  S  R  E  L  G  P  Q  E  P  T  G  S  Q  L  L  R  S  K  K
 421 CTGCCCCGAGTCCGTGAGCACCGAGGACCCCTGACCCAGCTTCGGGGCCACCCACCCCGG 480
     L  P  R  V  R  E  H  R  G  P  L  T  Q  L  R  G  H  P  P  R
 481 TGGCAGCCGATCTTCTGTGTTCTGCGTGGGGACGGCCGCCTAGAGTGGTTCAGCCACAAG 540
     W  Q  P  I  F  C  V  L  R  G  D  G  R  L  E  W  F  S  H  K
 541 GAGGAATATGAAAACGGGGGCCACTGCCTTGGCTCAACAGCCCTGACAGGATACACGCTC 600
     E  E  Y  E  N  G  G  H  C  L  G  S  T  A  L  T  G  Y  T  L
 601 CTGACTTCCCAGCGAGAATATCTCCGCCTTTTGGATGCTCTCTGCCCTGAATCCTTGGGA 660
     L  T  S  Q  R  E  Y  L  R  L  L  D  A  L  C  P  E  S  L  G
 661 GACCATACTCAGGAAGAGCCTGACTCCCTCTTGGAAGTGCCTGTGAGCTTCCCGCTGTTC 720
     D  H  T  Q  E  E  P  D  S  L  L  E  V  P  V  S  F  P  L  F
 721 CTGCAGCACCCCTTCCGCCGGCACCTCTGCTTCTCTGCAGCCACCAGGGAGGCACAGCAT 780
     L  Q  H  P  F  R  R  H  L  C  F  S  A  A  T  R  E  A  Q  H
 781 GCCTGGAGGCTGGCCCTGCAGGGTGGCATCCGGCTTCAGGGCACAGTCCTGCAGCGAAGC 840
     A  W  R  L  A  L  Q  G  G  I  R  L  Q  G  T  V  L  Q  R  S
 841 CAGGCCCCTGCTGCCCGGGCCTTCCTGGACGCCGTCCGACTCTACCGGCAGCACCAAGGC 900
     Q  A  P  A  A  R  A  F  L  D  A  V  R  L  Y  R  Q  H  Q  G
 901 CACTTTGGCGACGACGACGTGACCCTAGGCTCAGACGCCGAGGTGCTGACCGCGGTGCTG 960
     H  F  G  D  D  D  V  T  L  G  S  D  A  E  V  L  T  A  V  L
 961 ATGCGGGAGCAACTTCCCGCGCTGCGAGCCCAGACCCTTCCTGGCCTGCGGGGGGCAGGC 1020
     M  R  E  Q  L  P  A  L  R  A  Q  T  L  P  G  L  R  G  A  G
1021 CGCGCCCGCGCCTGGGCCTGGACCGAGCTTCTAGACGCCGTTCACGCAGCTGTCCTGGCC 1080
     R  A  R  A  W  A  W  T  E  L  L  D  A  V  H  A  A  V  L  A
1081 GGGGCCTCCGCCGGGCTCTGCGCCTTCCAGCCCGAAAAGGACGAGCTGCTTGCGTCGCTG 1140
     G  A  S  A  G  L  C  A  F  Q  P  E  K  D  E  L  L  A  S  L
1141 GAGAAGACGATCCGCCCGGACGTGGACCAGCTGCTGCGGCAGCGGGCGCGTGTGGCGGGG 1200
     E  K  T  I  R  P  D  V  D  Q  L  L  R  Q  R  A  R  V  A  G
1201 CGGCTGAGGACGGATATCAGGGGACCGCTCGAGTCGTGCCTGCGCCGGGAGGTGGACCCG 1260
     R  L  R  T  D  I  R  G  P  L  E  S  C  L  R  R  E  V  D  P
1261 CAGCTGCCCCGGGTCGTGCAGACCCTGCTGCGCACCGTGGAAGCCTCGCTCGAGGCGGTG 1320
     Q  L  P  R  V  V  Q  T  L  L  R  T  V  E  A  S  L  E  A  V
1321 CGGACCCTCCTGGCTCAAGGCATGGACCGACTGTCCCACCGCCTGCGCCAGAGCCCCTCA 1380
     R  T  L  L  A  Q  G  M  D  R  L  S  H  R  L  R  Q  S  P  S
1381 GGCACGCGGCTGCGCAGGGAGGTTTACTCATTTGGGGAGATGCCGTGGGACTTGGCGCTG 1440
     G  T  R  L  R  R  E  V  Y  S  F  G  E  M  P  W  D  L  A  L
1441 ATGCAGACATGCTACCGTGAGGCCGAGCGGAGCCGGGGGCGCTTGGGGCAGCTGGCAGCA 1500
     M  Q  T  C  Y  R  E  A  E  R  S  R  G  R  L  G  Q  L  A  A
1501 CCGTTTGGCTTTCTGGGGATGCAGAGCCTCGTGTTTGGGGCCCAAGATCTTGCACAGCAG 1560
     P  F  G  F  L  G  M  Q  S  L  V  F  G  A  Q  D  L  A  Q  Q
1561 CTCATGGCTGACGCCGTGGCCACCTTCCTGCAGCTGGCTGACCAGTGTCTGACGACGGCC 1620
     L  M  A  D  A  V  A  T  F  L  Q  L  A  D  Q  C  L  T  T  A
1621 CTCAACTGTGACCAGGCTGCCCAGAGGCTGGAGAGAGTCAGGGGGCGCGTGCTGAAGAAA 1680
