(19) IPOS INTELLECTUAL PROPERTY OFFICE OF SINGAPORE

(11) Publication number: **SG 193489 A1**
(43) Publication date: **30.10.2013**
(51) Int. Cl: **C12Q 1/68;**

(12) Patent Application

(21) Application number: **2013069646**
(22) Date of filing: **16.03.2012**
(30) Priority: **US 61/454,426 18.03.2011**

(71) Applicant: **EISAI R&D MANAGEMENT CO., LTD. 6-10 KOISHIKAWA, 4-CHOME, BUNKYO-KU, TOKYO, 112-8088 JP**

(72) Inventor: **AGOULNIK, Sergei, I. 62 Wildwood Road Andover, MA 01810 US**
BYRNE, Michael, Chapman 34 Hyslop Road Brookline, MA 02445 US
LITTLEFIELD, Bruce, A. 4 Twin Brook Circle Andover, MA 01810 US

(54) **Title:**
METHODS AND COMPOSITIONS FOR PREDICTING RESPONSE TO ERIBULIN

(57) **Abstract:**
The invention provides methods for predicting the efficacy of eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (e.g., eribulin mesylate), in the treatment of a subject suffering from breast cancer by determining the level of particular biomarkers in a sample derived from the subject.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 September 2012 (27.09.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/129100 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2012/029479

(22) International Filing Date:
16 March 2012 (16.03.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/454,426 18 March 2011 (18.03.2011) US

(71) Applicant *(for all designated States except US)*: **EISAI R&D MANAGEMENT CO., LTD.** [JP/JP]; 6-10 Koishikawa, 4-Chome, Bunkyo-ku, Tokyo, 112-8088 (JP).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **AGOULNIK, Sergei, I.** [US/US]; 62 Wildwood Road, Andover, MA 01810 (US). **BYRNE, Michael, Chapman** [US/US]; 34 Hyslop Road, Brookline, MA 02445 (US). **LITTLEFIELD, Bruce, A.** [US/US]; 4 Twin Brook Circle, Andover, MA 01810 (US).

(74) Agents: **ZACHARAKIS, Maria, Laccotripe** et al.; McCarter & English, LLP, 265 Franklin Street, Boston, MA 02110 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

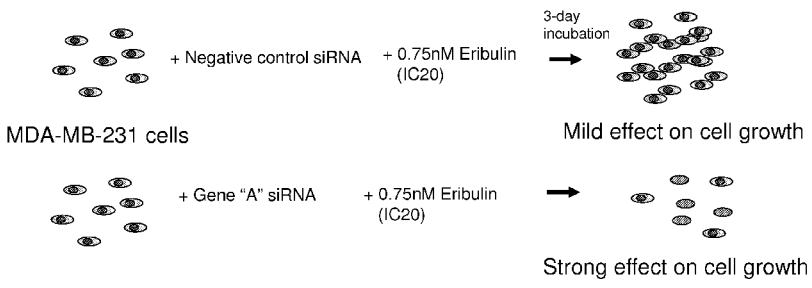
— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

— *with sequence listing part of description (Rule 5.2(a))*

(54) **Title**: METHODS AND COMPOSITIONS FOR PREDICTING RESPONSE TO ERIBULIN

FIGURE 1
Pre-Clinical Biomarker Studies
RNAi Screen In Breast Cancer Lines
Whole Human Genome RNAi Screen



(57) **Abstract**: The invention provides methods for predicting the efficacy of eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (e.g., eribulin mesylate), in the treatment of a subject suffering from breast cancer by determining the level of particular biomarkers in a sample derived from the subject.

WO 2012/129100 A1

METHODS AND COMPOSITIONS FOR PREDICTING RESPONSE TO ERIBULIN

RELATED APPLICATIONS

This application claims priority to U.S. Provisional Application No. 61/454426, filed March 18, 2011, the entire contents of which are hereby incorporated by reference herein.

BACKGROUND OF THE INVENTION

Cancer is a term used to describe a wide variety of diseases that are each characterized by the uncontrolled growth of a particular type of cell. It begins in a tissue containing such a cell and, if the cancer has not spread to any additional tissues at the time of diagnosis, may be treated by, for example, surgery, radiation, or another type of localized therapy. However, when there is evidence that cancer has metastasized from its tissue of origin, different approaches to treatment are typically used. Indeed, because it is not possible to determine the extent of metastasis, systemic approaches to therapy are usually undertaken when any evidence of spread is detected. These approaches involve the administration of chemotherapeutic drugs that interfere with the growth of rapidly dividing cells, such as cancer cells.

Halichondrin B is a structurally complex, macrocyclic compound that was originally isolated from the marine sponge *Halichondria okadai*, and subsequently was found in *Axinella sp.*, *Phakellia carteri*, and *Lissodendoryx sp.* A total synthesis of halichondrin B was published in 1992 (Aicher *et al.*, *J. Am. Chem. Soc.* 114:3162-3164, 1992). Halichondrin B has been shown to inhibit tubulin polymerization, microtubule assembly, beta-tubulin crosslinking, GTP and vinblastine binding to tubulin, and tubulin-dependent GTP hydrolysis *in vitro*. This molecule has also been shown to have anti-cancer properties *in vitro* and *in vivo*. Halichondrin B analogs having anti-cancer activities are described in U.S. Pat. No. 6,214,865 B1.

In particular, eribulin mesylate, a Halichondrin B analog, has been developed as an anticancer drug. Recently, eribulin mesylate was approved for the treatment of patients with metastatic breast cancer who have previously received at least two chemotherapeutic regimens for the treatment of metastatic disease, wherein prior therapy may have included an anthracycline and/or a taxane in either the adjuvant or metastatic setting. The ability to predict in advance of treatment whether a cancer patient is likely to be responsive to an anti-cancer agent can guide selection of appropriate treatment, and is beneficial to patients. Accordingly, there is a need for methods for and compositions

useful in, assessing or predicting responsiveness to eribulin in patients having cancer and, in particular, breast cancer.

SUMMARY OF THE INVENTION

The invention is based, at least in part, on the observation that a low level of expression, *e.g.*, the absence of expression, of the biomarkers identified herein is indicative of responsiveness to eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate). Specifically, the absence of expression or a low level of expression of these biomarkers, including, for example, one or more of those biomarkers set forth in Table 1, in a subject is indicative that the subject will be responsive to treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate).

Accordingly, in one aspect, the present invention provides a method for determining whether eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), may be used to treat a subject having breast cancer, by assaying a sample derived from the subject to determine the level of expression in the sample of a biomarker selected from the group of biomarkers listed in Table 1, wherein a low level of expression of the biomarker is indicative that eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), will be effective in treating the subject having breast cancer. In another aspect, the present invention provides a method for determining whether eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), may be used to treat a subject having breast cancer, by determining the level of expression of a biomarker selected from the group of biomarkers listed in Table 1 in a sample derived from the subject, wherein a low level of expression of the biomarker is indicative that eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), will be effective in treating the subject having breast cancer. In a further aspect, the present invention provides a method for predicting whether eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), may be used to treat a subject having breast cancer, by determining the level of expression of a biomarker selected from the group of biomarkers listed in Table 1 in a sample derived from the subject, and predicting that eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), will be effective in treating a subject having breast cancer when there is a low level of expression of the biomarker in the sample. In one embodiment of the foregoing aspects of the invention, the methods may further include obtaining a sample from a subject.

In yet another aspect, the present invention provides a method for determining the sensitivity of a breast tumor, for example, derived from a subject having breast cancer, to treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), by determining the level of expression of a biomarker selected from the group of biomarkers listed in Table 1 in the tumor, wherein a low level of expression of the biomarker in the tumor indicates that the tumor is sensitive to treatment with eribulin, or an analog thereof. In yet another aspect, the present invention provides a method for determining the sensitivity of a breast tumor, for example, derived from a subject having breast cancer, to treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), by determining the level of expression of a biomarker selected from the group of biomarkers listed in Table 1 in the tumor, and identifying the tumor as being sensitive to treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), when the biomarker is expressed in the tumor at a low level.

In further aspects of the invention, methods are provided for treating a subject having breast cancer. The methods include identifying a subject having breast cancer in which a biomarker selected from the group of biomarkers listed in Table 1 is expressed at a low level, and administering a therapeutically effective amount of eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), to the subject. In yet a further aspect, the present invention provides methods of treating a subject having breast cancer, by assaying a sample derived from the subject to determine the level of expression in the sample of a biomarker selected from the group of biomarkers listed in Table 1, and administering a therapeutically effective amount of eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), to the subject when a low level of expression of the biomarker is detected in the sample. In one embodiment of the foregoing aspects of the invention, the methods may further include obtaining a sample from a subject.

In further aspects of the invention, methods are provided for treating a subject having breast cancer that is sensitive to eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof. The methods include identifying a subject having breast cancer that is sensitive to eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof, (e.g., a breast cancer in which a biomarker selected from the group of biomarkers listed in Table 1 is expressed at a low level), and administering a therapeutically effective amount of eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), to the subject. In yet a further aspect, the present invention provides methods of treating a subject having breast cancer that is sensitive to eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof, by

assaying a sample derived from the subject to determine the level of expression in the sample of a biomarker selected from the group of biomarkers listed in Table 1, and administering a therapeutically effective amount of eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), to the subject when a low level of expression of the biomarker is detected in the sample. In one embodiment of the foregoing aspects of the invention, the methods may further include obtaining a sample from a subject.

In various embodiments, the subject has not been previously treated with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate). Alternatively, the subject has been previously treated with eribulin, an analog thereof, or a pharmaceutially acceptable salt thereof. In certain embodiments, the breast cancer is an Estrogen Receptor (ER) negative breast cancer and/or a Progesterone Receptor (PR) negative breast cancer and/or a Her-2 negative breast cancer.

In various embodiments, the level of expression of at least 2, at least 3, at least 4 or at least 5 biomarkers selected from the group of biomarkers listed in Table 1 is determined.

In particular embodiments, a predictive gene signature comprising a sub-combination of 2 or more biomarkers selected from the group of biomarkers listed in Table 1 is used. In various embodiments, the level of expression of at least 2, at least 3, at least 4 or at least 5 biomarkers selected from the group of biomarkers listed in Table 1 is determined. For example, the predictive gene signature may include at least 2 biomarkers, *e.g.,* DYSF and EDIL3; GNAT1 and ERGIC3; KRT24 and PAPLN; MANSC1 and PDGFB; PCDH1 and PDGFB; or PHOSPHO2 and PSENEN. In another embodiment, the predictive gene signature may include at least 3 biomarkers, *e.g.*, COL7A1, YTHDF1 and ZIC5; CKLF, IL10 and TUBB6; CDC20, CFL1 and TMEM79; HYAL2, NCBP1 and SNX11; or CEP152, NCBP1 and SATB1. In another embodiment, the predictive gene signature may include at least 4 biomarkers, *e.g.*, APBB2, CCL26, PSENEN and SATB1; ANG, JAM3, KLHL17 and PAPLN; ITFG3, MAD2L1BP, NMU and PDGFB; SPTA1, TYROBP, SNX11 and PSENEN; GRAMD4, GNAT1, TMIGD2 and YTHDF1; or GRAMD4, HYAL2, PHOSPHO2 and TUBB6. In another embodiment, the predictive gene signature may include at least 5 biomarkers, *e.g.*, CCL26, CDC20, ERGIC3, EDIL3 and PCDH1; DYSF, NMU, PHOSPHO2, PSENEN and SNX11; APBB2, CKLF, CYP4F3, TUBB6 and YTHDF1; or CEP152, MAD2L1BP, SPTA1, TMEM79 and ZIC5.

In particular embodiments, the predictive gene signature may include 2 or more of biomarkers ABI3, ANG, APBB2, CCL26, CDC20, CEP152, CFL1, CKLF, COL7A1,

CYP4F3, DYSF, GNAT1, GRAMD4, HYAL2, IL10, ITFG3, JAM3, KLHL17, KRT24, MAD2L1BP, MANSC1, MOBKL1B, NCBP1, NMU, PCDH1, PHOSPHO2, SPTA1, TMIGD2, TYROBP, ZIC5, ERGIC3, PDGFB, PSENEN, SATB1, SNX11, TMEM79 and YTHDF1, *e.g.*, ABI3 and ANG; APBB2 and CCL26; GNAT1 and GRAMD4; IL10 and ITFG3; MACSC1 and MOBKL1B; NMU and PCDH1; or TYROBP and ZIC5. In other embodiments, the predictive gene signature includes at least 3 of the previously recited biomarkers, *e.g.*, ABI3, ANG and APBB2; CCL26, CKLF and COL7A1; DYSF, GNAT1 and HYAL2; JAM3, KLHL17 and KRT24; NCBP1, NMU and PCDH1; SPTA1, TMIGD2 and TYROBP; or ZIC5, MAD2L1BP and CDC20. In other embodiments, the predictive gene signature includes at least 4 of the previously recited biomarkers, *e.g.*, ABI3, ANG, APBB2 and CCL26; CEP152, CFL1, CKLF and COL7A1; KRT24, MANSC1, MOBKL1B and SPTA1; TYROBP, TMIGD2, PHOSPHO2 and NMU; ABI3, GNATI, KLHL17 and SPTA1; or CEP152, HYAL2, PCDH1 and TMIGD2. In yet further embodiments, the predictive gene signature includes at least 5 of the previously recited biomarkers, *e.g.*, CKLF, COL7A1, GRAMD4, JAM3 and PCDH1; APBB2, CEP152, DYSF, IL10 and TYROBP; CYP4F3, HYAL2, ITFG3, KLHL17 and KRT24; NCBP1, SPTA1, TMIGD2, IL10 and JAM3; or CCL26, PHOSPHO2, SPTA1, TMIGD2 and ZIC5.

In other embodiments, the predictive gene signature may include 2 or more of biomarkers ERGIC3, PDGFB, PSENEN, SATB1, SNX11, TMEM79 or YTHDF1, or any sub-combination thereof, *e.g.*, ERGIC3 and PDGFB; ERGIC3 and PSENEN; ERGIC3 and SATB1; ERGIC3 and SNX11; ERGIC3 and TMEM79; ERGIC3 and YTHDF1; PDGFB and PSENEN; PDGFB and SATB1; PDGFB and SNX11; PDGFB and TMEM79; PDGFB and YTHDF1; PSENEN and SATB1; PSENEN and SNX11; PSENEN and TMEM79; PSENEN and YTHDF1; SATB1 and SNX11; SATB1 and TMEM79; SATB1 and YTHDF1; SNX11 and TMEM79; SNX11 and YTHDF1; or TMEM79 and YTHDF1. In other embodiments, the predictive gene signature includes at least 3 biomarkers, for example, ERGIC3, PDGFB and PSENEN; SATB1, SNX11 and TMEM79; SNX11, TMEM79 and YTHDF1; or ERGIC3, PDGFB and SATB1. In further embodiments, the predictive gene signature includes at least 4 biomarkers, for example, ERGIC3, PDGFB, PSENEN and SATB1; SNX11, TMEM79, YTHDF1 and ERGIC3; or ERGIC3, PDGFB, PSENEN and YTHDF1. In further embodiments, the predictive gene signature includes at least 5 biomarkers, for example, ERGIC3, PDGFB, PSENEN, SATB1 and SNX11; ERGIC3, PDGFB, PSENEN, SATB1 and TMEM79; or PSENEN, SATB1, SNX11, TMEM79 and YTHDF1. In yet further embodiments, the predictive gene signature includes at least 6 biomarkers, for example, ERGIC3, PDGFB, PSENEN, SATB1, SNX11 and TMEM79; PDGFB, PSENEN, SATB1, SNX11,

TMEM79 and YTHDF1; or ERGIC3, PSENEN, SATB1, SNX11, TMEM79 and YTHDF1. In yet another embodiment, the predictive gene signature includes 7 biomarkers, for example, ERGIC3, PDGFB, PSENEN, SATB1, SNX11, TMEM79 and YTHDF1.

In various embodiments, the biomarker is not one or more of SPTA1, PAPLN, PCDH1, TMIGD2 and/or KRT24. In a particular embodiment, the biomarker is not SPTA1, PAPLN, PCDH1, TMIGD2 and KRT24. In one embodiment, the biomarker is not SPTA1. In another embodiment, the biomarker is not PAPLN. In another embodiment, the biomarker is not PCDH1. In another embodiment, the biomarker is not TMIGD2. In yet another embodiment, the biomarker is not KRT24.

In particular embodiments, the predictive gene signature may include 2 or more of biomarkers ABI3, ANG, APBB2, CCL26, CDC20, CEP152, CFL1, CKLF, COL7A1, CYP4F3, DYSF, GNAT1, GRAMD4, HYAL2, IL10, ITFG3, JAM3, KLHL17, MAD2L1BP, MANSC1, MOBKL1B, NCBP1, NMU, PHOSPHO2, TYROBP, ZIC5, ERGIC3, PDGFB, PSENEN, SATB1, SNX11, TMEM79, YTHDF1, EDIL3 and TUBB6, *e.g.*, ABI3 and ANG; GRAMD4 and HYAL2; NMU and PHOSPHO2; ZIC5 and PSENEN; or SNX11 and MOBKL1B. In other embodiments, the predictive gene signature includes at least 3 of the previously recited biomarkers, *e.g.*, APBB2, CDC20 and CKLF; COL7A1, DYSF and GNAT1; NCBP1, SATB1 and EDIL3; PSENEN, DYSF and GNAT1; MANSC1, ZIC5 and CFL1; or CKLF, GRAMD4 and NMU. In other embodiments, the predictive gene signature includes at least 4 of the previously recited biomarkers, *e.g.*, ANG, CCL26, CEP152 and JAM3; APBB2, CYP4F3, ITFG3 and TYROBP; CYP4F3, MANSC1, PDGFB and YTHDF1; TUBB6, DYSF, PHOSPHO2 and CDC20; or CKLF, KLHL17, HYAL2 and ZIC5. In yet further embodiments, the predictive gene signature includes at least 5 of the previously recited biomarkers, *e.g.*, IL10, CEP152, COL7A1, TYROBP and ERGIC3; TMEM79, SNX11, PSENEN, GNAT1 and GRAMD4; JAM3, SNX11, KLHL17, MOBKL1B and ERGIC3; or NMU, PHOSPHO2, PDGFB, CFL1 and ANG.

In various methods and or kits of the invention, the biomarker is not ABI3, is not ANG, is not APBB2, is not CCL26, is not CDC20, is not CEP152, is not CFL1, is not CKLF, is not COL7A1, is not CYP4F3, is not DYSF, is not GNAT1, is not GRAMD4, is not HYAL2, is not IL10, is not ITFG3, is not JAM3, is not KLHL17, is not KRT24, is not MAD2L1BP, is not MANSC1, is not MOBKL1B, is not NCBP1, is not NMU, is not PCDH1, is not PHOSPHO2, is not SPTA1, is not TMIGD2, is not TYROBP, is not ZIC5, is not ERGIC3, is not PDGFB, is not PSENEN, is not SATB1, is not SNX11, is not TMEM79, is not EDIL3, is not PAPLN, is not TUBB6 and/or is not YTHDF1.

In certain embodiments, the biomarker is not expressed at a detectable level. In another embodiment, the biomarker is expressed at a low level as compared to a control. Expression can be determined directly or indirectly by any suitable method. In certain embodiments, the level of expression of the biomarker is determined at the nucleic acid level using any suitable method. For example, the level of expression of the biomarker can be determined by detecting cDNA, mRNA or DNA. In particular embodiments, the level of expression of the biomarker is determined by using a technique selected from the group consisting of polymerase chain reaction (PCR) amplification reaction, reverse-transcriptase PCR analysis, quantitative reverse-transcriptase PCR analysis, Northern blot analysis, RNAase protection assay, digital RNA detection/ quantitation (*e.g.*, nanoString) and combinations or sub-combinations thereof.

In certain embodiments, the level of expression of the biomarker can be determined by detecting miRNA. Specifically, mRNA expression can be assessed indirectly by assessing levels of miRNA, wherein an elevated level of an miRNA which controls the expression of an mRNA is indicative of a low level of expression of the mRNA encoding the biomarker.

