Supplementary Figure legends

Supplementary Figure 1: *Screening of a panel of normal tissues and tumor-derived cell lines by RT-PCR*
RT-PCR was performed on cDNAs from normal, non-cancerous tissues (lettered in black) and from tumor-derived cell lines (lettered in red). 2% agarose gels were run, and bands visualized by ethidium bromide staining. **N**: denotes the band corresponding to the novel splice form; **K**: denotes the band corresponding to the known splice form. **A**) *SCML1*; **B**) *EIF4G2*; **C**) *TPD52L2*; **D**) *NOS2A*; **E**) *HLA-DMB*; **F**) *SRRM1*; **G**) *PCBP2*; **H**) *BAT3*.

Supplementary Figure 2: *Production of a novel isoform of* IKBβ *by alternative splicing*.
A) Gene structure of *IKBβ*. Exons are shown as boxes. Translation of the known isoform of *IKBβ* (IKBβ-1) starts at exon Ib (shown in red). An alternative isoform of *IKBβ* (IKBβ-2) skips exon Ib, and uses alternative exon Ia (shown in black). This results in an isoform which is truncated at the N-terminus.
B) Predicted protein sequence of IKBβ. Sequence 1: the known isoform of IKBβ (IKBβ-1). Sequence 2: the novel isoform, IKBβ-2. The ankyrin domains (ANK) are underlined in bold. The inducible phosphorylation sites, S19 and S23, are shown in green lettering. The novel isoform, IKBβ-2, lacks the inducible phosphorylation sites, and also lacks the first ankyrin domain.

Supplementary tables

Table 1: *Classification of the original tissue sources of a random sample of mRNAs*
The tissue sources of a sample of 44 complete mRNA sequences (representing 20 genes) that were deposited in Genbank are listed below. The information below is from Genbank 2003 data. Genbank accession numbers are shown in brackets.

Gene Symbol	**known mRNAs cloned from: (Genbank data)**	
	Tumors	**Normal tissues**
BAT3	renal cell adenocarcinoma (1) (BC003133) T cell line HPB-All (leukemia) (2) (M33519)	
BCAS1	breast cancer cell line BT474 (3) (AF041260)	
CACNA2D4		testis (direct submission from MIPS; http://mips.gsf.de/proj/cDNA) (AL137658)
CAPSN1	uterus ieiomyosarcoma (1) (BC007779) muscle rhabdomyosarcoma (1) (BC011903) pancreas epithelioid carcinoma (1) (BC017308) Skin, melanotic melanoma (1) (BC000592) infant brain from patient with ALL (5) (AY007141)	Spleen (4) (X04106)
CHFR	placenta choriocarcinoma (1) (BC012072) NT2 neuronal cell line (teratocarcinoma) (6) (AK001658, AK027687)	
DCTN1		fetal brain (7) (8) (AF064205, X98801))
EIF4G2	K562 cells (chronic myeloid leukemia) (9) (X89713)	placenta (selected for longest clone) (10) (U73824)
GOLGA2	Brain, neuroblastoma (1) (BC069268)	placenta (selected for longest clone) (11) (AF248953)
GOSR2	large cell lung carcinoma (1) (BC009710)	
HLA-DMB	EBV transformed B cells (12) (U15085) B lymphoblastoid cell line JY (12) (Z23139)	
ITGAE		TGF-beta1 induced intra-epithelial lymphocytes (13) (L25851)
KHK	HepG2 liver hepatoblastoma (14) (X78678, Y09341) Lung, small cell carcinoma (1) (BC006233)	
LCK	Lymph, lymphoma (1) (BC013200)	Leukocytes (15) (X13529, M36881) PBL's from immunodeficient (CVID) patients