     L  N  C  D  Q  A  A  Q  R  L  E  R  V  R  G  R  V  L  K  K
1681 TTCAAATCGGACAGCGGGTTGGCGCAGAGGAGGTTCATCCGAGGCTGGGGTCTCTGCATC 1740
     F  K  S  D  S  G  L  A  Q  R  R  F  I  R  G  W  G  L  C  I
1741 TTTTTACCTTTTGTGCTGAGCCAACTCGAGCCAGGCTGCAAAAAGGAGCTGCCTGAGTTC 1800
     F  L  P  F  V  L  S  Q  L  E  P  G  C  K  K  E  L  P  E  F
1801 GAGGGGGATGTCCTTGCCGTGGGCAGCCAGGCTCTGACCACTGAGGGCATCTATGAGGAC 1860
     E  G  D  V  L  A  V  G  S  Q  A  L  T  T  E  G  I  Y  E  D
1861 GTCATCCGGGGGTGCTTGCTGCAGAGGATTGACCAAGAATTGAAAAAGACCCTTGGTGCC 1920
     V  I  R  G  C  L  L  Q  R  I  D  Q  E  L  K  K  T  L  G  A
1921 AATGATGTATCCTGCACTCTGGACGGCTGCTTGGAGGTCCCATGGGAACAGGAGGGAGCA 1980
     N  D  V  S  C  T  L  D  G  C  L  E  V  P  W  E  Q  E  G  A
1981 GGTGGGGAACTTCACAGGCTTCTGAAACATGCCAGTGATTGCTGGGGGTGGGGAGGTGGA 2040
     G  G  E  L  H  R  L  L  K  H  A  S  D  C  W  G  W  G  G  G
2041 GGCCCACATGGGGTGTATGGGATTGTCATGAGTGTGGGTTCATGGACGGAACAGAATGCA 2100
     G  P  H  G  V  Y  G  I  V  M  S  V  G  S  W  T  E  Q  N  A
2101 CTTTGTCTCCTTACCCAGCCAGCCTGGCCAAGCTCGTGAACAAGCTCAGTGGGAAGTTGT 2160
     L  C  L  L  T  Q  P  A  W  P  S  S  *
2161 GGAGTGCATCTGAGCCACAGAGGGACCCTCTAAACCCAGGCTCTGACATGCATGTGAGCT 2220
2221 CAGCCCTGGGGGGCTGAGATTGTGGTGAATTTGTGTCCCATCTCT 2265
```

Alternatively spliced final exon 1 (Isoform 2)

```
   1 GATGAGGAAACTGAGGCTGAGCGGGAAGGAGGGGCTTGTCCCAGGCAGCCAGACTCTGGT 60
     D  E  E  T  E  A  E  R  E  G  G  A  C  P  R  Q  P  D  S  G
  61 GCCCAGATCCAGCCACTCTGCCCACCGCCTTCTCCAGGAACATTCCGGAGCTGAATCTTC 120
     A  Q  I  Q  P  L  C  P  P  P  S  P  G  T  F  R  S  *
 121 ACCCACATCTATCTTGTTTCTATTGGATAAATGTCTACAAGTGGAATTTCTGGGCCAAAA 180
 181 CGGATGTGCCATCTTTAGGCTTTTGTAACCCCTGCAACTTCAGAAAACTGTACCATTTTA 240
 241 TAAAAAAAAAAAAAAAAA 218
```

Alternatively spliced final exon 2 (Isoform 3)

```
   1 GCTCCAAATCTTAACTTGGTGTCAAGTTTCCTGGCTGGGAGACAAGCTTTTACCGACTTC 60
     A  P  N  L  N  L  V  S  S  F  L  A  G  R  Q  A  F  T  D  F
  61 CTCTGCTTGCCAGCAAAGTCATCTGCTAACTGGATATTGGCAGCTTCTCTGCTGTCTTGC 120
     L  C  L  P  A  K  S  S  A  N  W  I  L  A  A  S  L  L  S  C
 121 AGCTGCTTCCGGAGTGGGTTCCACAGGGATTCCCGTGTGTTCTTGGTTCAGCTTGCAGAG 180
     S  C  F  R  S  G  F  H  R  D  S  R  V  F  L  V  Q  L  A  E
 181 GGACTTTCACACTCCCTGGAGACCGTTTCCTCCCATTCTGTCTGGAGTTTTCGGCCTACC 240
     G  L  S  H  S  L  E  T  V  S  S  H  S  V  W  S  F  R  P  T
 241 CCAAGACAATGAGATATTCCTGNCCTTTCCNCCTATTTCCCTCCAACCCCNCCTTCCGAA 300
     P  R  Q  *
 301 ATACATTTGCTCAATCATTTGCACTTCATAGGCCAAAAAAAAAAAAAAAAAAAAAAAAAA 360
```

Figure 2 DNA and protein sequence of *BCNP1*. There are three alternative final exons spliced at nucleotide 1981 (highlighted G) in the primary sequence and replaced by the sequences beneath. The three BCNP1 isoforms have been submitted to GenBank and their accession numbers are: isoform 1 AY254197, isoform 2 AY254198 and isoform 3 AY254199.