In other embodiments, the level of expression of the biomarker is determined at the protein level using any suitable method. For example, the presence or level of the protein can be detected using an antibody or antigen binding fragment thereof, which specifically binds to the protein. In particular embodiments, the antibody or antigen binding fragment thereof is selected from the group consisting of a murine antibody, a human antibody, a humanized antibody, a bispecific antibody, a chimeric antibody, a Fab, Fab', $F(ab')_2$, ScFv, SMIP, affibody, avimer, versabody, nanobody, and a domain antibody, or an antigen binding fragment of any of the foregoing. In particular embodiments, the antibody or antigen binding portion thereof is labeled, for example, with a label selected from the group consisting of a radio-label, a biotin-label, a chromophore-label, a fluorophore-label, and an enzyme-label. In certain embodiments, the level of expression of the biomarker is determined by using a technique selected from the group consisting of an immunoassay, a western blot analysis, a radioimmunoassay, immunofluorimetry, immunoprecipitation, equilibrium dialysis, immunodiffusion, electrochemiluminescence immunoassay (ECLIA), ELISA assay, immunopolymerase chain reaction and combinations or sub-combinations thereof. In particular embodiments, the immunoassay is a solution-based immunoassay selected from the group consisting of electrochemiluminescence, chemiluminescence, fluorogenic chemiluminescence, fluorescence polarization, and time-resolved fluorescence. In other embodiments, the immunoassay is a sandwich immunoassay selected from the group consisting of electrochemiluminescence, chemiluminescence,

and fluorogenic chemiluminescence. Other assays which rely on agents capable of detecting the protein, such as those relying upon a suitable binding partner or enzymatic activity, can also be used (*e.g.*, use of a ligand to detect a receptor molecule).

Samples can be obtained from a subject by any suitable method, and may optionally have undergone further processing step(s) (*e.g.*, freezing, fractionation, fixation, guanidine treatment, etc). Any suitable sample derived from a subject can be used, such as any tissue (*e.g.*, biopsy), cell, or fluid, as well as any component thereof, such as a fraction or extract. In various embodiments, the sample is a fluid obtained from the subject, or a component of such a fluid. For example, the fluid can be blood, plasma, serum, sputum, lymph, cystic fluid, nipple aspirate, urine, or fluid collected from a biopsy (*e.g.*, lump biopsy). In other embodiments, the sample is a tissue or component thereof obtained from the subject. For example, the tissue can be tissue obtained from a biopsy (*e.g.*, lump biopsy), breast tissue, connective tissue, and/or lymphatic tissue. In a particular embodiment, the tissue is breast tissue, or a component thereof (*e.g.*, cells collected from the breast tissue). In a particular embodiment, the component of the breast tissue are breast tissue cells. In another embodiment, the component of the breast tissue are circulating breast tumor cells.

In one embodiment, the subject is a human.

In another aspect, the present invention provides a kit for predicting whether eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), may be used to treat a subject having breast cancer, including reagents for determining the level of expression of a biomarker selected from the group of biomarkers listed in Table 1; and instructions for use of the kit to predict whether eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), may be used to treat a subject having breast cancer. For example, the reagent for determining the level of expression of the biomarker can be a probe for identifying a null mutation in the biomarker. The reagent for determining the level of expression of the biomarker can be a probe for amplifying and/or detecting the biomarker. In yet another embodiment, the reagent for determining the level of expression of the biomarker can be an antibody, for example, an antibody specific for the product of the expression of the wild type or null mutant version of the biomarker.

In a particular embodiment, the kit further includes reagents for obtaining a biological sample from a subject. In another embodiment, the kit includes a control sample.

In another aspect, the present invention provides methods for determining whether eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*,

eribulin mesylate), may be used to treat a subject having breast cancer by determining and/or identifying whether the subject carries at least one gene, selected from the group of biomarkers set forth in Table 1, which contains a null mutation. In another aspect, the present invention provides methods for predicting whether eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), may be used to treat a subject having breast cancer by assaying a sample derived from the subject to determine whether the subject carries at least one gene, selected from the group of biomarkers set forth in Table 1, which contains a null mutation. In a further aspect, the present invention provides methods for determining the sensitivity of a breast tumor to treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof by determining and/or identifying whether said tumor is characterized by at least one gene, selected from the group of biomarkers set forth in Table 1, which contains a null mutation. In yet another aspect, the present invention is directed to methods for treating a subject having breast cancer with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof by identifying whether at least one gene, selected from the group of biomarkers set forth in Table 1, which contains a null mutation and administering a therapeutically effective amount of eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate) to the subject.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 depicts the high throughput siRNA screening methods in Breast Cancer Cell Lines performed as described in Example 1.

Figure 2 depicts the identification and selection of certain genes for further consideration as biomarkers of efficacy of eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof.

Figure 3 depicts the confirmation assays performed as described in Example 1.

Figure 4 depicts the results of the QuantiTect SYBR Assays to determine the relative quantities of cDNAs after treatment with siRNA as set forth in Example 1.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides methods for determining whether eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), may be used to treat a subject having breast cancer, methods for predicting whether eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin

mesylate), may be used to treat a subject having breast cancer, methods for determining the sensitivity of a breast tumor to treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), and methods of treating a subject having breast cancer. Generally, the methods involve determining the level of expression of at least one biomarker as set forth in Table 1 in a sample derived from the subject, wherein a low level of expression of the biomarker is an indication that eribulin, or an analog thereof may be used to treat breast cancer and/or that the breast tumor is sensitive to treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate).

The invention is based, at least in part, on the observation that a low level of expression, *e.g.*, the absence of expression, of the biomarkers identified herein is indicative of responsiveness to eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate). As shown herein, siRNA techniques were employed to "knock down" expression of certain genes and assess the sensitivity of the resulting knock down cells to eribulin mesylate. Based on the findings from these studies, low levels of expression of each of the genes set forth in Table 1 was identified as being associated with the sensitivity of breast cancer cells to treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate).

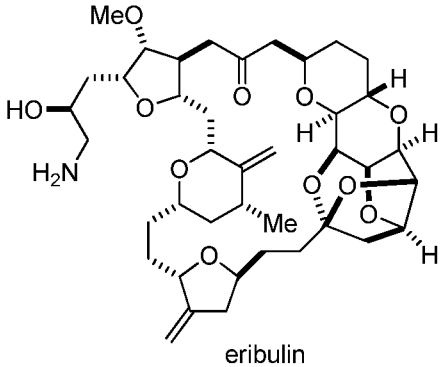
Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly understood by those of ordinary skill in the art. The meaning and scope of the terms should be clear, however, in the event of any latent ambiguity, definitions provided herein take precedent over any dictionary or extrinsic definition. Further, unless otherwise required by context, singular terms, for example, those characterized by "a" or "an", shall include pluralities, *e.g.*, one or more biomarkers. In this application, the use of "or" means "and/or", unless stated otherwise. Furthermore, the use of the term "including," as well as other forms of the term, such as "includes" and "included", is not limiting. Also, terms such as "element" or "component" encompass both elements and components comprising one unit and elements and components that comprise more than one unit unless specifically stated otherwise.

The phrase "determining whether eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), may be used to treat a subject having breast cancer" refers to assessing the likelihood that treatment of a subject with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate) will be effective (*e.g.*, provide a therapeutic benefit to the subject) or will not be effective in the subject. Assessment of the likelihood that treatment will or will not be effective typically can be performed before treatment with

eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), has begun or before treatment is resumed. Alternatively or in combination, assessment of the likelihood of efficacious treatment can be performed during treatment, for example, to determine whether treatment should be continued or discontinued. For example, such an assessment can be performed (a) by determining the level of expression of a biomarker, for example, a biomarker selected from the group of biomarkers listed in Table 1, in a sample derived from said subject, wherein a low level of expression of the biomarker indicates that eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof, may be used to treat said subject having breast cancer, or (b) by assaying a sample derived from said subject to determine the level of expression in said sample of a biomarker, for example, a biomarker selected from the group of biomarkers listed in Table 1, wherein a low level of expression of the biomarker indicates that eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof, may be used to treat said subject having breast cancer.

The phrase "determining the sensitivity of a breast tumor to treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof", as used herein, is intended to refer to assessing the susceptibility of a breast tumor, *e.g.*, breast cancer cells, to treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate). Sensitivity of a tumor can include the ability of eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), to kill tumor cells, to inhibit the spread and/or metastasis of tumor cells, and/or to inhibit the growth of tumor cells completely or partially (*e.g.*, slow down the growth of tumor cells by at least 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%). The assessment can be performed (i) before treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), is begun; (ii) before treatment is resumed in the subject; and/or during treatment, for example, to determine whether treatment should be continued or discontinued. For example, such a determination can be performed (a) by determining the level of expression of a biomarker, *e.g.*, a biomarker selected from the group of biomarkers listed in Table 1, in said tumor, wherein a low level of expression of the biomarker in said tumor indicates that said tumor is sensitive to treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof, or (b) by determining the level of expression of a biomarker *e.g.*, a biomarker selected from the group of biomarkers listed in Table 1, in said tumor, and identifying said tumor as being sensitive to treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof, when said biomarker is expressed in said tumor at a low level.

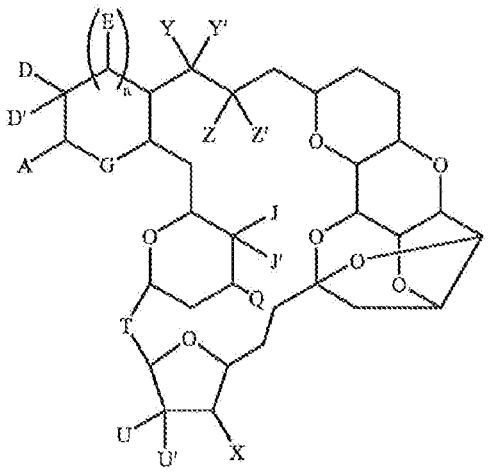
The term "eribulin" as used herein refers to the art-recognized fully synthetic macrocyclic ketone analog of halichondrin B. As set forth in U.S. Patent No. 6,214,865, the entire contents of which are incorporated herein by reference, eribulin has the following structure



eribulin

and can be generated using techniques as described therein or as described in Kim DS et al. (November 2009) *J. Am. Chem. Soc.* 131 (43): 15636–41, the entire contents of which are incorporated herein by reference. Eribulin is also known as ER-086526 and is identified by CAS number 253128-41-5. Eribulin mesylate is also known as E7389.

As used herein, the term "eribulin analog" includes compounds in which one or more atoms or functional groups of eribulin have been replaced with different atoms or functional groups. For example, eribulin analogs include compounds having the following formula (I), which also encompasses eribulin:



(I)

In formula (I), A is a C_{1-6} saturated or C_{2-6} unsaturated hydrocarbon skeleton, the skeleton being unsubstituted or having between 1 and 13 substituents, preferably between 1 and 10 substituents, *e.g.*, at least one substituent selected from cyano, halo,

azido, Q_1, and oxo. Each Q_1 is independently selected from OR_1, SR_1, SO_2R_1, OSO_2R_1, NR_2R_1, $NR_2(CO)R_1$, $NR_2(CO)(CO)R_1$, $NR_4(CO)NR_2R_1$, $NR_2(CO)OR_1$, $(CO)OR_1$, $O(CO)R_1$, $(CO)NR_2R_1$, and $O(CO)NR_2R_1$. The number of substituents can be, for example, from 1 to 6, from 1 to 8, from 2 to 5, or from 1 to 4. Throughout the disclosure, numerical ranges are understood to be inclusive.

Each of R_1, R_2, R_4, R_5, and R_6 is independently selected from H, C_{1-6} alkyl, C_{1-6} haloalkyl, C_{1-6} hydroxyalkyl, C_{1-6} aminoalkyl, C_{6-10} aryl, C_{6-10} haloaryl (*e.g.*, p-fluorophenyl or p-chlorophenyl), C_{6-10} hydroxyaryl, C_{1-4} alkoxy-C_6 aryl (*e.g.*, C_{1-3} alkoxy-C_6 aryl, p-methoxyphenyl, 3,4,5-trimethoxyphenyl, p-ethoxyphenyl, or 3,5-diethoxyphenyl), C_{6-10} aryl-C_{1-6} alkyl (*e.g.*, benzyl or phenethyl), C_{1-6} alkyl-C_{6-10} aryl, C_{6-10} haloaryl-C_{1-6} alkyl, C_{1-6} alkyl-C_{6-10} haloaryl, (C_{1-3} alkoxy-C_6 aryl)-C_{1-3} alkyl, C_{2-9} heterocyclic radical, C_{2-9} heterocyclic radical-C_{1-6} alkyl, C_{2-9} heteroaryl, and C_{2-9} heteroaryl-C_{1-6} alkyl. There may be more than one R_1, for example, if A is substituted with two different alkoxy (OR_1) groups such as butoxy and 2-aminoethoxy.

Examples of A include 2,3-dihydroxypropyl, 2-hydroxyethyl, 3-hydroxy-4-perfluorobutyl, 2,4,5-trihydroxypentyl, 3-amino-2-hydroxypropyl, 1,2-dihydroxyethyl, 2,3-dihyroxy-4-perfluorobutyl, 3-cyano-2-hydroxypropyl, 2-amino-1-hydroxy ethyl, 3-azido-2-hydroxypropyl, 3,3-difluoro-2,4-dihydroxybutyl, 2,4-dihydroxybutyl, 2-hydroxy-2(p-fluorophenyl)-ethyl, $-CH_2(CO)$(substituted or unsubstituted aryl), $-CH_2(CO)$(alkyl or substituted alkyl, such as haloalkyl or hydroxyalkyl) and 3,3-difluoro-2-hydroxypent-4-enyl.

Examples of Q_1 include -NH(CO)(CO)-(heterocyclic radical or heteroaryl), $-OSO_2$ -(aryl or substituted aryl), -O(CO)NH-(aryl or substituted aryl), aminoalkyl, hydroxyalkyl, -NH(CO)(CO)-(aryl or substituted aryl), -NH(CO)(alkyl)(heteroaryl or heterocyclic radical), O(substituted or unsubstituted alkyl)(substituted or unsubstituted aryl), and -NH(CO)(alkyl)(aryl or substituted aryl).

Each of D and D' is independently selected from R_3 and OR_3, wherein R_3 is H, C_{1-3} alkyl, or C_{1-3} haloalkyl. Examples of D and D' are methoxy, methyl, ethoxy, and ethyl. In some embodiments, one of D and D' is H.

The value for n is 1 or preferably 0, thereby forming either a six-membered or five-membered ring. This ring can be unsubstituted or substituted, *e.g.*, where E is R_5 or OR_5, and can be a heterocyclic radical or a cycloalkyl, *e.g.* where G is S, CH_2, NR_6, or preferably O.

Each of J and J' is independently H, C_{1-6} alkoxy, or C_{1-6} alkyl; or J and J' taken together are $=CH_2$ or -O-(straight or branched C_{1-5} alkylene)-O-, such as exocyclic methylidene, isopropylidene, methylene, or ethylene.

Q is C_{1-3} alkyl, and is preferably methyl.

T is ethylene or ethenylene, optionally substituted with $(CO)OR_7$, where R_7 is H or C_{1-6} alkyl.

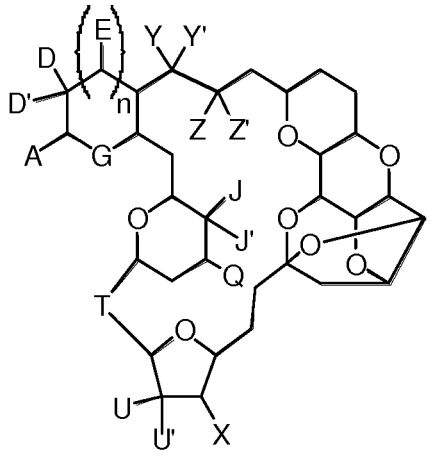
Each of U and U' is independently H, C_{1-6} alkoxy, or C_{1-6} alkyl; or U and U' taken together are $=CH_2$ or -O-(straight or branched C_{1-5} alkylene)-O-.

X is H or C_{1-6} alkoxy.

Each of Y and Y' is independently H or C_{1-6} alkoxy; or Y and Y' taken together are =O, $=CH_2$, or -O-(straight or branched C_{1-5} alkylene)-O-.

Each of Z and Z' is independently H or C_{1-6} alkoxy; or Z and Z' taken together are =O, $=CH_2$, or -O-(straight or branched C_{1-5} alkylene)-O-.

In certain embodiments, the eribulin analogs include compounds having the following formula (II):



(II)

In formula (II), the substitutions are defined as follows:

A is a C_{1-6} saturated or C_{2-6} unsaturated hydrocarbon skeleton, said skeleton being unsubstituted or having between 1 and 10 substituents, inclusive, independently selected from cyano, halo, azido, oxo, and Q_1.

Each Q_1 is independently selected from OR_1, SR_1, SO_2R_1, OSO_2R_1, NR_2R_1, $NR_2(CO)R_1$, $NR_2(CO)(CO)R_1$, $NR_4(CO)NR_2R_1$, $NR_2(CO)OR_1$, $(CO)OR_1$, $O(CO)R_1$, $(CO)NR_2R_1$, and $O(CO)NR_2R_1$.

Each of R_1, R_2 and R_4 is independently selected from H, C_{1-6} alkyl, C_{1-6} haloalkyl, C_{1-6} hydroxyalkyl, C_{1-6} aminoalkyl, C_{6-10} aryl, C_{6-10} haloaryl, C_{6-10} hydroxyaryl, C_{1-3} alkoxy-C_6 aryl, C_{6-10} aryl-C_{1-6} alkyl, C_{1-6} alkyl-C_{6-10} aryl, C_{6-10} haloaryl-C_{1-6} alkyl, C_{1-6} alkyl-C_{6-10} haloaryl, (C_{1-3} alkoxy-C_6 aryl)-C_{1-3} alkyl, C_{2-9}

heterocyclic radical, C_{2-9} heterocyclic radical-C_{1-6} alkyl, C_{2-9} heteroaryl, and C_{2-9} heteroaryl-C_{1-6} alkyl.

Each of D and D' is independently selected from R_3 and OR_3, wherein R_3 is H, C_{1-3} alkyl, or C_{1-3} haloalkyl.

n is 0 or 1.

E is R_5 or OR_5, wherein R_5 is independently selected from H, C_{1-6} alkyl, C_{1-6} haloalkyl, C_{1-6} hydroxyalkyl and C_{1-6} aminoalkyl.

G is O.

Each of J and J' is independently H, C_{1-6} alkoxy, or C_{1-6} alkyl; or J and J' taken together are $=CH_2$.

Q is C_{1-3} alkyl.

T is ethylene or ethenylene.

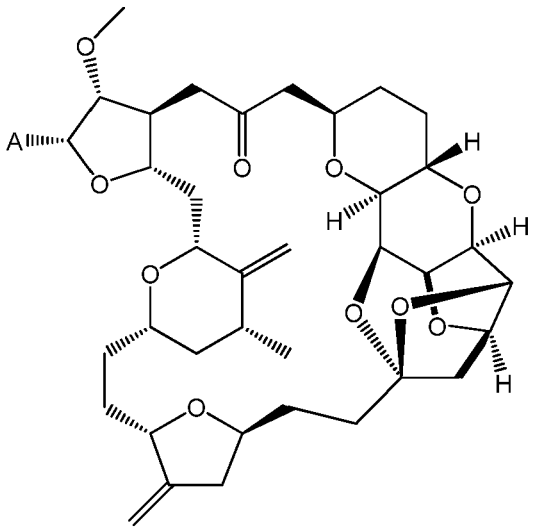
Each of U and U' is independently H, C_{1-6} alkoxy, or C_{1-6} alkyl; or U and U' taken together are $=CH_2$.

X is H or C_{1-6} alkoxy.

Each of Y and Y' is independently H or C_{1-6} alkoxy; or Y and Y' taken together are =O.

Each of Z and Z' is independently H or C_{1-6} alkoxy; or Z and Z' taken together are =O.

In some embodiments, the eribulin analogs include compounds having the following formula (III):



(III)

wherein A is a C_{1-6} saturated or C_{2-6} unsaturated hydrocarbon skeleton, the skeleton being unsubstituted or having between 1 and 13 substituents, e.g., between 1 and 10 substituents selected from cyano, halo, azido, Q_1, and oxo;

each Q_1 is independently selected from OR_1, SR_1, SO_2R_1, OSO_2R_1, NR_2R_1, $NR_2(CO)R_1$, $NR_2(CO)(CO)R_1$, $NR_4(CO)NR_2R_1$, $NR_2(CO)OR_1$, $(CO)OR_1$, $O(CO)R_1$, $(CO)NR_2R_1$, and $O(CO)NR_2R_1$; and

each of R_1, R_2, and R_4 is independently selected from H, C_{1-6} alkyl, C_{1-6} haloalkyl, C_{1-6} hydroxyalkyl, C_{1-6} aminoalkyl, C_{6-10} aryl, C_{6-10} haloaryl, C_{6-10} hydroxyaryl, C_{1-4} alkoxy-C_6 aryl, C_{6-10} aryl-C_{1-6} alkyl, C_{1-6} alkyl-C_{6-10} aryl, C_{6-10} haloaryl-C_{1-6} alkyl, C_{1-6} alkyl-C_{6-10} haloaryl, (C_{1-3} alkoxy-C_6 aryl)-C_{1-3} alkyl, C_{2-9} heterocyclic radical, C_{2-9} heterocyclic radical-C_{1-6} alkyl, C_{2-9} heteroaryl, and C_{2-9} heteroaryl-C_{1-6} alkyl.

