



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification <sup>7</sup> : C12N 15/12, C07K 14/705, 16/28, C12Q 1/68, A01K 67/027, C12N 15/62, A61K 38/00, G01N 33/566</p>	AI	<p>(11) International Publication Number: <b>WO 00/53747</b></p> <p>(43) International Publication Date: 14 September 2000 (14.09.00)</p>								
<p>(21) International Application Number: PCT/CA00/00259</p> <p>(22) International Filing Date: 9 March 2000 (09.03.00)</p> <p>(30) Priority Data:</p> <table border="0"> <tr> <td>60/124,260</td> <td>11 March 1999 (11.03.99)</td> <td>US</td> </tr> <tr> <td>60/127,386</td> <td>1 April 1999 (01.04.99)</td> <td>US</td> </tr> <tr> <td>60/144,919</td> <td>21 July 1999 (21.07.99)</td> <td>US</td> </tr> </table> <p>(71) Applicant (for all designated States except US): MOUNT SINAI HOSPITAL [CA/CA]; Samuel Lunenfeld Research Institute, Office of Technology Transfer &amp; Industrial Liaison, 600 University Avenue, Toronto, Ontario M5G 1X5 (CA).</p> <p>(72) Inventors; and</p> <p>(75) Inventors/Applicants (for US only): FOUSSIAS, George [CA/CA]; 18 Taylor Drive, Toronto, Ontario M4C 3B3 (CA). YOUSEF, George, M. [EG/CA]; Suite 1701, 50 Stephanie Street, Toronto, Ontario M5T 1B3 (CA). DIAMANDIS, Eleftherios, P. [CA/CA]; Suite 44, 1504 Gerrard Street W., Toronto, Ontario M5G 2X2 (CA).</p>	60/124,260	11 March 1999 (11.03.99)	US	60/127,386	1 April 1999 (01.04.99)	US	60/144,919	21 July 1999 (21.07.99)	US	<p>(74) Agents: VAN ZANT, Joan, M. et al.; Swabey Ogilvy Renault, Suite 1600, 1981 McGill College Avenue, Montréal, Québec H3A 2Y3 (CA).</p> <p>(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</p> <p><b>Published</b></p> <p><i>With international search report.</i></p> <p><i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p>
60/124,260	11 March 1999 (11.03.99)	US								
60/127,386	1 April 1999 (01.04.99)	US								
60/144,919	21 July 1999 (21.07.99)	US								
<p>(54) Title: SIALIC ACID-BINDING IG-LIKE LECTIN (SIGLEC) GENE; OB-BINDING PROTEIN LIKE (OB-BPL)</p>										
<p>(57) Abstract</p> <p>The invention relates to nucleic acid molecules, proteins encoded by such nucleic acid molecules; and use of the proteins and nucleic acid molecules.</p>										

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

<b>AL</b>	Albania	<b>ES</b>	Spain	<b>LS</b>	Lesotho	<b>SI</b>	Slovenia
<b>AM</b>	Armenia	<b>FI</b>	Finland	<b>LT</b>	Lithuania	<b>SK</b>	Slovakia
<b>AT</b>	Austria	<b>FR</b>	France	<b>LU</b>	Luxembourg	<b>SN</b>	Senegal
<b>AU</b>	Australia	<b>GA</b>	Gabon	<b>LV</b>	Latvia	<b>SZ</b>	Swaziland
<b>AZ</b>	Azerbaijan	<b>GB</b>	United Kingdom	<b>MC</b>	Monaco	<b>TD</b>	Chad
<b>BA</b>	Bosnia and Herzegovina	<b>GE</b>	Georgia	<b>MD</b>	Republic of Moldova	<b>TG</b>	Togo
<b>BB</b>	Barbados	<b>GH</b>	Ghana	<b>MG</b>	Madagascar	<b>TJ</b>	Tajikistan
<b>BE</b>	Belgium	<b>GN</b>	Guinea	<b>MK</b>	The former Yugoslav Republic of Macedonia	<b>TM</b>	Turkmenistan
<b>BF</b>	Burkina Faso	<b>GR</b>	Greece			<b>TR</b>	Turkey
<b>BG</b>	Bulgaria	<b>HU</b>	Hungary	<b>ML</b>	Mali	<b>TT</b>	Trinidad and Tobago
<b>BJ</b>	Benin	<b>IE</b>	Ireland	<b>MN</b>	Mongolia	<b>UA</b>	Ukraine
<b>BR</b>	Brazil	<b>IL</b>	Israel	<b>MR</b>	Mauritania	<b>UG</b>	Uganda
<b>BY</b>	Belarus	<b>IS</b>	Iceland	<b>MW</b>	Malawi	<b>US</b>	United States of America
<b>CA</b>	Canada	<b>IT</b>	Italy	<b>MX</b>	Mexico	<b>UZ</b>	Uzbekistan
<b>CF</b>	Central African Republic	<b>JP</b>	Japan	<b>NE</b>	Niger	<b>VN</b>	Viet Nam
<b>CG</b>	Congo	<b>KE</b>	Kenya	<b>NL</b>	Netherlands	<b>YU</b>	Yugoslavia
<b>CH</b>	Switzerland	<b>KG</b>	Kyrgyzstan	<b>NO</b>	Norway	<b>ZW</b>	Zimbabwe
<b>CI</b>	Côte d'Ivoire	<b>KP</b>	Democratic People's Republic of Korea	<b>NZ</b>	New Zealand		
<b>CM</b>	Cameroon			<b>PL</b>	Poland		
<b>CN</b>	China	<b>KR</b>	Republic of Korea	<b>PT</b>	Portugal		
<b>CU</b>	Cuba	<b>KZ</b>	Kazakstan	<b>RO</b>	Romania		
<b>CZ</b>	Czech Republic	<b>LC</b>	Saint Lucia	<b>RU</b>	Russian Federation		
<b>DE</b>	Germany	<b>LI</b>	Liechtenstein	<b>SD</b>	Sudan		
<b>DK</b>	Denmark	<b>LK</b>	Sri Lanka	<b>SE</b>	Sweden		
<b>EE</b>	Estonia	<b>LR</b>	Liberia	<b>SG</b>	Singapore		

SIALIC ACID-BINDING IG-LIKE LECTIN (SIGLEC) GENE; OB-BINDING PROTEIN LIKE (OB-BPL)

**FIELD OF THE INVENTION**

The invention relates to nucleic acid molecules, proteins encoded by such nucleic acid molecules; and use of the proteins and nucleic acid molecules

**5 BACKGROUND OF THE INVENTION**

The immunoglobulin superfamily (IgSF) encompasses a large number of cell surface molecules which play a vital role not only in immunity, but also in controlling the behaviour of cells in various tissues, through their ability to mediate cell surface recognition events. These molecules are characterized by the presence of at least one immunoglobulin (Ig) domain, a sandwich of two  $\beta$ -sheets stabilized by a conserved disulfide bond.

10 The core of this domain is composed of  $\beta$ -strands A,B,E in one sheet and G,F,C in the other, and arise from the ends of the domain sequence (Williams and Barclay 1988). In between, however, there is a great deal of sequence length variation. Such Ig domains occur in two types, the V-set and the C-set, and can be distinguished based on patterns of conserved amino acid residues responsible for forming the characteristic  $\beta$ -sheet sandwich. V-set domains consist of about 65-75 amino acid residues between conserved cysteines,  
15 whereas C-set domains have about 55-60 residues (reviewed in (Williams and Barclay 1988)). The C-set domains can be further divided into C1- and C2-sets, and are distinguished by the fact that, although showing signs of a C-set domain, the latter half of C2-set domains exhibit sequence patterns more homologous to V-set rather than C1-set domains (Williams *et al.*, 1989).

Recently, a novel family of structurally related IgSF molecules have been identified, which mediate  
20 protein-carbohydrate interactions through specific interactions with sialic acid-containing glycoproteins and glycolipids (Crocker *et al.*, 1996). This family was originally referred to as the sialoadhesins, but has recently been designated the sialic acid-binding Ig-like lectin (Siglec) family (Crocker *et al.*, 1998). These molecules are characterized by the presence of one N-terminal V-set domain, and a variable number of downstream C2-set domains, ranging from 16 in sialoadhesin to 1 in CD33 (Crocker *et al.*, 1996). Furthermore, these Ig-like  
25 domains possess some unique features. In the V-set domain, the conserved cysteine in  $\beta$ -strand F of classic V-set domains is absent, while a highly conserved cysteine is present in  $\beta$ -strand E in all siglecs identified so far. This results in the cysteines in  $\beta$ -strands B and E being next to each other in one  $\beta$ -sheet, which likely results in an intrasheet disulfide bond (Crocker *et al.*, 1996; Williams *et al.*, 1989). There is also an additional highly conserved cysteine residue in both the V-set and first C2-set domains of all siglecs. In the V-set domain  
30 it is located at the beginning of  $\beta$ -strand B, while in the C2-set domain it is found between  $\beta$ -strands B and C. These two additional cysteines have been found to form an interdomain disulfide bond, a feature unique to siglecs (Crocker *et al.*, 1996; Pedraza *et al.*, 1990).

Currently, the siglec family consists of sialoadhesin (Siglec-1), CD22 (Siglec-2), CD33 (Siglec-3), myelin-associated glycoprotein (MAG) (Siglec-4a), Schwann cell myelin protein (SMP) (Siglec-4b), OB-  
35 binding protein 2 (Siglec-5), OB-binding protein 1 (Siglec-6), and p75/AIRM1 (Siglec-7) (Cornish *et al.*, 1998; Crocker *et al.*, 1998; Falco *et al.*, 1999; Nicoll *et al.*, 1999; Patel *et al.*, 1999). Each member of the

Siglec family is expressed by specific cell types and exhibits a distinct function. Sialoadhesin is a macrophage-restricted adhesion molecule (Crocker *et al.*, 1994), CD22 is B lymphocyte-specific and regulates its activation (Stamenkovic and Seed 1990), CD33 is a myeloid-specific inhibitory receptor (Ulyanova *et al.*, 1999), and MAG functions in the formation and maintenance of axonal myelin structure (Li *et al.*, 1998). Siglec-5 and -6 (OB-BP2 and -BP1, respectively) are expressed in several tissues including placenta and peripheral blood leukocytes, and have shown an *in vitro* ability to bind leptin (Cornish *et al.*, 1998; Patel *et al.*, 1999), while OB-BPL (p75/AIRM1) is an inhibitory receptor expressed predominantly on human natural killer cells (Falco *et al.*, 1999; Nicoll *et al.*, 1999).

#### SUMMARY OF THE INVENTION

The present inventors have identified and characterized a gene encoding a novel member of the siglec family (OB-binding protein like or OB-BPL). The putative protein product displays a high degree of homology with siglec-7, as well as with siglec-5 and siglec-6. Further, it possesses all the structural features found in other siglecs. The gene was localized to 19q13.4, 43.19 Kb more telomeric than KLK-L6 (a member of the kallikrein gene family) through genomic sequencing data and restriction mapping with EcoRI. The novel siglec is encoded by 7 exons, with six intervening introns. In addition, it is highly expressed in bone marrow, placenta, spleen, and fetal liver, as well as other tissues at lower levels.

The OB-BPL protein described herein is referred to as "OB-BPL Protein". The gene encoding the protein is referred to as "*ob-bpl*".

Broadly stated the present invention relates to an isolated nucleic acid molecule which comprises:

- (i) a nucleic acid sequence encoding a protein having substantial sequence identity with an amino acid sequence of OB-BPL as shown in Table 5 or SEQ.ID.NO. 2 or 3;
- (ii) a nucleic acid sequence encoding a protein comprising an amino acid sequence of OB-BPL as shown in Table 5 or SEQ.ID.NO. 2 or 3;
- (iii) nucleic acid sequences complementary to (i);
- (iv) a degenerate form of a nucleic acid sequence of (i);
- (v) a nucleic acid sequence capable of hybridizing under stringent conditions to a nucleic acid sequence in (i), (ii) or (iii);
- (vi) a nucleic acid sequence encoding a truncation, an analog, an allelic or species variation of a protein comprising an amino acid sequence of OB-BPL as shown in Table 5 or SEQ.ID.NO. 2 or 3; or
- (vii) a fragment, or allelic or species variation of (i), (ii) or (iii).

Preferably, a purified and isolated nucleic acid molecule of the invention comprises:

- (i) a nucleic acid sequence comprising the sequence of SEQ.ID.NO. 1 wherein T can also be U;
- (ii) nucleic acid sequences complementary to (i), preferably complementary to the full nucleic acid sequence of SEQ.ID.NO. 1;
- (iii) a nucleic acid capable of hybridizing under stringent conditions to a nucleic acid of (i) or (ii) and preferably having at least 18 nucleotides; or

- (iv) a nucleic acid molecule differing from any of the nucleic acids of (i) to (iii) in codon sequences due to the degeneracy of the genetic code.

The invention also contemplates a nucleic acid molecule comprising a sequence encoding a truncation of an OB-BPL Protein, an analog, or a homolog of an OB-BPL Protein or a truncation thereof. (OB-BPL Protein and truncations, analogs and homologs of OB-BPL Protein are also collectively referred to herein as "OB-BPL Related Proteins").

The nucleic acid molecules of the invention may be inserted into an appropriate expression vector, i.e. a vector that contains the necessary elements for the transcription and translation of the inserted coding sequence. Accordingly, recombinant expression vectors adapted for transformation of a host cell may be constructed which comprise a nucleic acid molecule of the invention and one or more transcription and translation elements linked to the nucleic acid molecule.

The recombinant expression vector can be used to prepare transformed host cells expressing OB-BPL Related Proteins. Therefore, the invention further provides host cells containing a recombinant molecule of the invention. The invention also contemplates transgenic non-human mammals whose germ cells and somatic cells contain a recombinant molecule comprising a nucleic acid molecule of the invention, in particular one which encodes an analog of the OB-BPL Protein, or a truncation of the OB-BPL Protein.

The invention further provides a method for preparing OB-BPL Related Proteins utilizing the purified and isolated nucleic acid molecules of the invention. In an embodiment a method for preparing an OB-BPL Related Protein is provided comprising (a) transferring a recombinant expression vector of the invention into a host cell; (b) selecting transformed host cells from untransformed host cells; (c) culturing a selected transformed host cell under conditions which allow expression of the OB-BPL Related Protein; and (d) isolating the OB-BPL Related Protein.

The invention further broadly contemplates an isolated OB-BPL Protein comprising an amino acid sequence as shown in SEQ.ID.NO. 2 or 3.

The OB-BPL Related Proteins of the invention may be conjugated with other molecules, such as proteins, to prepare fusion proteins. This may be accomplished, for example, by the synthesis of N-terminal or C-terminal fusion proteins.

The invention further contemplates antibodies having specificity against an epitope of an OB-BPL Related Protein of the invention. Antibodies may be labeled with a detectable substance and used to detect proteins of the invention in tissues and cells.

The invention also permits the construction of nucleotide probes which are unique to the nucleic acid molecules of the invention and/or to proteins of the invention. Therefore, the invention also relates to a probe comprising a nucleic acid sequence of the invention, or a nucleic acid sequence encoding a protein of the invention, or a part thereof. The probe may be labeled, for example, with a detectable substance and it may be used to select from a mixture of nucleotide sequences a nucleic acid molecule of the invention including nucleic acid molecules coding for a protein which displays one or more of the properties of a protein of the invention.

The invention still further provides a method for identifying a substance which binds to a protein

of the invention comprising reacting the protein with at least one substance which potentially can bind with the protein, under conditions which permit the formation of complexes between the substance and protein and detecting binding. Binding may be detected by assaying for complexes, for free substance, or for non-complexed protein. The invention also contemplates methods for identifying substances that bind to other intracellular proteins that interact with an OB-BPL Related Protein. Methods can also be utilized which identify compounds which bind to OB-BPL gene regulatory sequences (e.g. promoter sequences).

Still further the invention provides a method for evaluating a compound for its ability to modulate the biological activity of an OB-BPL Related Protein of the invention. For example a substance which inhibits or enhances the interaction of the protein and a substance which binds to the protein may be evaluated. In an embodiment, the method comprises providing a known concentration of an OB-BPL Related Protein, with a substance which binds to the protein and a test compound under conditions which permit the formation of complexes between the substance and protein, and removing and/or detecting complexes.

Compounds which modulate the biological activity of a protein of the invention may also be identified using the methods of the invention by comparing the pattern and level of expression of the protein of the invention in tissues and cells, in the presence, and in the absence of the compounds.

The proteins of the invention and substances and compounds identified using the methods of the invention, and peptides of the invention may be used to modulate the biological activity of an OB-BPL Related Protein of the invention, and they may be used in the treatment of conditions such a disorders of the hematopoietic system and in particular leukemias. Accordingly, the substances and compounds may be formulated into compositions for administration to individuals suffering from a disorders of the hematopoietic system.

Therefore, the present invention also relates to a composition comprising one or more of a protein of the invention, a peptide of the invention, or a substance or compound identified using the methods of the invention, and a pharmaceutically acceptable carrier, excipient or diluent. A method for treating or preventing cancer or a disorder of the hematopoietic system is also provided comprising administering to a patient in need thereof, an OB-BPL Related Protein of the invention, or a composition of the invention.

Other objects, features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and the specific examples while indicating preferred embodiments of the invention are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from this detailed description.

#### **BRIEF DESCRIPTION OF THE DRAWINGS**

The invention will now be described in relation to the drawings in which:

Figure 1: Genomic Structure of a Novel Siglec. Shown are the exon/intron boundaries, as well as the predicted protein sequence. The single underlined region is the 5' untranslated region, and the double underlined region is the 3' untranslated region. In the shaded box is the putative polyadenylation signal.

Figure 2: Hydrophobicity Plot of the Novel Siglec. This shows the regions of the putative novel siglec protein which contain stretches of hydrophobic amino acid residues. As is evident, there are two such regions, the first corresponding to the signal peptide, and the second, at around residues 350-370, the putative transmembrane region.

5 Figure 3: Localization of the Novel Siglec Gene. The physical map of the genomic area around chromosome 19q13.3-q13.4 where the kallikrein gene family resides. Seven additional kallikreins map in the 132.1 Kb region (data not shown; see (Diamandis *et al.*, 1999)). Gene lengths are presented above each arrow, and distances between genes are shown below. Arrows denote the direction of transcription. The novel siglec gene resides 43.2 Kb telomeric to the KLK-L6 gene. KLK, kallikrein; PSA, prostate specific antigen; KLK-L, kallikrein-like; NES1, normal epithelial cell-specific 1 gene; TLSP, trypsin-like serine protease.

10 Figure 4: Siglec Family Multiple Alignment. The novel siglec was aligned with siglec-5 to -7 and CD33, using ClustalX (Jeanmougin *et al.*, 1998) (SEQ. ID. NOs. 10-13). The signal peptide was determined through computer prediction, and the Ig domain boundaries were assigned based on exon boundaries. The transmembrane domain was also predicted, while taking into consideration exon boundaries as well. The ITIM-like and SLAM-like motifs are indicated, as are the conserved cysteines (\*) which form the disulfide bonds of the Ig-like domains in siglecs, and the conserved arginine and aromatic residues ( $\alpha$ ) which are responsible for sialic acid binding and specificity.