		(16) (AF228313)
LEF1	Jurkat T cells, leukemia (17) (AF288571)	
NFKBIB/IKBβ	pancreas epitheloid carcinoma (1) (BC015528)	
PRP18	Hela cell line (18) (U51990) placenta choriocarcinoma (1) (BC000794)	
NOS2A	glioblastoma cell line A-172 (19) (D26525) osteoblastoma (direct submission by Ogawa et al; unpublished) (AB022318) colorectal adenocarcinoma (23) (L24553)	chondrocytes treated with IL1beta (20) (21) (X73029, U05810) respiratory epithelium (22) (U20141) cardiac myocytes (24) (AF068236) Hepatocyte (25) (L09210)
TPD52L2	breast carcinoma (26) (AF004430) placenta choriocarcinoma (1) (BC006804)	
SRRM1	U937 lymphoma cell line (27) (AF048977) Lung, small cell carcinoma (1) (BC017315)	Testis (1) (BC036187)
WBP2	Brain, anaplastic oligodendroglioma (1) (BC010616) Placenta, choriocarcinoma (1) (BC007452)	

Table 2: *Classification of 'novel normal' splices*

Number of 'novel normal' splices (LOD>1.0)	1308
Novel normal splices that are exon skips	472
Novel normal exon skips that are in coding region	349
Novel normal exon skips in coding region and inframe	104

Table 3: *Distribution of splice forms in normal tissues*

N: indicates the presence of the novel form; K: indicates the presence of the known form.
To indicate the relative abundance of the each splice form in each tissue, we adopted the following convention:
Boldface indicates that the splice form is the major form; *italics* indicates that the splice form is the minor form; $^{=}$ denotes that both the known and the novel forms were equally abundant; nd indicates not determined; - indicates that no expression observed.

gene	bone marrow	brain	breast	lung	skeletal muscle	placenta	spleen	testes	kidney
BAT3	K	K	K	K	K	**K**+N	K	**K**+*N*	**K**+*N*
BCAS1	K+N$^{=}$	*K***+N**	K	**K**+*N*	*K*+**N**	K	K	K+N$^{=}$	nd
CHFR	K+N$^{=}$	K+N$^{=}$	K+N$^{=}$	K+N$^{=}$	K	**K**+*N*	**K**+*N*	*K*+**N**	nd
EIF4G2	**K**+*N*	K	**K**+*N*	**K**+*N*	K	**K**+*N*	**K**+*N*	K	nd
HLA-DMB	**K**+*N*	**K**+*N*	**K**+*N*	**K**+*N*	**K**+*N*	**K**+*N*	**K**+*N*	**K**+*N*	nd
NOS2A	N	**K+***N*	K	K	K+N$^{=}$	*K*+**N**	K+N$^{=}$	N	nd
PCBP2	**K**+*N*	**K**+*N*	**K**+*N*	**K**+*N*	**K**+*N*	**K**+*N*	**K**+*N*	**K**+*N*	nd
SCML1	K	K	K	K	K	K	K	*K*+**N**	nd
SRRM1	*K*+**N**	N	N	*K*+**N**	*K*+**N**	N	*K*+**N**	*K*+**N**	*K*+**N**
TPD52L2	*K*+**N**	*K*+**N**	-	*K*+**N**	*K*+**N**	*K*+**N**	*K*+**N**	*K*+**N**	nd

BAT3 and *SCML1*: the novel splice form was strongly restricted to specific tissues (for *BAT3*, testes, kidney and placenta; for *SCML1*, testes).

HLA-DMB, *EIF4G2* and *PCBP2*: the novel splice form was ubiquitously expressed, but at a much lower level than the known splice form, i.e. as the minor form.

TPD52L2, *SRMM1*, *BCAS1* and *NOS2A*: the novel splice form was the dominant or major isoform in most normal tissues, or had approximately equal representation in normal tissues to that of the known splice form.

Table 4: *Distribution of splice forms in brain versus brain tumor cell lines*
N: indicates the presence of the novel form; K: indicates the presence of the known form; nd - not determined.
* denotes that this form is present greatly in excess of the other form.
DF, RW and U87: glioblastoma cell lines
SKNSH, SKNMC: neuroblastoma cell lines

gene	brain	DF	RW	U87	SKNSH	SKNMC
BAT3	K	K	K+N	K+N	K	K+N
BCAS1	N+K	nd	nd	K	K	K
CHFR	K+N	K	K	K	K	K
EIF4G2	K	K+N	K+N	K+N	K+N	K+N
HLA-DMB	K+N	K+N	K+N	K+N	K+N	K+N
NOS2A	K+N	N	N	N	K	K
PCBP2	K+N	K+N	K+N	K+N	K+N	K+N
SCML1	K	K+N	K+N	K+N	K+N	K+N
SRRM1	K+N	K+N	K+N	K+N	K+N*	K+N
TPD52L2	K+N	K+N	K+N	K+N	K+N	K+N