Hydrocarbon skeletons contain carbon and hydrogen atoms and may be linear, branched, or cyclic. Unsaturated hydrocarbons include one, two, three or more C--C double bonds (sp^2) or C--C triple bonds (sp). Examples of unsaturated hydrocarbon radicals include ethynyl, 2-propynyl, 1-propenyl, 2-butenyl, 1,3-butadienyl, 2-pentenyl, vinyl (ethenyl), allyl, and isopropenyl. Examples of bivalent unsaturated hydrocarbon radicals include alkenylenes and alkylidenes such as methylidyne, ethylidene, ethylidyne, vinylidene, and isopropylidene. In general, compounds of the invention have hydrocarbon skeletons ("A" in formula (I)) that are substituted, e.g., with hydroxy, amino, cyano, azido, heteroaryl, aryl, and other moieties described herein. Hydrocarbon skeletons may have two geminal hydrogen atoms replaced with oxo, a bivalent carbonyl oxygen atom (=O), or a ring-forming substituent, such as --O-(straight or branched alkylene or alkylidene)-O-- to form an acetal or ketal.

C_{1-6} alkyl includes linear, branched, and cyclic hydrocarbons, such as methyl, ethyl, propyl, isopropyl, butyl, isobutyl, sec-butyl, tert-butyl, pentyl, isopentyl, sec-pentyl, neo-pentyl, tert-pentyl, cyclopentyl, hexyl, isohexyl, sec-hexyl, cyclohexyl, 2-methylpentyl, tert-hexyl, 2,3-dimethylbutyl, 3,3-dimethylbutyl, 1,3-dimethylbutyl, and 2,3-dimethyl but-2-yl. Alkoxy (--OR), alkylthio (--SR), and other alkyl-derived moieties (substituted, unsaturated, or bivalent) are analogous to alkyl groups (R). Alkyl groups, and alkyl-derived groups such as the representative alkoxy, haloalkyl, hydroxyalkyl, alkenyl, alkylidene, and alkylene groups, can be C_{2-6}, C_{3-6}, C_{1-3} , or C_{2-4}.

Alkyls substituted with halo, hydroxy, amino, cyano, azido, and so on can have 1, 2, 3, 4, 5 or more substituents, which are independently selected (may or may not be the same) and may or may not be on the same carbon atom. For example, haloalkyls are alkyl groups with at least one substituent selected from fluoro, chloro, bromo, and iodo. Haloalkyls may have two or more halo substituents which may or may not be the same halogen and may or may not be on the same carbon atom. Examples include chloromethyl, periodomethyl, 3,3-dichloropropyl, 1,3-difluorobutyl, and 1-bromo-2-chloropropyl.

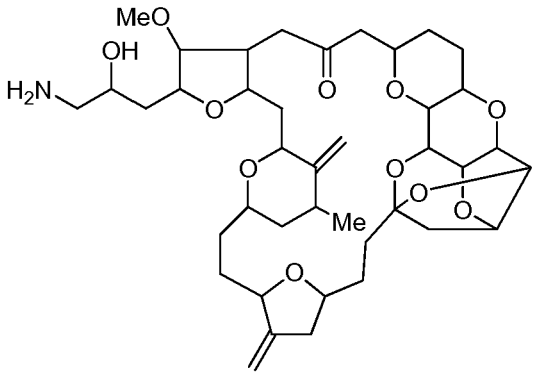
Heterocyclic radicals and heteroaryls include furyl, pyranyl, isobenzofuranyl, chromenyl, xanthenyl, phenoxathienyl, 2H-pyrrolyl, pyrrolyl, imidazolyl (e.g., 1-, 2- or 4-imidazolyl), pyrazolyl, isothiazolyl, isoxazolyl, pyridyl (e.g., 1-, 2-, or 3-pyridyl), pyrazinyl, pyrimidinyl, pyridazinyl, indolizinyl, isoindolyl, 3H-indolyl, indolyl (e.g., 1-, 2-, or 3-indolyl), indazolyl, purinyl, 4H-quinolizinyl, isoquinolyl, quinolyl, phthalazinyl, naphthyridinyl, quinoxalinyl, quinazolinyl, cinnolinyl, pteridinyl, pyrrolinyl, pyrrolidinyl, imidazolidinyl, pyrazolidinyl, pyrazolinyl, piperidyl, piperazinyl, indolinyl, isoindolinyl, and morpholinyl. Heterocyclic radicals and heteroaryls may be linked to the rest of the molecule at any position along the ring. Heterocyclic radicals and heteroaryls can be C_{2-9}, or smaller, such as C_{3-6}, C_{2-5}, or C_{3-7}.

Aryl groups include phenyl, benzyl, naphthyl, tolyl, mesityl, xylyl, and cumenyl.

It is understood that "heterocyclic radical", "aryl", and "heteroaryl" include those having 1, 2, 3, 4, or more substituents independently selected from lower alkyl, lower alkoxy, amino, halo, cyano, nitro, azido, and hydroxyl. Heterocyclic radicals, heteroaryls, and aryls may also be bivalent substituents of hydrocarbon skeleton "A" in formula (I).

The term “eribulin analog” includes eribulin prodrugs. The term “eribulin prodrugs” includes eribulin that has been chemically modified to be inactive or less active until bioactivation (*e.g.*, metabolism *in vivo*) by an enzyme which cleaves the chemically modified portion of the eribulin prodrug, thereby providing the active form of eribulin.

As used herein, the term "eribulin analog" includes all stereoisomers of eribulin and other compounds of formula (I), including diastereoisomers and enantiomers thereof. "Stereoisomers" refers to isomers that differ only in the arrangement of the atoms in space. "Diastereoisomers" refers to stereoisomers that are not mirror images of each other. "Enantiomers" refers to stereoisomers that are non-superimposable mirror images of one another. For example, Formula (IV) encompasses eribulin and such stereoisomers:

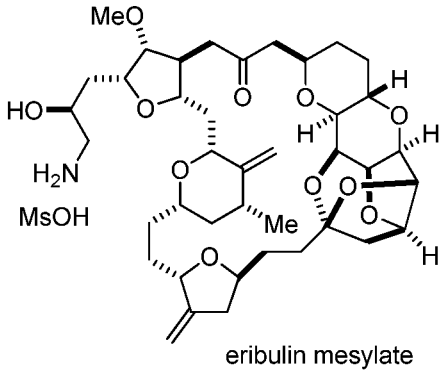


(IV)

The phrase "pharmaceutically acceptable salt," as used herein, is a salt formed from an acid and a basic nitrogen group of Eribulin or an eribulin analog. Examples of such salts include acid addition salts and base addition salts, such as inorganic acid salts or organic acid salts (e.g., hydrochloric acid salt, sulfuric acid salt, citrate, hydrobromic acid salt, hydroiodic acid salt, nitric acid salt, bisulfate, phosphoric acid salt, super phosphoric acid salt, isonicotinic acid salt, acetic acid salt, lactic acid salt, salicylic acid salt, tartaric acid salt, pantothenic acid salt, ascorbic acid salt, succinic acid salt, maleic acid salt, fumaric acid salt, gluconic acid salt, saccharinic acid salt, formic acid salt, benzoic acid salt, glutaminic acid salt, methanesulfonic acid salt, ethanesulfonic acid salt, benzenesulfonic acid salt, p-toluenesulfonic acid salt, pamoic acid salt (pamoate)), as well as salts of aluminum, calcium, lithium, magnesium, calcium, sodium, zinc, and diethanolamine. It will be understood that reference to eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof, includes pharmaceutically acceptable salts of eribulin as well as pharmaceutically acceptable salts of an analog thereof. Examples of such pharmaceutically acceptable salts include, but are not limited to, a pharmaceutically acceptable salt of Formula I, a pharmaceutically acceptable salt of Formula II, a pharmaceutically acceptable salt of Formula III or a pharmaceutically acceptable salt of Formula IV.

In a particular embodiment, eribulin mesylate, a pharmaceutically acceptable salt form of eribulin is utilized in the methods of the present invention. Eribulin mesylate is

sold under the trade name HALAVEN®. The chemical name for eribulin mesylate is 11,15:18,21:24,28-Triepoxy-7,9-ethano-12,15-methano-9H,15H-furo[3,2-i]furo[2',3':5,6]pyrano[4,3-b][1,4]dioxacyclopentacosin-5(4H)-one, 2-[(2S)-3-amino-2-hydroxypropyl]hexacosahydro-3-methoxy-26-methyl-20,27-bis(methylene)-,(2R,3R,3aS,7R,8aS,9S,10aR,11S,12R,13aR,13bS,15S,18S,21S,24S,26R,28R,29aS)-methanesulfonate (salt). Eribulin mesylate has the following structure



eribulin mesylate

Eribulin mesylate is label indicated for the treatment of patients with metastatic breast cancer who have previously received at least two chemotherapeutic regimens for the treatment of metastatic disease, including, for example, therapy with an anthracycline and a taxane in either the adjuvant or metastatic setting.

As used herein, the term "biomarker" is intended to encompass a substance that is used as an indicator of a biologic state and includes for example, genes (or portions thereof), mRNAs (or portions thereof), miRNAs (microRNAs), and proteins (or portions thereof). A "biomarker expression pattern" is intended to refer to a quantitative or qualitative summary of the expression of one or more biomarkers in a subject, such as in comparison to a standard or a control.

Various biomarkers which can be used in the methods described herein are summarized in Table 1. Table 1 provides gene abbreviations, Gene ID numbers and accession numbers for transcripts from which encoding nucleotide gene sequences can be identified. For example, gene ABI3 refers to a *Homo sapiens* ABI family, member 3. The nucleotide sequence for human ABI3 transcript variant 1 can be found at Accession Number NM_016428. Reference to a gene (e.g., ABI3) is intended to encompass naturally occurring or endogenous versions of the gene, including wild type, polymorphic or allelic variants or mutants (e.g., germline mutation, somatic mutation) of the gene, which can be found in a subject and/or tumor from a subject. In some embodiments, the sequence of the biomarker gene is at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at

least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to a sequence identified in Table 1 by Accession Number or Gene_ID number. For example, sequence identity can be determined by comparing sequences using NCBI BLAST (e.g., Megablast with default parameters).

As used herein, the phrase "predictive gene signature" refers to expression levels of two or more biomarkers of the present invention in a subject that are indicative of responsiveness to treatment with eribulin, or an eribulin analog. For example, the low level expression of at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9 or at least 10 biomarkers from Table 1 in a subject may constitute a gene signature that indicates that the subject will respond positively to treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate). Any sub-combination of 2 or more markers from Table 1 may constitute a predictive gene signature of the invention. In another example, the expression of at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9 or at least 10 biomarkers from Table 1 under particular threshold levels, or any sub-combination thereof, constitutes a gene signature that indicates that the subject will respond positively to treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate).

A "low level of expression" of the biomarker, for example, a biomarker selected from the group of biomarkers listed in Table 1, refers to a level of expression of the biomarker in a test sample (*e.g.*, a sample derived from a subject) that correlates with sensitivity to eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate). This can be determined by comparing the level of expression of the biomarker in the test sample with that of a suitable control. A "low level of expression" also includes a lack of detectable expression of the biomarker.

In some embodiments, the level of expression of the biomarker is determined relative to a control sample, such as the level of expression of the biomarker in normal tissue (*e.g.*, a range determined from the levels of expression of the biomarker observed in normal tissue samples). In these embodiments, a low level of expression will fall below or within the lower levels of this range. In some embodiments, the level of expression of the biomarker is determined relative to a control sample, such as the level of expression of the biomarker in samples (*e.g.*, tumor samples, circulating tumor cells) from other subjects . For example, the level of expression of the biomarker in samples from other subjects can be determined to define levels of expression which correlate with sensitivity to treatment with eribulin, an analog thereof, or a pharmaceutically

acceptable salt thereof (*e.g.*, eribulin mesylate), and the level of expression of the biomarker in the sample from the subject of interest is compared to these levels of expression, wherein a comparable or lower level of expression in the sample from the subject is indicative of a "low level of expression" of the biomarker in the sample. In another example, the level of expression of the biomarker in samples (*e.g.*, tumor samples, circulating tumor cells) from other subjects can be determined to define levels of expression which correlate with resistance or non-responsiveness to treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), and the level of expression of the biomarker in the sample from the subject of interest is compared to these levels of expression, wherein a lower level of expression in the sample from the subject is indicative of a "low level of expression" of the biomarker in the sample.

The term "known standard level" or "control level" can refer to an accepted or pre-determined expression level of the biomarker, for example, a biomarker selected from the group of biomarkers listed in Table 1 which is used to compare expression level of the biomarker in a sample derived from a subject. In one embodiment, the control expression level of the biomarker is the average expression level of the biomarker in samples derived from a population of subjects. For example, the control expression level can be the average expression level of the biomarker in breast cancer cells derived from a population of subjects with breast cancer. The population may be subjects who have not responded to treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), or the population may be a group of subjects who express the respective biomarker at high or normal levels. In some embodiments, the control level may constitute a range of expression of the biomarker in normal tissue, as described above. For example, the control level may constitute a range of expression of the biomarker in tumor samples from a variety of subjects having breast cancer, as described above.

As further information becomes available as a result of routine performance of the methods described herein, population-average values for "control" level of expression of the biomarkers of the present invention may be used. In other embodiments, the "control" level of expression of the biomarkers may be determined by determining expression level of the respective biomarker in a subject sample obtained from a subject before the suspected onset of breast cancer in the subject, from archived subject samples, and the like.

Control levels of expression of biomarkers of the invention may be available from publicly available databases. In addition, Universal Reference Total RNA

(Clontech Laboratories) and Universal Human Reference RNA (Stratagene) and the like can be used as controls. For example, qPCR can be used to determine the level of expression of a biomarker, and an increase in the number of cycles needed to detect expression of a biomarker in a sample from a subject, relative to the number of cycles needed for detection using such a control, is indicative of a low level of expression of the biomarker.

As used herein, the term "subject" or "patient" refers to human and non-human animals, *e.g.,* veterinary patients. The term "non-human animal" includes vertebrates, *e.g.*, mammals, such as non-human primates, mice, rabbits, sheep, dog, cat, horse, cow, or other rodent, ovine, canine, feline, equine or bovine species. In one embodiment, the subject is a human.

The term "sample" as used herein refers to cells, tissues or fluids isolated from a subject, as well as cells, tissues or fluids present within a subject. The term "sample" includes any body fluid (*e.g.*, blood, lymph, cystic fluid, nipple aspirates, urine and fluids collected from a biopsy (*e.g.*, lump biopsy)), tissue or a cell or collection of cells from a subject, as well as any component thereof, such as a fraction or extract. In one embodiment, the tissue or cell is removed from the subject. In another embodiment, the tissue or cell is present within the subject. Other samples include tears, plasma, serum, cerebrospinal fluid, feces, sputum and cell extracts. In one embodiment, the sample contains protein (*e.g.*, proteins or peptides) from the subject. In another embodiment, the sample contains RNA (*e.g.*, mRNA molecules) from the subject or DNA (*e.g.*, genomic DNA molecules) from the subject.

As used herein, the term "breast cancer" refers generally to the uncontrolled growth of breast tissue and, more specifically, to a condition characterized by anomalous rapid proliferation of abnormal cells in one or both breasts of a subject. The abnormal cells often are referred to as malignant or "neoplastic cells," which are transformed cells that can form a solid tumor. The term "tumor" refers to an abnormal mass or population of cells (*i.e.*, two or more cells) that result from excessive or abnormal cell division, whether malignant or benign, and pre-cancerous and cancerous cells. Malignant tumors are distinguished from benign growths or tumors in that, in addition to uncontrolled cellular proliferation, they can invade surrounding tissues and can metastasize. In breast cancer, neoplastic cells may be identified in one or both breasts only and not in another tissue or organ, in one or both breasts and one or more adjacent tissues or organs (*e.g.* lymph node), or in a breast and one or more non-adjacent tissues or organs to which the breast cancer cells have metastasized.

Eribulin, an analog thereof, or pharmaceutically acceptable salt thereof, can be used to treat breast cancer, and, accordingly, the methods of the present invention can be used in breast cancer and in subjects having breast cancer.

In one embodiment, the breast cancer is Estrogen Receptor (ER) negative breast cancer, Progesterone Receptor (PR) negative breast cancer and/or HER-2 negative breast cancer. For example, the breast cancer may be Estrogen Receptor (ER) negative and Progesterone Receptor (PR) negative breast cancer; Estrogen Receptor (ER) negative and HER-2 negative breast cancer; Progesterone Receptor (PR) negative and HER-2 negative breast cancer; or Estrogen Receptor (ER) negative breast cancer, Progesterone Receptor (PR) negative breast cancer and HER-2 negative (triple negative) breast cancer. Assessment of ER, PR and HER-2 status can be done using any suitable method. For example, HER-2 status can be assessed by immunohistochemistry (IHC) and/or gene amplification by fluorescence *in situ* hybridization (FISH), for example, according to National Comprehensive Cancer Network [NCCN] guidelines.

The breast cancer can be for example, adenocarcinoma, inflammatory breast cancer, recurrent (*e.g.*, locally recurrent) and/or metastatic breast cancer. In some embodiments, the breast cancer is endocrine refractory or hormone refractory. The terms "endocrine refractory" and "hormone refractory" refer to a cancer that is resistant to treatment with hormone therapy for breast cancer, *e.g.*, aromatase inhibitors or tamoxifen. Breast cancers arise most commonly in the lining of the milk ducts of the breast (ductal carcinoma), or in the lobules where breast milk is produced (lobular carcinoma). Accordingly, in various embodiments of the invention, the breast cancer can be ductal carcinoma or lobular carcinoma. Cancerous cells from the breast(s) may invade or metastasize to any other organ or tissue of the body. For example, cancer cells often invade lymph node cells and/or metastasize to the liver, brain and/or bone.

In various embodiments of the present invention, the subject may be suffering from Stage I, Stage II, Stage III or Stage IV breast cancer. The stage of a breast cancer can be classified as a range of stages from Stage 0 to Stage IV based on its size and the extent to which it has spread. The following table summarizes the stages, which are well known to clinicians:

STAGE	TUMOR SIZE	LYMPH NODE INVOLVEMENT	METASTASIS (SPREAD)
I	Less than 2 cm	No	No
II	Between 2-5 cm	No or in same side of breast	No
III	More than 5 cm	Yes, on same side of breast	No
IV	Not applicable	Not applicable	Yes

Various aspects of the invention are described in further detail in the following subsections.

I. Prediction of Responsiveness to Eribulin, an Analog thereof, or a Pharmaceutically Acceptable Salt thereof (*e.g.*, Eribulin Mesylate) in Subjects with Breast Cancer

In one aspect, the invention provides a method for determining whether eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), may be used to treat a subject having breast cancer and/or for determining the sensitivity of a breast tumor to treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate). The methods involve determining the expression level of at least one biomarker, for example, by assaying a sample derived from a subject having breast cancer. The identification of low levels of expression of at least one biomarker is indicative that eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate) may be used for treatment of the breast cancer and/or that the breast tumor is sensitive to the treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate).

In the methods of the invention, the expression level of at least one biomarker selected from the group of biomarkers set forth in Table 1 is assessed, which, as explained herein, can comprise determining the level of expression of one or more of these genes (e.g., ABI3, ANG) using various approaches, such as determining in a suitable sample the presence of certain DNA polymorphisms or null mutations, determining the level of RNA expressed from a gene, including an mRNA exemplified in Table 1 and/or other transcripts from the gene, or a protein product(s) of any of the foregoing.