15 Figure 5: Phylogenetic Analysis of the Siglec Family. The phylogenetic tree was created using ClustalX (Jeanmougin *et al.*, 1998) and TreeView (Page 1996). As is evident, siglec-7 and the novel siglec are very closely related, and they are both related to CD33, in addition to a more distant relation to the other siglecs.

20 Figure 6: Tissue Expression Profile of the Novel Siglec. RT-PCR was performed on 28 tissue total RNAs, for this novel siglec and actin (control gene). The novel siglec is highly expressed in bone marrow, placenta, spleen, and fetal liver. There is also a lower degree of expression in many of the other tissues, while it is absent in ovary, pancreas, skeletal muscle, and heart.

#### **DETAILED DESCRIPTION OF THE INVENTION**

In accordance with the present invention there may be employed conventional molecular biology, microbiology, and recombinant DNA techniques within the skill of the art. Such techniques are explained fully in the literature. See for example, Sambrook, Fritsch, & Maniatis, *Molecular Cloning: A Laboratory Manual*, Second Edition (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y); DNA Cloning: A Practical Approach, Volumes I and II (D.N. Glover ed. 1985); Oligonucleotide Synthesis (M.J. Gait ed. 1984); Nucleic Acid Hybridization B.D. Hames & S.J. Higgins eds. (1985); Transcription and Translation B.D. Hames & S.J. Higgins eds (1984); Animal Cell Culture R.I. Freshney, ed. (1986); Immobilized Cells and enzymes IRL Press, (1986); and B. Perbal, *A Practical Guide to Molecular Cloning* (1984).

#### **1. Nucleic Acid Molecules of the Invention**

As hereinbefore mentioned, the invention provides an isolated nucleic acid molecule having a

sequence encoding an OB-BPL Protein. The term "isolated" refers to a nucleic acid substantially free of cellular material or culture medium when produced by recombinant DNA techniques, or chemical reactants, or other chemicals when chemically synthesized. An "isolated" nucleic acid may also be free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid molecule) from which the nucleic acid is derived. The term "nucleic acid" is intended to include DNA and RNA and can be either double stranded or single stranded. In an embodiment, a nucleic acid molecule encodes an OB-BPL Protein comprising an amino acid sequence as shown in SEQ.ID.NO. 2 or 3, preferably a nucleic acid molecule comprising a nucleic acid sequence as shown in SEQ.ID.NO. 1.

The invention includes nucleic acid sequences complementary to a nucleic acid encoding an OB-BPL Protein comprising an amino acid sequence as shown in SEQ.ID.NO. 2 or 3, preferably the nucleic acid sequences complementary to a full nucleic acid sequence shown in SEQ.ID.NO. 1.

The invention includes nucleic acid molecules having substantial sequence identity or homology to nucleic acid sequences of the invention or encoding proteins having substantial identity or similarity to the amino acid sequence shown in SEQ.ID.NO. 2 or 3. Preferably, the nucleic acids have substantial sequence identity for example at least 65%, 70%, 75%, 80%, or 85% nucleic acid identity; more preferably 90% nucleic acid identity; and most preferably at least 95%, 96%, 97%, 98%, or 99% sequence identity. "Identity" as known in the art and used herein, is a relationship between two or more amino acid sequences or two or more nucleic acid sequences, as determined by comparing the sequences. It also refers to the degree of sequence relatedness between amino acid or nucleic acid sequences, as the case may be, as determined by the match between strings of such sequences. Identity and similarity are well known terms to skilled artisans and they can be calculated by conventional methods (for example see Computational Molecular Biology, Lesk, A.M. ed., Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D.W. ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part I, Griffin, A.M. and Griffin, H.G. eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G. Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J. eds. M. Stockton Press, New York, 1991, Carillo, H. and Lipman, D., SIAM J. Applied Math. 48:1073, 1988). Methods which are designed to give the largest match between the sequences are generally preferred. Methods to determine identity and similarity are codified in publicly available computer programs including the GCG program package (Devereux J. et al., Nucleic Acids Research 12(1): 387, 1984); BLASTP, BLASTN, and FASTA (Atschul, S.F. et al. J. Molec. Biol. 215: 403-410, 1990). The BLAST X program is publicly available from NCBI and other sources (BLAST Manual, Altschul, S. et al. NCBI NLM NIH Bethesda, Md. 20894; Altschul, S. et al. J. Mol. Biol. 215: 403-410, 1990).

Isolated nucleic acid molecules encoding an OB-BPL Protein, and having a sequence which differs from a nucleic acid sequence of the invention due to degeneracy in the genetic code are also within the scope of the invention. Such nucleic acids encode functionally equivalent proteins (e.g., an OB-BPL Protein) but differ in sequence from the sequence of an OB-BPL Protein due to degeneracy in the genetic code. As one example, DNA sequence polymorphisms within the nucleotide sequence of an OB-BPL Protein may result in silent mutations which do not affect the amino acid sequence. Variations in one or

more nucleotides may exist among individuals within a population due to natural allelic variation. Any and all such nucleic acid variations are within the scope of the invention. DNA sequence polymorphisms may also occur which lead to changes in the amino acid sequence of an OB-BPL Protein. These amino acid polymorphisms are also within the scope of the present invention.

5           Another aspect of the invention provides a nucleic acid molecule which hybridizes under stringent conditions, preferably high stringency conditions to a nucleic acid molecule which comprises a sequence which encodes an OB-BPL Protein having an amino acid sequence shown in SEQ.ID.NO. 2 or 3. Appropriate stringency conditions which promote DNA hybridization are known to those skilled in the art, or can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6.  
10 For example, 6.0 x sodium chloride/sodium citrate (SSC) at about 45°C, followed by a wash of 2.0 x SSC at 50°C may be employed. The stringency may be selected based on the conditions used in the wash step. By way of example, the salt concentration in the wash step can be selected from a high stringency of about 0.2 x SSC at 50°C. In addition, the temperature in the wash step can be at high stringency conditions, at about 65°C.

15           It will be appreciated that the invention includes nucleic acid molecules encoding an OB-BPL Related Protein including truncations of an OB-BPL Protein, and analogs of an OB-BPL Protein as described herein. The truncated nucleic acids or nucleic acid fragments may correspond to the transmembrane domain, cytoplasmic domain, IG domains, or ITIM-like or SLAM-like motifs as described in Table 4 and in Figure 4. It will further be appreciated that variant forms of the nucleic acid molecules  
20 of the invention which arise by alternative splicing of an mRNA corresponding to a cDNA of the invention are encompassed by the invention.

          An isolated nucleic acid molecule of the invention which comprises DNA can be isolated by preparing a labelled nucleic acid probe based on all or part of a nucleic acid sequence of the invention. The labeled nucleic acid probe is used to screen an appropriate DNA library (e.g. a cDNA or genomic DNA  
25 library). For example, a cDNA library can be used to isolate a cDNA encoding an OB-BPL Related Protein by screening the library with the labeled probe using standard techniques. Alternatively, a genomic DNA library can be similarly screened to isolate a genomic clone encompassing a gene encoding an OB-BPL Related Protein. Nucleic acids isolated by screening of a cDNA or genomic DNA library can be sequenced by standard techniques.

30           An isolated nucleic acid molecule of the invention which is DNA can also be isolated by selectively amplifying a nucleic acid encoding an OB-BPL Related Protein using the polymerase chain reaction (PCR) methods and cDNA or genomic DNA. It is possible to design synthetic oligonucleotide primers from the nucleotide sequence of the invention for use in PCR. A nucleic acid can be amplified from cDNA or genomic DNA using these oligonucleotide primers and standard PCR amplification techniques.  
35 The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. cDNA may be prepared from mRNA, by isolating total cellular mRNA by a variety of techniques, for example, by using the guanidinium-thiocyanate extraction procedure of Chirgwin et al., Biochemistry, 18, 5294-5299 (1979). cDNA is then synthesized from the mRNA using reverse transcriptase (for example,

Moloney MLV reverse transcriptase available from Gibco/BRL, Bethesda, MD, or AMV reverse transcriptase available from Seikagaku America, Inc., St. Petersburg, FL).

An isolated nucleic acid molecule of the invention which is RNA can be isolated by cloning a cDNA encoding an OB-BPL Related Protein into an appropriate vector which allows for transcription of the cDNA to produce an RNA molecule which encodes an OB-BPL Related Protein. For example, a cDNA can be cloned downstream of a bacteriophage promoter, (e.g. a T7 promoter) in a vector, cDNA can be transcribed *in vitro* with T7 polymerase, and the resultant RNA can be isolated by conventional techniques.

Nucleic acid molecules of the invention may be chemically synthesized using standard techniques. Methods of chemically synthesizing polydeoxynucleotides are known, including but not limited to solid-phase synthesis which, like peptide synthesis, has been fully automated in commercially available DNA synthesizers (See e.g., Itakura et al. U.S. Patent No. 4,598,049; Caruthers et al. U.S. Patent No. 4,458,066; and Itakura U.S. Patent Nos. 4,401,796 and 4,373,071).

Determination of whether a particular nucleic acid molecule encodes an OB-BPL Related Protein can be accomplished by expressing the cDNA in an appropriate host cell by standard techniques, and testing the expressed protein in the methods described herein. A cDNA encoding an OB-BPL Related Protein can be sequenced by standard techniques, such as dideoxynucleotide chain termination or Maxam-Gilbert chemical sequencing, to determine the nucleic acid sequence and the predicted amino acid sequence of the encoded protein.

The initiation codon and untranslated sequences of an OB-BPL Related Protein may be determined using computer software designed for the purpose, such as PC/Gene (IntelliGenetics Inc., Calif.). The intron-exon structure and the transcription regulatory sequences of a gene encoding an OB-BPL Related Protein may be confirmed by using a nucleic acid molecule of the invention encoding an OB-BPL Related Protein to probe a genomic DNA clone library. Regulatory elements can be identified using standard techniques. The function of the elements can be confirmed by using these elements to express a reporter gene such as the lacZ gene which is operatively linked to the elements. These constructs may be introduced into cultured cells using conventional procedures or into non-human transgenic animal models. In addition to identifying regulatory elements in DNA, such constructs may also be used to identify nuclear proteins interacting with the elements, using techniques known in the art.

In a particular embodiment of the invention, the nucleic acid molecules isolated using the methods described herein are mutant *OB-BPL* gene alleles. The mutant alleles may be isolated from individuals either known or proposed to have a genotype which contributes to the symptoms of a disorder of the hematopoietic system (e.g. leukemias). Mutant alleles and mutant allele products may be used in therapeutic and diagnostic methods described herein. For example, a cDNA of a mutant *OB-BPL* gene may be isolated using PCR as described herein, and the DNA sequence of the mutant allele may be compared to the normal allele to ascertain the mutation(s) responsible for the loss or alteration of function of the mutant gene product. A genomic library can also be constructed using DNA from an individual suspected of or known to carry a mutant allele, or a cDNA library can be constructed using RNA from tissue known,

or suspected to express the mutant allele. A nucleic acid encoding a normal *OB-BPL* gene or any suitable fragment thereof, may then be labeled and used as a probe to identify the corresponding mutant allele in such libraries. Clones containing mutant sequences can be purified and subjected to sequence analysis. In addition, an expression library can be constructed using cDNA from RNA isolated from a tissue of an individual known or suspected to express a mutant *OB-BPL* allele. Gene products made by the putatively mutant tissue may be expressed and screened, for example using antibodies specific for an OB-BPL Related Protein as described herein. Library clones identified using the antibodies can be purified and subjected to sequence analysis.

The sequence of a nucleic acid molecule of the invention, or a fragment of the molecule, may be inverted relative to its normal presentation for transcription to produce an antisense nucleic acid molecule. An antisense nucleic acid molecule may be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art.

## 2. Proteins of the Invention

An amino acid sequence of an OB-BPL Protein comprises a sequence as shown in SEQ.ID.NO. 2 or 3. The protein is highly expressed in bone marrow, placenta, spleen, and fetal liver.

In addition to proteins comprising an amino acid sequence as shown in SEQ.ID.NO. 2 or 3, the proteins of the present invention include truncations of an OB-BPL Protein, analogs of an OB-BPL Protein, and proteins having sequence identity or similarity to an OB-BPL Protein, and truncations thereof as described herein (i.e. OB-BPL Related Proteins). Truncated proteins may comprise peptides of between 3 and 70 amino acid residues, ranging in size from a tripeptide to a 70 mer polypeptide.

The truncated proteins may have an amino group (-NH<sub>2</sub>), a hydrophobic group (for example, carbobenzoxy, dansyl, or T-butyloxycarbonyl), an acetyl group, a 9-fluorenylmethoxy-carbonyl (PMOC) group, or a macromolecule including but not limited to lipid-fatty acid conjugates, polyethylene glycol, or carbohydrates at the amino terminal end. The truncated proteins may have a carboxyl group, an amido group, a T-butyloxycarbonyl group, or a macromolecule including but not limited to lipid-fatty acid conjugates, polyethylene glycol, or carbohydrates at the carboxy terminal end.

The proteins of the invention may also include analogs of an OB-BPL Protein, and/or truncations thereof as described herein, which may include, but are not limited to an OB-BPL Protein, containing one or more amino acid substitutions, insertions, and/or deletions. Amino acid substitutions may be of a conserved or non-conserved nature. Conserved amino acid substitutions involve replacing one or more amino acids of an OB-BPL Protein amino acid sequence with amino acids of similar charge, size, and/or hydrophobicity characteristics. When only conserved substitutions are made the resulting analog is preferably functionally equivalent to an OB-BPL Protein. Non-conserved substitutions involve replacing one or more amino acids of the OB-BPL Protein amino acid sequence with one or more amino acids which possess dissimilar charge, size, and/or hydrophobicity characteristics.

One or more amino acid insertions may be introduced into an OB-BPL Protein. Amino acid insertions may consist of single amino acid residues or sequential amino acids ranging from 2 to 15 amino acids in length.

Deletions may consist of the removal of one or more amino acids, or discrete portions from an OB-BPL Protein sequence. The deleted amino acids may or may not be contiguous. The lower limit length of the resulting analog with a deletion mutation is about 10 amino acids, preferably 20 to 40 amino acids.

5 The proteins of the invention include proteins with sequence identity or similarity to an OB-BPL Protein and/or truncations thereof as described herein. Such OB-BPL Proteins include proteins whose amino acid sequences are comprised of the amino acid sequences of OB-BPL Protein regions from other species that hybridize under selected hybridization conditions (see discussion of stringent hybridization conditions herein) with a probe used to obtain an OB-BPL Protein. These proteins will generally have the same regions which are characteristic of an OB-BPL Protein. Preferably a protein will have substantial  
10 sequence identity for example, about 65%, 70%, 75%, 80%, or 85% identity, preferably 90% identity, more preferably at least 95%, 96%, 97%, 98%, or 99% identity, and most preferably 98% identity with an amino acid sequence shown in SEQ.ID.NO. 2 or 3. A percent amino acid sequence homology, similarity or identity is calculated as the percentage of aligned amino acids that match the reference sequence using known methods as described herein.

15 The invention also contemplates isoforms of the proteins of the invention. An isoform contains the same number and kinds of amino acids as a protein of the invention, but the isoform has a different molecular structure. Isoforms contemplated by the present invention preferably have the same properties as a protein of the invention as described herein.

The present invention also includes OB-BPL Related Proteins conjugated with a selected protein,  
20 or a marker protein (see below) to produce fusion proteins. Additionally, immunogenic portions of an OB-BPL Protein and an OB-BPL Protein Related Protein are within the scope of the invention.

AN OB-BPL Related Protein of the invention may be prepared using recombinant DNA methods. Accordingly, the nucleic acid molecules of the present invention having a sequence which encodes an OB-BPL Related Protein of the invention may be incorporated in a known manner into an appropriate  
25 expression vector which ensures good expression of the protein. Possible expression vectors include but are not limited to cosmids, plasmids, or modified viruses (e.g. replication defective retroviruses, adenoviruses and adeno-associated viruses), so long as the vector is compatible with the host cell used.

The invention therefore contemplates a recombinant expression vector of the invention containing a nucleic acid molecule of the invention, and the necessary regulatory sequences for the transcription and  
30 translation of the inserted protein-sequence. Suitable regulatory sequences may be derived from a variety of sources, including bacterial, fungal, viral, mammalian, or insect genes [For example, see the regulatory sequences described in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990)]. Selection of appropriate regulatory sequences is dependent on the host cell chosen as discussed below, and may be readily accomplished by one of ordinary skill in the art. The  
35 necessary regulatory sequences may be supplied by the native OB-BPL Protein and/or its flanking regions.

The invention further provides a recombinant expression vector comprising a DNA nucleic acid molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is linked to a regulatory sequence in a manner which allows for expression, by transcription of

the DNA molecule, of an RNA molecule which is antisense to the nucleic acid sequence of a protein of the invention or a fragment thereof. Regulatory sequences linked to the antisense nucleic acid can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance a viral promoter and/or enhancer, or regulatory sequences can be chosen which direct tissue or cell type specific expression of antisense RNA.

The recombinant expression vectors of the invention may also contain a marker gene which facilitates the selection of host cells transformed or transfected with a recombinant molecule of the invention. Examples of marker genes are genes encoding a protein such as G418 and hygromycin which confer resistance to certain drugs,  $\beta$ -galactosidase, chloramphenicol acetyltransferase, firefly luciferase, or an immunoglobulin or portion thereof such as the Fc portion of an immunoglobulin preferably IgG. The markers can be introduced on a separate vector from the nucleic acid of interest.

The recombinant expression vectors may also contain genes which encode a fusion moiety which provides increased expression of the recombinant protein; increased solubility of the recombinant protein; and aid in the purification of the target recombinant protein by acting as a ligand in affinity purification. For example, a proteolytic cleavage site may be added to the target recombinant protein to allow separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Typical fusion expression vectors include pGEX (Amrad Corp., Melbourne, Australia), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the recombinant protein.

The recombinant expression vectors may be introduced into host cells to produce a transformant host cell. "Transformant host cells" include host cells which have been transformed or transfected with a recombinant expression vector of the invention. The terms "transformed with", "transfected with", "transformation" and "transfection" encompass the introduction of a nucleic acid (e.g. a vector) into a cell by one of many standard techniques. Prokaryotic cells can be transformed with a nucleic acid by, for example, electroporation or calcium-chloride mediated transformation. A nucleic acid can be introduced into mammalian cells via conventional techniques such as calcium phosphate or calcium chloride coprecipitation, DEAE-dextran-mediated transfection, lipofectin, electroporation or microinjection. Suitable methods for transforming and transfecting host cells can be found in Sambrook et al. (Molecular Cloning: A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory press (1989)), and other laboratory textbooks.