TPD52L2: All samples have both forms, with the novel form being the major form. Tumor cell lines have less of the novel form.
SRRM1: All samples have both forms, with novel form being the major form.
BAT3: expressed at low levels in all samples.
PCBP2: novel is minor form in all samples.
HLA-DMB: all samples have both forms. Tumor cell lines have lower amounts of major form.

Table 5: *Primer sequences used for RT-PCR*

Gene Symbol	Forward primer sequence	Reverse primer sequence
ACAA1	gaatctgaggccggaaca	gctcttgctgccttctgc
BAT3	tccctccttggtgagctg	tgagagaagcagctgggg
BCAS1	ccagaaggactggagactgc	tgggtcctgatttctccttg
CACNA2D4	cttgccctctctctttctgc	ttgcacatggcctgatagtc
CAPNS1	gcctggtttgagagttctgc	atgctgcgacatgtgtcaat
CCT6A	tgaagaccctgaacccca	tgctttttaatggccaaaatg
CCND3	gagctgctgtgttgcgaa	gcagaatgaaggccagga
CHFR	taccagcaccagtggaacag	cttcctttaggggagatgcc
COCH	cggcccagcctttatagc	gcccctctaggtctggga
DARS	gccgcaagagtcaggaga	cttgcatggtctcccacc
DCTN1	gatgaagggcatggcatc	tcacctgctgacgctgag
DCTN3	aaggtgcaggtggctttg	gggatggggagcagctat
EIF4G2/p97/DAP5	ccatgggggacacatcat	tcctggataagcggtgga
ELN	gtcggagggcttggagtt	gcatgggatggggttaca
EPB49	acgctgtcctctccctc	agctccacgtgttccagg
ESRRA	gcctccaacgagtgtgagat	gtcagacagcgacagcgat
FXR1	tgccagcgaatctcatca	gggccgtttattgccttt
GOLGA2	gacagagcaaattggccg	gtgagactggcaccaggg
GOSR2	tttgtccagcaaggagcc	gtcagcagcatcccacct
HLA-DMB	ccatctgtgcaagtagccaa	aagatgttggagaggcatgg
ITGAE	aaatttgcgtcccaacca	ccagaagtccaccaacgc
KHK	gctggacgtcatcagcct	atccgctgcagcatcttc
LCK	tggacaacggtggcttct	ttgcagctcagggtgtca
LEF1	ccgtcacacatcccatca	ctgaggcttcacgtgcatt
LMAN2L/VIPL	gatggcttggcaatctgg	attgcggacaatggctgt
MAPT/TAU	gccaccttctcctcctcc	ttccagtcccgtctttgc
MEF2B	gaggggactaaacacctcc	agacttgatgctgactgggg
NFKBIB/IKBb	cagacgcgctttcgtacaaca	tctggcctaggtcattctgc
NOS2A	ccaggaggagatgctgga	ccctcagagcgctgacat
PCBP2	gctggcattccacaatcc	gatgcatccaaacctgcc
PDHA	aaggagacttgggggcac	cctccctggtgagcactg
PPT1	cctgatggccctgaagaa	cgggtttcatccaaggaa
PRP18	ttgggcttgttgttccgtat	ggtcctcctcttttggctg