TABLE 1: BIOMARKERS OF RESPONSIVENESS TO ERIBULIN, AN ANALOG THEREOF, OR A PHARMACEUTICALLY ACCEPTABLE SALT THEREOF (e.g., ERIBULIN MESYLATE)

GENE NAME	GENE_ID	ACCESSION NO.	SEQ ID NO:	NAME
ABI3	51225	NM_016428	1 & 2	Homo sapiens ABI family, member 3 (ABI3), transcript variant 1, mRNA
ANG	283	NM_001097577	3 & 4	Homo sapiens angiogenin, ribonuclease, RNase A family, 5 (ANG), transcript variant 2, mRNA
APBB2	323	NM_173075	5 & 6	Homo sapiens amyloid beta (A4) precursor protein-binding, family B, member 2 (APBB2), mRNA
CCL26	10344	NM_006072	7 & 8	Homo sapiens chemokine (C-C motif) ligand 26 (CCL26), mRNA
CDC20	991	NM_001255	9 & 10	Homo sapiens cell division cycle 20 homolog (S. cerevisiae) (CDC20), mRNA
CEP152	22995	NM_014985	11 & 12	Homo sapiens centrosomal protein 152kDa (CEP152), mRNA
CFL1	1072	NM_005507	13 & 14	Homo sapiens cofilin 1 (non-muscle) (CFL1), mRNA
CKLF	51192	NM_016326	15 & 16	Homo sapiens chemokine-like factor (CKLF), transcript variant 3, mRNA
COL7A1	1294	NM_000094	17 & 18	Homo sapiens collagen, type VII, alpha 1 (COL7A1), mRNA
CYP4F3	4051	NM_000896	19 & 20	Homo sapiens cytochrome P450, family 4, subfamily F, polypeptide 3 (CYP4F3), mRNA
DYSF	8291	NM_003494	21 & 22	Homo sapiens dysferlin, limb girdle muscular dystrophy 2B (autosomal recessive) (DYSF), transcript variant 8, mRNA
EDIL3	10085	NM_005711	23 & 24	Homo sapiens EGF-like repeats and discoidin I-like domains 3 (EDIL3), mRNA
ERGIC3	51614	NM_015966	25 & 26	Homo sapiens ERGIC and golgi 3 (ERGIC3), transcript variant 2, mRNA
GNAT1	2779	NM_000172	27 & 28	Homo sapiens guanine nucleotide binding protein (G protein), alpha transducing activity polypeptide 1 (GNAT1), transcript variant 2, mRNA
GRAMD4	23151	NM_015124	29 & 30	Homo sapiens GRAM domain containing 4 (GRAMD4), mRNA
HYAL2	8692	NM_003773	31 & 32	Homo sapiens hyaluronoglucosaminidase 2 (HYAL2), transcript variant 1, mRNA
IL10	3586	NM_000572	33 & 34	Homo sapiens interleukin 10 (IL10), mRNA
ITFG3	83986	NM_032039	35 & 36	Homo sapiens integrin alpha FG-GAP repeat containing 3 (ITFG3), mRNA
JAM3	83700	NM_032801	37 & 38	Homo sapiens junctional adhesion molecule 3 (JAM3), mRNA
KLHL17	339451	NM_198317	39 & 40	Homo sapiens kelch-like 17 (Drosophila) (KLHL17), mRNA
KRT24	192666	NM_019016	41 & 42	Homo sapiens keratin 24 (KRT24), mRNA
MAD2L1BP	9587	NM_014628	43 & 44	Homo sapiens MAD2L1 binding protein (MAD2L1BP), transcript variant 2, mRNA
MANSC1	54682	NM_018050	45 & 46	Homo sapiens MANSC domain containing 1 (MANSC1), mRNA
MOBKL1B	55233	NM_018221	47 & 48	Homo sapiens MOB1, Mps One Binder kinase activator-like 1B (yeast) (MOBKL1B), mRNA
NCBP1	4686	NM_002486	49 & 50	Homo sapiens nuclear cap binding protein

				subunit 1, 80kDa (NCBP1), mRNA
NMU	10874	NM_006681	51 & 52	Homo sapiens neuromedin U (NMU), mRNA
PAPLN	89932	NM_173462	53 & 54	Homo sapiens papilin, proteoglycan-like sulfated glycoprotein (PAPLN), mRNA
PCDH1	5097	NM_002587	55 & 56	Homo sapiens protocadherin 1 (PCDH1), transcript variant 1, mRNA
PDGFB	5155	NM_002608	57 & 58	Homo sapiens platelet-derived growth factor beta polypeptide (simian sarcoma viral (v-sis) oncogene homolog) (PDGFB), transcript variant 1, mRNA
PHOSPHO2	493911	NM_001008489	59 & 60	Homo sapiens phosphatase, orphan 2 (PHOSPHO2), mRNA
PSENEN	55851	NM_172341	61 & 62	Homo sapiens presenilin enhancer 2 homolog (C. elegans) (PSENEN), mRNA
SATB1	6304	NM_002971	63 & 64	Homo sapiens SATB homeobox 1 (SATB1), transcript variant 1, mRNA
SNX11	29916	NM_013323	65 & 66	Homo sapiens sorting nexin 11 (SNX11), transcript variant 2, mRNA
SPTA1	6708	NM_003126	67 & 68	Homo sapiens spectrin, alpha, erythrocytic 1 (elliptocytosis 2) (SPTA1), mRNA
TMEM79	84283	NM_032323	69 & 70	Homo sapiens transmembrane protein 79 (TMEM79), transcript variant 1, mRNA
TMIGD2	126259	NM_144615	71 & 72	Homo sapiens transmembrane and immunoglobulin domain containing 2 (TMIGD2), mRNA
TUBB6	84617	NM_032525	73 & 74	Homo sapiens tubulin, beta 6 (TUBB6), mRNA
TYROBP	7305	NM_198125	75 & 76	Homo sapiens TYRO protein tyrosine kinase binding protein (TYROBP), transcript variant 2, mRNA
YTHDF1	54915	NM_017798	77 & 78	Homo sapiens YTH domain family, member 1 (YTHDF1), mRNA
ZIC5	85416	NM_033132	79 & 80	Homo sapiens Zic family member 5 (odd-paired homolog, Drosophila) (ZIC5), mRNA

Each of the accession numbers identified in Table 1, and their corresponding sequences, are hereby incorporated herein by reference.

In various embodiments, the level of expression of at least 2, 3, 4, 5, 6, 7, 8, 9 or 10 biomarkers selected from the group of biomarkers listed in Table 1 is determined.

In particular embodiments, a predictive gene signature comprising a sub-combination of 2 or more biomarkers selected from the group of biomarkers listed in Table 1 is used. In various embodiments, the level of expression of at least 2, at least 3, at least 4 or at least 5 biomarkers selected from the group of biomarkers listed in Table 1 is determined. For example, the predictive gene signature may include at least 2 biomarkers, *e.g.*, DYSF and EDIL3; GNAT1 and ERGIC3; KRT24 and PAPLN; MANSC1 and PDGFB; PCDH1 and PDGFB; or PHOSPHO2 and PSENEN. In another embodiment, the predictive gene signature may include at least 3 biomarkers, *e.g.*, COL7A1, YTHDF1 and ZIC5; CKLF, IL10 and TUBB6; CDC20, CFL1 and TMEM79; HYAL2, NCBP1 and SNX11; or CEP152, NCBP1 and SATB1. In another embodiment, the predictive gene signature may include at least 4 biomarkers, *e.g.*,

APBB2, CCL26, PSENEN and SATB1; ANG, JAM3, KLHL17 and PAPLN; ITFG3, MAD2L1BP, NMU and PDGFB; SPTA1, TYROBP, SNX11 and PSENEN; GRAMD4, GNAT1, TMIGD2 and YTHDF1; or GRAMD4, HYAL2, PHOSPHO2 and TUBB6. In another embodiment, the predictive gene signature may include at least 5 biomarkers, *e.g.*, CCL26, CDC20, ERGIC3, EDIL3 and PCDH1; DYSF, NMU, PHOSPHO2, PSENEN and SNX11; APBB2, CKLF, CYP4F3, TUBB6 and YTHDF1; or CEP152, MAD2L1BP, SPTA1, TMEM79 and ZIC5.

In particular embodiments, the predictive gene signature may include 2 or more of biomarkers ABI3, ANG, APBB2, CCL26, CDC20, CEP152, CFL1, CKLF, COL7A1, CYP4F3, DYSF, GNAT1, GRAMD4, HYAL2, IL10, ITFG3, JAM3, KLHL17, KRT24, MAD2L1BP, MANSC1, MOBKL1B, NCBP1, NMU, PCDH1, PHOSPHO2, SPTA1, TMIGD2, TYROBP, ZIC5, ERGIC3, PDGFB, PSENEN, SATB1, SNX11, TMEM79 and YTHDF1, *e.g.*, ABI3 and ANG; APBB2 and CCL26; GNAT1 and GRAMD4; IL10 and ITFG3; MACSC1 and MOBKL1B; NMU and PCDH1; or TYROBP and ZIC5. In other embodiments, the predictive gene signature includes at least 3 of the previously recited biomarkers, *e.g.*, ABI3, ANG and APBB2; CCL26, CKLF and COL7A1; DYSF, GNAT1 and HYAL2; JAM3, KLHL17 and KRT24; NCBP1, NMU and PCDH1; SPTA1, TMIGD2 and TYROBP; or ZIC5, MAD2L1BP and CDC20. In other embodiments, the predictive gene signature includes at least 4 of the previously recited biomarkers, *e.g.*, ABI3, ANG, APBB2 and CCL26; CEP152, CFL1, CKLF and COL7A1; KRT24, MANSC1, MOBKL1B and SPTA1; TYROBP, TMIGD2, PHOSPHO2 and NMU; ABI3, GNATI, KLHL17 and SPTA1; or CEP152, HYAL2, PCDH1 and TMIGD2. In yet further embodiments, the predictive gene signature includes at least 5 of the previously recited biomarkers, *e.g.*, CKLF, COL7A1, GRAMD4, JAM3 and PCDH1; APBB2, CEP152, DYSF, IL10 and TYROBP; CYP4F3, HYAL2, ITFG3, KLHL17 and KRT24; NCBP1, SPTA1, TMIGD2, IL10 and JAM3; or CCL26, PHOSPHO2, SPTA1, TMIGD2 and ZIC5.

In other embodiments, the predictive gene signature may include 2 or more of biomarkers ERGIC3, PDGFB, PSENEN, SATB1, SNX11, TMEM79 or YTHDF1, or any sub-combination thereof, *e.g.*, ERGIC3 and PDGFB; ERGIC3 and PSENEN; ERGIC3 and SATB1; ERGIC3 and SNX11; ERGIC3 and TMEM79; ERGIC3 and YTHDF1; PDGFB and PSENEN; PDGFB and SATB1; PDGFB and SNX11; PDGFB and TMEM79; PDGFB and YTHDF1; PSENEN and SATB1; PSENEN and SNX11; PSENEN and TMEM79; PSENEN and YTHDF1; SATB1 and SNX11; SATB1 and TMEM79; SATB1 and YTHDF1; SNX11 and TMEM79; SNX11 and YTHDF1; or TMEM79 and YTHDF1. In other embodiments, the predictive gene signature includes at least 3 biomarkers, for example, ERGIC3, PDGFB and PSENEN; SATB1, SNX11

and TMEM79; SNX11, TMEM79 and YTHDF1; or ERGIC3, PDGFB and SATB1. In further embodiments, the predictive gene signature includes at least 4 biomarkers, for example, ERGIC3, PDGFB, PSENEN and SATB1; SNX11, TMEM79, YTHDF1 and ERGIC3; or ERGIC3, PDGFB, PSENEN and YTHDF1. In further embodiments, the predictive gene signature includes at least 5 biomarkers, for example, ERGIC3, PDGFB, PSENEN, SATB1 and SNX11; ERGIC3, PDGFB, PSENEN, SATB1 and TMEM79; or PSENEN, SATB1, SNX11, TMEM79 and YTHDF1. In yet further embodiments, the predictive gene signature includes at least 6 biomarkers, for example, ERGIC3, PDGFB, PSENEN, SATB1, SNX11 and TMEM79; PDGFB, PSENEN, SATB1, SNX11, TMEM79 and YTHDF1; or ERGIC3, PSENEN, SATB1, SNX11, TMEM79 and YTHDF1. In yet another embodiment, the predictive gene signature includes 7 biomarkers, for example, ERGIC3, PDGFB, PSENEN, SATB1, SNX11, TMEM79 and YTHDF1.

In various embodiments, the biomarker is not one or more of SPTA1, PAPLN, PCDH1, TMIGD2 and/or KRT24. In a particular embodiment, the biomarker is not SPTA1, PAPLN, PCDH1, TMIGD2 and KRT24. In one embodiment, the biomarker is not SPTA1. In another embodiment, the biomarker is not PAPLN. In another embodiment, the biomarker is not PCDH1. In an alternative embodiment, the biomarker is not TMIGD2. In yet another embodiment, the biomarker is not KRT24.

In particular embodiments, the predictive gene signature may include 2 or more of biomarkers ABI3, ANG, APBB2, CCL26, CDC20, CEP152, CFL1, CKLF, COL7A1, CYP4F3, DYSF, GNAT1, GRAMD4, HYAL2, IL10, ITFG3, JAM3, KLHL17, MAD2L1BP, MANSC1, MOBKL1B, NCBP1, NMU, PHOSPHO2, TYROBP, ZIC5, ERGIC3, PDGFB, PSENEN, SATB1, SNX11, TMEM79, YTHDF1, EDIL3 and TUBB6, *e.g.*, ABI3 and ANG; GRAMD4 and HYAL2; NMU and PHOSPHO2; ZIC5 and PSENEN; or SNX11 and MOBKL1B. In other embodiments, the predictive gene signature includes at least 3 of the previously recited biomarkers, *e.g.*, APBB2, CDC20 and CKLF; COL7A1, DYSF and GNAT1; NCBP1, SATB1 and EDIL3; PSENEN, DYSF and GNAT1; MANSC1, ZIC5 and CFL1; or CKLF, GRAMD4 and NMU. In other embodiments, the predictive gene signature includes at least 4 of the previously recited biomarkers, *e.g.*, ANG, CCL26, CEP152 and JAM3; APBB2, CYP4F3, ITFG3 and TYROBP; CYP4F3, MANSC1, PDGFB and YTHDF1; TUBB6, DYSF, PHOSPHO2 and CDC20; or CKLF, KLHL17, HYAL2 and ZIC5. In yet further embodiments, the predictive gene signature includes at least 5 of the previously recited biomarkers, *e.g.*, IL10, CEP152, COL7A1, TYROBP and ERGIC3; TMEM79, SNX11, PSENEN, GNAT1 and GRAMD4; JAM3, SNX11, KLHL17, MOBKL1B and ERGIC3; or NMU, PHOSPHO2, PDGFB, CFL1 and ANG.

In various methods and or kits of the invention, the biomarker is not ABI3, is not ANG, is not APBB2, is not CCL26, is not CDC20, is not CEP152, is not CFL1, is not CKLF, is not COL7A1, is not CYP4F3, is not DYSF, is not GNAT1, is not GRAMD4, is not HYAL2, is not IL10, is not ITFG3, is not JAM3, is not KLHL17, is not KRT24, is not MAD2L1BP, is not MANSC1, is not MOBKL1B, is not NCBP1, is not NMU, is not PCDH1, is not PHOSPHO2, is not SPTA1, is not TMIGD2, is not TYROBP, is not ZIC5, is not ERGIC3, is not PDGFB, is not PSENEN, is not SATB1, is not SNX11, is not TMEM79, is not EDIL3, is not PAPLN, is not TUBB6 and/or is not YTHDF1.

Any suitable analytical method, can be utilized in the methods of the invention to assess (directly or indirectly) the level of expression of a biomarker in a sample. In some embodiments, a difference is observed between the level of expression of a biomarker, as compared to the control level of expression of the biomarker. In one embodiment, the difference is greater than the limit of detection of the method for determining the expression level of the biomarker. In further embodiments, the difference is greater than or equal to the standard error of the assessment method, and preferably the difference is at least about 2-, about 3-, about 4-, about 5-, about 6-, about 7-, about 8-, about 9-, about 10-, about 15-, about 20-, about 25-, about 100-, about 500- or about 1000-fold greater than the standard error of the assessment method. In some embodiments, the level of expression of the biomarker in a sample as compared to a control level of expression is assessed using parametric or nonparametric descriptive statistics, comparisons, regression analyses, and the like.

In some embodiments, a difference in the level of expression of the biomarker in the sample derived from the subject is detected relative to the control, and the difference is about 5%, about 10%, about 15%, about 20%, about 25%, about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, or about 90% less than the expression level of the biomarker in the control sample.

The level of expression of a biomarker, for example, as set forth in Table 1, in a sample obtained from a subject may be assayed by any of a wide variety of techniques and methods, which transform the biomarker within the sample into a moiety that can be detected and/or quantified. Non-limiting examples of such methods include analyzing the sample using immunological methods for detection of proteins, protein purification methods, protein function or activity assays, nucleic acid hybridization methods, nucleic acid reverse transcription methods, and nucleic acid amplification methods, immunoblotting, Western blotting, Northern blotting, electron microscopy, mass spectrometry, *e.g.,* MALDI-TOF and SELDI-TOF, immunoprecipitations, immunofluorescence, immunohistochemistry, enzyme linked immunosorbent assays

(ELISAs), *e.g.*, amplified ELISA, quantitative blood based assays, *e.g.*, serum ELISA, quantitative urine based assays, flow cytometry, Southern hybridizations, array analysis, and the like, and combinations or sub-combinations thereof.

In one embodiment, the level of expression of the biomarker in a sample is determined by detecting a transcribed polynucleotide, or portion thereof, *e.g.*, mRNA, or cDNA, of the biomarker gene. RNA may be extracted from cells using RNA extraction techniques including, for example, using acid phenol/guanidine isothiocyanate extraction (RNAzol B; Biogenesis), RNeasy RNA preparation kits (Qiagen) or PAXgene (PreAnalytix, Switzerland). Typical assay formats utilizing ribonucleic acid hybridization include nuclear run-on assays, RT-PCR, quantitative PCR analysis, RNase protection assays (Melton *et al.*, *Nuc. Acids Res.* 12:7035), Northern blotting and *in situ* hybridization. Other suitable systems for mRNA sample analysis include microarray analysis (*e.g.*, using Affymetrix's microarray system or Illumina's BeadArray Technology).

In one embodiment, the level of expression of the biomarker is determined using a nucleic acid probe. The term "probe", as used herein, refers to any molecule that is capable of selectively binding to a specific biomarker. Probes can be synthesized by one of skill in the art, or derived from appropriate biological preparations. Probes can be specifically designed to be labeled, by addition or incorporation of a label. Examples of molecules that can be utilized as probes include, but are not limited to, RNA, DNA, proteins, antibodies, and organic molecules.

As indicated above, isolated mRNA can be used in hybridization or amplification assays that include, but are not limited to, Southern or Northern analyses, polymerase chain reaction (PCR) analyses and probe arrays. One method for the determination of mRNA levels involves contacting the isolated mRNA with a nucleic acid molecule (probe) that can hybridize to the biomarker mRNA. The nucleic acid probe can be, for example, a full-length cDNA, or a portion thereof, such as an oligonucleotide of at least about 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 250 or about 500 nucleotides in length and sufficient to specifically hybridize under appropriate hybridization conditions to the biomarker genomic DNA. In a particular embodiment the probe will bind the biomarker genomic DNA under stringent conditions. Such stringent conditions, for example, hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45° C., followed by one or more washes in 0.2X SSC, 0.1% SDS at 50-65° C., are known to those skilled in the art and can be found in Current Protocols in Molecular Biology, Ausubel et al., eds., John Wiley & Sons, Inc. (1995), sections 2, 4, and 6, the teachings of which are hereby incorporated by reference herein. Additional stringent conditions can be found in

Molecular Cloning: A Laboratory Manual, Sambrook et al., Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (1989), chapters 7, 9, and 11, the teachings of which are hereby incorporated by reference herein.

In one embodiment, the mRNA is immobilized on a solid surface and contacted with a probe, for example by running the isolated mRNA on an agarose gel and transferring the mRNA from the gel to a membrane, such as nitrocellulose. In an alternative embodiment, the probe(s) are immobilized on a solid surface, for example, in an Affymetrix gene chip array, and the probe(s) are contacted with mRNA. A skilled artisan can readily adapt mRNA detection methods for use in determining the level of the biomarker mRNA.

The level of expression of the biomarker in a sample can also be determined using methods that involve the use of nucleic acid amplification and/or reverse transcriptase (to prepare cDNA) of for example mRNA in the sample, *e.g.*, by RT-PCR (the experimental embodiment set forth in Mullis, 1987, U.S. Pat. No. 4,683,202), ligase chain reaction (Barany (1991) *Proc. Natl. Acad. Sci. USA* 88:189-193), self sustained sequence replication (Guatelli *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi *et al.* (1988) *Bio/Technology* 6:1197), rolling circle replication (Lizardi *et al.* , U.S. Pat. No. 5,854,033) or any other nucleic acid amplification method, followed by the detection of the amplified molecules. These approaches are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers. In particular aspects of the invention, the level of expression of the biomarker is determined by quantitative fluorogenic RT-PCR (*e.g.*, the TaqMan™ System). Such methods typically utilize pairs of oligonucleotide primers that are specific for the biomarker. Methods for designing oligonucleotide primers specific for a known sequence are well known in the art.

The expression levels of biomarker mRNA can be monitored using a membrane blot (such as used in hybridization analysis such as Northern, Southern, dot, and the like), or microwells, sample tubes, gels, beads or fibers (or any solid support comprising bound nucleic acids). See, for example, U.S. Pat. Nos. 5,770,722, 5,874,219, 5,744,305, 5,677,195 and 5,445,934, the contents of which as they relate to these assays are incorporated herein by reference. The determination of biomarker expression level may also comprise using nucleic acid probes in solution.