membrane and cytosol fractions. Western blotting of CLL cytosol and membrane preparations with anti-BLK antibody revealed that BLK immunoreactivity was high in membrane preparations but barely detectable in cytosol preparations (Figure 1). This indicates that in CLL, BLK is localized almost exclusively to the cell-surface membranes. In support of this result, no BLK was detected in the cytosol fraction of the same CLL samples by tandem MS (data not shown).

Identification and cloning of BCNP1

Microsequencing of a peptide from a gel slice at approximately 80 kDa identified a GenScan predicted protein containing 21 exons (http://www.ensembl.org). Sequence database searching identified several public domain ESTs that showed strong homology to different regions of the predicted transcript. Alignment of the predicted protein with genomic DNA allowed the design of multiple PCR primers covering the putative open-reading frame. These primers were then used to clone the entire gene from normal spleen and Daudi cell line cDNA preparations. We have termed this gene *BCNP1*. We identified three *BCNP1* isoforms each with a different alternately spliced final exon encoding three proteins of 667, 652 and 698 amino acids (GenBank accessions AY254197, AY254198, AY254199 respectively) (Figure 2). The full-length *BCNP1* sequences are predicted to contain three transmembrane domains and showed no significant homologies to known protein classes and contained no known protein motifs (Figure 2). The protein was however 31% identical to a gene of unknown function, NIBAN, which is upregulated in renal carcinoma.[18]

All the matching ESTs on the public databases came from B-cell cDNA libraries, and thus *BCNP1* was a clear candidate for more detailed examination as a potential new B-cell antigen. Quantitative RT-PCR analysis showed that *BCNP1* was highly expressed in CLL and lymphoma samples and restricted to B-cell containing tissues (Figure 3). Furthermore, there was far more *BCNP1* mRNA expressed in B-cell malignancies than in normal B-cell containing tissues.

Identification and characterization of MIG2B

In a gel slice from CLL membranes cut from approximately 60 kDa, three tandem peptides were found that matched the conceptual translations of cDNA clones derived from normal thymus (clone #AK093719) and from a Burkitt's lymphoma cell line (clone #BC004347). The latter is a 2502 bp cDNA clone predicted to encode a 236 amino-acid protein with a molecular weight of 27 kDa. However, there are many alternative methionines available in the proposed 5′ UTR before an inframe stop codon is found for this clone. The thymus cDNA has a conceptual translation of 663 amino acids. Supporting the presence of the larger predicted protein, two of these tandem