Suitable host cells include a wide variety of prokaryotic and eukaryotic host cells. For example, the proteins of the invention may be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus), yeast cells, or mammalian cells. Other suitable host cells can be found in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1991).

A host cell may also be chosen which modulates the expression of an inserted nucleic acid sequence, or modifies (e.g. glycosylation or phosphorylation) and processes (e.g. cleaves) the protein in a desired fashion. Host systems or cell lines may be selected which have specific and characteristic mechanisms for post-translational processing and modification of proteins. For example, eukaryotic host

cells including CHO, VERO, BHK, HeLA, COS, MDCK, 293, 3T3, and WI38 may be used. For long-term high-yield stable expression of the protein, cell lines and host systems which stably express the gene product may be engineered.

5 Host cells and in particular cell lines produced using the methods described herein may be particularly useful in screening and evaluating compounds that modulate the activity of an OB-BPL Related Protein.

10 The proteins of the invention may also be expressed in non-human transgenic animals including but not limited to mice, rats, rabbits, guinea pigs, micro-pigs, goats, sheep, pigs, non-human primates (e.g. baboons, monkeys, and chimpanzees) [see Hammer et al. (Nature 315:680-683, 1985), Palmiter et al. (Science 222:809-814, 1983), Brinster et al. (Proc Natl. Acad. Sci USA 82:44384442, 1985), Palmiter and Brinster (Cell. 41:343-345, 1985) and U.S. Patent No. 4,736,866]. Procedures known in the art may be used to introduce a nucleic acid molecule of the invention encoding an OB-BPL Related Protein into animals to produce the founder lines of transgenic animals. Such procedures include pronuclear microinjection, retrovirus mediated gene transfer into germ lines, gene targeting in embryonic stem cells, 15 electroporation of embryos, and sperm-mediated gene transfer.

The present invention contemplates a transgenic animal that carries the *OB-BPL* gene in all their cells, and animals which carry the transgene in some but not all their cells. The transgene may be integrated as a single transgene or in concatamers. The transgene may be selectively introduced into and activated in specific cell types (See for example, Lasko et al, 1992 Proc. Natl. Acad. Sci. USA 89: 6236). The transgene 20 may be integrated into the chromosomal site of the endogenous gene by gene targeting. The transgene may be selectively introduced into a particular cell type inactivating the endogenous gene in that cell type (See Gu et al Science 265: 103-106).

The expression of a recombinant OB-BPL Related Protein in a transgenic animal may be assayed using standard techniques. Initial screening may be conducted by Southern Blot analysis, or PCR methods 25 to analyze whether the transgene has been integrated. The level of mRNA expression in the tissues of transgenic animals may also be assessed using techniques including Northern blot analysis of tissue samples, *in situ* hybridization, and RT-PCR. Tissue may also be evaluated immunocytochemically using antibodies against OB-BPL Protein.

30 Proteins of the invention may also be prepared by chemical synthesis using techniques well known in the chemistry of proteins such as solid phase synthesis (Merrifield, 1964, J. Am. Chem. Assoc. 85:2149-2154) or synthesis in homogenous solution (Houbenweyl, 1987, Methods of Organic Chemistry, ed. E. Wansch, Vol. 15 I and II, Thieme, Stuttgart).

35 N-terminal or C-terminal fusion proteins comprising an OB-BPL Related Protein of the invention conjugated with other molecules, such as proteins, may be prepared by fusing, through recombinant techniques, the N-terminal or C-terminal of an OB-BPL Related Protein, and the sequence of a selected protein or marker protein with a desired biological function. The resultant fusion proteins contain OB-BPL Protein fused to the selected protein or marker protein as described herein. Examples of proteins which may be used to prepare fusion proteins include immunoglobulins, glutathione-S-transferase (GST),

hemagglutinin (HA), and truncated myc.

### 3. Antibodies

OB-BPL Related Proteins of the invention can be used to prepare antibodies specific for the proteins. Antibodies can be prepared which bind a distinct epitope in an unconserved region of the protein.

5 An unconserved region of the protein is one that does not have substantial sequence homology to other proteins. A region from a conserved region such as a well-characterized domain can also be used to prepare an antibody to a conserved region of an OB-BPL Related Protein. Antibodies having specificity for an OB-BPL Related Protein may also be raised from fusion proteins created by expressing fusion proteins in bacteria as described herein.

10 The invention can employ intact monoclonal or polyclonal antibodies, and immunologically active fragments (e.g. a Fab, (Fab)<sub>2</sub> fragment, or Fab expression library fragments and epitope-binding fragments thereof), an antibody heavy chain, and antibody light chain, a genetically engineered single chain Fv molecule (Ladner et al, U.S. Pat. No. 4,946,778), or a chimeric antibody, for example, an antibody which contains the binding specificity of a murine antibody, but in which the remaining portions are of human  
15 origin. Antibodies including monoclonal and polyclonal antibodies, fragments and chimeras, may be prepared using methods known to those skilled in the art.

### 4. Applications of the Nucleic Acid Molecules, OB-BPL Related Proteins, and Antibodies of the Invention

The nucleic acid molecules, OB-BPL Related Proteins, and antibodies of the invention may be  
20 used in the prognostic and diagnostic evaluation of cancer or disorders of the hematopoietic system, and the identification of subjects with a predisposition to cancer or hematopoietic disorders (Section 4.1.1 and 4.1.2). Methods for detecting nucleic acid molecules and OB-BPL Related Proteins of the invention, can be used to monitor cancer or hematopoietic disorders by detecting OB-BPL Related Proteins and nucleic acid molecules encoding OB-BPL Related Proteins. It would also be apparent to one skilled in the art that  
25 the methods described herein may be used to study the developmental expression of OB-BPL Related Proteins and, accordingly, will provide further insight into the role of OB-BPL Related Proteins. The applications of the present invention also include methods for the identification of compounds that modulate the biological activity of *OB-BPL* or OB-BPL Related Proteins (Section 4.2). The compounds, antibodies etc. may be used for the treatment of cancer or hematopoietic disorders (Section 4.3).

#### 30 4.1 Diagnostic Methods

A variety of methods can be employed for the diagnostic and prognostic evaluation of cancer or disorders of the hematopoietic system (e.g. leukemias), and the identification of subjects with a predisposition to cancer or hematopoietic disorders. Such methods may, for example, utilize nucleic acid molecules of the invention, and fragments thereof, and antibodies directed against OB-BPL Related  
35 Proteins, including peptide fragments. In particular, the nucleic acids and antibodies may be used, for example, for: (1) the detection of the presence of *OB-BPL* mutations, or the detection of either over- or under-expression of *OB-BPL* mRNA relative to a non-disorder state or the qualitative or quantitative detection of alternatively spliced forms of *OB-BPL* transcripts which may correlate with certain conditions

or susceptibility toward such conditions; and (2) the detection of either an over- or an under-abundance of OB-BPL Related Proteins relative to a non-disorder state or the presence of a modified (e.g., less than full length) OB-BPL Protein which correlates with a disorder state, or a progression toward a disorder state.

5 The methods described herein may be performed by utilizing pre-packaged diagnostic kits comprising at least one specific *OB-BPL* nucleic acid or antibody described herein, which may be conveniently used, e.g., in clinical settings, to screen and diagnose patients and to screen and identify those individuals exhibiting a predisposition to developing a disorder.

10 Nucleic acid-based detection techniques are described, below, in Section 4.1.1. Peptide detection techniques are described, below, in Section 4.1.2. The samples that may be analyzed using the methods of the invention include those which are known or suspected to express *OB-BPL* or contain OB-BPL Related Proteins. The samples may be derived from a patient or a cell culture, and include but are not limited to biological fluids, tissue extracts, freshly harvested cells, and lysates of cells which have been incubated in cell cultures.

15 Oligonucleotides or longer fragments derived from any of the nucleic acid molecules of the invention may be used as targets in a microarray. The microarray can be used to simultaneously monitor the expression levels of large numbers of genes and to identify genetic variants, mutations, and polymorphisms. The information from the microarray may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

20 The preparation, use, and analysis of microarrays are well known to a person skilled in the art. (See, for example, Brennan, T. M. et al. (1995) U.S. Pat. No. 5,474,796; Schena, et al. (1996) Proc. Natl. Acad. Sci. 93:10614-10619; Baldeschweiler et al. (1995), PCT Application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R. A. et al. (1997) Proc. Natl. Acad. Sci. 94:2150-2155; and Heller, M. J. et al. (1997) U.S. Pat. No. 5,605,662.)

#### 25 **4.1.1 Methods for Detecting Nucleic Acid Molecules of the Invention**

30 The nucleic acid molecules of the invention allow those skilled in the art to construct nucleotide probes for use in the detection of nucleic acid sequences of the invention in samples. Suitable probes include nucleic acid molecules based on nucleic acid sequences encoding at least 5 sequential amino acids from regions of the OB-BPL Protein, preferably they comprise 15 to 30 nucleotides. A nucleotide probe may be labeled with a detectable substance such as a radioactive label which provides for an adequate signal and has sufficient half-life such as  $^{32}\text{P}$ ,  $^3\text{H}$ ,  $^{14}\text{C}$  or the like. Other detectable substances which may be used include antigens that are recognized by a specific labeled antibody, fluorescent compounds, enzymes, antibodies specific for a labeled antigen, and luminescent compounds. An appropriate label may be selected having regard to the rate of hybridization and binding of the probe to the nucleotide to be  
35 detected and the amount of nucleotide available for hybridization. Labeled probes may be hybridized to nucleic acids on solid supports such as nitrocellulose filters or nylon membranes as generally described in Sambrook et al, 1989, Molecular Cloning, A Laboratory Manual (2nd ed.). The nucleic acid probes may be used to detect genes, preferably in human cells, that encode OB-BPL Related Proteins. The nucleotide

probes may also be useful in the diagnosis of disorders of the hematopoietic system or cancer; in monitoring the progression of such disorders; or monitoring a therapeutic treatment.

The probe may be used in hybridization techniques to detect genes that encode OB-BPL Related Proteins. The technique generally involves contacting and incubating nucleic acids (e.g. recombinant DNA molecules, cloned genes) obtained from a sample from a patient or other cellular source with a probe of the present invention under conditions favorable for the specific annealing of the probes to complementary sequences in the nucleic acids. After incubation, the non-annealed nucleic acids are removed, and the presence of nucleic acids that have hybridized to the probe if any are detected.

The detection of nucleic acid molecules of the invention may involve the amplification of specific gene sequences using an amplification method such as PCR, followed by the analysis of the amplified molecules using techniques known to those skilled in the art. Suitable primers can be routinely designed by one of skill in the art.

Genomic DNA may be used in hybridization or amplification assays of biological samples to detect abnormalities involving *ob-bpl* structure, including point mutations, insertions, deletions, and chromosomal rearrangements. For example, direct sequencing, single stranded conformational polymorphism analyses, heteroduplex analysis, denaturing gradient gel electrophoresis, chemical mismatch cleavage, and oligonucleotide hybridization may be utilized.

Genotyping techniques known to one skilled in the art can be used to type polymorphisms that are in close proximity to the mutations an OB-BPL gene. The polymorphisms may be used to identify individuals in families that are likely to carry mutations. If a polymorphism exhibits linkage disequilibrium with mutations in an OB-BPL gene, it can also be used to screen for individuals in the general population likely to carry mutations. Polymorphisms which may be used include restriction fragment length polymorphisms (RFLPs), single-base polymorphisms, and simple sequence repeat polymorphisms (SSLPs).

A probe of the invention may be used to directly identify RFLPs. A probe or primer of the invention can additionally be used to isolate genomic clones such as YACs, BACs, PACs, cosmids, phage or plasmids. The DNA in the clones can be screened for SSLPs using hybridization or sequencing procedures.

Hybridization and amplification techniques described herein may be used to assay qualitative and quantitative aspects of *OB-BPL* expression. For example, RNA may be isolated from a cell type or tissue known to express *OB-BPL* and tested utilizing the hybridization (e.g. standard Northern analyses) or PCR techniques referred to herein. The techniques may be used to detect differences in transcript size which may be due to normal or abnormal alternative splicing. The techniques may be used to detect quantitative differences between levels of full length and/or alternatively splice transcripts detected in normal individuals relative to those individuals exhibiting symptoms of a hematopoietic disorder or other disease conditions.

The primers and probes may be used in the above described methods *in situ* i.e directly on tissue sections (fixed and/or frozen) of patient tissue obtained from biopsies or resections.

#### **4.1.2 Methods for Detecting OB-BPL Related Proteins**

Antibodies specifically reactive with an OB-BPL Related Protein, or derivatives, such as enzyme conjugates or labeled derivatives, may be used to detect OB-BPL Related Proteins in various samples (e.g. biological materials). They may be used as diagnostic or prognostic reagents and they may be used to detect abnormalities in the level of OB-BPL Related Proteins expression, or abnormalities in the structure, and/or temporal, tissue, cellular, or subcellular location of an OB-BPL Related Protein. Antibodies may also be used to screen potentially therapeutic compounds *in vitro* to determine their effects on disorders of the hematopoietic system, and other conditions. *In vitro* immunoassays may also be used to assess or monitor the efficacy of particular therapies. The antibodies of the invention may also be used *in vitro* to determine the level of *OB-BPL* expression in cells genetically engineered to produce an OB-BPL Related Protein.

The antibodies may be used in any known immunoassays which rely on the binding interaction between an antigenic determinant of an OB-BPL Related Protein and the antibodies. Examples of such assays are radioimmunoassays, enzyme immunoassays (e.g. ELISA), immunofluorescence, immunoprecipitation, latex agglutination, hemagglutination, and histochemical tests. The antibodies may be used to detect and quantify OB-BPL Related Proteins in a sample in order to determine its role in particular cellular events or pathological states, and to diagnose and treat such pathological states.

In particular, the antibodies of the invention may be used in immuno-histochemical analyses, for example, at the cellular and sub-subcellular level, to detect an OB-BPL Related Protein, to localize it to particular cells and tissues, and to specific subcellular locations, and to quantitate the level of expression.

Cytochemical techniques known in the art for localizing antigens using light and electron microscopy may be used to detect an OB-BPL Related Protein. Generally, an antibody of the invention may be labeled with a detectable substance and an OB-BPL Related Protein may be localised in tissues and cells based upon the presence of the detectable substance. Examples of detectable substances include, but are not limited to, the following: radioisotopes (e.g.,  $^3\text{H}$ ,  $^{14}\text{C}$ ,  $^{35}\text{S}$ ,  $^{125}\text{I}$ ,  $^{131}\text{I}$ ), fluorescent labels (e.g., FITC, rhodamine, lanthanide phosphors), luminescent labels such as luminol; enzymatic labels (e.g., horseradish peroxidase, beta-galactosidase, luciferase, alkaline phosphatase, acetylcholinesterase), biotinyl groups (which can be detected by marked avidin e.g., streptavidin containing a fluorescent marker or enzymatic activity that can be detected by optical or calorimetric methods), predetermined polypeptide epitopes recognized by a secondary reporter (e.g., leucine zipper pair sequences, binding sites for secondary antibodies, metal binding domains, epitope tags). In some embodiments, labels are attached via spacer arms of various lengths to reduce potential steric hindrance. Antibodies may also be coupled to electron dense substances, such as ferritin or colloidal gold, which are readily visualised by electron microscopy.

The antibody or sample may be immobilized on a carrier or solid support which is capable of immobilizing cells, antibodies etc. For example, the carrier or support may be nitrocellulose, or glass, polyacrylamides, gabbros, and magnetite. The support material may have any possible configuration including spherical (e.g. bead), cylindrical (e.g. inside surface of a test tube or well, or the external surface of a rod), or flat (e.g. sheet, test strip). Indirect methods may also be employed in which the primary antigen-antibody reaction is amplified by the introduction of a second antibody, having specificity for the antibody reactive against OB-BPL Related Protein. By way of example, if the antibody having specificity

against an OB-BPL Related Protein is a rabbit IgG antibody, the second antibody may be goat anti-rabbit gamma-globulin labeled with a detectable substance as described herein.

Where a radioactive label is used as a detectable substance, an OB-BPL Related Protein may be localized by radioautography. The results of radioautography may be quantitated by determining the density of particles in the radioautographs by various optical methods, or by counting the grains.

#### 4.2 Methods for Identifying or Evaluating Substances/Compounds

The methods described herein are designed to identify substances that modulate the biological activity of an OB-BPL Related Protein including substances that bind to OB-BPL Related Proteins, or bind to other proteins that interact with an OB-BPL Related Protein, to compounds that interfere with, or enhance the interaction of an OB-BPL Related Protein and substances that bind to the OB-BPL Related Protein or other proteins that interact with an OB-BPL Related Protein. Methods are also utilized that identify compounds that bind to *OB-BPL* regulatory sequences.

The substances and compounds identified using the methods of the invention include but are not limited to peptides such as soluble peptides including Ig-tailed fusion peptides, members of random peptide libraries and combinatorial chemistry-derived molecular libraries made of D- and/or L-configuration amino acids, phosphopeptides (including members of random or partially degenerate, directed phosphopeptide libraries), antibodies [e.g. polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, single chain antibodies, fragments, (e.g. Fab, F(ab)<sub>2</sub>, and Fab expression library fragments, and epitope-binding fragments thereof)], and small organic or inorganic molecules. The substance or compound may be an endogenous physiological compound or it may be a natural or synthetic compound.

Substances which modulate an OB-BPL Related Protein can be identified based on their ability to bind to an OB-BPL Related Protein. Therefore, the invention also provides methods for identifying substances which bind to an OB-BPL Related Protein. Substances identified using the methods of the invention may be isolated, cloned and sequenced using conventional techniques. A substance that associates with a polypeptide of the invention may be an agonist or antagonist of the biological or immunological activity of a polypeptide of the invention.

The term "agonist", refers to a molecule that increases the amount of, or prolongs the duration of, the activity of the polypeptide. The term "antagonist" refers to a molecule which decreases the biological or immunological activity of the polypeptide. Agonists and antagonists may include proteins, nucleic acids, carbohydrates, or any other molecules that associate with a polypeptide of the invention.

Substances which can bind with an OB-BPL Related Protein may be identified by reacting an OB-BPL Related Protein with a test substance which potentially binds to an OB-BPL Related Protein, under conditions which permit the formation of substance-OB-BPL Related Protein complexes and removing and/or detecting the complexes. The complexes can be detected by assaying for substance-OB-BPL Related Protein complexes, for free substance, or for non-complexed OB-BPL Related Protein. Conditions which permit the formation of substance-OB-BPL Related Protein complexes may be selected having regard to factors such as the nature and amounts of the substance and the protein.

The substance-protein complex, free substance or non-complexed proteins may be isolated by

conventional isolation techniques, for example, salting out, chromatography, electrophoresis, gel filtration, fractionation, absorption, polyacrylamide gel electrophoresis, agglutination, or combinations thereof. To facilitate the assay of the components, antibody against OB-BPL Related Protein or the substance, or labeled OB-BPL Related Protein, or a labeled substance may be utilized. The antibodies, proteins, or substances may be labeled with a detectable substance as described above.