In one embodiment of the invention, microarrays are used to detect the level of expression of a biomarker. Microarrays are particularly well suited for this purpose because of the reproducibility between different experiments. DNA microarrays provide

one method for the simultaneous measurement of the expression levels of large numbers of genes. Each array consists of a reproducible pattern of capture probes attached to a solid support. Labeled RNA or DNA is hybridized to complementary probes on the array and then detected by laser scanning. Hybridization intensities for each probe on the array are determined and converted to a quantitative value representing relative gene expression levels. See, *e.g.*, U.S. Pat. Nos. 6,040,138, 5,800,992 and 6,020,135, 6,033,860, and 6,344,316, the contents of which as they relate to these assays are incorporated herein by reference. High-density oligonucleotide arrays are particularly useful for determining the gene expression profile for a large number of RNA's in a sample.

Expression of a biomarker can also be assessed at the protein level, using a detection reagent that detects the protein product encoded by the mRNA of the biomarker, directly or indirectly. For example, if an antibody reagent is available that binds specifically to a biomarker protein product to be detected, then such an antibody reagent can be used to detect the expression of the biomarker in a sample from the subject, using techniques, such as immunohistochemistry, ELISA, FACS analysis, and the like.

Other known methods for detecting the biomarker at the protein level include methods such as electrophoresis, capillary electrophoresis, high performance liquid chromatography (HPLC), thin layer chromatography (TLC), hyperdiffusion chromatography, and the like, or various immunological methods such as fluid or gel precipitation reactions, immunodiffusion (single or double), immunoelectrophoresis, radioimmunoassay (RIA), enzyme-linked immunosorbent assays (ELISAs), immunofluorescent assays, and Western blotting.

Proteins from samples can be isolated using a variety of techniques, including those well known to those of skill in the art. The protein isolation methods employed can, for example, be those described in Harlow and Lane (Harlow and Lane, 1988, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York).

In one embodiment, antibodies, or antibody fragments, are used in methods such as Western blots or immunofluorescence techniques to detect the expressed proteins. Antibodies for determining the expression of the biomarkers of the invention are commercially available. For example, ERGIC-3 specific antibodies are commercially available from Santa Cruz Biotechnology, Inc. (ERGIC-3 (P-16) Antibody and ERGIC-3 (Y-23) Antibody) and Sigma Aldrich (HPA015968, AV47209, HPA015242, SAB4502151). PDGFB specific antibodies are commercially available from Santa Cruz

Biotechnology, Inc. (*e.g.*, PDGF-B (C-5) Antibody and PDGF-B (H-55) Antibody) and Sigma Aldrich (*e.g.*, HPA011972, SAB2101755 and SAB2900226). PSENEN specific antibodies are commercially available from Origene (*e.g.*, Catalog No. TA306367) and Sigma Aldrich (*e.g.*, PRS3981, WH0055851M1, PRS3979 and P5622). Further by way of example, SATB1 antibodies are commercially available from, for example, Abcam (Catalog No. ab49061, ab92307 and ab70004), Abnova Corporation (Catalog No. PAB13379), and Aviva Systems Biology (Catalog No. ARP33362_P050). SNX11 antibodies are commercially available from Abcam (Catalog Nos. ab4128, ab67578, ab76816 and ab76762) and Abnova Corporation (Catalog Nos. PAB6362 and H00029916-B01). TMEM79 Antibodies are commercially available from, for example, Abcam (Catalog No. ab81539) and Sigma Aldrich (Catalog No. SAB2102475). Finally, YTHDF1 antibodies are commercially available from, for example, Abnova Corporation (Catalog No. PAB17446), Aviva Systems Biology (Catalog No. ARP57032_P050) and Santa Cruz Biotechnology (Catalog No. sc-86026).

It is generally preferable to immobilize either the antibody or proteins on a solid support for Western blots and immunofluorescence techniques. Suitable solid phase supports or carriers include any support capable of binding an antigen or an antibody. Well-known supports or carriers include glass, polystyrene, polypropylene, polyethylene, dextran, nylon, amylases, natural and modified celluloses, polyacrylamides, gabbros, and magnetite.

One skilled in the art will know many other suitable carriers for binding antibody or antigen, and will be able to adapt such support for use with the present invention. For example, protein isolated from cells can be run on a polyacrylamide gel electrophoresis and immobilized onto a solid phase support such as nitrocellulose. The support can then be washed with suitable buffers followed by treatment with the detectably labeled antibody. The solid phase support can then be washed with the buffer a second time to remove unbound antibody. The amount of bound label on the solid support can then be detected by conventional means. Means of detecting proteins using electrophoretic techniques are well known to those of skill in the art (see generally, R. Scopes (1982) *Protein Purification*, Springer-Verlag, N.Y.; Deutscher, (1990) *Methods in Enzymology* Vol. 182: *Guide to Protein Purification*, Academic Press, Inc., N.Y.).

Other standard methods include immunoassay techniques which are well known to one of ordinary skill in the art and may be found in Principles And Practice Of Immunoassay, 2nd Edition, Price and Newman, eds., MacMillan (1997) and Antibodies, A Laboratory Manual, Harlow and Lane, eds., Cold Spring Harbor Laboratory, Ch. 9 (1988), each of which is incorporated herein by reference in its entirety.

Antibodies used in immunoassays to determine the level of expression of the biomarker, may be labeled with a detectable label. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by incorporation of a label (e.g., a radioactive atom), coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin.

In one embodiment, the antibody is labeled, *e.g.* a radio-labeled, chromophore-labeled, fluorophore-labeled, or enzyme-labeled antibody. In another embodiment, an antibody derivative (*e.g.*, an antibody conjugated with a substrate or with the protein or ligand of a protein-ligand pair (*e.g.*, biotin-streptavidin), or an antibody fragment (*e.g.* a single-chain antibody, or an isolated antibody hypervariable domain) which binds specifically with the biomarker is used.

In one embodiment of the invention, proteomic methods, *e.g.,* mass spectrometry, are used. Mass spectrometry is an analytical technique that consists of ionizing chemical compounds to generate charged molecules (or fragments thereof) and measuring their mass-to-charge ratios. In a typical mass spectrometry procedure, a sample is obtained from a subject, loaded onto the mass spectrometry, and its components (*e.g.*, the biomarker) are ionized by different methods (*e.g.*, by impacting them with an electron beam), resulting in the formation of charged particles (ions). The mass-to-charge ratio of the particles is then calculated from the motion of the ions as they transit through electromagnetic fields.

For example, matrix-associated laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) or surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF MS) which involves the application of a biological sample, such as serum, to a protein-binding chip (Wright, G.L., Jr., *et al.* (2002) *Expert Rev Mol Diagn* 2:549; Li, J., *et al.* (2002) *Clin Chem* 48:1296; Laronga, C., *et al.* (2003) *Dis biomarkers* 19:229; Petricoin, E.F., *et al.* (2002) 359:572; Adam, B.L., *et al.* (2002) *Cancer Res* 62:3609; Tolson, J., *et al.* (2004) *Lab Invest* 84:845; Xiao, Z., *et al.* (2001) *Cancer Res* 61:6029) can be used to determine the expression level of a biomarker at the protein level.

Furthermore, *in vivo* techniques for determination of the expression level of the biomarker include introducing into a subject a labeled antibody directed against the biomarker, which binds to and transforms the biomarker into a detectable molecule. As

discussed above, the presence, level, or even location of the detectable biomarker in a subject may be detected by standard imaging techniques.

In general, where a difference in the level of expression of a biomarker and the control is to be detected, it is preferable that the difference between the level of expression of the biomarker in a sample from a subject having breast cancer and being treated with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), or being considered for treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), and the amount of the biomarker in a control sample, is as great as possible. Although this difference can be as small as the limit of detection of the method for determining the level of expression, it is preferred that the difference be greater than the limit of detection of the method or greater than the standard error of the assessment method, and preferably a difference of at least about 2-, about 3-, about 4-, about 5-, about 6-, about 7-, about 8-, about 9-, about 10-, about 15-, about 20-, about 25-, about 100-, about 500-, 1000-fold greater than the standard error of the assessment method.

In another aspect, the present invention provides methods for determining whether eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), may be used to treat a subject having breast cancer by determining and/or identifying whether the subject carries at least one gene, selected from the group of biomarkers set forth in Table 1, which contains a polymorphism, for example, a mutation, that results in decreased expression and/or reduced function of the encoded protein, wherein the presence of a polymorphism in at least one gene is indicative that eribulin, an analog thereof, or a pharmaceutically acceptable salt there, will be effective in treating a subject. In another aspect, the present invention provides methods for predicting whether eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), will be effective in treating a subject having breast cancer by assaying a sample derived from the subject to determine whether the subject carries at least one gene, selected from the group of biomarkers set forth in Table 1, which contains a polymorphism, for example, a mutation, that results in decreased expression and/or reduced function of the encoded protein, wherein the presence of the polymorphism in at least one gene in said sample is indicative that eribulin, an analog thereof, or a pharmaceutically acceptable salt there, may be used to treat said subject. In a further aspect, a method is provided for predicting whether eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof, will be effective in treating a subject having breast cancer, the method comprising determining the presence of a polymorphism that results in reduced expression and/or function in a gene encoding a biomarker selected from the group of biomarkers listed in Table 1 in a sample derived

from said subject, and predicting that eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof, may be used to treat said subject based on the presence of the polymorphism.

In a further aspect, the present invention provides methods for determining the sensitivity of a breast tumor to treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof by determining and/or identifying whether said tumor contains a polymorphism in at least one gene resulting in reduced expression and/or function of the encoded protein, wherein the gene is selected from the group of biomarkers set forth in Table 1. Identification and/or determination that the tumor contains such a polymorphism is indicative of sensitivity of said tumor to treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof. In yet another aspect, the present invention is directed to methods for treating a subject having breast cancer with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof by identifying whether a sample derived from said subject has at least one gene, selected from the group of biomarkers set forth in Table 1, which contains a polymorphism resulting in reduced expression and/or function of the encoded protein and administering a therapeutically effective amount of eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate) to the subject when the polymorphism is identified.

In another aspect, the present invention provides methods for determining whether eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), may be used to treat a subject having breast cancer by determining and/or identifying whether the subject carries at least one gene, selected from the group of biomarkers set forth in Table 1, which contains a null mutation, wherein the presence of a null mutation in at least one gene is indicative that eribulin, an analog thereof, or a pharmaceutically acceptable salt there, will be effective in treating the subject. In another aspect, the present invention provides methods for predicting whether eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), may be used to treat a subject having breast cancer by assaying a sample derived from the subject to determine whether the subject carries at least one gene, selected from the group of biomarkers set forth in Table 1, which contains a null mutation, wherein the presence of a null mutation in at least one gene in said sample is indicative that eribulin, an analog thereof, or a pharmaceutically acceptable salt there, will be effective in treating the subject. In a further aspect, a method is provided for predicting whether eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof, may be used to treat a subject having breast cancer, the method comprising determining the presence of a null mutation in a biomarker selected from the group of biomarkers listed in Table 1 in

a sample derived from said subject, and predicting that eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof, may be used to treat said subject based on the presence of said null mutation.

In a further aspect, the present invention provides methods for determining the sensitivity of a breast tumor to treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof by determining and/or identifying whether said tumor contains a null mutation in at least one gene, selected from the group of biomarkers set forth in Table 1. Identification and/or determination that the tumor contains a null mutation is indicative of sensitivity of said tumor to treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof. In yet another aspect, the present invention is directed to methods for treating a subject having breast cancer with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof by identifying whether a sample derived from said subject has at least one gene, selected from the group of biomarkers set forth in Table 1, which contains a null mutation and administering a therapeutically effective amount of eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate) to the subject when a null mutation is identified.

As used herein, the term "null mutation" refers to a mutation in a genomic DNA sequence that causes the product of the gene to be non-functional or largely absent. Such mutations may occur in the coding and/or regulatory regions of the gene, and may be changes of individual residues, or insertions or deletions of regions of nucleic acids. These mutations may also occur in the coding and/or regulatory regions of other genes which may regulate or control the gene and/or the product of the gene so as to cause the gene product to be non-functional or largely absent. The null mutation may be a deletion of the native gene or a portion thereof. These sequence disruptions or modifications may include insertions, missense, frameshift, deletion, or substitutions, or replacements of DNA sequence, or any combination thereof. For example, the null mutation may result in the insertion of a premature stop codon. The null mutation results in functional inactivation of the gene product by, for example, inhibiting its production partially or completely; disrupting, inhibiting or curtailing the translation of the protein product, or resulting in a nonfunctional protein. The null mutation may be a pre-existing mutation in the subject or a mutation which arose in a tumor. The presence of a null mutation in a biomarker can be indicated by a lack of detectable expression of the biomarker. However, the presence of a null mutation can be determined by other methods. For example, in such embodiments, a sample of the subject's DNA may be sequenced in order to identify the presence of a null mutation. Any of the well-known methods for sequencing the biomarkers may be used in the methods of the invention,

such as the methods described in, for example, U.S. Patent No. 5,075,216, Engelke *et al.* (1988) *Proc. Natl. Acad. Sci. U.S.A.* 85, 544-548 and Wong *et al.* (1987) *Nature* 330, 384-386; Maxim and Gilbert (1977) *Proc. Natl. Acad. Sci. U.S.A.* 74:560; or Sanger (1977) *Proc. Natl. Acad. Sci. U.S.A.* 74:5463. In addition, any of a variety of automated sequencing procedures can be utilized see, *e.g.*, Naeve, C.W *et al.* (1995) *Biotechniques* 19:448, including sequencing by mass spectrometry (see, *e.g.*, PCT International Publication No. WO 94/16101; Cohen *et al.* (1996) *Adv. Chromatogr.* 36:127-162; and Griffin *et al.* (1993) *Appl. Biochem. Biotechnol.* 38:147-159.

Determining the presence or absence of a null mutation in the sample may also be accomplished using various techniques such as polymerase chain reaction (PCR) amplification reaction, reverse-transcriptase PCR analysis, single-strand conformation polymorphism analysis (SSCP), mismatch cleavage detection, heteroduplex analysis, Southern blot analysis, Western blot analysis, deoxyribonucleic acid sequencing, restriction fragment length polymorphism analysis, haplotype analysis, serotyping, and combinations or sub-combinations thereof.

Any suitable sample obtained from a subject having breast cancer can be used to assess the level of expression, including a lack of expression, of the biomarker, for example, a biomarker provided in Table 1. For example, the sample may be any fluid or component thereof, such as a fraction or extract, *e.g.*, blood, plasma, lymph, cystic fluid, urine, nipple aspirates, or fluids collected from a biopsy (*e.g.*, lump biopsy), obtained from the subject. In a typical situation, the fluid may be blood, or a component thereof, obtained from the subject, including whole blood or components thereof, including, plasma, serum, and blood cells, such as red blood cells, white blood cells and platelets. The sample may also be any tissue or fragment or component thereof, *e.g.*, breast tissue, connective tissue, lymph tissue or muscle tissue obtained from the subject.

Techniques or methods for obtaining samples from a subject are well known in the art and include, for example, obtaining samples by a mouth swab or a mouth wash; drawing blood; or obtaining a biopsy. Isolating components of fluid or tissue samples (*e.g.*, cells or RNA or DNA) may be accomplished using a variety of techniques.

The sample from the cancer may be obtained by biopsy of the patient's cancer. In certain embodiments, more than one sample from the patient's tumor is obtained in order to acquire a representative sample of cells for further study. For example, a patient with breast cancer may have a needle biopsy to obtain a sample of cancer cells. Several biopsies of the tumor may be used to obtain a sample of cancer cells. In other embodiments, the sample may be obtained from surgical excision of the tumor. In this

case, one or more samples may be taken from the excised tumor for analysis using the methods of the invention.

After the sample is obtained, it may be further processed. The cancer cells may be cultured, washed, or otherwise selected to remove normal tissue. The cells may be trypsinized to remove the cells from the tumor sample. The cells may be sorted by fluorescence activated cell sorting (FACS) or other cell sorting technique. The cells may be cultured to obtain a greater number of cells for study. In certain instances the cells may be immortalized. For some applications, the cells may be frozen or the cells may be embedded in paraffin.

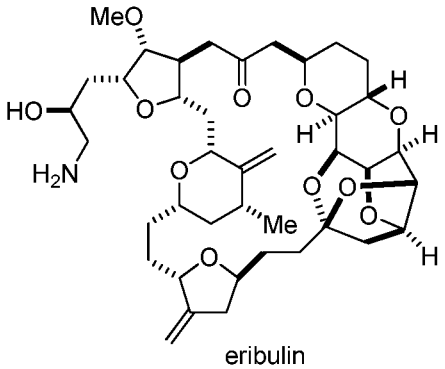
II. Treatment with Eribulin, Analogs Thereof, or Pharmaceutically Acceptable Salts Thereof (*e.g.*, Eribulin Mesylate)

Given the observation that the expression levels of certain biomarkers, for example, those set forth in Table 1, in a subject having breast cancer influences the responsiveness of the subject to eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), a skilled artisan can select an appropriate treatment regimen for the subject based on the expression levels of the biomarkers in the subject. Accordingly, the present invention provides methods for treating a subject having breast cancer by (i) identifying a subject having breast cancer in which at least one biomarker selected from the group of biomarkers listed in Table 1 has a low level of expression and (ii) administering a therapeutically effective amount of eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), to the subject . In another aspect, the present invention provides methods for treating a subject having breast cancer by (i) assaying a sample derived from the subject to determine the level of expression of at least one biomarker selected from the group of biomarkers listed in Table 1 and (ii) administering a therapeutically effective amount of eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate) to the subject when a low level of expression of the at least one biomarker is detected in the sample.

In various embodiments, the subject may have been previously treated with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate). In other embodiments, the subject may be under consideration for treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), for the first time. The level of expression of one or more, *e.g.*, at least 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10, biomarkers identified in Table 1 is determined. If level of expression of at least one biomarker (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10

biomarkers) is determined to be a low level of expression, treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), is likely to be efficacious. However, it is not necessary that all of the biomarkers assayed have a low level of expression as compared to the respective control. For example, while certain biomarkers may be present at normal or high levels of expression, treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), may be indicated when, for example, a low level of expression is present for at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9 or at least 10 biomarkers.

When a low level of expression of one or more of the biomarkers of the invention is found (*e.g.*, due to the presence of a null mutation in the biomarker gene) in a sample derived from a subject having breast cancer, the subject may be treated with eribulin, having the following the structure, with a pharmaceutically acceptable salt of eribulin, or with an eribulin analog or pharmaceutically acceptable thereof.



eribulin

In some embodiments, a pharmaceutically acceptable salt of eribulin is administered to the subject, such as eribulin mesylate.

The treatment regimen for eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), that is selected typically includes at least one of the following parameters and more typically includes many or all of the following parameters: the dosage, the formulation, the route of administration and/or the frequency of administration. Selection of the particular parameters of the treatment regimen can be based on known treatment parameters for eribulin previously established in the art such as those described in the Dosage and Administration protocols set forth in the FDA Approved Label for HALAVEN®, the entire contents of which are incorporated herein by reference. For example, eribulin mesylate can be administered intravenously

on Days 1 and 8 of a 21 day cycle, for example at a dose of 1.4 mg/m^2, or if a dose reduction is indicated (*e.g.*, for hepatic or renal impairment), at a dose of 0.7 mg/m^2 or 1.1 mg/m^2. Various modifications to dosage, formulation, route of administration and/or frequency of administration can be made based on various factors including, for example, the disease, age, sex, and weight of the patient, as well as the severity or stage of cancer (see, for example, U.S. Patent No. 6,653,341 and U.S. Patent No. 6,469,182, the entire contents of each of which are hereby incorporated herein by reference).

As used herein, the term "therapeutically effective amount" means an amount of eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate) as described herein, that is capable of treating breast cancer. The dose of a compound to be administered according to this invention will, of course, be determined in light of the particular circumstances surrounding the case including, for example, the compound administered, the route of administration, condition of the patient, and the pathological condition being treated, for example, the stage of breast cancer.

For administration to a subject, eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), typically is formulated into a pharmaceutical composition comprising eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof, and a pharmaceutically acceptable carrier. Therapeutic compositions typically should be sterile and adequately stable under the conditions of manufacture and storage. Pharmaceutical compositions also can be administered in a combination therapy, *i.e.*, combined with other agents, such as those agents set forth below (see, for example, U.S. Patent No. 6,214,865 and U.S. Patent No. 6,653,341, the entire contents of each of which are hereby incorporated herein by reference).

As used herein, "pharmaceutically acceptable carrier" includes any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like that are physiologically compatible. Preferably, the carrier is suitable for parenteral (*e.g.*, intravenous, intramuscular, subcutaneous, intrathecal) administration (*e.g.*, by injection or infusion). Depending on the route of administration, the active compound may be coated in a material to protect the compound from the action of acids and other natural conditions that may inactivate the compound.

The pharmaceutical compositions may include one or more pharmaceutically acceptable salts, as defined above.