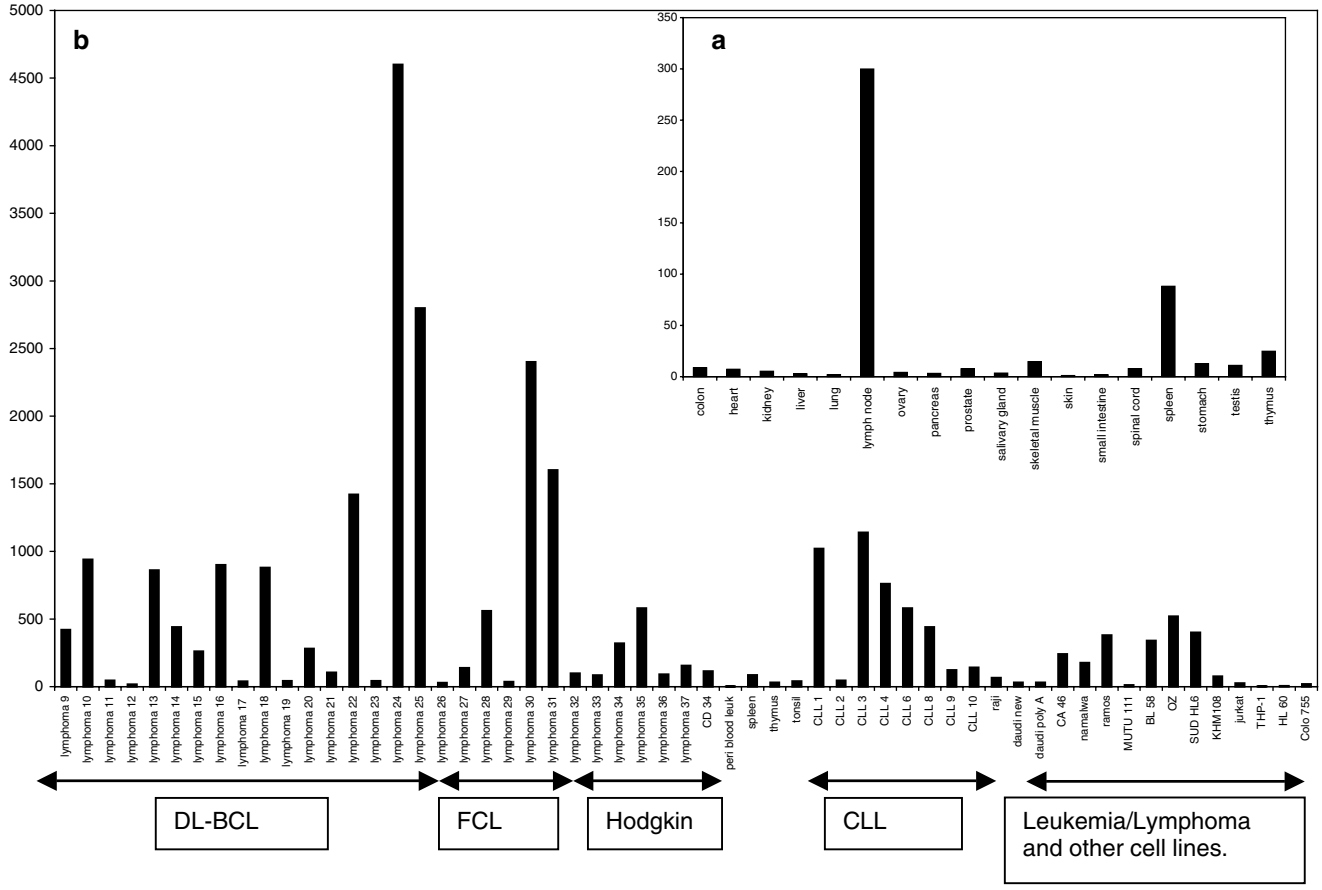

Figure 3 Quantitative mRNA expression of *BCNP1* in multiple normal tissues (**a**) as well as B-cell malignancies (**b**): CLL – chronic lymphocytic leukemia, DLBCL – diffuse, large B-cell lymphoma, FCL – follicular B-cell lymphoma and Hodgkin's lymphoma. Note that the only normal tissues showing significant *BCNP1* expression are the B-cell-rich tissues including spleen, lymph node and tonsil, whereas multiple leukemia and lymphoma samples show increased *BCNP1* mRNA expression. Values on the *y*-axis are mRNA copy no. per nanogram of cDNA. *y*-Axis values in (**a**) are lower than (**b**) due to the increased expression of BCNP1 in malignant B cells.

peptides extended in the 5′ direction the methionine start codon proposed in BC004347. The third tandem peptide was identified within the BC004347 original translation, as were other mass matched peptides throughout the newly defined protein. These data thus allowed confirmation of the correct open-reading frame of this gene. The nucleotide and protein sequences are shown in Figure 4a. This 663 amino-acid protein now contained significant homology with MIG2, (S69890; Locus Link 10979) a

a

MAGMK***TASGDYIDSSWELR***VFVGEEDPEAESVTLRVTGESHIGGVLLK***IVEQINR***KQDWSDHAIWWEQKRQ***WLLQTHWTLDK***YGIL
ADARLFFGPQHRPVILRLPNRRALRLRASFSQPLFQAVAAICRLLSIR***HPEELSLLR***APEKKEKKKKEKEPEEELYDLSK***VVLAGG***
VAPALFRGMPAHFSDSAQTEACYHMLSRPQPPPDPLLLQR***LPRPSSLSDK***TQLHSR***WLDSSR***CLMQQGIK***AGDALWLR***FKYYSFFD
LDPKTDPVR***LTQLYEQAR***WDLLLEEIDCTEELMMVFAALQYHINKLSQSGEVGEPAGTDPGLDDLDVALSNLEVKLEGSAPTDVLD
SLTTIPELKDHLRIFRPRKLTLKGYRQHWVVFKETTLSYYKSQDEAPGDPIQQLNLKGCEVVPDVNVSGQKFCIK**LLVPSPEGMSE**
IYLRCQDEQQYARWMAGCRLASKGRTMADSSYTSEVQAILAFLSLQRTGSGGPGNHPHGPDASAEGLNPYGLVAPRFQRKFKAKQL
TPRILEAHQNVAQLSLAEAQLRFIQAWQSLPDFGISYVMVRFKGSRK**DEILGIANNR**LIRIDLAVGDVVKTWRFSNMRQWNVNWDI
RQVAIEFDEHINVAFSCVSASCRIVHEYIGGYIFLSTRERARGEELDEDLFLQLTGGHEAF

b

```
MIG2B: 3   GMKTASGDYIDSSWELRVFVGEEDPEAESVTLRVTGESHIGGVLLKIVEQINRKQDWSDH 62
           G++   G Y D +WEL V V + +     +TLRVTGE HIGGV+LK+VE+++ K+DWSDH
MIG-2: 45  GIRMPDGCYADGTWELSVHVTDLN---RDITLRVTGEVHIGGVMLKLVEKLDVKKDWSDH 101

MIG2B: 63  AIWWEQKRQWLLQTHWTLDKYGILADARLFFGPQHRPVILRLPNRRALRLRASFSQPLFQ 122
           A+WWE+KR WLL+THWTLDKYGI ADA+L F PQH+ + L+LPN + ++++ +FS  +F+
MIG-2: 102 ALWWEKKRTWLLKTHWTLDKYGIQADAKLQFTPQHKLLRLQLPNMKYVKVKVNFSDRVFK 161

MIG2B: 123 AVAAICRLLSIRHPEELSLLRAPEKKEKKKK---EKEPEEELYDLSKVVLAGG-----VA 174
           AV+ IC+  +IRHPEELSLL+ P    KKKK   + + E+E  +L   ++  G      +
MIG-2: 162 AVSDICKTFNIRHPEELSLLKKPRDPTKKKKKKLDDQSEDEALELEGPLITPGSGSIYSS 221