AN OB-BPL Related Protein, or the substance used in the method of the invention may be insolubilized. For example, an OB-BPL Related Protein, or substance may be bound to a suitable carrier such as agarose, cellulose, dextran, Sephadex, Sepharose, carboxymethyl cellulose polystyrene, filter paper, ion-exchange resin, plastic film, plastic tube, glass beads, polyamine-methyl vinyl-ether-maleic acid copolymer, amino acid copolymer, ethylene-maleic acid copolymer, nylon, silk, etc. The carrier may be in the shape of, for example, a tube, test plate, beads, disc, sphere etc. The insolubilized protein or substance may be prepared by reacting the material with a suitable insoluble carrier using known chemical or physical methods, for example, cyanogen bromide coupling.

The invention also contemplates a method for evaluating a compound for its ability to modulate the biological activity of an OB-BPL Related Protein of the invention, by assaying for an agonist or antagonist (i.e. enhancer or inhibitor) of the binding of an OB-BPL Related Protein with a substance which binds with an OB-BPL Related Protein. The basic method for evaluating if a compound is an agonist or antagonist of the binding of an OB-BPL Related Protein and a substance that binds to the protein, is to prepare a reaction mixture containing the OB-BPL Related Protein and the substance under conditions which permit the formation of substance-OB-BPL Related Protein complexes, in the presence of a test compound. The test compound may be initially added to the mixture, or may be added subsequent to the addition of the OB-BPL Related Protein and substance. Control reaction mixtures without the test compound or with a placebo are also prepared. The formation of complexes is detected and the formation of complexes in the control reaction but not in the reaction mixture indicates that the test compound interferes with the interaction of the OB-BPL Related Protein and substance. The reactions may be carried out in the liquid phase or the OB-BPL Related Protein, substance, or test compound may be immobilized as described herein. The ability of a compound to modulate the biological activity of an OB-BPL Related Protein of the invention may be tested by determining the biological effects on cells.

It will be understood that the agonists and antagonists i.e. inhibitors and enhancers that can be assayed using the methods of the invention may act on one or more of the binding sites on the protein or substance including agonist binding sites, competitive antagonist binding sites, non-competitive antagonist binding sites or allosteric sites.

The invention also makes it possible to screen for antagonists that inhibit the effects of an agonist of the interaction of OB-BPL Related Protein with a substance which is capable of binding to the OB-BPL Related Protein. Thus, the invention may be used to assay for a compound that competes for the same binding site of an OB-BPL Related Protein.

The invention also contemplates methods for identifying compounds that bind to proteins that interact with an OB-BPL Related Protein. Protein-protein interactions may be identified using conventional

methods such as co-immunoprecipitation, crosslinking and co-purification through gradients or chromatographic columns. Methods may also be employed that result in the simultaneous identification of genes which encode proteins interacting with an OB-BPL Related Protein. These methods include probing expression libraries with labeled OB-BPL Related Protein.

5 Two-hybrid systems may also be used to detect protein interactions *in vivo*. Generally, plasmids are constructed that encode two hybrid proteins. A first hybrid protein consists of the DNA-binding domain of a transcription activator protein fused to an OB-BPL Related Protein, and the second hybrid protein consists of the transcription activator protein's activator domain fused to an unknown protein encoded by a cDNA which has been recombined into the plasmid as part of a cDNA library. The plasmids are  
10 transformed into a strain of yeast (e.g. *S. cerevisiae*) that contains a reporter gene (e.g. lacZ, luciferase, alkaline phosphatase, horseradish peroxidase) whose regulatory region contains the transcription activator's binding site. The hybrid proteins alone cannot activate the transcription of the reporter gene. However, interaction of the two hybrid proteins reconstitutes the functional activator protein and results in expression of the reporter gene, which is detected by an assay for the reporter gene product.

15 It will be appreciated that fusion proteins may be used in the above-described methods. In particular, OB-BPL Related Proteins fused to a glutathione-S-transferase may be used in the methods.

The reagents suitable for applying the methods of the invention to evaluate compounds that modulate an OB-BPL Related Protein may be packaged into convenient kits providing the necessary materials packaged into suitable containers. The kits may also include suitable supports useful in  
20 performing the methods of the invention.

#### **4.3 Compositions and Treatments**

The proteins of the invention, substances or compounds identified by the methods described herein, antibodies, and antisense nucleic acid molecules of the invention may be used for modulating the biological activity of an OB-BPL Related Protein, and they may be used in the treatment of conditions such  
25 as cancer and disorders of the hematopoietic system, in particular leukemias.

Hematopoietic disorders include but are not limited to myeloproliferative or other proliferative disorders of blood forming organs such as thromocythemias, polycythemias, and leukemias (acute myelogenous leukemia, chronic myelogenous leukemia). The proteins, substances, compounds, antibodies, and antisense nucleic acid molecules of the invention may be used in conjunction with bone marrow  
30 transplant, or in the treatment of aplasia or myelosuppression caused by radiation, chemical treatment, or chemotherapy. They may also be used to treat hematopoietic disorders associated with viral or bacterial infections.

Accordingly, the substances, antibodies, peptides, and compounds may be formulated into pharmaceutical compositions for administration to subjects in a biologically compatible form suitable for  
35 administration *in vivo*. By "biologically compatible form suitable for administration *in vivo*" is meant a form of the active substance to be administered in which any toxic effects are outweighed by the therapeutic effects. The active substances may be administered to living organisms including humans, and animals. Administration of a therapeutically active amount of a pharmaceutical composition of the present invention

is defined as an amount effective, at dosages and for periods of time necessary to achieve the desired result. For example, a therapeutically active amount of a substance may vary according to factors such as the disease state, age, sex, and weight of the individual, and the ability of antibody to elicit a desired response in the individual. Dosage regima may be adjusted to provide the optimum therapeutic response. For  
5 example, several divided doses may be administered daily or the dose may be proportionally reduced as indicated by the exigencies of the therapeutic situation.

The active substance may be administered in a convenient manner such as by injection (subcutaneous, intravenous, etc.), oral administration, inhalation, transdermal application, or rectal  
10 administration. Depending on the route of administration, the active substance may be coated in a material to protect the substance from the action of enzymes, acids and other natural conditions that may inactivate the substance.

The compositions described herein can be prepared by per se known methods for the preparation of pharmaceutically acceptable compositions which can be administered to subjects, such that an effective quantity of the active substance is combined in a mixture with a pharmaceutically acceptable vehicle.  
15 Suitable vehicles are described, for example, in Remington's Pharmaceutical Sciences (Remington's Pharmaceutical Sciences, Mack Publishing Company, Easton, Pa., USA 1985). On this basis, the compositions include, albeit not exclusively, solutions of the active substances in association with one or more pharmaceutically acceptable vehicles or diluents, and contained in buffered solutions with a suitable pH and iso-osmotic with the physiological fluids.

20 Vectors derived from retroviruses, adenovirus, herpes or vaccinia viruses, or from various bacterial plasmids, may be used to deliver nucleic acid molecules to a targeted organ, tissue, or cell population. Methods well known to those skilled in the art may be used to construct recombinant vectors which will express antisense nucleic acid molecules of the invention. (See, for example, the techniques described in Sambrook et al (supra) and Ausubel et al (supra)).

25 The nucleic acid molecules comprising full length cDNA sequences and/or their regulatory elements enable a skilled artisan to use sequences encoding a protein of the invention as an investigative tool in sense (Yousoufian H and H F Lodish 1993 Mol Cell Biol 13:98-104) or antisense (Eguchi et al (1991) Annu Rev Biochem 60:631-652) regulation of gene function. Such technology is well known in the art, and sense or antisense oligomers, or larger fragments, can be designed from various locations along the  
30 coding or control regions.

Genes encoding a protein of the invention can be turned off by transfecting a cell or tissue with vectors which express high levels of a desired OB-BPL-encoding fragment. Such constructs can inundate cells with untranslatable sense or antisense sequences. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until all copies are disabled by endogenous  
35 nucleases.

Modifications of gene expression can be obtained by designing antisense molecules, DNA, RNA or PNA, to the regulatory regions of a gene encoding a protein of the invention, ie, the promoters, enhancers, and introns. Preferably, oligonucleotides are derived from the transcription initiation site, eg,

between -10 and +10 regions of the leader sequence. The antisense molecules may also be designed so that they block translation of mRNA by preventing the transcript from binding to ribosomes. Inhibition may also be achieved using "triple helix" base-pairing methodology. Triple helix pairing compromises the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Therapeutic advances using triplex DNA were reviewed by Gee J E et al (In: Huber B E and B I Carr (1994) Molecular and Immunologic Approaches, Futura Publishing Co, Mt Kisco N.Y.).

Ribozymes are enzymatic RNA molecules that catalyze the specific cleavage of RNA. Ribozymes act by sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. The invention therefore contemplates engineered hammerhead motif ribozyme molecules that can specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding a protein of the invention.

Specific ribozyme cleavage sites within any potential RNA target may initially be identified by scanning the target molecule for ribozyme cleavage sites which include the following sequences, GUA, GUU and GUC. Once the sites are identified, short RNA sequences of between 15 and 20 ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be determined by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Methods for introducing vectors into cells or tissues include those methods discussed herein and which are suitable for *in vivo*, *in vitro* and *ex vivo* therapy. For *ex vivo* therapy, vectors may be introduced into stem cells obtained from a patient and clonally propagated for autologous transplant into the same patient (See U.S. Pat. Nos. 5,399,493 and 5,437,994). Delivery by transfection and by liposome are well known in the art.

The nucleic acid molecules disclosed herein may also be used in molecular biology techniques that have not yet been developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including but not limited to such properties as the triplet genetic code and specific base pair interactions.

The invention also provides methods for studying the function of a polypeptide of the invention. Cells, tissues, and non-human animals lacking in expression or partially lacking in expression of a nucleic acid molecule or gene of the invention may be developed using recombinant expression vectors of the invention having specific deletion or insertion mutations in the gene. A recombinant expression vector may be used to inactivate or alter the endogenous gene by homologous recombination, and thereby create a deficient cell, tissue, or animal.

Null alleles may be generated in cells, such as embryonic stem cells by deletion mutation. A recombinant gene may also be engineered to contain an insertion mutation that inactivates the gene. Such a construct may then be introduced into a cell, such as an embryonic stem cell, by a technique such as transfection, electroporation, injection etc. Cells lacking an intact gene may then be identified, for example by Southern blotting, Northern Blotting, or by assaying for expression of the encoded polypeptide using

the methods described herein. Such cells may then be fused to embryonic stem cells to generate transgenic non-human animals deficient in a polypeptide of the invention. Germline transmission of the mutation may be achieved, for example, by aggregating the embryonic stem cells with early stage embryos, such as 8 cell embryos, *in vitro*; transferring the resulting blastocysts into recipient females and; generating germline transmission of the resulting aggregation chimeras. Such a mutant animal may be used to define specific cell populations, developmental patterns and *in vivo* processes, normally dependent on gene expression.

The invention thus provides a transgenic non-human mammal all of whose germ cells and somatic cells contain a recombinant expression vector that inactivates or alters a gene encoding a OB-BPL Related Protein. In an embodiment the invention provides a transgenic non-human mammal all of whose germ cells and somatic cells contain a recombinant expression vector that inactivates or alters a gene encoding an OB-BPL Related Protein resulting in an OB-BPL Related Protein associated pathology. Further the invention provides a transgenic non-human mammal which does not express an OB-BPL Related Protein of the invention. In an embodiment, the invention provides a transgenic non-human mammal which does not express an OB-BPL Related Protein of the invention resulting in an OB-BPL Related Protein associated pathology. An OB-BPL Related Protein pathology refers to a phenotype observed for an OB-BPL Related Protein homozygous mutant.

A transgenic non-human animal includes but is not limited to mouse, rat, rabbit, sheep, hamster, dog, cat, goat, and monkey, preferably mouse.

The invention also provides a transgenic non-human animal assay system which provides a model system for testing for an agent that reduces or inhibits a pathology associated with an OB-BPL Related Protein, preferably an OB-BPL Related Protein associated pathology, comprising:

- (a) administering the agent to a transgenic non-human animal of the invention; and
- (b) determining whether said agent reduces or inhibits the pathology (e.g. OB-BPL Related Protein associated pathology) in the transgenic non-human animal relative to a transgenic non-human animal of step (a) which has not been administered the agent.

The agent may be useful in the treatment and prophylaxis of conditions such as cancer or hematopoietic disorders as discussed herein. The agents may also be incorporated in a pharmaceutical composition as described herein.

The activity of the proteins, substances, compounds, antibodies, nucleic acid molecules, agents, and compositions of the invention may be confirmed in animal experimental model systems. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED<sub>50</sub> (the dose therapeutically effective in 50% of the population) or LD<sub>50</sub> (the dose lethal to 50% of the population) statistics. The therapeutic index is the dose ratio of therapeutic to toxic effects and it can be expressed as the ED<sub>50</sub>/LD<sub>50</sub> ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred.

The following non-limiting examples are illustrative of the present invention:

#### Example

#### MATERIALS AND METHODS

### New Gene Identification

Nucleotide sequencing data of approximately 130 Kb on chromosome 19q13.4 was obtained from the Lawrence Livermore National Laboratory (LLNL) web site (<http://www-bio.llnl.gov/genome/genome.html>), in the form of one contig. This genomic sequence was subjected to a number of computer algorithms (gene prediction programs) designed to predict the presence of putative new genes. All programs used were previously thoroughly evaluated using a large number of known genes (Yousef *et al.*, 1999a). Based on these results, the most reliable algorithms - GeneBuilder (gene prediction) (<http://125.itba.mi.cnr.it/~webgene/genebuilder.html>) and GeneBuilder (exon prediction) (<http://125.itba.mi.cnr.it/~webgene/genebuilder.html>); Grail 2 (<http://compbio.ornl.gov>); and GENEID-3 (<http://apolo.imim.es/geneid.html>) – were selected for further use.

### Expressed sequence tag (EST) identification

The genomic sequence of the putative new gene was subjected to a homology search against the human EST database using the BLASTN algorithm (<http://www.ncbi.nlm.nih.gov/BLAST>) (Altschul *et al.*, 1997). Clones showing >95% homology were obtained from the I.M.A.G.E. consortium through Research Genetics Inc. (Huntsville, AL). The clone obtained was then propagated according to the suppliers instructions, purified, and sequenced from both directions with an automated sequencer, using the insert-flanking vector primers T3 and T7.

### Molecular Characterization of a Novel Siglec

The sequence derived from the computer predicted exons of the putative new gene was also used to search the non-redundant protein sequence database, using the BLASTP algorithm (Altschul *et al.*, 1997). Several proteins showing a high degree of homology were selected, and their nucleotide coding sequences were aligned with predicted coding sequence using the ClustalX multiple alignment program (Jeanmougin *et al.*, 1998). From this, regions on the putative gene were selected which showed the least amount of homology to the others and PCR primers were designed: F1 (TCACCGGCTCTCTGTGAATG – SEQ.ID.NO. 4) and R1 (GTCTTCTGCCCAAGGTTTCAG - SEQ.ID.NO. 5). Using these primers, PCR was performed on bone marrow cDNA, prepared as discussed below, and chosen based on the tissue expression results. The PCR conditions were as follows: 2.5 units HotStarTaq polymerase (Qiagen, Valencia, CA), 1X PCR buffer with 1.5 mM MgCl<sub>2</sub> (Qiagen), 1 µl cDNA, 200 µM dNTPs (deoxynucleoside triphosphates), and 250 ng of primers, using the Mastercycler® gradient thermocycler (Eppendorf Scientific, Inc., Westbury, NY). The temperature profile was: denaturation at 95°C for 15 min. followed by 94°C for 30 s., annealing at 58°C for 30 s., and extension at 72°C for 1 min. for a total of 35 cycles, followed by a final extension at 72°C for 10 min. The PCR product was subjected to electrophoresis on a 2% agarose gel and stained with ethidium bromide. Aliquots of the PCR products were subsequently extracted from the gel and the purified DNA was directly sequenced using an automated sequencer.

In order to verify the sequence surrounding the proposed start codon, another set of primers were designed, again derived from regions showing low homology with other known genes: F3 (TCCTCTAAGTCTTGAGCCCG - SEQ.ID.NO. 6) and R3 (CAGACGTTGAGATGGACGGT - SEQ.ID.NO. 7). PCR was performed using bone marrow cDNA, prepared as described below. The

conditions used for the PCR reaction were identical to those discussed previously, with electrophoresis of the PCR product on a 2% agarose gel, gel extraction, and automated sequencing as before.

Following final characterization of the genomic structure of this novel siglec, the putative protein product was aligned with the protein sequences of the other siglec family members using the ClustalX multiple sequence alignment tool. Further, phylogenetic analysis was performed using ClustalX in combination with TreeView (Page 1996).

Sequence analysis tools, available through the internet, were also utilized to detect the presence of possible sites of post-translational modification on the putative protein. The analysis programs PROSITE motif search (<http://www.expasy.ch/prosite/>) (Bairoch *et al.*, 1997), and NetOGlyc 2.0 (Hansen *et al.*, 1995; Hansen *et al.*, 1998) were used to detect N- and O-glycosylation, as well as the presence of kinase phosphorylation motifs. Further, the putative protein was assessed for the presence of a possible signal peptide, using SignalP v1.1 (<http://www.cbs.dtu.dk/>) (Nielsen *et al.*, 1997). For the prediction of transmembrane domains, two independent algorithms were used, TMPred ([http://www.ch.embnet.org/software/TMPRED\\_form.html](http://www.ch.embnet.org/software/TMPRED_form.html)) and DAS (<http://www.biokemi.su.se/~server/>). In addition, the hydropathic profile of this novel siglec was determined, using the Kyte-Doolittle method ([http://bioinformatics.weizmann.ac.il/hydroph/plot\\_hydroph.html](http://bioinformatics.weizmann.ac.il/hydroph/plot_hydroph.html)).

#### Mapping and Chromosomal Localization of a Novel Siglec

As mentioned previously, the contig on which the novel siglec gene was identified was obtained from the LLNL. EcoRI restriction maps were obtained from LLNL, and also generated using the Webcutter restriction analysis tool (<http://www.firstmarker.com/cutter/cutter2.html>), for both this contig, as well as the adjacent more centromeric contigs, containing the recently identified kallikrein gene family (Diamandis *et al.*, 1999; Yousef *et al.*, 1999a). Overlapping restriction fragments were identified and used to order the contigs and determine the distance between KLK-L6, the most telomeric member of the kallikrein gene family, and this novel siglec.