There are numerous types of anti-cancer approaches that can be used in conjunction with eribulin, an analog thereof, or a pharmaceutically acceptable salt

thereof (*e.g.*, eribulin mesylate) treatment, according to the invention. These include, for example, treatment with chemotherapeutic agents, biological agents (*e.g.*, hormonal agents, cytokines (such as interleukins, interferons, granulocyte colony stimulating factor (G-CSF), macrophage colony stimulating factor (M-CSF), and granulocyte macrophage colony stimulating factor (GM-CSF)), chemokines, vaccine antigens, and antibodies), anti-angiogenic agents (*e.g.*, angiostatin and endostatin), radiation, and surgery, as described in more detail in U.S. Pat. No. 6,653,341 B1 and U.S. Publ. No. 2006/0104984 A1, the teachings of which are incorporated herein by reference in their entirety.

The methods of the invention can employ these approaches to treat the same types of cancers as those for which they are known in the art to be used, as well as others, as can be determined by those of skill in this art. Also, these approaches can be carried out according to parameters (*e.g.*, regimens and doses) that are similar to those that are known in the art for their use. However, as is understood in the art, it may be desirable to adjust some of these parameters, due to the additional use of eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), with these approaches. For example, if a drug is normally administered as a sole therapeutic agent, when combined with eribulin, according to the invention, it may be desirable to decrease the dosage of the drug, as can be determined by those of skill in this art. Examples of the methods of the invention, as well as compositions that can be used in these methods, are provided below.

Chemotherapeutic drugs of several different types including, for example, antimetabolites, antibiotics, alkylating agents, plant alkaloids, hormonal agents, anticoagulants, antithrombotics, and other natural products, among others, can be used in conjunction with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof, according to the invention. Specific, non-limiting examples of these classes of drugs, as well as cancers that can be treated by their use, are as follows.

Numerous approaches for administering anti-cancer drugs are known in the art, and can readily be adapted for use in the present invention. In the case that one or more drugs are to be administered in conjunction with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), for example, the drugs can be administered together, in a single composition, or separately, as part of a comprehensive treatment regimen. For systemic administration, the drugs can be administered by, for example, intravenous infusion (continuous or bolus). Appropriate scheduling and dosing of such administration can readily be determined by those of skill

in this art based on, for example, preclinical studies in animals and clinical studies (*e.g.*, phase I studies) in humans.

Many regimens used to administer chemotherapeutic drugs involve, for example, intravenous administration of a drug (or drugs) followed by repetition of this treatment after a period (*e.g.*, 1-4 weeks) during which the patient recovers from any adverse side effects of the treatment. It may be desirable to use both drugs at each administration or, alternatively, to have some (or all) of the treatments include only one drug (or a subset of drugs).

Kits of the Invention

The invention also provides compositions and kits for predicting whether eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), may be used to treat a subject having breast cancer. These kits include reagents for determining the level of expression of at least one, for example, 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10, biomarker(s) selected form the group of biomarkers listed in Table 1 and instructions for use of the kit to predict whether eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), may be used to treat a subject having breast cancer.

The kits of the invention may optionally comprise additional components useful for performing the methods of the invention. By way of example, the kits may comprise reagents for obtaining a biological sample from a subject, a control sample, and/or eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate).

In one embodiment, the reagents for determining the expression level of at least one biomarker in a biological sample from the subject comprises a nucleic acid preparation sufficient to detect expression of a nucleic acid, *e.g.*, mRNA, encoding the biomarker. This nucleic acid preparation includes at least one, and may include more than one, nucleic acid probe or primer, the sequence(s) of which is designed such that the nucleic acid preparation can detect the expression of nucleic acid, *e.g.*, mRNA, encoding the biomarker in the sample from the subject. A preferred nucleic acid preparation includes two or more PCR primers that allow for PCR amplification of a segment of the mRNA encoding the biomarker of interest. In other embodiments, the kit includes a nucleic acid preparation for each of at least 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 biomarkers provided in Table 1.

Alternatively, the reagents for detecting expression levels in the subject of one or more biomarkers predictive of responsiveness to eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), can comprise a reagent that detects the gene product of the nucleic acid encoding the biomarker(s) of interest sufficient to distinguish it from other gene products in a sample from the subject. A non-limiting example of such a reagent is a monoclonal antibody preparation (comprising one or more monoclonal antibodies) sufficient to detect protein expression of at least one biomarker in a sample from the subject, such as a peripheral blood mononuclear cell sample.

The means for determining the expression level of the biomarkers of Table 1 can also include, for example, buffers or other reagents for use in an assay for evaluating expression (*e.g.*, at either the nucleic acid or protein level).

In another embodiment, the kit can further comprise eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof (*e.g.*, eribulin mesylate), for treating breast cancer or another cancer, as described herein in the subject.

Preferably, the kit is designed for use with a human subject.

The present invention is further illustrated by the following examples which should not be construed as further limiting. The contents of all references, patents and published patent applications cited throughout this application, as well as the Figures and the Appendix of sequences provided herein, are expressly incorporated herein by reference in their entirety.

EXAMPLES

EXAMPLE 1: IDENTIFICATION OF RESISTANCE BIOMARKERS FOR TREATMENT WITH ERIBULIN

siRNA techniques were employed to "knock down" expression of certain genes and assess the sensitivity of the resulting knock down cells to eribulin. Based on these studies, the expression of those genes set forth in Table 1 were identified as being associated with the sensitivity of breast cancer cells to treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof.

siRNA Transfection Optimization and Assay Development

Transfection conditions for human breast cancer cell lines MDA-MB-231 (ATCC Catalog No. HTB26™) and BT-549 (ATCC Catalog No. HTB122™) were optimized using transfection reagent DharmaFect 1 from Dharmacon. The MDA-MB-231 and BT-549 cell lines have been reported to be Estrogen Receptor (ER) negative, Progesterone Receptor (PR) negative and HER-2 negative (triple negative). As a non-targeting negative control we used Silencer Negative Control #1 siRNA from Applied Biosystems. The siGENOME TOX (siTOX) Transfection Control (Dharmacon), an RNA duplex designed to induce cell death, was used as a positive control for cell proliferation assays. A reverse transfection procedure in which siRNA was first mixed with transfection reagent and then cells were added to the well, was used in all experiments. Cell viability was compared in cells treated with medium, negative control siRNA and siTOX reagent combined with different amount of DharmaFect 1. Final selected transfection conditions were as follows: MDA-MB-231 cells at 0.035 µl of DharmaFect 1 per well, while BT-549 cells at 0.05 µl of DharmaFect 1 per well. Assays and library screening were performed at 50 nM final concentration of siRNAs. Efficacy of transfection was further confirmed by qPCR with control SMARTpool siRNA reagents targeting PPIA and GAPDH genes (Dharmacon).

High-throughput siRNA Screening

The whole human genome siRNA library was purchased from Dharmacon. The library was diluted to 5 µM. Each well contained 4 SMARTpool siRNA reagents, each directed against a particular gene (four siRNAs targeting the same gene in a single well). More then 18,500 human genes were targeted with this library. 4 µl of each set of siRNAs from the library plates were transferred to 384-well master plates containing 36 µl OPTI-MEM medium per well. 40 µl of diluted DharmaFect 1 reagent were added to each well of the master plate and mixed. 10 µl of siRNA and transfection reagent mixture per well were distributed into five screening plates. After 10 min incubation, cells in 40µl of growth medium were added to each well. Each screening plate included several replicates of negative control siRNA, positive control siTOX as well as medium plus transfection reagent (no siRNA) containing wells. After a 24 hour incubation, 3 screening plates received 10 µl of DMSO diluted in cell growth medium, while 2 repeats were treated with 10 µl of eribulin mesylate (E7389) in growth medium, yielding a final concentration corresponding to the IC_{20} for E7389 for the cell line tested (0.75 nM E7389 for MDA-MB-231; eribulin was provided in a stock solution of DMSO and diluted in growth medium) (see Figure 1). Cell viability was determined at 96 hours after

transfection by CellTiter-Glo luminescent assay from Promega. 10 µl of CellTiter-Glo solution per well were used. Plates were mixed for 2 minutes on a horizontal shaker, incubated for 10 minutes and read on an EnVision® multilabel plate reader from PerkinElmer.

Identification of Primary Hits

Identification of genes with a significant effect on cell sensitivity to E7389 was performed by a method that was similar to the method described for a paclitaxel siRNA screen (Whitehurst *et al.* (2007) *Nature* 446:815-819). Briefly, measurement of each well was normalized by average of medium plus transfection reagent containing reference wells on a plate (32 wells/plate). The biological replicates were averaged for DMSO and E7389-treated plates. For each gene, a two sample t-test was performed to identify significantly different values for wells treated with two different conditions. To narrow down the hit list, the magnitude of response was taken into account by arranging all data according to fold change ratio (average E7389/average DMSO) in ascending order. 364 genes with a fold change among the lowest 5 percentile of the distribution passed the cut-off level. Then, hypothetical open reading frames and genes encoding hypothetical proteins were excluded and the analysis was focused on the 240 remaining genes (see Figure 2).

Confirmation Assays

siRNA SMARTpools for the 240 selected genes were ordered in ON-TARGET*plus* format from Dharmacon. These reagents contain a modified sense strand to prevent interaction with RISC and favor antisense strand uptake. The antisense strand seed region is modified to decrease off-target activity and enhance target specificity. These reagents were used for confirmation secondary screening. To identify common genes influencing sensitivity of cancer cells to E7389, BT-549 breast cancer cells were screened with the 240 selected siRNA pools. The screening was performed using the same protocol as the primary screen with MDA-MB-231 cells, with the final concentration of eribulin mesylate (E7389) in each well corresponding to the IC_{20} for BT-549 cells (0.25 nM E7389).

Data analysis showed that the treatment with 40 out of 240 siRNA pools caused significant differences when comparing E7389-treated wells to carrier-treated wells in both cell lines (Table 2).

To confirm specific down-regulation of 40 genes with siRNA pools, quantitative PCR analysis of targeted mRNAs was performed. MDA-MB-231 and BT-549 cells were transfected with 40 ON-TARGET*plus* siRNAs or non-targeting negative control siRNA according to the above protocol. 48 hours later cells were lysed and cDNAs were synthesized according to the manufacturer's instructions for use of the TaqMan® Gene Expression Cells-to-CT™ Kit (Applied Biosystems). Relative quantities of remaining cDNAs after the treatment with siRNAs were evaluated using QuantiTect SYBR Green PCR Kit with the gene-specific QuantiTect Primer Assays (Qiagen). Results of the analysis are shown in Figure 4. The following 18 genes were down-regulated more than 50 percent in both tested cell lines: CFL1, NMU, MOBKL1B, HYAL2, PSENEN, CYP4F3, ITFG3, EDIL3, YTHDF1, CDC20, CCL26, TMEM79, MANSC1, DYSF, ERGIC3, GRAMD4, NCBP1, SNX11. 14 genes were down-regulated more than 50 percent in at least one cell line (PDGFB, APBB2, SATB1, MAD2L1BP, TUBB6, CEP152, KLH17, COL7A1, CKLF, PHOSPHO2, GNAT1, ABI3, TYROBP, IL10), while other 3 genes were down-regulated more than 35 percent in at least one cell line (ANG, ZIC5, JAM3). Expression of SPTA1, PAPLN, PCDH1, TMIGD2, and KRT24 was either not detectable by this method in MDA-MB-231 cells or didn't change after siRNA treatment in BT-549 cells. The foregoing results indicate that down-regulation of the 40 genes can lead to increased sensitivity to eribulin, an analog thereof, or pharmaceutically acceptable salt thereof.

Table 2. List of 40 overlapping genes from the screening of MDA-MB-231 and BT-549 cells. Fold changes (FC) compared to control and associated p-values are depicted.

gene	Gene ID	MDA-MB-231		BT-549	
		FC	t-test	FC	t-test
PSENEN	55851	0.70	0.03	0.39	0.00
PHOSPHO2	493911	0.56	0.04	0.44	0.02
CCL26	10344	0.71	0.02	0.45	0.03
CDC20	991	0.47	0.03	0.47	0.03
MAD2L1BP	9587	0.47	0.00	0.48	0.01
JAM3	83700	0.63	0.04	0.55	0.00
KLHL17	339451	0.65	0.04	0.57	0.00
PCDH1	5097	0.71	0.02	0.61	0.06
ABI3	51225	0.66	0.04	0.62	0.04
TMIGD2	126259	0.66	0.05	0.62	0.04
NCBP1	4686	0.72	0.04	0.63	0.02
IL10	3586	0.72	0.02	0.64	0.05
ANG	283	0.74	0.02	0.64	0.00
KRT24	192666	0.69	0.01	0.65	0.00
TMEM79	84283	0.66	0.01	0.67	0.02
PDGFB	5155	0.65	0.00	0.68	0.03
SNX11	29916	0.68	0.05	0.68	0.05
CFL1	1072	0.69	0.02	0.68	0.05
CKLF	51192	0.62	0.03	0.70	0.01
TUBB6	84617	0.73	0.01	0.71	0.01
HYAL2	8692	0.67	0.02	0.71	0.05
TYROBP	7305	0.73	0.04	0.71	0.01
APBB2	323	0.70	0.00	0.71	0.02
YTHDF1	54915	0.61	0.02	0.73	0.01
CEP152	22995	0.46	0.03	0.74	0.05
COL7A1	1294	0.73	0.03	0.75	0.05
NMU	10874	0.67	0.00	0.76	0.05
SPTA1	6708	0.38	0.03	0.76	0.02
ERGIC3	51614	0.70	0.05	0.76	0.05
SATB1	6304	0.67	0.02	0.77	0.04
MOBKL1B	55233	0.74	0.03	0.77	0.01
GNAT1	2779	0.66	0.04	0.78	0.03
ITFG3	83986	0.73	0.05	0.78	0.05
DYSF	8291	0.26	0.01	0.78	0.05
MANSC1	54682	0.65	0.03	0.78	0.05
EDIL3	10085	0.65	0.01	0.79	0.05
GRAMD4	23151	0.72	0.00	0.79	0.01
ZIC5	85416	0.69	0.03	0.80	0.05
PAPLN	89932	0.71	0.03	0.80	0.01
CYP4F3	4051	0.74	0.02	0.81	0.01

5

APPENDIX

NUCLEIC ACID AND AMINO ACID SEQUENCES OF BIOMARKERS

Gener Name	Gene ID	Accession NO.	Name
ABI3	51225	NM_016428	Homo sapiens ABI family, member 3 (ABI3), transcript variant 1, mRNA

mRNA Sequence

TCCTATCCACCCTCCACTCCCCTGTCCCTTGGTGACTCATCCCTGAGCTTCCCAAGGAAGCCCCCACCCT
CTGCCCTTTCCTCCCGCCTTCCATGAGTGGAAAATCCACCTCCGCCCCCTATAGCAGGCCAGCCCCCTTC
CTCCCCAGTCTCCGACCCCATCCCCCAGCCGACCAGTTTCCTCTCCAGGACCAGGGAGCAATCACAGCTG
CCCCGACCTTGGCTTCCTCTGCTGGGTGGGATTGGGGGCTGGGCCCCCAAATGGGCCCCTGGCTTCCCCC
TTCCTCTGGGCAGGGGACAGAGAGACACAGGCTCGGGGAGCAGGACTGACTTCCTCTTGTCCCGGAATGA
GCATGCCTGCCCTTTGCAAGCAGGTTTGGGTCTCACGCAGAGGAAACCAAAAGCAATAAGAGGGAGGGAA
GGCAGAGCAACCAATCAAGGGCAGGGTGAGACTCAAAACGAGCGGGCTCCCTGGGGAGCCAGACAGAGGC
TGGGGGTGATGGCGGAGCTACAGCAGCTGCAGGAGTTTGAGATCCCCACTGGCCGGGAGGCTCTGAGGGG
CAACCACAGTGCCCTGCTGCGGGTCGCTGACTACTGCGAGGACAACTATGTGCAGGCCACAGACAAGCGG
AAGGCGCTGGAGGAGACCATGGCCTTCACTACCCAGGCACTGGCCAGCGTGGCCTACCAGGTGGGCAACC
TGGCCGGGCACACTCTGCGCATGTTGGACCTGCAGGGGGCCGCCCTGCGGCAGGTGGAAGCCCGTGTAAG
CACGCTGGGCCAGATGGTGAACATGCATATGGAGAAGGTGGCCCGAAGGGAGATCGGCACCTTAGCCACT
GTCCAGCGGCTGCCCCCCGGCCAGAAGGTCATCGCCCCAGAGAACCTACCCCCTCTCACGCCCTACTGCA
GGAGACCCCTCAACTTTGGCTGCCTGGACGACATTGGCCATGGGATCAAGGACCTCAGCACGCAGCTGTC
AAGAACAGGCACCCTGTCTCGAAAGAGCATCAAGGCCCCTGCCACACCCGCCTCCGCCACCTTGGGGAGA
CCACCCCGGATTCCCGAGCCAGTGCACCTGCCGGTGGTGCCCGACGGCAGACTCTCCGCCGCCTCCTCTG
CGTCTTCCCTGGCCTCGGCCGGCAGCGCCGAAGGTGTCGGTGGGGCCCCCACGCCCAAGGGGCAGGCAGC
ACCTCCAGCCCCACCTCTCCCCAGCTCCTTGGACCCACCTCCTCCACCAGCAGCCGTCGAGGTGTTCCAG
CGGCCTCCCACGCTGGAGGAGTTGTCCCCACCCCCACCGGACGAAGAGCTGCCCCTGCCACTGGACCTGC
CTCCTCCTCCACCCCTGGATGGAGATGAATTGGGGCTGCCTCCACCCCCACCAGGATTTGGGCCTGATGA
GCCCAGCTGGGTGCCTGCCTCATACTTGGAGAAAGTGGTGACACTGTACCCATACACCAGCCAGAAGGAC
AATGAGCTCTCCTTCTCTGAGGGCACTGTCATCTGTGTCACTCGCCGCTACTCCGATGGCTGGTGCGAGG
GCGTCAGCTCAGAGGGGACTGGATTCTTCCCTGGGAACTATGTGGAGCCCAGCTGCTGACAGCCCAGGGC
TCTCTGGGCAGCTGATGTCTGCACTGAGTGGGTTTCATGAGCCCCAAGCCAAAACCAGCTCCAGTCACAG
CTGGACTGGGTCTGCCCACCTCTTGGGCTGTGAGCTGTGTTCTGTCCTTCCTCCCATCGGAGGGAGAAGG
GGTCCTGGGGAGAGAGAATTTATCCAGAGGCCTGCTGCAGATGGGGAAGAGCTGGAAACCAAGAAGTTTG
TCAACAGAGGACCCCTACTCCATGCAGGACAGGGTCTCCTGCTGCAAGTCCCAACTTTGAATAAAACAGA
TGATGTCCTGTGACTGCCCCACAGAGATAAGGGGCCAGGAGGGATTGAAAGGCATCCCAGTTCTAAGGCT
GCTGCTAATTACAGCCCCCAACCTCCAACCCACCAGCTGACCTAGAAGCAGCATCTTCCCATTTCCTCAG
TACCCACAAAGTGCAGCCCACATTGGACCCCAGACACCCCTCTGCAGCCATTGACTGCAACTTGTTCTTT
TGCCCATTGAAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 1)

Translated protein sequence

MAELQQLQEFEIPTGREALRGNHSALLRVADYCEDNYVQATDKRKALEETMAFTTQALASVAYQVGNLAG
HTLRMLDLQGAALRQVEARVSTLGQMVNMHMEKVARREIGTLATVQRLPPGQKVIAPENLPPLTPYCRRP
LNFGCLDDIGHGIKDLSTQLSRTGTLSRKSIKAPATPASATLGRPPRIPEPVHLPVVPDGRLSAASSASS
LASAGSAEGVGGAPTPKGQAAPPAPPLPSSLDPPPPPAAVEVFQRPPTLEELSPPPPDEELPLPLDLPPP
PPLDGDELGLPPPPPGFGPDEPSWVPASYLEKVVTLYPYTSQKDNELSFSEGTVICVTRRYSDGWCEGVS
SEGTGFFPGNYVEPSC (SEQ ID NO: 2)

ANG	283	NM_001097577	Homo sapiens angiogenin, ribonuclease, RNase A family, 5 (ANG), transcript variant 2, mRNA

mRNA Sequence

TCCAGGTTCACACAACTGGAACCCATCTCCAGGAACAAACAGCTGGAACCCATCTCCCGTTGAAGGGAAA
CTGCCAGATTTTTGTAAGATTCTTCCTCCTGGGAGCCTGTGTTGGAAGAGATGGTGATGGGCCTGGGCGT