MIG2B: 175 PALFR-----------GMP-----AHFSDSAQTEACYHMLSRPQPPPDPLLLQRLPRPSS 218
           P L+            G P     A F DSA +E    +L+  QP   P +L ++ +P +
MIG-2: 222 PGLYSKTMTPTYDAHDGSPLSPTSAWFGDSALSEGNPGILAVSQPITSPEILAKMFKPQA 281

MIG2B: 219 LSDKTQLHSRWLDSSRCLMQQGIKAGDALWLRFKYYSFFDLDPKTDPVRLTQLYEQARWD 278
           L DK +++  WLDSSR LM+Q +K  +AL LRFKYYSFFDL+PK D +R+ QLYEQA+W
MIG-2: 282 LLDKAKINQGWLDSSRSLMEQDVKENEALLLRFKYYSFFDLNPKYDAIRINQLYEQAKWA 341

MIG2B: 279 LLLEEIDCTEEEMMVFAALQYHINKLSQSGEVGEPAGTDPGLDDLDVALSNLEVKLEGSA 338
           +LLEEI+CTEEEMM+FAALQYHINKLS          +D  +D++D ALS+LE+ LEG
MIG-2: 342 ILLEEIECTEEEMMMFAALQYHINKLSIMTSENHLNNSDKEVDEVDAALSDLEITLEGGK 401

MIG2B: 339 PTDVLDSLTTIPELKDHLRIFRPRKLTLKGYRQHWVVFKETTLSYYKSQDEAPGDPIQQL 398
            + +L  +T+IPEL D++++F+P+KLTLKGY+Q+W  FK+T++S YKS++E+ G P  Q+
MIG-2: 402 TSTILGDITSIPELADYIKVFKPKKLTLKGYKQYWCTFKDTSISCYKSKEESSGTPAHQM 461

MIG2B: 399 NLKGCEVVPDVNVSGQKFCIKLLVPSPEGMSEIYLRCQDEQQYARWMAGCRLASKGRTMA 458
           NL+GCEV PDVN+SGQKF IKLL+P  EGM+EI+LRC +E+QYA WMA CRLASKG+TMA
MIG-2: 462 NLRGCEVTPDVNISGQKFNIKLLIPVAEGMNEIWLRCDNEKQYAHWMAACRLASKGKTMA 521

MIG2B: 459 DSSYTSEVQAILAFLSLQRTGSGGPGNHPHGPDASAEGLNPYGLVAPRFQRKFKAKQLTP 518
           DSSY  EVQ IL+FL +Q        +    P+     + P  LV+PR+ +K+K KQ+T
MIG-2: 522 DSSYNLEVQNILSFLKMQHLNP----DPQLIPEQITTDITPECLVSPRYLKKYKNKQITA 577

MIG2B: 519 RILEAHQNVAQLSLAEAQLRFIQAWQSLPDFGISYVMVRFKGSRKDEILGIANNRLIRID 578
           RILEAHQNVAQ+SL EA++RFIQAWQSLP+FGI++ + RF+G +K+E++GIA NRLIR+D
MIG-2: 578 RILEAHQNVAQMSLIEAKMRFIQAWQSLPEFGITHFIARFQGGKKEELIGIAYNRLIRMD 637

MIG2B: 579 LAVGDVVKTWRFSNMRQWNVNWDIRQVAIEFDEHINVAFSCVSASCRIVHEYIGGYIFLS 638
            + GD +KTWRFSNM+QWNVNW+I+ V +EF + + ++F C    C++VHE+IGGYIFLS
MIG-2: 638 ASTGDAIKTWRFSNMKQWNVNWEIKMVTVEFADEVRLSFICTEVDCKVVHEFIGGYIFLS 697

MIG2B: 639 TRERARGEELDEDLFLQLTGG 659
           TR + + E LDE++F +LT G
MIG-2: 698 TRAKDQNESLDEEMFYKLTSG 718
```

Figure 4 (**a**) Protein sequence of MIG2B. Tandem peptides used to identify this sequence are underlined and in bold. Mass matched peptides are in bold italics. The methionine start codon proposed from BC004347 is indicated with an arrow. (**b**) Blast alignment of MIG2B with MIG2. (**c**) Quantitative RT-PCR analysis of MIG2B mRNA expression. The normal tissue distribution (**a**) shows low expression in most normal tissues except the B-cell-rich tissues, lymph node and spleen. (**b**) shows consistent high expression in different B-cell malignancies. Values on the *y*-axis are mRNA copy no. per nanogram of cDNA. *y*-axis values in (**a**) are lower than (**b**) due to the increased expression of MIG2B in malignant B cells.