RPN2	gtgcccttactgctcgtctc	ttaatggatggctcagtccc
SCAMP2	cggacccagtggatgtaaac	gcagctttcctgtccagttc
SCML1	gcagttttaccgcgacct	gaaggttgctcggggaat
SH3BGR	tttggctcctcctccaga	aaccaggtcaggcagcat
SLC3A2	gtggccaccaaggtgaag	caggagccttgcctgaga
SUPT5H	gcgggcagtgagaaagaa	tccagaggttttgcaggc
SRRM1/SRM160	cgccaatacagacgacaaaa	ggatggacttctcctccgtc
TAF2	tccagagtcagtagcaggca	tttacctgttcccttccgtg
TEX27/Hs.6120	acctcgataaccacgcca	gcaccgacgtcgactctt
TIE1	gtgcccaggcatatcagg	tagctctatgctgccccg
TIM17b	cctccagctccaaaagagtg	ctgcgaagctacctccaatc
TPD52L2	aaggagaggcactgtggaga	tgaggaagggaggttgtcac
UBE1C	ggtgtggggacactggag	tctgcagcaacttcagcct
WARS	aagaccacaccacttcctgc	tgaggaaactgtgtaaccactga
WBP2	ggaacagtgaaggcggaa	tgtgaggattgcctgggt
Z391G	cctggagatggatgaccg	atggcaaagacaggcagg
TBC1D7	gcttctcccggaccactt	agcctgaggtgtggcatc

Supplementary Data References:

1. Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F. *et al*. (2002) Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proc Natl Acad Sci U S A*, **99,** 16899-16903.
2. Banerji, J., Sands, J., Strominger, J.L. and Spies, T. (1990) A gene pair from the human major histocompatibility complex encodes large proline-rich proteins with multiple repeated motifs and a single ubiquitin-like domain. *Proc Natl Acad Sci U S A*, **87,** 2374-2378.
3. Collins, C., Rommens, J.M., Kowbel, D., Godfrey, T., Tanner, M., Hwang, S.I., Polikoff, D., Nonet, G., Cochran, J., Myambo, K. *et al*. (1998) Positional cloning of ZNF217 and NABC1: genes amplified at 20q13.2 and overexpressed in breast carcinoma. *Proc Natl Acad Sci U S A*, **95,** 8703-8708.
4. Ohno, S., Emori, Y. and Suzuki, K. (1986) Nucleotide sequence of a cDNA coding for the small subunit of human calcium-dependent protease. *Nucleic Acids Res*, **14,** 5559.
5. Yu, W., Andersson, B., Worley, K.C., Muzny, D.M., Ding, Y., Liu, W., Ricafrente, J.Y., Wentland, M.A., Lennon, G. and Gibbs, R.A. (1997) Large-scale concatenation cDNA sequencing. *Genome Res*, **7,** 353-358.
6. Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K. *et al*. (2004) Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nat Genet*, **36,** 40-45.
7. Collin, G.B., Nishina, P.M., Marshall, J.D. and Naggert, J.K. (1998) Human DCTN1: genomic structure and evaluation as a candidate for Alstrom syndrome. *Genomics*, **53,** 359-364.
8. Tokito, M.K., Howland, D.S., Lee, V.M. and Holzbaur, E.L. (1996) Functionally distinct isoforms of dynactin are expressed in human neurons. *Mol Biol Cell*, **7,** 1167-1180.
9. Levy-Strumpf, N., Deiss, L.P., Berissi, H. and Kimchi, A. (1997) DAP-5, a novel homolog of eukaryotic translation initiation factor 4G isolated as a putative modulator of gamma interferon-induced programmed cell death. *Mol Cell Biol*, **17,** 1615-1625.
10. Imataka, H., Olsen, H.S. and Sonenberg, N. (1997) A new translational regulator with homology to eukaryotic translation initiation factor 4G. *Embo J*, **16,** 817-825.
11. Weide, T., Bayer, M., Koster, M., Siebrasse, J.P., Peters, R. and Barnekow, A. (2001) The Golgi matrix protein GM130: a specific interacting partner of the small GTPase rab1b. *EMBO Rep*, **2,** 336-341.
12. Kelly, A.P., Monaco, J.J., Cho, S.G. and Trowsdale, J. (1991) A new human HLA class II-related locus, DM. *Nature*, **353,** 571-573.
13. Shaw, S.K., Cepek, K.L., Murphy, E.A., Russell, G.J., Brenner, M.B. and Parker, C.M. (1994) Molecular cloning of the human mucosal lymphocyte integrin alpha E subunit. Unusual structure and restricted RNA distribution. *J Biol Chem*, **269,** 6016-6025.
14. Bonthron, D.T., Brady, N., Donaldson, I.A. and Steinmann, B. (1994) Molecular basis of essential fructosuria: molecular cloning and mutational analysis of human ketohexokinase (fructokinase). *Hum Mol Genet*, **3,** 1627-1631.
15. Perlmutter, R.M., Marth, J.D., Lewis, D.B., Peet, R., Ziegler, S.F. and Wilson, C.B. (1988) Structure and expression of lck transcripts in human lymphoid cells. *J Cell Biochem*, **38,** 117-126.