TTTGTTGTTGGTCTTCGTGCTGGGTCTGGGTCTGACCCCACCGACCCTGGCTCAGGATAACTCCAGGTAC
ACACACTTCCTGACCCAGCACTATGATGCCAAACCACAGGGCCGGGATGACAGATACTGTGAAAGCATCA
TGAGGAGACGGGGCCTGACCTCACCCTGCAAAGACATCAACACATTTATTCATGGCAACAAGCGCAGCAT
CAAGGCCATCTGTGAAAACAAGAATGGAAACCCTCACAGAGAAAACCTAAGAATAAGCAAGTCTTCTTTC
CAGGTCACCACTTGCAAGCTACATGGAGGTTCCCCCTGGCCTCCATGCCAGTACCGAGCCACAGCGGGGT
TCAGAAACGTTGTTGTTGCTTGTGAAAATGGCTTACCTGTCCACTTGGATCAGTCAATTTTCCGTCGTCC
GTAACCAGCGGGCCCCTGGTCAAGTGCTGGCTCTGCTGTCCTTGCCTTCCATTTCCCCTCTGCACCCAGA
ACAGTGGTGGCAACATTCATTGCCAAGGGCCCAAAGAAAGAGCTACCTGGACCTTTTGTTTTCTGTTTGA
CAACATGTTTAATAAATAAAAATGTCTTGATATCAGTAAGAA (SEQ ID NO: 3)

Translated protein sequence

MVMGLGVLLLVFVLGLGLTPPTLAQDNSRYTHFLTQHYDAKPQG
RDDRYCESIMRRRGLTSPCKDINTFIHGNKRSIKAICENKNGNPHRENLRISKSSFQV
TTCKLHGGSPWPPCQYRATAGFRNVVVACENGLPVHLDQSIFRRP (SEQ ID NO: 4)

APBB2	323	NM_173075	Homo sapiens amyloid beta (A4) precursor protein-binding, family B, member 2 (APBB2), mRNA

mRNA Sequence

GCCAAAGCCTGGAGAAGTGGAATCTCGTCAGCGCCGCTCCCTGCGCGGGACTCGCGGAACGGCACTGAGC
ATGCTCAGTTGCCGGAGCCCGTTCTGGTCTCAAGTAGGAAGCTAGTGCGCTGTAACCGCATCTGATCTGG
GCGCTCCGGGAAGGGCGAGACTGGAGCAGAGCCGCTGGGCGCCGGAGCCGAGGCGAGCGCCGCGCGCACC
ACTGGTTGGAGTTGCTGTGGGTGAGCTGCTGTGGTCTGTAGCCAAGCATGCTGTGGTCGGATCTGCCCAG
CCGTGGAACAGAAACATTTGCTGGATGGAAAATCCATAAAAGAAAGCTCCTGTGAAAAGCTGAGGCTGAC
AATAATTTAAGCAAAATCAGATCGATCTCTTTGGGCTGCCTGACCTCCTTGGGTGCTTGCTATTAATTAA
CAGACTTTGTGGGGAAAAAAAGGAGCTTGCCTTCTGAGCTTTGTACCAAAGACCTGGGAAAACTAACCAT
CTCAGTCTTTCCTGAGGACTTGGGAACTGCCGAGGCCTCTGCCAATGTGTTGACTGTCGCTATGGGCTCA
CTGTTGTCCAGGCAGCTCATATTTCAAATTATAACCTATTTCCTGCACCATTGCTGACGCCTGGTGATCC
ATGTCAGAAGTACTTCCAGCTGACTCAGGTGTTGACACCTTGGCAGTGTTTATGGCCAGCAGCGGAACTA
CAGACGTCACAAATCGGAACAGCCCAGCCACACCACCAAACACCCTTAACCTCCGATCCTCCCACAATGA
ACTGTTGAACGCTGAAATAAAACACACAGAAACCAAGAACAGCACACCTCCCAAATGCAGGAAAAAATAT
GCACTAACTAACATCCAGGCGGCCATGGGCCTCTCGGATCCAGCTGCACAGCCCCTGCTGGGAAATGGCT
CTGCCAACATCAAGCTGGTGAAAAATGGGGAGAACCAGCTCCGTAAGGCTGCAGAGCAAGGGCAGCAGGA
CCCCAACAAAAACCTGAGCCCCACTGCAGTCATCAACATAACTTCTGAGAAGTTAGAGGGTAAAGAGCCC
CACCCACAGGATTCCTCGAGCTGTGAGATTTTACCCTCCCAGCCCAGGAGAACTAAGAGCTTCCTAAATT
ACTATGCAGATCTGGAAACCTCAGCCAGAGAACTAGAGCAGAACCGAGGCAATCACCATGGGACTGCGGA
AGAGAAATCCCAGCCAGTCCAGGGCCAGGCCTCCACCATCATTGGGAATGGCGATTTGCTGCTGCAGAAA
CCAAACAGACCCCAGTCCAGCCCTGAAGACGGCCAAGTAGCCACAGTGTCATCCAGCCCAGAAACCAAGA
AGGATCATCCGAAAACAGGGGCCAAAACCGACTGTGCACTGCACCGGATCCAGAACCTGGCACCGAGCGA
TGAGGAGTCCAGCTGGACAACGTTGTCCCAAGACAGTGCCTCACCCAGCTCCCCGGATGAAACAGATATA
TGGAGTGATCACTCATTTCAGACTGATCCAGATTTGCCGCCTGGCTGGAAAAGAGTCAGTGACATTGCCG
GGACCTATTATTGGCACATCCCAACAGGAACGACTCAGTGGGAACGGCCCGTCTCCATCCCAGCAGATCT
CCAGGGTTCTAGGAAAGGGTCACTTAGTTCTGTAACGCCATCTCCCACCCCAGAGAACGAGGATTTGCAT
GCAGCCACTGTTAACCCGGACCCCAGTTTAAAAGAGTTTGAAGGAGCAACCCTACGCTATGCATCTTTGA
AACTCAGAAATGCCCCACACCCTGATGATGATGATTCTTGTAGTATCAACAGTGACCCAGAAGCCAAGTG
TTTTGCTGTGCGTTCTCTGGGATGGGTAGAGATGGCAGAAGAGGACCTCGCCCCCGGTAAAAGTAGTGTT
GCGGTCAACAACTGCATCAGGCAACTTTCCTACTGCAAAAATGACATCCGAGACACAGTCGGGATTTGGG
GAGAGGGGAAAGACATGTACCTGATCCTGGAGAATGACATGCTCAGCCTGGTGGACCCCATGGACCGCAG
CGTGCTGCACTCGCAGCCCATCGTCAGCATCCGCGTGTGGGGCGTGGGCCGCGACAATGGCCGGGATTTT
GCTTATGTAGCAAGAGATAAAGATACAAGAATTTTGAAATGTCATGTATTTCGATGTGACACACCAGCAA
AAGCCATTGCCACAAGTCTCCACGAGATCTGCTCCAAGATTATGGCTGAACGGAAGAATGCCAAAGCGCT
GGCCTGCAGCTCCTTACAGGAAAGGGCCAATGTGAACCTCGATGTCCCTTTGCAAGATTTTCCAACACCA
AAGACTGAGCTGGTCCAGAAGTTCCACGTGCAGTACTTGGGCATGTTACCTGTAGACAAACCAGTCGGAA
TGGATATTTTGAACAGTGCCATAGAAAATCTTATGACCTCATCCAACAAGGAGGACTGGCTGTCAGTGAA
CATGAACGTGGCTGATGCCACTGTGACTGTCATCAGTGAAAAGAATGAAGAGGAAGTCTTAGTGGAATGT
CGTGTGCGATTCCTGTCCTTCATGGGTGTTGGGAAGGACGTCCACACATTTGCCTTCATCATGGACACGG
GGAACCAGCGCTTTGAGTGCCACGTTTTCTGGTGCGAGCCTAATGCTGGTAACGTGTCTGAGGCGGTGCA
GGCCGCCTGCATGTTACGATATCAGAAGTGCTTGGTAGCCAGGCCGCCTTCTCAGAAAGTTCGACCACCT
CCACCGCCAGCAGATTCAGTAACCAGAAGAGTCACAACCAATGTAAAACGAGGGGTCTTATCCCTCATTG
ACACTTTGAAACAGAAACGCCCTGTCACCGAAATGCCATAGCTGCACATGCAAAAGGACTCGGCTATTTA

CCTGAAGATTGACTAGCTACACTAAAGAAAATGAACTCCGCCATCCGACCTTCCATCCAGTTGCTGATGC
TTTGTCTTCAGAGAATTTACCCTTAACCAAGCAGTGTTAGACAAGCATGTTCTCTCGTCTTGCCACCATC
ATGTGATATGAAAAGAAGCATGAATAATTTTTTTTGCTGTAAGTTACATCATGCGCAGTGGAAGGTCTTT
TTCTTATTGTAAATATTGTGAACATTACTTAACTTCACACACACACAGAGAAGAGTGTGGCCCCACCCCT
CCTAGTGAACTAACGCTGCGTCCTTGGAATGAATGATGCGTGAGTTAGTTTCACTGTCTTCTTGGCTGGA
CCTGTCACAAGCAACCTTTAAGTCCTACAGCACTTTGCCCTGTTTTCAACATTGGAGTAGGCACTGCATA
GCAGATACCATTGAATTGCTGTAAAAATAGGATGGCGAGTTTGTGTTTTAATTTTTCATAAAATTGAACC
TGTTGGTTGACAAAATTGGCTGTTGGCATCAGTATAGAAACCAACTGGCAGCTTTCCCTGACAAGCTCTT
TGACACATGGACACCATTTCATGTCTACAGCTGTTTGTGGGATGTTGGAAAAAAATGAAACTTCAAAATT
GATGAAAAACTAAATTCGAGGAATTAAAATCGAACAAAACATAGCCTTTCTTTTCCGATGGTTTTCAAAC
TGATTATTTTTAAAAGAGATTAATAAAATCATAATGCATTTTGGGTGGGACATATTTCAAACTTCTGCCT
TATATTGTACGGTGCAGCTAGAGAATTATAGTTCACTATGGCCATTCTCTACATAAACATTAAGATGAAA
TACTCCTCATCAGCCTTTCATCCTTAGTTTGAGAATTAGCTGATATGCAATTTGAAGTTGAGGAAATATC
ATTGATATTTCTATCATGCACGATTATTTTAGATTTCTACCACCGTGTGATTTTTGCTAGTCCATGTGCT
AGAGGTAAACGTTCTGCTGGAATTCTGCATCCAGCTCTATCCCCCTCTGATGCTTTTTGCCCAGAAAGCT
GTCTGTCCATCATGTATTGTCCATGGCAACAAATTACATTAGGTTGAACCTTTCCTTGATTTTATGTATT
TAATATTAGAATTTGTTGGACTCAACTAGATATATTTTTTAATTTATATTTTTTCCATTTTACTTTGAAG
ATTTGAAATGTTCATACCTGAGCAAAGTCTACACAGGAGTAATGGACTGTTTAACAAGTTTCCCAAAACA
GCATTTTCCTGCTCCTTCGTATGTAGGTGAGAAACTTAGCTGGAAAGACATACAAATTTAGACTCTCGTT
GACATTGTCGTTTTAAAAGGAAGTTGCTAAGGCGATCAATCTCAATATTAGTCTTGTTTACTTCTTCTTA
ATGTCAAAATTAACATTTACAACATCCAATTATAAAAGTAATGCTTTATGTTTATACACTGCTATGTACT
TGTCAAAATGGTTTCCACATTCTTATCACATCTGAGCCTTACCAGGTAGAGAAGGTACTAAATACACTTT
AGAAGTAAAAATATGAAGTACCGAGAGGCTAAACCCACTGGCCTAAGATCTCACCAAAGTTCATGAAAAC
CAGGACTAGGACCCACGGCTCCCAAAGCCCGTTCTTGCTGTGTTGTGCTGCCTCCATATCCGTCAGGAAG
AGCCTTTCCAGAATGATTCTGGGCATATACTAAGAAGAGCAGGTATGGAAAGATCTATTGTCAGGGAATC
TTAGAATTCCCTACACGAGTGGGAGAAAGATGTCCAAATTCCTTACGCAGTGGTATTCATGATGGTGCCC
TATCTAAGTCCAGGACTGTTTTCCTACAGCGTGCCTCAAAAGTGTTGTAGAGGGCAGGATTCTACATTCA
CAGCCTGTTCCATCTACGAGATTTTCCAGATGCTACTTGTGGTAGACATTCCTAACTCATGGTACTTAGC
CACCAGAGATCATGATGGAATGAGTGGGTGGCTTTTCTACCTGCCATTCCCTCAGAATTCATGAGGGGTG
GGGGACAGGGGGACCGGAATTGTCTTAGCACCCCAATGTTATGACAAAACTATGCTACTTTAGAAACGCA
GTCTGTTTTTCACCAATTGACATACTACTGATCTGAAGTAACCAGTGCCATCATAAGAAATTACTGCATT
AAGAAAATCCTTGCTGTGCCCTTTGAAAAGCTGTTCAGAAATCATTTACAGTGATCTTTCATCTCGGTCG
CTGTAGTGAAACATTTTAGTGTGATAAATTTCAAAATTCTAAACAAATTACCCACTTTTATATTGGAAAT
CTCTACCAGAACTCCCTCTTCATTTTTTAAGGCATACATTTGCTTGTTTTCAAGATCAAGAATTCTGAGC
TAGCTTTAAGTAGCAAACTGATTTATATGTGCAATTATAGGATGCATTAAGATGAATGATAGCCTTTACA
TATTGAAAACTTTGCAGACGTTTTGTTTTGAAAATGGCATTGTATAGTAAATGCAAATTAATTTTGTAAA
ATTATGTTAAAGAGTATGTTCAGACACTTTCTGCCATGGCCAAAAAGTATGTATGAAAGTATGTGTGTAT
TTGTTTGTAAAAGGATGCCAATGTTTTACCTGATATCTTAGTGACACTTCAGTTATCTATGCATTCTTTA
GATCTGTGATTCGGTAAACAGGCAGCCATGTTCACGATGCCTTCTATGTCTTACCATATTTTTAATTAAC
CTGTTAAATACAGCTTAAAATATTTTTATTTTATTTATTCTATTTTTACTGAAATATACTGCATTATTGT
GTTAATGTATTATCTTTCCTGGATATTATCTCCCAGTGTATCCAGATCTAAGTAATCTCAGTGAACTATA
CATTGCCTAAAAAGTGGTTTTGTAATGATTTGTAGTCACATTTCTATTGGGATATGTAGAAGAAAAGGCA
AAATGCTTAAAGTTCCTTTTATTTTTTAAAAGCAGCTAGATAGACACAGACTTGCCACCTCATACATCTG
CTCCTTGGCAACATCAAGGGGAACGACTAGCCAACATGCCTATGGCTAAAAACTTTCCTTTGCAGACTAA
AGCACTGCTTGGTGCTTCGTTTTTCTACCCTTCACAACATGTGTGATTTCATCTAAGAGATATATACATG
TACACATGCCCTTTGTTTCCACCTGGATACAAGATCACTCATAGCTAATTAGGACCATTGTTTTTTGTTC
ATCTGTCTTGTTGCATGAAGGGACATTAGACCCATTTCCATTAAAATAAGTTCTTGGTGATAAACTGTGG
CACTGCTACTTCTTTTTAAATCCACTTTATGATTTCAAGATGGACACTTGTAAGATGACTCGACACAAGG
CCATTGCCTGGAAGCCCCAGAGCTTTCCTCTGTTTGTATGGCCCGTTCATGTCCCAGGCATTGCAACACA
AACTCCTCAAGATTTCACCACAACATGACAAGCATTTTCCTAACTGATATTAGCACAATTTAACTAATAA
GCCCCTTCGCTCTCTAGTTGGCCAGGCTTAACCTAATACACATCTAACGTGTGTGCCACACGGCCAGTAG
AAAGTTTAACTTCAGCTTCAGGGCAAAGATACCCACTCACACCGTGTCAACGCAAGCAGTAGTTCCTGGC
CTCCAGAGCAGCTTACTTCCCCTGAAAGAACGCTTTGTTTTCCTTTATGCCCTTTTCCTGTTGACCACTT
TTACACATTTAAATGTAATTTGTTGTGAGAATAAATTTAGCTGCATAAAACGTTCGGCTCATTTATCTGA
CATCTTAGTCACATATACAAGGAATAGAAATAGAAACTCGGTGTCTCTAGTTATTTTTAAATTATTCTTA
CCTCAGACTTCTTAGAAATCACTTTAGTAATGGAGCATTTTGCTTTGATTAGTTACTACATATTTCTGCC
TGGTAAGAACTAGGAAGTAACTTCAAATTTTGAGTAATCACCCTGTACTTATTTGGTGATCAGGAAGGCC
AGCTGGCCTTCCGGACATAGAAGCCTATTTAGTCACCAACTCGAGTCTTTTGTAAGCGGTCTTGCTAGGA
TTGTGATATTTTAGCACGAAGAAGTTTATCACTTCCTTTAAGAACCTGACATCAAAGAATAAAGAATAGA
GGTGTACACACACTAAATCCAAAATGAAAGGTAACTAGAGAAATCAGTTGAATCTGGTTTAGCTTAACTG
TTAGGCGCAGGAAGGCAGATAAACAGAATTTAAAGTATGTCCCCGCTTTTTGTTCATCTTGCACTTCCAC

AGTGGTTTCTCTCTAGTCAGTAACAAAATTTCATTATGGTTTCAGGCATTATATGGTGGTAAATAATTTC
AGATTAAAAATGTGTTTGCTATTGGAGTATCTGAATACTAGTAATTTCATTATTTAGAATTTTGCAGCAC
TTTTATCTCAAGAAGAAGTCCAAGAATGTAAAATGCCAAATGAAACATGTCAGTGGAATCAATATTCTCC
TTCATTAGAATTCCCTCATATTGCTTTTTTTTTTTTTCTTCAGACAGAGGAGTCTTACTCTGGAGTGCAG
TGGTGTAATTTCAGCTCACCACAACCTCCACCTCCCAAGTTCAAGCAATTCTCGCCTCAGCCTCCTGAGT
AGCTGGGATTACAGGCATGCACCGCCACGCCTGGCCAATTTGTATATTTTTAGTAGAGACAGGGTTTCGC
CACGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCCGCCCGCCCCAACCTCCCAAAGTATGT
GAGCCACCACGTCCGGCCTCATATTGCTTTTATCCAAAATTCTTTTCCCTTTTCACTCTACCAAAGTATT
TAAATAATCCTGTCCTTCATAGAAGATTCTCAAAGAAGAAAACTGCAGTGTAATTAATGAATGGTTTAAT
TCAGAATCTTCATATACTTCTAAAGAGAAAAATAATTTAGTGCCAAATGCATGTTAGGAGATAATCAATG
TAAGTGGCAACAAATTGTGACTTCACATGCTACTGTAGAGATCAGAAAATTATCCTAAACTATTCCATAA
CAATGAGACAACATCACAGAAAATACACTTGAAAATAAAAATCTCAAGACCAGCTACTTCTGGACAATGG
AATACTTTTCAGTCTGGTATGGTGGAGGGCCCGAAAAGGATAAGGGATTCTTATGATACACAATGGGATT
CTTTACTGAACAATATGTTAAATTAAGCTGCACCGCCTTCCTTGAGGCATGGACTACCCTAACCAACCAG
ATAGAAATCTGGGTGGGATAAGAGGATGAGCCACACGCTATAATTTTAGGGCAAGGAGATAGTGTTTGAT
TTTCAAAATCAGCAAAATAAGCTGAGCACTTTATATCTTTCTGTACAAGAGTGATAACATGAAGAATTCT
TCTTCAGGGATTTAAAATACAATAAGCCTGGTTCAACTATAAAAAGTCTTGTTTCCTTTCTTCATTGACA
CTTTTTTTTTTTTTTTTTTTTTTTTGAGGCAAGGTCTCACTCTGCTTCCCAGGCTGGAGTGCAGTGGGGC
AATATTGGCTCACTGCAACCTGCACCTCCTGGACTCAAGAGATCCTCGTACCTCAGCCTCCTAAGTAGCT
GGGACTACAGGCGTGTCCCACCACACCCAGCTAATTTTTGTATTTTTTGTAGAGATGGGGTTTTGGGGTT
TCGCCATGTTGTCCAGGCTCGTCTGGAACTCCGGTGCTCAAGTGGCGTGCCCACCTCAGCCTCCCAAACT
GCTGAGATTACAGATGTGAGCCACTGCACCCAGCCCACTGACACGTTTTACTGATAAATGTAAATCTAAG
CTAAAATAAAAATAATGTATTACCGCTATAATACAATTCACCATTCTCTTTTCTCACTTCAAGTAAGAAA
GTAAAAATAGAATATCAGAGCTGAAGTAGACCTAAGTATTCATCTTGAAGAAGATAATATTCTAAAAATC
ATGCCACCTGAATTGAGCATTTAGGAATTTATGTAACATTTCTATACAACTGAATTGCAAAAATAAAACT
TTAAATTCAAACTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 5)