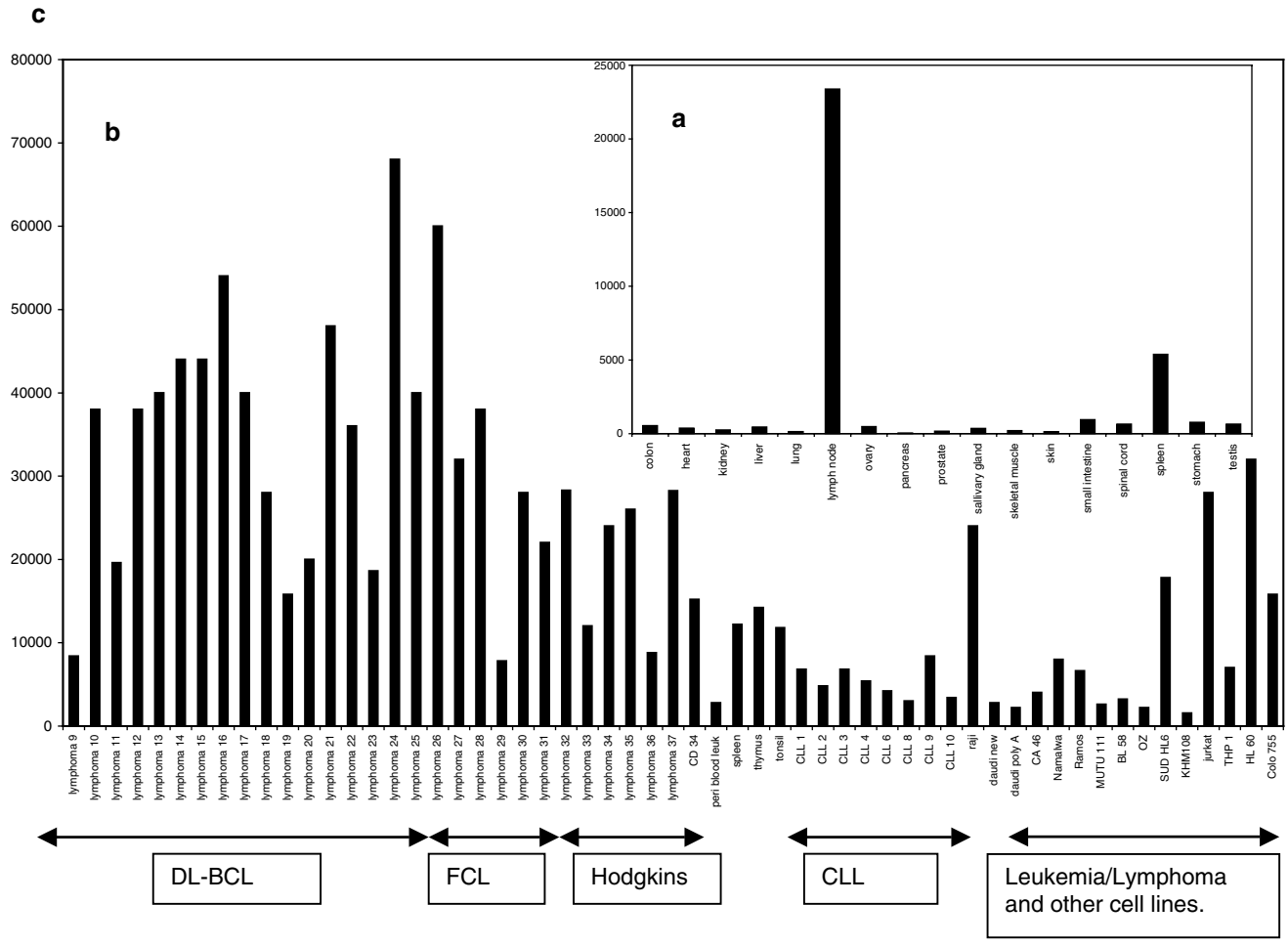

Figure 4 Continued

predicted protein of greater than 720 amino acids and matched completely the larger predicted protein, AK093719 (Figure 4b).