16. Boncristiano, M., Majolini, M.B., D'Elios, M.M., Pacini, S., Valensin, S., Ulivieri, C., Amedei, A., Falini, B., Del Prete, G., Telford, J.L. *et al*. (2000) Defective recruitment and activation of ZAP-70 in common variable immunodeficiency patients with T cell defects. *Eur J Immunol*, **30,** 2632-2638.
17. Hovanes, K., Li, T.W. and Waterman, M.L. (2000) The human LEF-1 gene contains a promoter preferentially active in lymphocytes and encodes multiple isoforms derived from alternative splicing. *Nucleic Acids Res*, **28,** 1994-2003.
18. Horowitz, D.S. and Krainer, A.R. (1997) A human protein required for the second step of pre-mRNA splicing is functionally related to a yeast splicing factor. *Genes Dev*, **11,** 139-151.
19. Hokari, A., Zeniya, M. and Esumi, H. (1994) Cloning and functional expression of human inducible nitric oxide synthase (NOS) cDNA from a glioblastoma cell line A-172. *J Biochem (Tokyo)*, **116,** 575-581.
20. Charles, I.G., Palmer, R.M., Hickery, M.S., Bayliss, M.T., Chubb, A.P., Hall, V.S., Moss, D.W. and Moncada, S. (1993) Cloning, characterization, and expression of a cDNA encoding an inducible nitric oxide synthase from the human chondrocyte. *Proc Natl Acad Sci U S A*, **90,** 11419-11423.
21. Maier, R., Bilbe, G., Rediske, J. and Lotz, M. (1994) Inducible nitric oxide synthase from human articular chondrocytes: cDNA cloning and analysis of mRNA expression. *Biochim Biophys Acta*, **1208,** 145-150.
22. Guo, F.H., De Raeve, H.R., Rice, T.W., Stuehr, D.J., Thunnissen, F.B. and Erzurum, S.C. (1995) Continuous nitric oxide synthesis by inducible nitric oxide synthase in normal human airway epithelium in vivo. *Proc Natl Acad Sci U S A*, **92,** 7809-7813.
23. Sherman, P.A., Laubach, V.E., Reep, B.R. and Wood, E.R. (1993) Purification and cDNA sequence of an inducible nitric oxide synthase from a human tumor cell line. *Biochemistry*, **32,** 11600-11605.
24. Luss, H., Li, R.K., Shapiro, R.A., Tzeng, E., McGowan, F.X., Yoneyama, T., Hatakeyama, K., Geller, D.A., Mickle, D.A., Simmons, R.L. *et al*. (1997) Dedifferentiated human ventricular cardiac myocytes express inducible nitric oxide synthase mRNA but not protein in response to IL-1, TNF, IFNgamma, and LPS. *J Mol Cell Cardiol*, **29,** 1153-1165.
25. Geller, D.A., Lowenstein, C.J., Shapiro, R.A., Nussler, A.K., Di Silvio, M., Wang, S.C., Nakayama, D.K., Simmons, R.L., Snyder, S.H. and Billiar, T.R. (1993) Molecular cloning and expression of inducible nitric oxide synthase from human hepatocytes. *Proc Natl Acad Sci U S A*, **90,** 3491-3495.
26. Nourse, C.R., Mattei, M.G., Gunning, P. and Byrne, J.A. (1998) Cloning of a third member of the D52 gene family indicates alternative coding sequence usage in D52-like transcripts. *Biochim Biophys Acta*, **1443,** 155-168.
27. Blencowe, B.J., Issner, R., Nickerson, J.A. and Sharp, P.A. (1998) A coactivator of pre-mRNA splicing. *Genes Dev*, **12,** 996-1009.