#### Tissue Expression

Total RNA from 28 normal human tissues was obtained (Clontech, Palo Alto, CA, USA), and reverse transcription was performed using SuperScript II™, according to the manufacturer's instructions (Gibco BRL, Gaithersburg, MD, USA). PCR was then performed using primers F2 (CGTGGGAGATACGGGCATAG - SEQ.ID.NO.8) and R2 (AAAAGGGAGGGCAGTGTG - SEQ.ID.NO.9), using the same PCR conditions described previously. PCR for actin was also performed as describe elsewhere (Yousef *et al.*, 1999b), as a control for cDNA quality.

## RESULTS

### Identification of a Novel Siglec on 19q13.4

Computer analysis of the approximately 130 Kb contig predicted a putative new gene consisting of six exons. Five of these were predicted by at least three programs, with only one exon being predicted by two of the four programs (Table 1). Homology search for the putative new gene against the human EST database revealed the presence of one unique EST (GenBank accession # AA936059) which showed 98% identity to the sixth predicted exon.

The entire insert of this EST was sequenced, followed by alignment of this nucleotide sequence with the genomic sequence of the putative gene, using the "BLAST 2 sequences" program. This revealed the presence of an additional area, between predicted exons 5 and 6, with 98% identity to the EST. This suggested that there was an additional exon in this area which was not detected by the prediction algorithms used.

#### Characterization of the Genomic Structure of the Novel Siglec Gene and its Protein Product

With the aid of unique primers, designed as discussed in the experimental section, RT-PCR was performed on bone marrow cDNA and two additional products were isolated, both encompassing multiple predicted exons. Upon sequencing of these PCR products, the presence of all six predicted exons, as well as the newly identified exon, found from the EST sequence were confirmed. With both cDNA and genomic sequence at hand, the genomic organization of this new gene was determined (Figure 1). The gene encoding this novel siglec encompasses a genomic area of 5,421 bp. It is composed of seven exons, with six intervening introns. The lengths of the exons are 509, 279, 48, 267, 91, 97, and 417 bp, respectively. All the intron/exon splice sites and their flanking sequences are closely related to the consensus splice sites (-mGTAAGT...CAGm-, where m is any base) (Iida 1990).

The proposed protein coding region of the novel siglec gene consists of 1,392 nucleotides, producing a 463 amino acid protein, with a predicted molecular mass of 50.1 kDa, excluding any post-translational modifications. The translation initiation codon (ATG) at position 1171 of the first exon (according to the numbering of SEQ. ID. NO. 1 and GenBank Accession No. AF135027), was chosen because: 1) the flanking region surrounding that codon closely matches the Kozak consensus sequence for translational initiation, particularly at position -3 (a purine), which appears to be the most highly conserved (Kozak 1991); 2) using this initiation codon, the proposed protein contains an N-terminal signal sequence which shows a high degree of homology to other similar proteins (see below). The 3' terminus of the novel siglec gene was verified by the presence of a poly dA tail present in the EST sequence. Further, it is evident from Figure 1 that this gene possesses a 5' untranslated region of at least 88 nucleotides, as well as a 3' untranslated region of 228 nucleotides.

Examination of the hydrophobicity profile of the novel siglec protein revealed two regions with long stretches of hydrophobic residues. The first of these occurs at the N-terminus, suggesting the presence of a signal peptide (Figure 2). This is consistent with findings from a signal sequence prediction program (Nielsen *et al.*, 1997), which predicts a 17 amino acid residue signal sequence. The second region occurs between residues 349 and 370, suggestive of a transmembrane domain, and is consistent with results from transmembrane region prediction programs. Based on this information, the protein product of this novel gene is likely a type I transmembrane protein, after cleavage of the 17 residue signal sequence.

Through the use of sequence analysis tools, the various putative post-translational modification sites were identified (Table 2). There are numerous potential sites in this novel siglec where there could be either O- or N-glycosylation. Furthermore, several possible sites of phosphorylation have been identified for cAMP-dependent protein kinase, protein kinase C, and casein kinase 2.

#### Mapping and Chromosomal Localization of a Novel Siglec

The contig in which the gene encoding a novel siglec was identified is located at 19q13.4, telomeric to the kallikrein gene KLK3 (PSA). Previous studies have identified and mapped the kallikrein gene family locus on this region of chromosome 19 (Diamandis *et al.*, 1999; Yousef *et al.*, 1999a). The contig containing the novel siglec gene was found, through EcoRI restriction mapping, to be located adjacent to this kallikrein gene family. The novel siglec gene is located 43.19 Kb more telomeric than KLK-L6, at 19q13.4. A detailed physical map of the area which contains some known genes and the newly identified siglec gene is shown in Figure 3. By computer analysis, no other genes were predicted between KLK-L6 and this novel siglec.

#### Homology with other Siglec Family Members

Using the predicted protein sequence, a homology search was performed against the GenBank database using the BLASTP program. The novel siglec showed a high degree of homology to other known members of the Siglec family (Table 3). A multiple alignment of this novel siglec with the other family members was also performed using the ClustalX alignment program. As is evident in Figure 4, the N-terminal signal sequence is highly conserved within this family of proteins. Furthermore, the protein contains Ig domains typically found in Siglec family members: an N-terminal V-set domain, followed by multiple C2-set domains (Crocker *et al.*, 1996). This novel siglec contains a total of 3 Ig domains, a V-set and two C2-set domains, based on homology with known Ig domains. As shown in Table 4, the V-set and first C2-set domains are highly similar to Siglec-7 and CD33, with the second C2-set showing highest homologies with Siglec-7 and Siglec-6. The novel siglec exhibits conservation of the cysteine residues in the V-set and first C2-set domains, which form the two characteristic disulfide bridges in other Siglec family members. In the V-set domain, Cys 41 and Cys 102 form an intrasheet disulfide bond, whereas Cys 36 and Cys 170 of the first C2-set domain are likely to form the interdomain disulfide bond, based on findings for other siglecs (Crocker *et al.*, 1996; Williams *et al.*, 1989). The V-set domain also possesses a conserved arginine which has been found to be essential for sialic-acid binding (van der Merwe *et al.*, 1996), as well as two conserved aromatic residues in  $\beta$ -strands A and G which have been found to make hydrophobic contacts with the N-acetyl and glycerol side groups of N-acetyl neuraminic acid (May *et al.*, 1998). As is evident from Figure 4, this novel siglec also possesses the critical arginine, at position 120, as well as the aromatic residue in  $\beta$ -strand G; however it lacks the aromatic residue in the A  $\beta$ -strand. The domain boundaries were determined based on the one domain: one exon rule (Williams and Barclay 1988), while taking into consideration the domain assignments of others (Cornish *et al.*, 1998; Crocker *et al.*, 1998; Falco *et al.*, 1999; Nicoll *et al.*, 1999; Patel *et al.*, 1999).

Examination of the transmembrane and intracellular domains of Siglec family members reveals that it is more variable than the extracellular domain. However, there are regions that show a high level of conservation. As shown in Figure 4, all the Siglecs possess a single transmembrane domain, consisting of approximately 25 residues. In addition, within the cytoplasmic domain, there are two highly conserved motifs. The first of these, L(HQ)YA(SV)L, exhibits similarity to an immunoreceptor tyrosine kinase inhibitory motif (ITIM), which has a 6 amino acid consensus sequence (ILV)xYxx(LV) (Burshtyn *et al.*, 1997; Vivier and Daeron 1997). The second motif, TEYSE(IV), is homologous to a sequence (TxYxx(IV))

recently found in the signaling lymphocyte activation molecule (SLAM) which is responsible for the binding of the SLAM-associated protein (SAP) (Coffey *et al.*, 1998; Sayos *et al.*, 1998).

Phylogenetic analysis of the entire siglec family was performed using ClustalX and TreeView. This revealed that the novel siglec is very closely related to Siglec-7, followed by CD33 (Figure 5). It is evident that this novel gene, which encodes a putative siglec protein is the newest member of the siglec family. It possesses all the necessary features, including the Ig-like domains, the type I transmembrane topology, as well as the conserved cytoplasmic motifs, and shows a close phylogenetic relationship to the other siglec family members.

#### Tissue Expression Profile of a Novel Siglec

RT-PCR was performed on a panel of tissue-specific total RNA preparations (Figure 6). High levels of expression of the novel siglec were found in bone marrow, placenta, spleen, and fetal liver. Lower levels of expression were also evident in fetal brain, stomach, lung, thymus, prostate, brain, mammary, adrenal gland, colon, trachea, cerebellum, testis, small intestine, and spinal cord. Expression of this novel siglec was absent in heart, skeletal muscle, pancreas, and ovary. All PCR products obtained were of equal length, and corresponded to the length of the product obtained from overlapping EST (accession # AA936059). Sequencing of the PCR products ensured specificity.

#### DISCUSSION

Using the positional candidate gene approach a novel gene belonging to the siglec family was identified. This gene is comprised of 7 exons, with 6 intervening introns. The coding region of this gene is composed of 1,392 nucleotides, producing a 463 amino acid protein, with a predicted molecular mass of 50.1 kDa. This gene is located at 19q13.4, 43.19 Kb telomeric to the newly identified kallikrein KLK-L6. The high degree of homology between this novel siglec and other siglecs provides strong evidence that this protein also plays a role in sialic acid-dependent protein-glycoprotein or -glycolipid interactions. It possesses the unique pattern of conserved cysteine residues in its Ig-like domains, which are found only in members of the siglec family. Further, this novel siglec possesses the conserved arginine residue, which has been found to be essential for sialic acid binding (van der Merwe *et al.*, 1996). Of note, however, is that it only possesses one of the two conserved aromatic residues in the V-set domain, which may be suggestive of a unique sialic acid specificity, differing from that of previously identified siglecs.

The tissue expression profile of the novel siglec was examined and it was found to be highly expressed in bone marrow, placenta, spleen, and fetal liver. The high level of expression of this novel siglec in bone marrow is consistent with findings from groups investigating the other siglec family members. All currently known siglecs have been found to be expressed in some type of bone marrow stem cell-derived cell, ranging from myeloid progenitor cells for CD33 to natural killer cells for Siglec-7 and B lymphocytes for CD22. It is likely that this novel siglec is predominantly expressed on a distinct subset of immune cells, where it plays an intercellular signaling role. This is supported by the presence of ITIM-like and SLAM-like motifs in the cytoplasmic domain of this novel siglec, with similar domains in other siglecs. ITIM motifs are consensus binding sites for the SH2 (src homology 2) domains of the phosphatases SHP-1 and SHP-2 (Borges *et al.*, 1997; Le Drean *et al.*, 1998). It has been reported that the

phosphorylation of the ITIM-like motif in CD22, the phosphatase SHP-1 is recruited, suggesting a possible function of this siglec as a B cell receptor-associated negative co-receptor (Vivier and Daeron 1997). The second cytoplasmic motif has been identified in SLAM and several SLAM-like proteins, a family of immunoregulatory molecules of the IgSF, and is responsible for the binding of a new SH2-containing molecule, SAP (Coffey *et al.*, 1998; Sayos *et al.*, 1998). The binding of SAP was shown to inhibit the binding of SHP-2 to its respective binding site on these SLAM proteins. The presence of such a motif in the novel siglec, and other siglecs, suggests that there may be a similar regulatory mechanism present in the cytoplasmic domains of siglecs, with SAP inhibiting the binding of SHP-1 and SHP-2 to the ITIM-like motif.

10           The regulation of SHP-1 and SHP-2 binding to ITIM motifs, and thus their activation, very likely affects downstream tyrosine-kinase dependent pathways by regulating the phosphorylation state of components in these pathways. Thus, the siglec family of ITIM and SLAM-bearing receptors probably play a role in controlling the activation of a number of cell types. By extension, it is possible that these siglecs may be involved in the regulation of tumour growth. CD33 has already been identified as an important marker for the diagnosis of acute myelogenous leukemia (AML), particularly for the undifferentiated form, and serves to distinguish AML from lymphoid leukemias (Bernstein *et al.*, 1992; Dinndorf *et al.*, 1986; Griffin *et al.*, 1984). Recently, Kossman *et. al.* and Sievers *et. al.* have reported the use of anti-CD33 monoclonal antibodies in phase I studies for the treatment of AML, and have shown selective ablation of malignant hematopoiesis (Kossman *et al.*, 1999; Sievers *et al.*, 1999). The newly identified member of the siglec family may have utility as a target for immunological antineoplastic therapy.

25           Having illustrated and described the principles of the invention in a preferred embodiment, it should be appreciated to those skilled in the art that the invention can be modified in arrangement and detail without departure from such principles. All modifications coming within the scope of the following claims are claimed.

          All publications, patents and patent applications referred to herein are incorporated by reference in their entirety to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference in its entirety.

**FULL CITATIONS FOR REFERENCES REFERRED TO IN THE SPECIFICATION**

- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389-402.
- 5 Bairoch, A., Bucher, P. and Hofmann, K. (1997). The PROSITE database, its status in 1997. *Nucleic Acids Res* **25**: 217-21.
- Bernstein, I. D., Singer, J. W., Smith, F. O., Andrews, R. G., Flowers, D. A., Petersens, J., Steinmann, L., Najfeld, V., Savage, D., Fruchtman, S. and et al. (1992). Differences in the frequency of normal and clonal precursors of colony-forming cells in chronic myelogenous leukemia and acute myelogenous leukemia.
- 10 *Blood* **79**: 1811-6.
- Borges, L., Hsu, M. L., Fanger, N., Kubin, M. and Cosman, D. (1997). A family of human lymphoid and myeloid Ig-like receptors, some of which bind to MHC class I molecules. *J Immunol* **159**: 5192-6.
- Burshtyn, D. N., Yang, W., Yi, T. and Long, E. O. (1997). A novel phosphotyrosine motif with a critical amino acid at position -2 for the SH2 domain-mediated activation of the tyrosine phosphatase SHP- 1. *J Biol Chem* **272**: 13066-72.
- 15 Coffey, A. J., Brooksbank, R. A., Brandau, O., Oohashi, T., Howell, G. R., Bye, J. M., Cahn, A. P., Durham, J., Heath, P., Wray, P. *et al.*, (1998). Host response to EBV infection in X-linked lymphoproliferative disease results from mutations in an SH2-domain encoding gene [see comments]. *Nat Genet* **20**: 129-35.
- 20 Cornish, A. L., Freeman, S., Forbes, G., Ni, J., Zhang, M., Cepeda, M., Gentz, R., Augustus, M., Carter, K. C. and Crocker, P. R. (1998). Characterization of siglec-5, a novel glycoprotein expressed on myeloid cells related to CD33. *Blood* **92**: 2123-32.
- Crocker, P. R., Clark, E. A., Filbin, M., Gordon, S., Jones, Y., Kehrl, J. H., Kelm, S., Le Douarin, N., Powell, L., Roder, J. *et al.*, (1998). Siglecs: a family of sialic-acid binding lectins [letter]. *Glycobiology* **8**: v.
- 25 Crocker, P. R., Kelm, S., Hartnell, A., Freeman, S., Nath, D., Vinson, M. and Mucklow, S. (1996). Sialoadhesin and related cellular recognition molecules of the immunoglobulin superfamily. *Biochem Soc Trans* **24**: 150-6.
- 30 Crocker, P. R., Mucklow, S., Bouckson, V., McWilliam, A., Willis, A. C., Gordon, S., Milon, G., Kelm, S. and Bradfield, P. (1994). Sialoadhesin, a macrophage sialic acid binding receptor for haemopoietic cells with 17 immunoglobulin-like domains. *Embo J* **13**: 4490-503.
- Diamandis, E. P., Yousef, G. M., Luo, L.-Y., Magklara, A. and Obiezu, C. V. (1999). The new human kallikrein gene family: implications in carcinogenesis. *Trends Endocrinol Metab.*, 2000, 11:54-60.
- 35 Dinndorf, P. A., Andrews, R. G., Benjamin, D., Ridgway, D., Wolff, L. and Bernstein, I. D. (1986). Expression of normal myeloid-associated antigens by acute leukemia cells. *Blood* **67**: 1048-53.
- Falco, M., Biassoni, R., Bottino, C., Vitale, M., Sivori, S., Augugliaro, R., Moretta, L. and Moretta, A. (1999). Identification and molecular cloning of p75/AIRM1, a novel member of the sialoadhesin family

- that functions as an inhibitory receptor in human natural killer cells. *J Exp Med* **190**: 793-802.
- Griffin, J. D., Linch, D., Sabbath, K., Larcom, P. and Schlossman, S. F. (1984). A monoclonal antibody reactive with normal and leukemic human myeloid progenitor cells. *Leuk Res* **8**: 521-34.
- Hansen, J. E., Lund, O., Engelbrecht, J., Bohr, H. and Nielsen, J. O. (1995). Prediction of O-glycosylation  
5 of mammalian proteins: specificity patterns of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase. *Biochem J* **308**: 801-13.
- Hansen, J. E., Lund, O., Tolstrup, N., Gooley, A. A., Williams, K. L. and Brunak, S. (1998). NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconj J* **15**: 115-30.
- 10 Iida, Y. (1990). Quantification analysis of 5'-splice signal sequences in mRNA precursors. Mutations in 5'-splice signal sequence of human beta-globin gene and beta-thalassemia. *J Theor Biol* **145**: 523-33.
- Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G. and Gibson, T. J. (1998). Multiple sequence alignment with Clustal X. *Trends Biochem Sci* **23**: 403-5.
- Kossmann, S. E., Scheinberg, D. A., Jurcic, J. G., Jimenez, J. and Caron, P. C. (1999). A phase I trial of  
15 humanized monoclonal antibody HuM195 (anti-CD33) with low-dose interleukin 2 in acute myelogenous leukemia. *Clin Cancer Res* **5**: 2748-55.
- Kozak, M. (1991). An analysis of vertebrate mRNA sequences: intimations of translational control. *J Cell Biol* **115**: 887-903.
- Le Drean, E., Vely, F., Olcese, L., Cambiaggi, A., Guia, S., Krystal, G., Gervois, N., Moretta, A., Jotereau,  
20 F. and Vivier, E. (1998). Inhibition of antigen-induced T cell response and antibody-induced NK cell cytotoxicity by NKG2A: association of NKG2A with SHP-1 and SHP-2 protein-tyrosine phosphatases [published erratum appears in Eur J Immunol 1998 Mar;28(3):1122]. *Eur J Immunol* **28**: 264-76.
- Li, C., Trapp, B., Ludwin, S., Peterson, A. and Roder, J. (1998). Myelin associated glycoprotein modulates  
glia-axon contact in vivo. *J Neurosci Res* **51**: 210-7.
- 25 May, A. P., Robinson, R. C., Vinson, M., Crocker, P. R. and Jones, E. Y. (1998). Crystal structure of the N-terminal domain of sialoadhesin in complex with 3' sialyllactose at 1.85 Å resolution. *Mol Cell* **1**: 719-28.
- Nicoll, G., Ni, J., Liu, D., Klenerman, P., Munday, J., Dubock, S., Mattei, M. G. and Crocker, P. R. (1999). Identification and Characterization of a Novel Siglec, Siglec-7, Expressed by Human Natural Killer Cells  
30 and Monocytes. *J Biol Chem* **274**: 34089-34095.
- Nielsen, H., Engelbrecht, J., Brunak, S. and von Heijne, G. (1997). A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Int J Neural Syst* **8**: 581-99.
- Page, R. D. (1996). TreeView: an application to display phylogenetic trees on personal computers. *Comput  
35 Appl Biosci* **12**: 357-8.
- Patel, N., Brinkman-Van der Linden, E. C., Altmann, S. W., Gish, K., Balasubramanian, S., Timans, J. C., Peterson, D., Bell, M. P., Bazan, J. F., Varki, A. and Kastelein, R. A. (1999). OB-BP1/Siglec-6, a leptin- and sialic acid-binding protein of the immunoglobulin superfamily. *J Biol Chem* **274**: 22729-38.