MIG2 is a gene whose expression is upregulated by mitogens in fibroblasts.[19] Its functions remain unknown. The *Caenorhabditis elegans* homolog (UNC-112), which shows 59% homology to MIG2B at the protein level, is involved in integrin localization at the plasma membrane and the formation of cell–matrix adhesion structures.[20] Interestingly, UNC-112 also interacts with the homolog of integrin-linked kinase or ILK.[21]

In view of the similarity between this protein and MIG2, we have named this protein *MIG2B*. *MIG2B* was 52% identical and 76% homologous to MIG2 at the protein level. *MIG2B* contained two ezrin/radixin/moesin (ERM) domains indicative of a cytoskeletal association, separated by a pleckstrin homology domain, commonly found in eukaryotic signaling proteins. There were no obvious transmembrane domains in the *MIG2B* sequence. Although this protein was identified in CLL-plasma-membrane fractions its further identification in cytosol fractions (data not shown) suggests that this protein is also intracellular in CLL.

The *MIG2B* Unigene cluster (Hs.180535) appeared to be chimeric, but showed a strong hemopoietic bias with >40% of clones deriving from this lineage including clones from T- and B-cell leukemic cell lines, spleen, tonsil and bone marrow tissues.

Quantitative RT-PCR analysis showed extremely specific hemopoietic expression as well as increased expression in B-cell malignancies (Figure 4c). *MIG2B* unlike *BCNP1* had a more widespread expression in non-B-cell hemopoietic cell types, such as HL60, THP1 and Jurkat cell lines. We have however only observed this protein in CLL samples and not in the many other cell types such as endothelial, neuronal and epithelial we have examined by the methods described here (data not shown).

Discussion

The aim of this study was to derive a preliminary 'map' of the cell-surface proteins expressed in CLL in order to identify new potential therapeutic targets. CLL provides relatively easy access to pure intact malignant cells. We describe an efficient and reproducible method for the generation of plasma membranes from these primary cells. Cell-surface antigens are generally of high molecular weight and hydrophobic, and 1D gels allow these types of proteins to both enter the gel and to be resolved. The inability of previous proteomics studies to identify these proteins lies partly in the inability of 2D gels to resolve proteins of this type. Use of the MOWSE algorithm to target masses for sequence data generation has also played a part in biasing protein identification towards sequences of specific biological significance, and greater coverage of lower abundance proteins such as those involved in transmembrane signal transduction.

This study identified known proteins and proteins not previously recognized to be resident in the CLL-plasma-

membrane fraction. Also, a number of previously unidentified, novel proteins were detected (Table 1). *BCNP1* and *MIG2B* represent the type of novel protein with probable plasma membrane localization that this approach is capable of discovering. We cannot assign plasma membrane localization for these proteins with complete certainty due to contamination of the CLL-plasma-membrane fraction with intracellular membrane proteins and cytosolic proteins; confirmation of this point will require the development of specific antibody reagents. Nevertheless, the presence of three predicted transmembrane domains in *BCNP1* and plasma membrane-binding ERM domains in *MIG2B*, are consistent with these proteins being localized to membranes. *BCNP1* is a B-cell-specific protein, whose expression is retained and often upregulated in B-cell malignancies. The biological basis for this variability of expression remains to be determined. Whether *BCNP1* or any of the other novel proteins identified in this project (Table 1b) play a role in the pathogenesis of CLL remains to be determined.

The importance of integrating proteomic and genomic technologies at many different levels is highlighted with the definition of *MIG2B*. Initial proteomic findings can be rapidly profiled at the mRNA level in large numbers of normal and disease samples and the cellular localization can be confirmed through protein tagging technologies in relevant models. If protein detection reagents and samples are available: the abundance and localization of protein in normal and disease samples can also be determined. If protein reagents or samples are not available, genomic-based approaches together with information from proteomic expression databases in which the relative levels and localization of proteins in different tissues or cells are known helps decisions to be made on further studies.

The study shows that it is possible to use proteomics, in combination with genomic analysis, to identify novel membrane proteins in CLL that may have disease relevance. Changes in protein localization or modification in normal and disease cells may provide novel prognostic markers and even drug targets. Further analysis of plasma membrane fractions from patients with CLL will allow the recognition of other potential targets. The challenge for proteomics will be to identify and exploit such changes therapeutically.