- Pedraza, L., Owens, G. C., Green, L. A. and Salzer, J. L. (1990). The myelin-associated glycoproteins: membrane disposition, evidence of a novel disulfide linkage between immunoglobulin-like domains, and posttranslational palmitoylation. *J Cell Biol* **111**: 2651-61.
- Sayos, J., Wu, C., Morra, M., Wang, N., Zhang, X., Allen, D., van Schaik, S., Notarangelo, L., Geha, R.,  
5 Roncarolo, M. G., Oettgen, H., De Vries, J. E., Aversa, G. and Terhorst, C. (1998). The X-linked lymphoproliferative-disease gene product SAP regulates signals induced through the co-receptor SLAM [see comments]. *Nature* **395**: 462-9.
- Sievers, E. L., Appelbaum, F. R., Spielberger, R. T., Forman, S. J., Flowers, D., Smith, F. O., Shannon-Dorcy, K., Berger, M. S. and Bernstein, I. D. (1999). Selective ablation of acute myeloid leukemia using  
10 antibody-targeted chemotherapy: a phase I study of an anti-CD33 calicheamicin immunoconjugate. *Blood* **93**: 3678-84.
- Stamenkovic, I. and Seed, B. (1990). The B-cell antigen CD22 mediates monocyte and erythrocyte adhesion. *Nature* **345**: 74-7.
- Ulyanova, T., Blasioli, J., Woodford-Thomas, T. A. and Thomas, M. L. (1999). The sialoadhesin CD33  
15 is a myeloid-specific inhibitory receptor. *Eur J Immunol* **29**: 3440-9.
- van der Merwe, P. A., Crocker, P. R., Vinson, M., Barclay, A. N., Schauer, R. and Kelm, S. (1996). Localization of the putative sialic acid-binding site on the immunoglobulin superfamily cell-surface molecule CD22. *J Biol Chem* **271**: 9273-80.
- Vivier, E. and Daeron, M. (1997). Immunoreceptor tyrosine-based inhibition motifs. *Immunol Today* **18**:  
20 286-91.
- Williams, A. F. and Barclay, A. N. (1988). The immunoglobulin superfamily--domains for cell surface recognition. *Annu Rev Immunol* **6**: 381-405.
- Williams, A. F., Davis, S. J., He, Q. and Barclay, A. N. (1989). Structural diversity in domains of the immunoglobulin superfamily. *Cold Spring Harb Symp Quant Biol* **54**: 637-47.
- 25 Yousef, G. F., Luo, L. and Diamandis, E. P. (1999a). Identification of novel human kallikrein-like genes on chromosome 19q13.3-q13.4. *Anticancer Res* **79**: 2843-2852.
- Yousef, G. M., Obiezu, C. V., Luo, L. Y., Black, M. H. and Diamandis, E. P. (1999b). Prostase/KLK-L1 is a new member of the human kallikrein gene family, is expressed in prostate and breast tissues, and is hormonally regulated. *Cancer Res* **59**: 4252-6.

**Table 1: Genomic organization of a novel siglec.**

Exon No.	Coding Region <sup>1</sup>		No. of base pairs	EST Match <sup>2</sup>	Intron Phase	Exon Predicted <sup>3</sup>
	From (bp)	To (bp)				
1	1083	1591	509	-	I	B,C
2	1793	2071	279	-	I	A,B,C,D
3	2277	2324	48	-	I	A,B,D
4	3226	3492	267	-	I	A,B,C
5	4145	4235	91	-	0	A,B,C,D
6	4610	4706	97	+	0	-
7	6087	6503	417	+	-	A,B,C

1. The coding region shown includes the 5' untranslated region in exon 1, and the 3' untranslated region in exon 7. Numbers refer to GenBank accession no. AF135027.
2. EST; GenBank accession no. AA936059
3. The exon prediction programs are as follows: A) GeneBuilder (gene prediction); B) GeneBuilder (exon prediction); C) Grail 2; D) GENEID-3.

5

**Table 2: Putative post-translational modification sites in the novel siglec.**

<b>Modification<sup>1</sup></b>	<b>Residue</b>	<b>Position<sup>2</sup></b>
O-glycosylation	Thr	76, 192, 193
	Ser	184, 186, 195
N-glycosylation	Asn	101, 138, 161, 225, 231, 238, 256, 334
cAMP-dependent Protein Kinase phosphorylation	Ser/Thr	374
Protein Kinase C phosphorylation	Ser/Thr	372, 377, 421
Casein Kinase 2 phosphorylation	Ser/Thr	387, 412, 425, 452

10

1. The proposed O-glycosylation sites were determined through NetOGlyc 2.0 (Hansen *et al.*, 1998). The remainder of the post-translational modifications were predicted by PROSITE (Hansen *et al.*, 1995).
2. The residue numbering is according to the numbering of the novel siglec, as shown in Figure 4.

5

**Table 3: Overall homology of this novel siglec with other known siglecs.**

Siglec Family Member <sup>1</sup>	Homology to the Novel Siglec <sup>2</sup>	
	% identity	% similarity
Siglec-7 (p75/AIRM1) (AF170485)	75	80
Siglec-5 (OB-BP2) (U71383)	52	65
CD33 (M23197)	52	64
Siglec-6 (OB-BP1) (U71382)	49	60
Sialoadhesin (Z36293)	27	43
CD22 (X52785)	26	42
Myelin associated glycoprotein (MAG) (M29273)	25	42

1. GenBank accession numbers for each of the siglec family members is also shown, in brackets.

10 2. Homology was determined using the BLASTP algorithm (Altschul *et al.*, 1997).

5

**Table 4: Ig-like domain homology between the novel siglec and other siglec family members<sup>1</sup>**

	<b>Homologous Protein</b>	<b>Domain</b>	<b>% identity</b>	<b>% similarity</b>
<b>Novel Siglec Ig 1 (V-set)</b>	Siglec-7 (p75/AIRM1)	1	75	78
	CD33	1	61	71
	Siglec-5 (OB-BP2)	1	54	67
	Siglec-6 (OB-BP1)	1	54	62
	MAG	1	32	48
	Sialoadhesin	1	29	48
	CD22	1	28	44
<b>Novel Siglec Ig 2 (C2-set)</b>	Siglec-7 (p75/AIRM1)	2	89	93
	CD33	2	63	75
	Siglec-6 (OB-BP1)	2	58	70
	Siglec-5 (OB-BP2)	2	58	71
	Sialoadhesin	2	30	46
		12	31	44
	MAG	2	25	46
CD22	2	27	43	
<b>Novel Siglec Ig 3 (C2-set)</b>	Siglec-7 (p75/AIRM1)	3	76	79
	Siglec-6 (OB-BP1)	3	52	67
	Siglec-5 (OB-BP2)	3	48	62
	Sialoadhesin	13	33	48
		7	31	42
		15	28	40
	MAG	3	27	49

1. GenBank accession numbers for the listed siglecs are the same as those shown in Table 3.

10

Table 5. Predicted exons of the unknown gene UG . The translated protein sequences of each exon (open reading frame) are shown

Exon No.	Putative coding region From(bp) To(bp)	No. of bases	Translated protein sequence	EST match <sup>2</sup>	Intron phase <sup>3</sup>	Stop codon <sup>4</sup>	Exon prediction program <sup>5</sup>
1	44,129 44,641	513	PPLSLEPAVPERRTLRNRRLAALAPLTPDMLLLLPLL WGRERAEGQTSKLLTMQSSVTVQEGLCVHVPCSFYSYPS HGWIYFPGVPVHGYWFREGANTDQDAPVATNNPARAV WEETDRFHLGDPHFKNCTLSIRDARRSDAAGRYFFRM EKGSIKWYKHH RLSVNVV	+	I	-	B,C
2	44,843 45,121	279	ALTRPNLIPGLESGCPQNLTCSPWACEQGTTPMIS WIGTSVPLDFSTTRSSVLTLPQPQDHGTSLTQVTFPG ASVTNKTVHLNVS	+	I	-	A,B,C,D
3	45,327 45,374	48	YPPQNLTMVTFQGDGT	-	I	-	A,B,D
4	46,318 46,542	225	EGQSLRLVCAVDAVDSNPPARLSWRGLTLCPSQPSN PGVLELPWVHLRDAAEFTCRANPLGQQVYLNVS LQ	+	I	-	A,B,C
5	47,195 47,283	186	SKATSGVTQGVGGAGATLVFLSFCVIFV	+	0	-	A,B,C,D
6	49,136 49,554	186	GPLTEPWAEDSPDQPPASARSSVGEQELQYASLSFQ MVKPWDS RGQEAIDTEYSEIKIHR	+	-	+	A,B,C

\* All footnotes same as table 2.

1. Conventional numbering of exons in comparison to the five coding exons of PSA. Nucleotide numbers refer to the related contig (see text).
2. (+) = >95% homology with published human EST sequences.
3. Intron phase: 0=the intron occurs between codons; I=the intron occurs after the first nucleotide of the codon; II=the intron occurs after the second nucleotide of the codon.
4. (+) denotes the exon containing the stop codon.
5. H=histidine, D=aspartic acid, S=serine. The aminoacids of the catalytic triad are bold and underlined. A = GeneBuilder (gene analysis), B = GeneBuilder (exon analysis), C = Grail 2, D = GENEID-3

**We Claim:**

1. An isolated OB-BPL nucleic acid molecule of at least 30 nucleotides which hybridizes to SEQ ID NO. 1, or the complement of SEQ ID NO. 1, under stringent hybridization conditions.
2. An isolated nucleic acid molecule which comprises:
  - (i) a nucleic acid sequence encoding a polypeptide having substantial sequence identity with the amino acid sequence shown in Table 5 or SEQ. ID. NO.2 or SEQ. ID. NO 3.;
  - (ii) nucleic acid sequences complementary to (i);
  - (iii) a degenerate form of a nucleic acid sequence of (i);
  - (iv) a nucleic acid sequence comprising at least 18 nucleotides and capable of hybridizing to a nucleic acid sequence in (i), (ii), or (iii);
  - (v) a nucleic acid sequence encoding a truncation, an analog, an allelic or species variation of a polypeptide comprising the amino acid sequence shown in Table 5 or SEQ. ID. NO.2, or SEQ. ID. NO 3; or
  - (vi) a fragment, or allelic or species variation of (i), (ii) or (iii).
3. An isolated nucleic acid molecule which comprises:
  - (a) a nucleic acid sequence having substantial sequence identity or sequence similarity with a nucleic acid sequence of SEQ. ID. NO. 1;
  - (b) nucleic acid sequences complementary to (i), preferably complementary to the full nucleic acid sequence of SEQ. ID. NO. 1;
  - (c) nucleic acid sequences differing from any of the nucleic acid sequences of (i) or (ii) in codon sequences due to the degeneracy of the genetic code; or
  - (d) a fragment, or allelic or species variation of (i), (ii) or (iii).
4. An isolated nucleic acid molecule which encodes a protein which binds an antibody of a OB-BPL polypeptide.
5. A regulatory sequence of an isolated nucleic acid molecule as claimed in any of the preceding claims fused to a nucleic acid which encodes a heterologous protein.
6. A vector comprising a nucleic acid molecule of any of the preceding claims.
7. A host cell comprising a nucleic acid molecule of any of the preceding claims.
8. An isolated OB-BPL protein comprising an amino acid sequence of SEQ. ID. NO. 2 or 3.
9. An isolated protein having at least 65% amino acid sequence identity to an amino acid sequence of SEQ. ID. NO. 2 or 3.
10. A method for preparing a protein as claimed in claim 9 comprising:
  - (a) transferring a vector as claimed in claim 7 into a host cell;

- (b) selecting transformed host cells from untransformed host cells;
  - (c) culturing a selected transformed host cell under conditions which allow expression of the protein; and
  - (d) isolating the protein.
11. A protein prepared in accordance with the method of claim 11.
  12. An antibody having specificity against an epitope of a polypeptide as claimed in claim 9.
  13. An antibody as claimed in claim 13 labeled with a detectable substance and used to detect the polypeptide in biological samples, tissues, and cells.
  14. A probe comprising a sequence encoding a protein as claimed in claim 9, or a part thereof.
  15. A method of diagnosing and monitoring conditions mediated by a protein as claimed in claim 9 by determining the presence of a nucleic acid molecule as claimed in any of the preceding claims or a polypeptide as claimed in any of the preceding claims.
  16. A method as claimed in claim 16 wherein the condition is cancer or a disorder of the hematopoietic system.
  17. A method for identifying a substance which associates with a protein as claimed in claim 9 comprising (a) reacting the protein with at least one substance which potentially can associate with the protein, under conditions which permit the association between the substance and protein, and (b) removing or detecting protein associated with the substance, wherein detection of associated protein and substance indicates the substance associates with the protein.
  18. A method for evaluating a compound for its ability to modulate the biological activity of a protein as claimed in claim 9 comprising providing a known concentration of the protein with a substance which associates with the protein and a test compound under conditions which permit the formation of complexes between the substance and protein, and removing and/or detecting complexes.
  19. A method for detecting a nucleic acid molecule encoding a protein comprising an amino acid sequence of SEQ. ID. NO. 2 or 3 in a biological sample comprising the steps of:
    - (a) hybridizing a nucleic acid molecule of claim 2 to nucleic acids of the biological sample, thereby forming a hybridization complex; and
    - (b) detecting the hybridization complex wherein the presence of the hybridization complex correlates with the presence of a nucleic acid molecule encoding the protein in the biological sample.

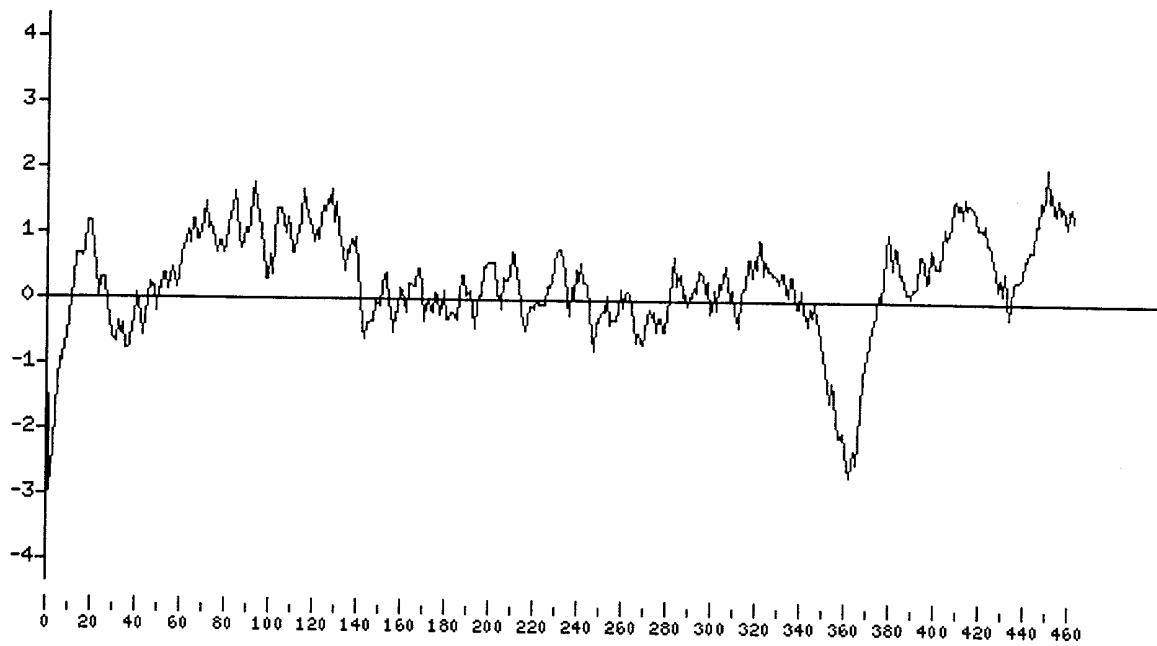
20. A method as claimed in claim 20 wherein nucleic acids of the biological sample are amplified by the polymerase chain reaction prior to the hybridizing step.
21. A method for treating a condition mediated by a protein as claimed in claim 9 comprising administering an effective amount of an antibody as claimed in claim 13 or a substance or compound identified in accordance with a method claimed in claim 18 or 19.
22. A method as claimed in claim 22 wherein the condition is a disorder of the hematopoietic system.
23. A composition comprising one or more of a nucleic acid molecule or protein claimed in any of the preceding claims, or a substance or compound identified using a method as claimed in any of the preceding claims, and a pharmaceutically acceptable carrier, excipient or diluent.
24. Use of one or more of a nucleic acid molecule or protein claimed in any of the preceding claims, or a substance or compound identified using a method as claimed in any of the preceding claims in the preparation of a pharmaceutical composition for treating a condition mediated by a polypeptide as claimed in claim 9.
25. A transgenic non-human mammal which does not express an OB-BPL protein as claimed in claim 9 resulting in an OB-BPL associated pathology.
26. A transgenic animal assay system which provides a model system for testing for an agent that reduces or inhibits an OB-BPL associated pathology
  - (a) administering the agent to a transgenic non-human animal as claimed in claim 26; and
  - (b) determining whether said agent reduces or inhibits an OB-BPL associated pathology in the transgenic non-human animal relative to a transgenic non-human animal of step (a) which has not been administered the agent.

1/6  
Figure 1

TCCTCTAAGTCTTGAGCCCGCAGTTCCTGAGAGAAGAACCCTGAGGAACAGACGTTCCCTCGCGGCCCTGGCACCTCTAACCCAGAC  
 ATG CTG CTG CTG CTG CTG CCC CTG CTC TGG GGG AGG GAG AGG GCG GAA GGA CAG ACA AGT AAA CTG  
 M L L L L L L P L L W G R E R A E G Q T S K L  
 CTG ACG ATG CAG AGT TCC GTG ACG GTG CAG GAA GGC CTG TGT GTC CAT GTG CCC TGC TCC TTC TCC  
 L T M Q S S V T V Q E G L C V H V P C S F S  
 TAC CCC TCG CAT GGC TGG ATT TAC CCT GGC CCA GTA GTT CAT GGC TAC TGG TTC CGG GAA GGG GCC  
 Y P S H G W I Y P G P V V H G Y W F R E G A  
 AAT ACA GAC CAG GAT GCT CCA GTG GCC ACA AAC AAC CCA GCT CGG GCA GTG TGG GAG GAG ACT CGG  
 N T D Q D A P V A T N N P A R A V W E E T R  
 GAC CGA TTC CAC CTC CTT GGG GAC CCA CAT ACC AAG AAT TGC ACC CTG AGC ATC AGA GAT GCC AGA  
 D R F H L L G D P H T K N C T L S I R D A R  
 AGA AGT GAT GCG GGG AGA TAC TTC TTT CGT ATG GAG AAA GGA AGT ATA AAA TGG AAT TAT AAA CAT  
 R S D A G R Y F F R M E K G S I K W N Y K H  
 CAC CGG CTC TCT GTG AAT GTG ACA Ggtaa...INTRON1...cagCC TTG ACC CAC AGG CCC AAC ATC CTC  
 H R L S V N V T A L T H R P N I L  
 ATC CCA GGC ACC CTG GAG TCC GGC TGC CCC CAG AAT CTG ACC TGC TCT GTG CCC TGG GCC TGT GAG  
 I P G T L E S G C P Q N L T C S V P W A C E  
 CAG GGG ACA CCC CCT ATG ATC TCC TGG ATA GGG ACC TCC GTG TCC CCC CTG GAC CCC TCC ACC ACC  
 Q G T P P M I S W I G T S V S P L D P S T T  
 CGC TCC TCG GTG CTC ACC CTC ATC CCA CAG CCC CAG GAC CAT GGC ACC AGC CTC ACC TGT CAG GTG  
 R S S V L T L I P Q P Q D H G T S L T C Q V  
 ACC TTC CCT GGG GCC AGC GTG ACC ACG AAC AAG ACC GTC CAT CTC AAC GTG TCC Tgtga..INTRON2..  
 T F P G A S V T T N K T V H L N V S  
 cagAC CCG CCT CAG AAC TTG ACC ATG ACT GTC TTC CAA GGA GAC GGC ACA Ggtag..INTRON3..cagTA  
 Y P P Q N L T M T V F Q G D G T V  
 TCC ACA GTC TTG GGA AAT GGC TCA TCT CTG TCA CTC CCA GAG GGC CAG TCT CTG CGC CTG GTC TGT  
 S T V L G N G S S L S L P E G Q S L R L V C  
 GCA GTT GAT GCA GTT GAC AGC AAT CCC CCT GCC AGG CTG AGC CTG AGC TGG AGA GGC CTG ACC CTG  
 A V D A V D S N P P A R L S L S W R G L T L  
 TGC CCC TCA CAG CCC TCA AAC CCG GGG GTG CTG GAG CTG CCT TGG GTG CAC CTG AGG GAT GCA GCT  
 C P S Q P S N P G V L E L P W V H L R D A A  
 GAA TTC ACC TGC AGA GCT CAG AAC CCT CTC GGC TCT CAG CAG GTC TAC CTG AAC GTC TCC CTG CAG  
 E F T C R A Q N P L G S Q Q V Y L N V S L Q  
 Agtag.. INTRON4 ..cagGC AAA GCC ACA TCA GGA GTG ACT CAG GGG GTG GTC GGG GGA GCT GGA GCC  
 S K A T S G V T Q G V V G G A G A  
 ACA GCC CTG GTC TTC CTG TCC TTC TGC GTC ATC TTC GTT GTgtaa.. INTRON5 ..cagA GTG AGG  
 T A L V F L S F C V I F V V V R  
 TCC TGC AGG AAG AAA TCG GCA AGG CCA GCA GCG GGC GTG GGA GAT ACG GGC ATA GAG GAT GCA AAC  
 S C R K K S A R P A A G V G D T G I E D A N  
 GCT GTC AGG GGT TCA GCC TCT CAG gtga.. INTRON6 ..cag GGG CCC CTG ACT GAA CCT TGG GCA  
 A V R G S A S Q G P L T E P W A  
 GAA GAC AGT CCC CCA GAC CAG CCT CCC CCA GCT TCT GCC CGC TCC TCA GTG GGG GAA GGA GAG CTC  
 E D S P P D Q P P P A S A R S S V G E G E L  
 CAG TAT GCA TCC CTC AGC TTC CAG ATG GTG AAG CCT TGG GAC TCG CGG GGA CAG GAG GCC ACT GAC  
 Q Y A S L S F Q M V K P W D S R G Q E A T D  
 ACC GAG TAC TCG GAG ATC AAG ATC CAC AGA TGA GAAACTGCAGAGACTCACCCCTGATTGAGGGATCACAGCCCTC  
 T E Y S E I K I H R \*

CAGGCAAGGGAGAAGTCAGAGGCTGATTCTTGTAGAATTAACAGCCCTCAACCGTGATGAGCTATGATAACACTATGAATTATGTGCA  
GAGTGAAAAGCACACAGGCTTTAGAGTCAAAGTATCTCAAACCTGAATCCACACTGTGCCCTCCCTTTTATTTTTTAACTAAAAGA  
CAGACAAATTCCT

Figure 2



3/6  
Figure 3

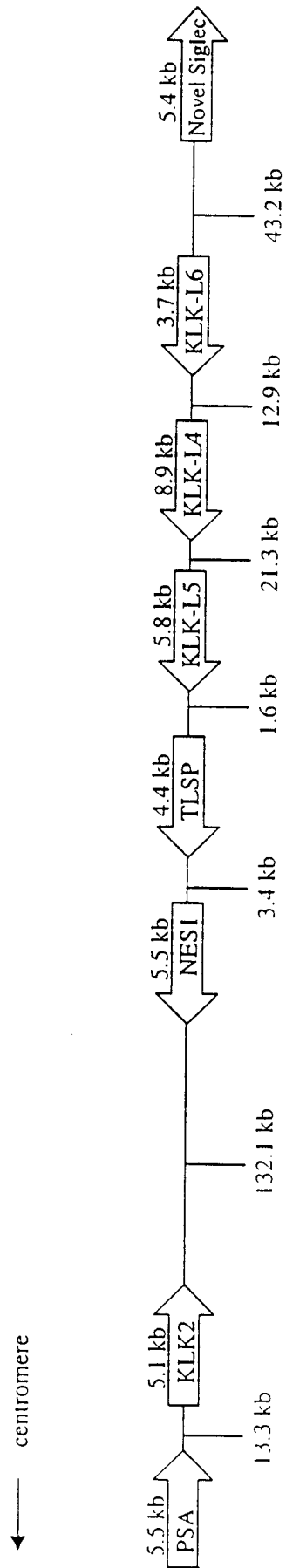


Figure 4

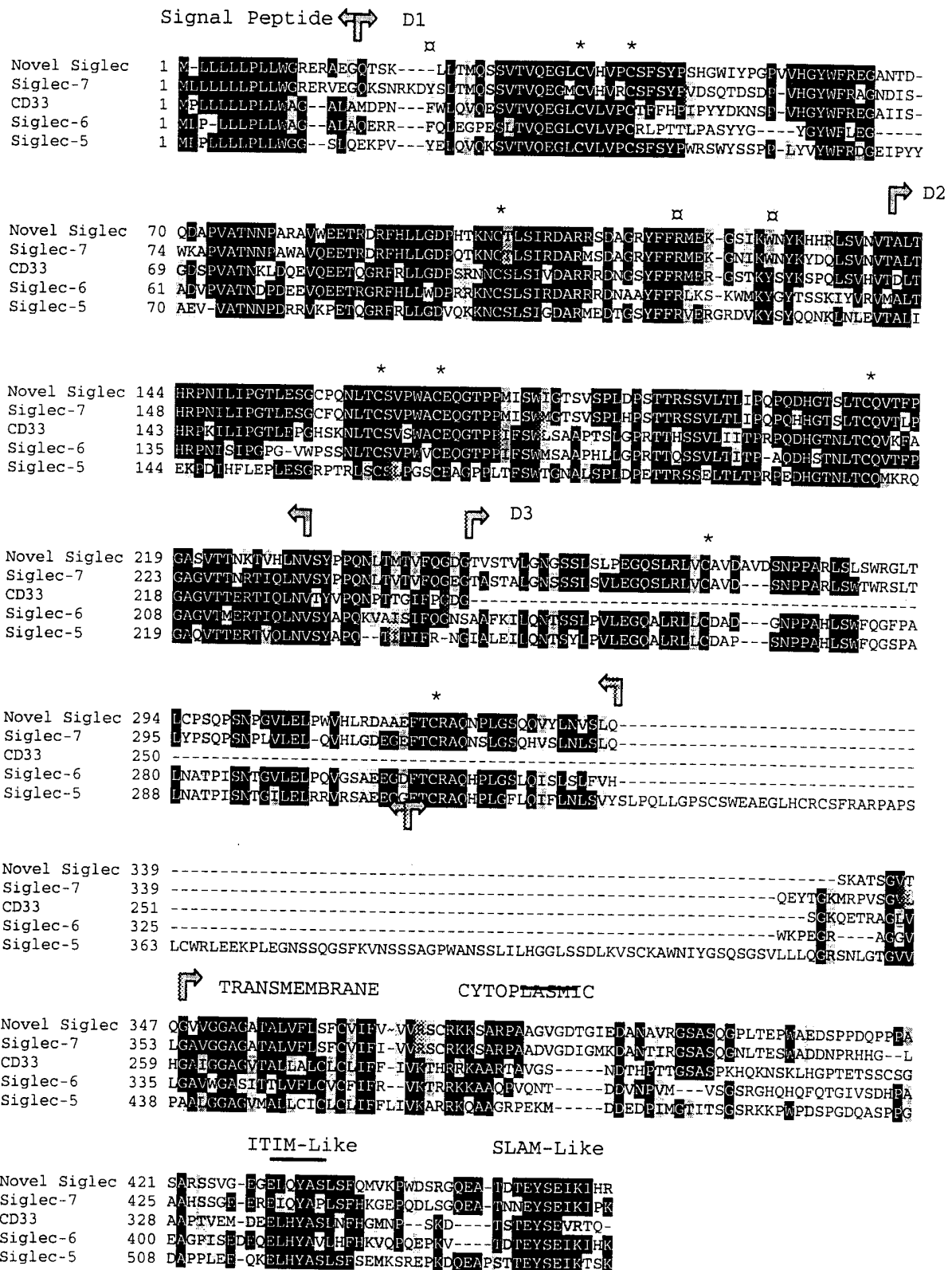
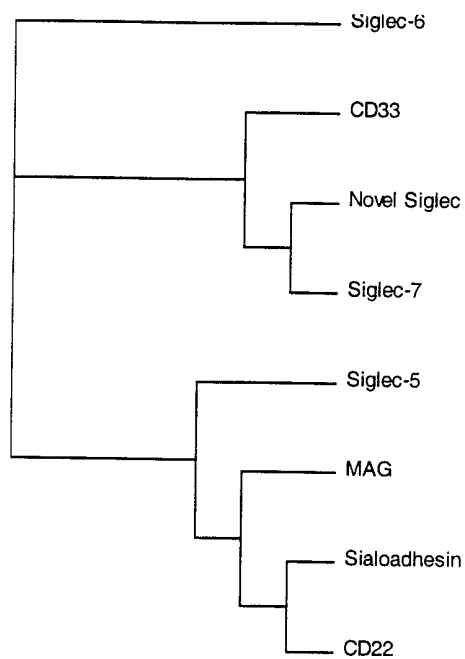
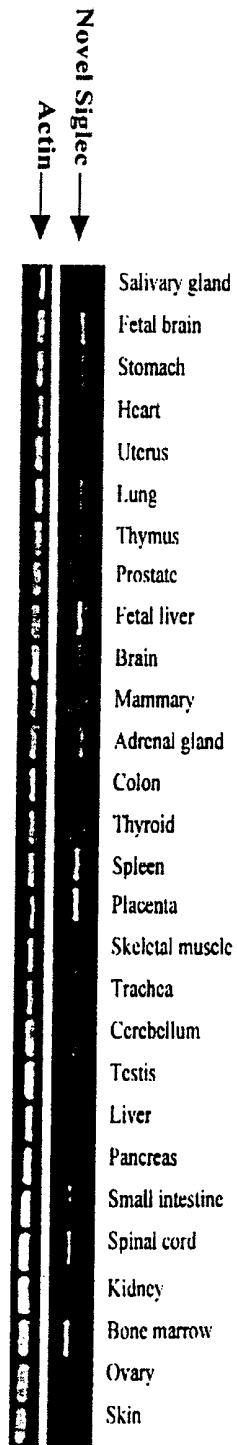


Figure 5



6/6  
Figure 6



## Sequence Listing

SEQ.ID.NO. 1

5 OB-BPL NA

1 ctgacatcgg tctaaggagg ctccaccaca caggatactc ccagttcctc ccagcccagc  
61 cacacagtta caaaaaatct tctgctgaat tcaggtgcat gtttcttctc  
ccatgaaatc  
10 121 ctgcccacg cggaggggtgc atgccccggg gctgggtgctc tactccaca  
tctcctgttc  
181 accgggacct ctgggagcag agcttgcaga agcaaagccc cagggccagc  
tgatecctca  
15 241 tgttctggag taaccaggga agtgtggctg agcgagacat cggtggtgaa  
gaaacccttc  
301 gtgggtcagt gaggaaagga gaaatatctt cccttttgaa atctgccctc  
tttcttccga  
361 atttcctccc ttccaagccc cacagtaca cagtcacagc ctcagtttcc  
cagacctcct  
20 421 gcgagccagg ctcccctctg tgtccttggc gtgtatcaac acatagaatc  
ctcacctcca  
481 cagccccatg tccctctgct cagtctcctc gagattgaa ccctgaccta  
cgggagagtg  
25 541 accccatcct gcctgtgccc cccaactgaa gctcctgccg tggacagctc  
agacgtcagg  
601 agcctccatg tctcccctcc gcacagtgc cccttgggga cagtgtccag  
ctctgctgtg  
661 ttgacctcgt tgtgcagaaa ggggggtgcc acctcgtgct tgcttggggg  
ggatgagagg  
30 721 ctgcctgctc accccacctg cccccatc tctcatggcc ccaagtccca  
cctctgagct  
781 acgcagatat ccacagctc tgactcaggg gtctggccag atgggactca  
ttttcacctc  
841 gcagaaaatg tcacctaagg ggaagggcat tgagagggag gcaggagggtg  
35 ggggtgggccc  
901 tgcattgggt gggaaatctgg gtgagtctgt ctcccgtctt ggctcaggg  
accagaggat  
961 gaacattggg tgggtggacg tggatctccc agggctgacc cgggctgac  
agtgtctggg  
40 1021 tgtaaagtcc ctctctgag gaggctactg ttccgacctc gccctgtct  
tctgttaggg  
1081 cctcctctaa gtcttgagcc cgcagttcct gagagaagaa ccctgaggaa  
cagacgttcc  
1141 ctgcgggccc tggcacctct aaccacagac atgctgctgc tgetgctgcc  
45 cctgctctgg  
1201 gggagggaga gggcggaagg acagacaagt aaactgctga cgatgcagag  
ttccgtgacg  
1261 gtgcaggaag gcctgtgtgt ccatgtgccc tgetccttct cctaccctc  
gcatggctgg  
50 1321 atttaccctg gccagtagt tcatggctac tggttccggg aaggggcca  
tacagaccag  
1381 gatgctccag tggccacaaa caaccagct cgggcaggtg gggaggagac  
tcgggaccga  
1441 ttccacctc ttggggacct acataccaag aattgcacct tgagcatcag  
55 agatgccaga  
1501 agaagtgatg cggggagata cttctttcgt atggagaaag gaagtataaa  
atggaattat  
1561 aaacatcacc ggctctctgt gaatgtgaca ggtaaggcac aggctccagg  
aaaggccaca  
60 1621 gggaaagtc atgggggagg caggaaagg ctgggatgga gccctgccc

caggagaggg  
 1681 cttaggggtga agcgagttag ctcagggcag gagctggacc agagcctgag  
 ctccccccag  
 5 1741 ggctgcacca tggatcctct gacctgatcc tgagtcccc tctcttcacc  
 agccttgacc  
 1801 cacaggccca acatcctcat cccaggcacc ctggagtccg gctgccccca  
 gaatctgacc  
 1861 tgctctgtgc cctgggcctg tgagcagggg acacccccca tgatctcctg  
 gatagggacc  
 10 1921 tccgtgtccc ccctggaccc ctccaccacc cgctcctcgg tgctcacct  
 catcccacag  
 1981 cccaggacc atggcaccag cctcaactgt caggtgacct tccctggggc  
 cagcgtgacc  
 15 2041 acgaacaaga ccgccatct caacgtgtcc tgtgagtgtc gggccgggac  
 gcctgggtcc  
 2101 ctgatggggg gagcgtcaag cctggacact ggggtgctggg tcccggaatc  
 tgggctgggtg  
 2161 gtggggtcag gaggacactg gctctgcctt ccctgtttat gcggctcctg  
 gggacagaca  
 20 2221 gggccagtgt cccagccct cacagtgatg cgggtctcca tgtctttctg  
 tcccagaccc  
 2281 gcctcagaac ttgaccatga ctgtcttcca aggagacggc acaggtagga  
 tggagctccc  
 2341 tccctggggc tggaggagca gggccttcag gtcaggatgg ggctggctta  
 25 ttcctcaacc  
 2401 tggactcact ttggcaaca gggatgtcct tgtgggtgaa ctcagggccc  
 ctctgtatcc  
 2461 ttaggccccca aggccacttg tcccacct cccatcacct cccttggact  
 cccccacaca  
 30 2521 cccccctc agcctcaaac aagaagaggg tggcattcac acagcaggac  
 caggccttga  
 2581 ggctccttct catgtatctc ctgaatacat ctccaccctt atctgtttat  
 ttctgatagt  
 2641 tctgatctaa gtacttctgg acaggtgata aatgtccatg ggcaaaaatt  
 35 caaattgcag  
 2701 agcaaaggct ctctccgat gcctgcccc ctcccagaa ccaaccactg  
 tccatccagg  
 2761 ctgccctgag tctcggtttg tacacctgga ggatctcaga ggtggttga  
 cgtecgtagt  
 40 2821 gagactgtcc gcacctcct ctagggtgtg gtgtgagtcc actgcatgga  
 tggactctga  
 2881 ttttgtggca tctcctaag gaagatcacg gcaactaatt catcctacgg  
 caggatagaa  
 2941 caatcttgta tctacttcca caggaatctc taagcctgtg ggttaagttc  
 45 ctaaaagcaa  
 3001 aatgtagcta cattatatgt tctttcttat tttgaaagat aagcccaaac  
 tgttctcgat  
 3061 gaagcgggga gaagtttaca tcccagcag tgagtgggtga aagtgtgtgt  
 tccagaact  
 50 3121 tcagtctatg tctgtgtgtc agttgtgtc atcagtctct ttctgtatcc  
 ttccttttc  
 3181 tccagatcta tgtatctctc tgacctctg tctcttttc tacagtatcc  
 acagtcttgg  
 3241 gaaatggctc atctctgtca ctcccagagg gccagtctct gcgcctggtc  
 55 tgtgcagttg  
 3301 atgcagttga cagcaatccc cctgccaggc tgagcctgag ctggagaggc  
 ctgacctgt  
 3361 gccctcaca gccctcaaac cgggggtgctc tggagctgcc ttgggtgcac  
 ctgagggatg  
 60 3421 cagctgaatt cacctgcaga gctcagaacc ctctcggctc tcagcaggtc  
 tacctgaacg  
 3481 tctccctgca gagtgagtgc accagtatgc tggggagggg ctggagagga  
 gaacacacct  
 3541 cctccacct tagtaactgc tgagcgtgga ccttcagaga ggagctccgc