Acknowledgements

This work was supported in part by grants from the Kay Kendall Leukemia Fund and the Leukemia Research Fund.

References

1 Rozman C, Montserrat E. Current concepts: chronic lymphocytic leukemia. *N Engl J Med* 1995; **333**: 1052–1057.
2 Kalil N, Cheson BD. Chronic lymphocytic leukemia. *Oncologist* 1999; **4**: 352–369.
3 Aalto Y, El-Rifai W, Vilpo L, Ollila J, Nagy B, Vihinen M *et al*. Distinct gene expression profiling in chronic lymphocytic leukemia with 11q23 deletion. *Leukemia* 2001; **15**: 1721–1728.
4 Stratowa C, Loffler G, Lichter P, Stilgenbauer S, Haberl P, Schweifer N *et al*. CDNA microarray gene expression analysis of B-cell chronic lymphocytic leukemia proposes potential new prognostic markers involved in lymphocyte trafficking. *Int J Cancer* 2001; **91**: 474–480.
5 Korz C, Pscherer A, Benner A, Mertens D, Schaffner C, Leupolt E *et al*. Evidence for distinct pathomechanisms in B-cell chronic lymphocytic leukemia and mantle cell lymphoma by quantitative expression analysis of cell cycle and apoptosis-associated genes. *Blood* 2002; **99**: 4554–4561.
6 Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A *et al*. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000; **403**: 503–511.
7 Voss T, Ahorn H, Harbel P, Wilgenbus K. Correlation of clinical data with proteomics profiles in 24 patients with B-cell chronic lymphocytic leukemia. *Int J Cancer* 2000; **91**: 180–186.
8 Pasquali C, Fialka I, Huber L-A. Subcellular fractionation, electromigration analysis and mapping of organelles. *J Chromatogr B Biomed Sci Appl* 1999; **722**: 89–102.
9 Eng J-K, McCormack A-L, Yates III J-R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein. *J Am Soc Mass Spectrom* 1994; **5**: 976–989.
10 Boyd RS, Duggan MJ, Shone CC, Foster KA. The effect of botulinum neurotoxins on the release of insulin from the insulinoma cell lines HIT-15 and RINm5F. *J Biol Chem* 1995; **270**: 18216–18218.
11 Pappin DJC, Hojrup P,Bleasby AJ. Rapid identification of proteins by peptide mass fingerprinting. *Curr Biol* 1993; **3**: 327–332, also http://www.hgmp.mrc.ac.uk/Bioinformatics/Webapp/mowse.
12 Adam PJ, Boyd R, Tyson KL, Fletcher GC, Stamps A, Hudson L *et al*. Comprehensive proteomic analysis of breast cancer cell membranes reveals unique proteins with potential roles in clinical cancer. *J Biol Chem* 2003; **278**: 6482–6489.
13 Belov L, de la Vega O, dos Remedios CG, Mulligan SP, Christopherson RI. Immunotyping of leukemias using a cluster of differentiation antibody microarray. *Cancer Res* 2001; **61**: 4483–4489.
14 Hulkkonen J, Vilpo L, Hurme M, Vilpo J. Surface antigen expression in chronic lymphocytic leukemia: clustering analysis, interrelationships and effects of chromosomal abnormalities. *Leukemia* 2002; **16**: 178–185.
15 Treumann A, Lifely MR, Schneider P, Ferguson MA. Primary structure of CD52. *J Biol Chem* 1995; **270**: 6088–6099.
16 Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. *J Mol Biol* 1997; **268**: 78–94.
17 Venkataraman C, Muthusamy N, Muthukkumar S, Bondada S. Activation of lyn, blk, and btk but not syk in CD72-stimulated B lymphocytes. *J Immunol* 1998; **160**: 3322–3329.
18 Majima S, Kajino K, Fukuda T, Otsuka F, Hino O. A novel gene 'Niban' upregulated in renal carcinogenesis cloning by the cDNA-amplified fragment length polymorphis approach. *Jpn J Cancer Res* 2000; **91**: 869–874.
19 Wick M, Burger C, Brusselbach S, Lucibello FC, Muller R. Identification of serum-inducible genes: different patterns of gene regulation during G0→S and G1→S progression. *J Cell Sci* 1994; **107**: 227–239.
20 Rogalski TM, Mullen GP, Gilbert MM, Williams BD, Moerman DG. The UNC-112 gene in *Caenorhabditis elegans* encodes a novel component of the cell–matrix adhesion structures required for integrin localization in the muscle membrane. *J Cell Biol* 2000; **150**: 253–264.
21 Mackinnon AC, Qadota H, Norman KR, Moerman DG, Williams BD. *C. elegans* PAT-4/ILK functions as an adaptor protein within integrin adhesion complexes. *Curr Biol* 2002; **14**: 787–797.