tctggctctgt  
 3601 gctcagctgt gaggtctgga acttccctgg gacccacagc accactgtcc  
 tcttctctgcc  
 3661 agggaagggt tgtgggggtgg ggagagggca ggagtggatc tcagagggga  
 5 caggatgggg  
 3721 cgggacaggt gtgttttaggg agacaagcgc ctttctttgc agggctgaac  
 tggagtcaca  
 3781 caactgagat acttgctttg agcatcaaat taaaaaaaaag aaaaagccca  
 gcaagtccagc  
 10 3841 aatcaaatga aatcatattg caatgcaata atcttttaaa aaaagtaaaa  
 attgaatgca  
 3901 aaacaaattc attaatggat aaaatattaa aattgtgaaa aaaaacccca  
 aaaggaatgg  
 3961 ctggcacttg cacgcctcac tggcctcagg aagagtctct ccatgtcctg  
 15 ctctctctca  
 4021 ttcctgttct ttgtgtctgg aaaggggaag tggaaataga agtctaggac  
 cctacaggaa  
 4081 gtgggaggag aagagaccca attctctatg atatatcaca aaaataactc  
 ccatctgtca  
 20 4141 acaggcaaag ccacatcagg agtgactcag ggggtggctg ggggagctgg  
 agccacagcc  
 4201 ctggctctcc tgccttctg cgctcatctc gttgtgtaag catggaccct  
 agagaggag  
 4261 ggaggagag ccctggggga ggacaggctg gaagctggat ccctgaagcc  
 25 agagctggag  
 4321 ggacctggat gggcaagag cttggggcaa gaaggaggtc acaggtgcat  
 ggtgagaatt  
 4381 ccatgtgggc ctgtgtttga ggagctttga gtctgtggca aaccttgga  
 cccactgtcc  
 30 4441 aggagaagag agcctctggt ctcaaccttg gggctctctaa gactggacca  
 ctgctttccc  
 4501 acctcagtca cccctgcagt cccttaatag gaaacacatg ggggtacctg  
 gtctgcccac  
 4561 cgcaccccaa tctgaccaca ctgaaaggct ctctgggtctc ttcactcaga  
 35 gtgaggtcct  
 4621 gcaggaagaa atcggcaagg ccagcagcgg gcgtgggaga tacgggcata  
 gaggatgcaa  
 4681 acgctgtcag gggttcagcc tctcaggtga gtgatgtgga ctctccacag  
 ccagcatgta  
 40 4741 gcctggacac ctcccacagg atgacccca ggactaatca gctgggcgta  
 gccaaagtta  
 4801 cctcctctct gttcttctct tcttctctgt agcccaaat cacaatgttt  
 ggttggtttc  
 4861 ctcccctaag aacagctttt attgtctctg ctccctatcc tgaccctca  
 45 ttgctgaggc  
 4921 ctgaggatct ctgtcttttg ttccctcacc tgtctgctg tctcctctcc  
 tttctgcct  
 4981 ggggggactg tccagaagac atcatcgtcc agttcctctg catttgaaca  
 gctgttcccc  
 50 5041 caccctcaa taccgtttag agcagaagcc agcaatact atctgtcagg  
 gacagataga  
 5101 aactattttc ggcttcatgg gccacacagt ctctattgag ctctcaaat  
 ctgctgttgt  
 5161 agcaagaaag aagccatata ccctgtgtaa acaaatgaat atggctgtgt  
 55 gccaaataaaa  
 5221 ctattcacia acataaagag tgggctggat atgactcaga tactgtagtt  
 tgacaacccc  
 5281 tgatctagag taaaaatccc aaactctata gcctgcagca gtgcacattc  
 tgactttttt  
 60 5341 tgtttttttt tttttttggt gttgttgttt ttgagacaga gtcttgtctt  
 gtcgcccagg  
 5401 ctggagtgca gtggtgcgat ctctgctcac tgcaacttcc accttccggg  
 ttcaagccat  
 5461 tctcctgcct cagcctccgg agtagctggg actacaggcg cctgccacca

cgcccagcta  
 5521 atttttttgt attttttagta gagacggggt ttcactgtgt tagccaggat  
 ggtctcagtc  
 5581 tcctgacctt gtgatctgcc caccttggct tcccgaagtg ctgggattac  
 5 aggcgtgagc  
 5641 cactgtgacc ggccacattc tgacctttta agcacctacc tctccactag  
 ggcaagaaca  
 5701 agggggaagt gagtgaggct gttgcctcaa gtgcattttt tcgtttgttt  
 gtttttgttt  
 10 5761 tttgagatgg agtctcgtc tgtcacccag gatgtagtgc agtggcacia  
 tcttggctta  
 5821 ctgcaacctc tgctcctag gttcaagcga ttctcctgcc tcagcctcct  
 gagtagctgg  
 5881 gattaaaggt gcacaccacc acacctggct aattttgtat ttttagtaga  
 15 gacagggttt  
 5941 caccatggtg gccaggctgg tctcaaactc ctgacctcag gtgatccgcc  
 tacctcagcc  
 6001 tcctgaagag ctgggattac agatgtgagc caccgcgcc catcctcact  
 gtctgctctg  
 20 6061 actcacttct ctctcccatg tctcaggggc cctgactga accttgggca  
 gaagacagtc  
 6121 ccccagacca gcctccccca gcttctgccc gctcctcagt gggggaagga  
 gagctccagt  
 6181 atgcatccct cagcttccag atggggaagc cttgggactc gcggggacag  
 25 gaggccactg  
 6241 acaccgagta ctgggagatc aagatccaca gatgagaaac tgacagagact  
 caccctgatt  
 6301 gagggatcac agcccccca ggcaaggag aagtcagagg ctgattcttg  
 tagaattaac  
 30 6361 agccctcaac gtgatgagct atgataacac tatgaattat gtgcagagtg  
 aaaagcacac  
 6421 aggcttttaga gtcaaagtat ctcaaactg aatccacact gtgccctccc  
 ttttattttt  
 6481 ttaactaaaa gacagacaaa ttct

SEQ.ID.NO. 2

OB-BPL AA

40

PPLSLEPAVPERRTLNRNRRSLAALAPLTPDMLLLLLLPLWGRERAEGQTSKLLTMQSSV  
 TVQEGLCVHVPCSFSPSHGWIYPGPVVHGYWFREGANTDQDAPVATNNPARAVWEETR  
 DRFHLLGDPHTKNCTLSIRDARRSDAGRYFFRMEKGSIKWNYKHHRLSVNVTALTHRPN  
 ILIPGTLESGCPQNLTCVWPWACEQGTTPPMISWIGTSVSPPLDPSTTRSSVLTLPQPQD  
 45 HGTSLTCQVTFPGASVTTNKTVHLNVSYPQNLMTVVFQGDGTGQSLRLVCAVDAVDSN  
 PPARLSLSWRGLTLCPSQPSNPGVLELPWVHLRDAEFTCRAQNPLGSQQVYLVNLSLQK  
 ATSGVTQGVVGGAGATALVFLSFCVIFVGPLTEPWAEDSPPDQPPPASARSSVGEDELQ  
 YASLSFQMKPWDSRGQEATDTEYSEIKIHR

SEQ.ID.NO. 3

OB-BPL AA

5 MLLLLLPLLWGRERAEGQTSKLLTMQSSVTVQEGLCVHVPCSFSPSHGWIYPGPVVH  
GYWFREGANTDQDAPVATNNPARAVWEETRDRFHLLGDPHTKNCLSI RDARRSDAGR  
YFFRMEKGSIKWNYKHHRLSVNVTALTHRPNILIPGTLESGCPQNLTCVWPWACEQG  
TPPMISWIGTSVSPLDPSTTRSSVLTLLIPQPQDHGTSLTCQVTFPGASVTTNKTVHL  
NVSYPQNLTMTVFQGDGTVSTVLGNGSSLSPPEGQSLRLVCAVDAVDSNPPARLSL  
10 SWRGLTLCPSQPSNPGVLELPWVHLRDAAEFTCRAQNPLGSQQVYLNVSLSKATSG  
VTQGVVGGAGATALVFLSFCVIFVVVRSCRKKSARPAAGVGDGTGIEDANAVRGSASQ  
GPLTEPWAEDSPDQPPPASARSSVGEDELQYASLSFQMVKPWDSRGQEATDTEYSE  
IKIHR

SEQ.ID.NO. 4

15

TCACCGGCTCTCTGTGAATG

SEQ.ID.NO. 5

20 GTCTTCTGCCCAAGGTCAG

SEQ.ID.NO. 6

25

TCCTCTAAGTCTTGAGCCCG

SEQ.ID.NO. 7

CAGACGTTGAGATGGACGGT

30 SEQ.ID.NO.8

CGTGGGAGATACGGGCATAG

SEQ.ID.NO. 9

35 AAAAGGGAGGGCACAGTGTG

SEQ.ID.NO. 10

Siglec-7

5 MLLLLLLPLLWGRERVEGQKSNRKDYSLTMQSSVTVQEGMCVHVRCFSFSYPVDSQTDSDPVHGY  
 WFRAGNDISWKAPVATNNPAWAVQEETRDRFHLLGDPQTKNCTLSIRDARMSDAGRYFFRMEKG  
 NIKWNYKYDQLSVNVTALTHRPNILIPGTLESQCFQNLTCVWPWACEQGTTPMISWMGTSVSPLHP  
 STTRSSVLTLPQPQHHGTSLTCQVTLPGAGVTTNRTIQLNVSYPPQNLTVTVFQEGTASTALGNS  
 SSVLEGLQSLRLVCAVDSNPPARLSWTWRSLTYPSQPSNPLVLELQVHLGDEGEFTCRAQNSLG  
 10 SQHVSLNLSLQQEYTGKMRPVSGVLLGAVGGAGATALVFLSFCVIFIVRSCRKKSARPAADVGD  
 GMKDANTIRGSASQGNLTESWADDNPRHHGLAAHSSGEEREIQYAPLSFHKGEPQDLSGQEATNN  
 EYSEIKIPK

SEQ.ID.NO. 11

15

CD33

MPLLLLLPLLWAGALAMPNFWLQVQESVTVQEGLCVLVPCTFFHPIPYDKNSPVHGYWFREG  
 AIIIGDSPVATNKLDQEVQEETQGRFRLGDPNRNCSLSIVDARRRDNGSYFFRMERGSTKYSSYSS  
 20 PQLSVHVTDLTHRPKILIPGTLEPGHNSKNTCSVSWACEQGTPIFSWLSAAPTSLGPRTHSSVLIIT  
 PRPQDHGTNLTCQVKFAGAGVTTERTIQLNVTYVPQNPTTGIFPGDGSQKQETRAGLVHGAIGGA  
 GVTALLALCLCLIFFIVKTHRKAARTAVGSNDTHPTTGSASPKHQKNSKLHGPTETSSCSGAAPT  
 VEM-DEELHYASLNFGMNPSKDTSTEYSEVRTQ

25 SEQ.ID.NO. 12

Siglec-6

MLP-LLLPLLWAGALAQERRFQLEGPELTVQEGLCVLVPCLPTTLPASYYGYGYWFLEG  
 30 ADVPVATNDPDEEVQEETRGRFHLLWDPKRNCSLSIRDARRRDNAAYFFRLKSKWMKYGYTSS  
 KIYVRVMALTHRPNISIPGGVWPSSNLTCVWPVWCEQGTPIFSWMSAAPHLLGPRTTQSSVLTIT  
 PAQDHSTNLTCQVTFPGAGVTMERTIQLNVSYAPQKVAISIFQNSAAFKILQNTSSLPVLEGQALR  
 LLCADGNPPAHLNFWQGFALNATPISNTGVLELPQVGSAAEGDFTCRAQHPLGSLQISLSLFVH  
 WKPEGRAGGVLGAVWGASITTLVFLCVCFIFR.VKTRRKKAAQPVQNTDDVNPVMVSGSRGHQHQ  
 35 FQTGIVSDHPA EAGPISEDEQELHYAVLHFHKVQPQEPKVTDTTEYSEIKIHK

SEQ.ID.NO. 13

Siglec-5

5 MLPLLLLPLLWGGSLQEKPVYELQVQKSVTVQEGLCVLPVPCSFSPWRSWYSSPPLYVY  
WFRDGEIPYYAEVVATNNPDRRVKPETQGRFRLLDVQKKNCSLSIGDARMEDTGSYF  
RVERGRDVKYSYQQNKLNLEVTALIEKPDIFLEPLESGRPTRLSCSLPGSCEAGPPLT  
FSWTGNALSPLDPETTRSELTLTPRPEDHGTLNLTQMKRQGAQVTTERTVQLNVSYAPQT  
ITIFRNGIALEILQNTSYLPVLEGQALRLLCDAPSNPPAHLNWFQSPALNATPISNTGILELRRVRS  
10 EEGGFTCRAQHPLGFLQIFLNLSVYSLPQLLGPSCSWEAEGLHCRCFRARPAPSLCWRLEEKPLEG  
NSSQGSFKVNSSSAGPWANSSLILHGGLSSDLKVSCKAWNIYGSQGSVLLLQGRSNLGTGVVPA  
LGGAGVMALLCICLCLIFFLIVKARRKQAAGRPEKMDDEDPIMGTITSGSRKKPWPDPSPGDQASPP  
GDAPPLEE-QKELHYASLSFSEMKSREPKDQEAPSTTEYSEIKTSK

15

# INTERNATIONAL SEARCH REPORT

International Application No <b>PCT/CA 00/00259</b>
--

**A. CLASSIFICATION OF SUBJECT MATTER**  
 IPC 7 C12N15/12 C07K14/705 C07K16/28 C12Q1/68 A01K67/027  
 C12N15/62 A61K38/00 G01N33/566

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
 IPC 7 C12N C07K C12Q A01K A61K G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, EMBASE, MEDLINE, BIOSIS, STRAND, CAB Data, PAJ, WPI Data

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0 869 178 A (SMITHKLINE BEECHAM CORP) 7 October 1998 (1998-10-07) the whole document ---	2,6,7,10
A	RAZI N. ET AL: "Masking and unmasking of the sialic acid-binding lectin activity of CD22 Siglec -2) on B lymphocytes." PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, (23 JUN 1998) 95/13 (7469-7474). , XP000913900 the whole document --- -/--	

Further documents are listed in the continuation of box C.       Patent family members are listed in annex.

\* Special categories of cited documents :

*A* document defining the general state of the art which is not considered to be of particular relevance *E* earlier document but published on or after the international filing date *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) *O* document referring to an oral disclosure, use, exhibition or other means *P* document published prior to the international filing date but later than the priority date claimed	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. *&* document member of the same patent family
---	---

Date of the actual completion of the international search  <h2 style="text-align: center;">10 August 2000</h2>	Date of mailing of the international search report  <h2 style="text-align: center;">17/08/2000</h2>
--	---

Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer  <h2 style="text-align: center;">Hix, R</h2>
--	---

INTERNATIONAL SEARCH REPORT

International Application No

PCT/CA 00/00259

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CORNISH A.L. ET AL: "Characterization of siglec -5, a novel glycoprotein expressed on myeloid cells related to CD33." BLOOD, (15 SEP 1998) 92/6 (2123-2132). , XP000913901 cited in the application the whole document	
A	CROCKER P R ET AL: "Siglecs: a family of sialic-acid binding lectins 'letter!'" GLYCOBIOLOGY, (1998 FEB) 8 (2) , pages v-vi, XP000913934 cited in the application the whole document	
A	KELM S. ET AL: "Functional groups of sialic acids involved in binding to siglecs (sialoadhesins) deduced from interactions with synthetic analogues." EUROPEAN JOURNAL OF BIOCHEMISTRY, (1 AUG 1998) 255/3 (663-672). , XP000913907 the whole document	
A	CROCKER P R ET AL: "SIALOADHESIN AND RELATED CELLULAR RECOGNITION MOLECULES OF THE IMMUNOGLOBULIN SUPERFAMILY" BIOCHEMICAL SOCIETY TRANSACTIONS, GB, COLCHESTER, ESSEX, vol. 24, 1996, pages 150-156, XP000913910 ISSN: 0300-5127 cited in the application the whole document	
P, X	YOUSEF G M ET AL: "IDENTIFICATION OF NOVEL HUMAN KALLIKREIN-LIKE GENES ON CHROMOSOME 19Q13.3 - Q13.4" ANTICANCER RESEARCH, GR, HELENIC ANTICANCER INSTITUTE, ATHENS, , vol. 19, July 1999 (1999-07), pages 2843-2852, XP000911217 ISSN: 0250-7005 the whole document	1-15
P, X	EP 0 924 297 A (SMITHKLINE BEECHAM CORP) 23 June 1999 (1999-06-23) the whole document	2,6,7,10
E	WO 00 32221 A (FERRARA NAPOLEONE ; WILLIAMS P MICKEY (US); BAKER KEVIN P (US); GEN) 8 June 2000 (2000-06-08) the whole document	2,6,7,10
	-/--	

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/CA 00/00259

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
E	WO 00 12708 A (BAKER KEVIN ; GENENTECH INC (US); GODDARD AUDREY (US); GURNEY AUSTI) 9 March 2000 (2000-03-09) the whole document -----	2, 6, 7, 10

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/CA 00/00259

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0869178 A	07-10-1998	JP 11042093 A	16-02-1999
		JP 2000125888 A	09-05-2000
EP 0924297 A	23-06-1999	CA 2226910 A	02-10-1998
		JP 11028095 A	02-02-1999
		JP 2000083689 A	28-03-2000
WO 0032221 A	08-06-2000	AU 1602999 A	16-06-1999
		WO 0032776 A	08-06-2000
		WO 0032778 A	08-06-2000
		AU 3072199 A	27-09-1999
		WO 0037638 A	29-06-2000
		WO 0037640 A	29-06-2000
		AU 4328699 A	20-12-1999
		WO 0036102 A	22-06-2000
		AU 5590899 A	21-03-2000
		AU 5816799 A	03-04-2000
		AU 5920099 A	03-04-2000
		AU 5922999 A	03-04-2000
		AU 6498499 A	03-04-2000
AU 1200500 A	01-05-2000		
WO 0012708 A	09-03-2000	AU 5590899 A	21-03-2000
		AU 6041399 A	10-04-2000
		WO 0017353 A	30-03-2000