Translated protein sequence

MSEVLPADSGVDTLAVFMASSGTTDVTNRNSPATPPNTLNLRSS
HNELLNAEIKHTETKNSTPPKCRKKYALTNIQAAMGLSDPAAQPLLGNGSANIKLVKN
GENQLRKAAEQGQQDPNKNLSPTAVINITSEKLEGKEPHPQDSSSCEILPSQPRRTKS
FLNYYADLETSARELEQNRGNHHGTAEEKSQPVQGQASTIIGNGDLLLQKPNRPQSSP
EDGQVATVSSSPETKKDHPKTGAKTDCALHRIQNLAPSDEESSWTTLSQDSASPSSPD
ETDIWSDHSFQTDPDLPPGWKRVSDIAGTYYWHIPTGTTQWERPVSIPADLQGSRKGS
LSSVTPSPTPENEDLHAATVNPDPSLKEFEGATLRYASLKLRNAPHPDDDDSCSINSD
PEAKCFAVRSLGWVEMAEEDLAPGKSSVAVNNCIRQLSYCKNDIRDTVGIWGEGKDMY
LILENDMLSLVDPMDRSVLHSQPIVSIRVWGVGRDNGRDFAYVARDKDTRILKCHVFR
CDTPAKAIATSLHEICSKIMAERKNAKALACSSLQERANVNLDVPLQDFPTPKTELVQ
KFHVQYLGMLPVDKPVGMDILNSAIENLMTSSNKEDWLSVNMNVADATVTVISEKNEE
EVLVECRVRFLSFMGVGKDVHTFAFIMDTGNQRFECHVFWCEPNAGNVSEAVQAACML
RYQKCLVARPPSQKVRPPPPPADSVTRRVTTNVKRGVLSLIDILKQKRPVTEMP (SEQ ID NO: 6)

CCL26	10344	NM_006072	Homo sapiens chemokine (C-C motif) ligand 26 (CCL26), mRNA

mRNA Sequence

CTGGAATTGAGGCTGAGCCAAAGACCCCAGGGCCGTCTCAGTCTCATAAAAGGGGATCAGGCAGGAGGAG
TTTGGGAGAAACCTGAGAAGGGCCTGATTTGCAGCATCATGATGGGCCTCTCCTTGGCCTCTGCTGTGCT
CCTGGCCTCCCTCCTGAGTCTCCACCTTGGAACTGCCACACGTGGGAGTGACATATCCAAGACCTGCTGC
TTCCAATACAGCCACAAGCCCCTTCCCTGGACCTGGGTGCGAAGCTATGAATTCACCAGTAACAGCTGCT
CCCAGCGGGCTGTGATATTCACTACCAAAAGAGGCAAGAAAGTCTGTACCCATCCAAGGAAAAAATGGGT
GCAAAAATACATTTCTTTACTGAAAACTCCGAAACAATTGTGACTCAGCTGAATTTTCATCCGAGGACGC
TTGGACCCCGCTCTTGGCTCTGCAGCCCTCTGGGGAGCCTGCGGAATCTTTTCTGAAGGCTACATGGACC
CGCTGGGGAGGAGAGGGTGTTTCCTCCCAGAGTTACTTTAATAAAGGTTGTTCATAGAGTTGACTTGTTC
AT (SEQ ID NO: 7)

Translated protein sequence

MMGLSLASAVLLASLLSLHLGTATRGSDISKTCCFQYSHKPLPW
TWVRSYEFTSNSCSQRAVIFTTKRGKKVCTHPRKKWVQKYISLLKTPKQL (SEQ ID NO: 8)

CDC20	991	NM_001255	Homo sapiens cell division cycle 20 homolog

			(S. cerevisiae) (CDC20), mRNA

mRNA Sequence

GAGGCGTAAGCCAGGCGTGTTAAAGCCGGTCGGAACTGCTCCGGAGGGCACGGGCTCCGTAGGCACCAAC
TGCAAGGACCCCTCCCCCTGCGGGCGCTCCCATGGCACAGTTCGCGTTCGAGAGTGACCTGCACTCGCTG
CTTCAGCTGGATGCACCCATCCCCAATGCACCCCCTGCGCGCTGGCAGCGCAAAGCCAAGGAAGCCGCAG
GCCCGGCCCCCTCACCCATGCGGGCCGCCAACCGATCCCACAGCGCCGGCAGGACTCCGGGCCGAACTCC
TGGCAAATCCAGTTCCAAGGTTCAGACCACTCCTAGCAAACCTGGCGGTGACCGCTATATCCCCCATCGC
AGTGCTGCCCAGATGGAGGTGGCCAGCTTCCTCCTGAGCAAGGAGAACCAGCCTGAAAACAGCCAGACGC
CCACCAAGAAGGAACATCAGAAAGCCTGGGCTTTGAACCTGAACGGTTTTGATGTAGAGGAAGCCAAGAT
CCTTCGGCTCAGTGGAAAACCACAAAATGCGCCAGAGGGTTATCAGAACAGACTGAAAGTACTCTACAGC
CAAAAGGCCACTCCTGGCTCCAGCCGGAAGACCTGCCGTTACATTCCTTCCCTGCCAGACCGTATCCTGG
ATGCGCCTGAAATCCGAAATGACTATTACCTGAACCTTGTGGATTGGAGTTCTGGGAATGTACTGGCCGT
GGCACTGGACAACAGTGTGTACCTGTGGAGTGCAAGCTCTGGTGACATCCTGCAGCTTTTGCAAATGGAG
CAGCCTGGGGAATATATATCCTCTGTGGCCTGGATCAAAGAGGGCAACTACTTGGCTGTGGGCACCAGCA
GTGCTGAGGTGCAGCTATGGGATGTGCAGCAGCAGAAACGGCTTCGAAATATGACCAGTCACTCTGCCCG
AGTGGGCTCCCTAAGCTGGAACAGCTATATCCTGTCCAGTGGTTCACGTTCTGGCCACATCCACCACCAT
GATGTTCGGGTAGCAGAACACCATGTGGCCACACTGAGTGGCCACAGCCAGGAAGTGTGTGGGCTGCGCT
GGGCCCCAGATGGACGACATTTGGCCAGTGGTGGTAATGATAACTTGGTCAATGTGTGGCCTAGTGCTCC
TGGAGAGGGTGGCTGGGTTCCTCTGCAGACATTCACCCAGCATCAAGGGGCTGTCAAGGCCGTAGCATGG
TGTCCCTGGCAGTCCAATGTCCTGGCAACAGGAGGGGGCACCAGTGATCGACACATTCGCATCTGGAATG
TGTGCTCTGGGGCCTGTCTGAGTGCCGTGGATGCCCATTCCCAGGTGTGCTCCATCCTCTGGTCTCCCCA
TTACAAGGAGCTCATCTCAGGCCATGGCTTTGCACAGAACCAGCTAGTTATTTGGAAGTACCCAACCATG
GCCAAGGTGGCTGAACTCAAAGGTCACACATCCCGGGTCCTGAGTCTGACCATGAGCCCAGATGGGGCCA
CAGTGGCATCCGCAGCAGCAGATGAGACCCTGAGGCTATGGCGCTGTTTTGAGTTGGACCCTGCGCGGCG
GCGGGAGCGGGAGAAGGCCAGTGCAGCCAAAAGCAGCCTCATCCACCAAGGCATCCGCTGAAGACCAACC
CATCACCTCAGTTGTTTTTTATTTTTCTAATAAAGTCATGTCTCCCTTCATGTTTTTTTTTTAAAAAAAA
AAAAAAAAAAAAAAAA (SEQ ID NO: 9)

Translated protein sequence

MAQFAFESDLHSLLQLDAPIPNAPPARWQRKAKEAAGPAPSPMR
AANRSHSAGRTPGRTPGKSSSKVQTTPSKPGGDRYIPHRSAAQMEVASFLLSKENQPE
NSQTPTKKEHQKAWALNLNGFDVEEAKILRLSGKPQNAPEGYQNRLKVLYSQKATPGS
SRKTCRYIPSLPDRILDAPEIRNDYYLNLVDWSSGNVLAVALDNSVYLWSASSGDILQ
LLQMEQPGEYISSVAWIKEGNYLAVGTSSAEVQLWDVQQQKRLRNMTSHSARVGSLSW
NSYILSSGSRSGHIHHHDVRVAEHHVATLSGHSQEVCGLRWAPDGRHLASGGNDNLVN
VWPSAPGEGGWVPLQTFTQHQGAVKAVAWCPWQSNVLATGGGTSDRHIRIWNVCSGAC
LSAVDAHSQVCSILWSPHYKELISGHGFAQNQLVIWKYPTMAKVAELKGHTSRVLSLT
MSPDGATVASAAADETLRLWRCFELDPARRREREKASAAKSSLIHQGIR (SEQ ID NO: 10)

CEP152	22995	NM_014985	Homo sapiens centrosomal protein 152kDa (CEP152), mRNA

mRNA Sequence

GCCCACCGGGCGAGCTTCTAGTCGGCGATTGAAGGATGCGAGTGCTCCTTAAGGGCCTCCGCCCCGTGAG
TTCGGTTGTGACTAGGAAGGAGCTAGTGGACTAGAGCCAGGGTAAGGGGATCTGCTAGAAGTTGGTCTTC
CGCCAGGACTAGAGTTTCCTCGCGGTAACAGCCTCCGTGGCCTCCGGAGGACCATGTCATTAGACTTTGG
CAGTGTGGCACTACCAGTGCAAAATGAAGATGAAGAGTATGACGAAGAGGACTATGAAAGAGAGAAAGAG
TTGCAGCAGTTACTCACAGACCTTCCCCATGACATGCTGGATGACGACCTCTCCTCTCCAGAGCTCCAGT
ATTCGGACTGCAGCGAGGATGGCACAGACGGACAACCACATCATCCTGAGCAATTGGAGATGAGCTGGAA
TGAGCAAATGCTGCCCAAATCTCAAAGTGTAAATGGCTATAATGAAATTCAGAGTTTATATGCTGGAGAA
AAATGTGGTAATGTCTGGGAAGAAAATAGAAGTAAAACTGAAGACCGACATCCTGTGTACCATCCTGAAG
AAGGTGGAGATGAAGGTGGAAGTGGTTATAGTCCTCCAAGTAAATGTGAACAGACTGATTTATATCACCT
TCCTGAAAACTTTAGGCCATATACCAATGGTCAGAAGCAGGAATTTAATAACCAAGCAACCAATGTAATT
AAATTTTCAGATCCTCAATGGAACCATTTTCAGGGTCCCAGTTGTCAAGGTTTGGAACCGTATAATAAAG
TGACATATAAACCTTATCAGTCTTCTGCCCAGAATAATGGCTCACCAGCCCAGGAGATAACAGGAAGTGA
CACATTCGAAGGCCTGCAACAACAATTTTTAGGAGCTAATGAGAACTCTGCAGAAAATATGCAGATTATT
CAACTTCAGGTTCTTAACAAAGCAAAAGAGAGACAACTGGAGAACTTAATTGAAAAGTTAAATGAAAGTG
AACGTCAAATTCGATATCTGAATCACCAGCTTGTAATAATAAAAGATGAAAAGGATGGTTTGACTCTCAG

CCTTCGAGAATCACAGAAACTCTTTCAGAATGGAAAAGAAAGAGAGATACAGCTTGAAGCTCAAATAAAA
GCACTGGAGACTCAGATACAAGCATTAAAAGTCAATGAAGAACAGATGATCAAGAAGTCCAGAACAACTG
AAATGGCTCTGGAAAGCTTGAAGCAGCAGCTGGTGGACCTTCATCATTCTGAATCACTTCAACGAGCTAG
AGAACAGCATGAGAGCATTGTTATGGGCCTCACAAAGAAGTACGAAGAGCAAGTATTGTCCTTACAAAAG
AATTTGGATGCCACAGTCACCGCACTTAAAGAACAGGAAGACATTTGCTCTCGTCTGAAAGATCACGTGA
AACAACTGGAAAGGAATCAAGAAGCAATCAAGTTAGAAAAGACTGAGATCATTAATAAGTTGACAAGAAG
TCTAGAGGAGAGTCAAAAGCAGTGTGCCCACTTGTTGCAGTCCGGGTCAGTACAAGAGGTGGCTCAGCTA
CAGTTCCAGCTGCAGCAAGCACAGAAGGCACATGCTATGAGTGCAAACATGAACAAGGCTTTGCAAGAAG
AATTAACAGAACTAAAAGATGAAATTTCTCTCTATGAATCTGCTGCAAAACTAGGAATACATCCAAGTGA
CTCAGAAGGAGAATTAAATATAGAACTCACTGAATCGTATGTGGATTTGGGTATTAAAAAGGTCAACTGG
AAAAAATCCAAAGTTACCAGCATTGTACAAGAAGAAGACCCAAATGAAGAGCTTTCAAAAGATGAGTTCA
TTCTGAAGTTAAAGGCAGAAGTACAGCGTTTGCTGGGTAGCAACTCAATGAAGCGTCATCTGGTGTCTCA
GTTACAAAATGACCTCAAAGACTGTCATAAGAAAATTGAAGATCTCCACCAAGTGAAGAAGGATGAAAAA
AGCATTGAGGTTGAGACTAAAACAGATACCTCAGAAAAACCAAAGAATCAATTATGGCCTGAGTCTTCTA
CTTCTGATGTTGTCAGAGATGATATTCTGCTGCTTAAAAATGAAATTCAAGTTTTACAACAACAAAATCA
GGAACTTAAAGAAACTGAAGGAAAACTGAGAAATACAAATCAAGACTTATGTAATCAAATGAGACAAATG
GTACAAGATTTTGACCATGACAAACAAGAAGCTGTGGATAGGTGTGAAAGGACTTATCAGCAGCACCATG
AAGCCATGAAAACTCAAATACGTGAAAGCCTATTAGCAAAGCATGCTTTGGAGAAGCAGCAGCTCTTTGA
GGCTTATGAGAGAACTCATTTGCAACTGAGGTCTGAGTTGGATAAGTTGAATAAGGAGGTGACTGCTGTG
CAGGAATGTTACCTAGAAGTGTGCAGAGAGAAGGATAATCTAGAATTGACTCTCAGGAAGACCACTGAAA
AGGAGCAACAGACTCAGGAGAAGATCAAAGAAAAACTCATTCAACAGCTTGAAAAGGAGTGGCAGTCTAA
GCTGGATCAAACTATAAAGGCAATGAAAAAGAAGACCTTAGATTGTGGCAGCCAAACTGACCAAGTAACC
ACCAGTGATGTTATTTCCAAGAAAGAGATGGCAATTATGATAGAAGAGCAGAAGTGCACAATCCAGCAAA
ACTTAGAACAAGAGAAGGACATAGCCATCAAGGGGGCTATGAAGAAACTCGAAATTGAATTGGAACTCAA
ACATTGTGAAAATATTACCAAACAGGTAGAAATAGCTGTGCAAAATGCTCATCAGCGATGGCTGGGAGAA
CTACCAGAGCTGGCAGAGTATCAAGCACTTGTGAAGGCAGAACAGAAAAAGTGGGAAGAACAGCATGAGG
TCTCTGTGAACAAAAGGATATCATTTGCTGTTTCTGAAGCTAAAGAGAAATGGAAGAGTGAGCTTGAAAA
TATGAGGAAAAATATACTTCCTGGAAAGGAATTGGAAGAGAAGATTCATTCTCTTCAGAAGGAACTTGAG
TTAAAGAACGAAGAAGTCCCTGTGGTCATCAGGGCTGAGTTAGCTAAGGCTCGGAGTGAATGGAACAAAG
AAAAGCAAGAAGAAATCCACAGAATCCAAGAACAAAATGAGCAAGATTACCGGCAATTTTTAGATGATCA
CCGAAATAAAATTAATGAGGTGCTTGCGGCAGCTAAAGAAGACTTTATGAAACAAAAAACTGAACTACTT
CTTCAGAAGGAGACAGAATTACAAACTTGTCTAGACCAGAGTCGTAGAGAATGGACTATGCAGGAAGCCA
AGCGGATCCAACTGGAAATCTATCAGTATGAGGAAGACATCCTGACTGTACTTGGGGTTCTTTTAAGTGA
TACCCAAAAGGAGCACATCAGTGATTCTGAGGACAAGCAGCTTTTGGAAATCATGTCGACTTGTTCTTCA
AAATGGATGTCTGTGCAATATTTTGAAAAACTAAAGGGCTGCATACAGAAAGCATTTCAAGATACACTTC
CTCTGCTTGTAGAAAACGCTGACCCAGAATGGAAAAAGAGAAATATGGCCGAGCTCTCTAAGGATTCTGC
CAGCCAGGGCACTGGCCAAGGAGACCCTGGACCTGCTGCTGGACACCATGCTCAGCCCTTGGCCTTACAA
GCAACAGAAGCAGAAGCTGAAGAGAATAATAAAGTTGTTGAAGAATTAATAGAAGAAAACAACGACATGA
AGAATAAATTGGAAGAATTGCAAACACTTTGTAAAACACCACCAAGGTCATTGTCAGCAGGGGCCATTGA
AAATGCTTGCCTGCCATGCAGTGGGGGAGCCTTGGAAGAACTTCGTGGGCAGTACATTAAAGCTGTAAAA
AAAATTAAATGTGACATGCTTCGTTATATTCAGGAGAGTAAGGAACGAGCTGCAGAAATGGTAAAAGCAG
AGGTACTGCGAGAACGTCAAGAAACCGCCCGAAAGATGCGCAAATATTATTTGATTTGCCTCCAACAGAT
TTTGCAGGATGATGGAAAAGAAGGGGCTGAGAAAAAGATTATGAATGCTGCTAGCAAACTTGCTACAATG
GCAAAATTACTGGAAACACCTATTTCTAGTAAGTCCCAAAGCAAAACTACACAGTCAGCACTGCCCCTAA
CTTCAGAGATGCTGATTGCAGTTAAAAAATCAAAAAGAAATGATGTGAATCAGAAAATACCATGTTGTAT
TGAAAGCAAATCAAATAGTGTAAACACCATCACCAGAACTCTGTGCGAACAAGCTCCCAAGAGGAGGGCA
GCTTGTAACTTACAAAGGCTGTTAGAGAACTCAGAGCATCAGAGCATAAAGCATGTGGGATCCAAAGAGA
CACATTTGGAATTCCAGTTTGGGGATGGTAGTTGCAAGCACCTAAACAGTTTGCCAAGGAATGTTTCTCC
TGAGTTTGTTCCTTGTGAAGGTGAAGGAGGCTTTGGTTTGCACAAGAAGAAAGACCTACTCAGTGATAAT
GGTTCTGAATCACTTCCGCATTCAGCTGCATACCCCTTTCTTGGAACCTTAGGAAATAAACCCTCACCTA
GATGTACCCCTGGTCCTTCTGAATCAGGATGCATGCATATAACCTTTCGCGATTCTAATGAAAGACTTGG
TTTAAAAGTATATAAATGCAATCCACTAATGGAAAGTGAAAATGCTGCATCTGAGAAAAGTCAAGGTTTG
GATGTTCAGGAACCTCCAGTAAAAGATGGAGGGGACCTTAGTGACTGCTTGGGCTGGCCTTCCAGCAGTG
CAACCTTATCCTTTGACAGTCGTGAAGCATCATTTGTACATGGTAGGCCACAAGGAACTTTGGAAATACC
AAGTGAATCTGTTAAATCCAAACAGTTTTCACCATCCGGTTATCTTTCAGATACAGAGGAAAGTAATATG
ATTTGTCAAACAATGAAATGTCAGCGTTATCAAACTCCATACCTGTCAGAAGAAACCACGTATTTGGAGC
CAGGAAAGATCAGTGTGAATTGTGGACACCCATCTCGTCATAAGGCTGATAGATTAAAGTCAGATTTCAA
AAAACTGAGCAGTACATTACCATCTTCAGTGTGTCAGCAGCCTTCAAGAAAATTAATTGTTCCGCTATCT
AGCCAACAAGATAGTGGCTTTGATAGCCCATTTGTTAATCTAGACTAATTATGGTACAGTATTTAAGAAG
AATCATTAATATATTAACAAAAATGGAAGGGAAGACCTCATACTGAAAAAAATTGTGAGCCCTGCCTCTT
TTGAGATGTTTTAATAACATCTGTTATATAAGTAAAGCATTCTTCTAAAATTGCTTGAGATATTTATGTT

GCCTTAATATTCCAAAGGCCTGATGGTGTATGTATAATCTGCTTTTGTGTGGTGCTTATTTTTGGTTTCT
AAACCATCTATTTTTATACTTATAAATTGACTCACTCTGCAGTGTTAACTTATTTAAATAAACTTGCATA
TGGTCTGTAAAAAAAAAA (SEQ ID NO: 11)

Translated protein sequence

MSLDFGSVALPVQNEDEEYDEEDYEREKELQQLLTDLPHDMLDD
DLSSPELQYSDCSEDGTDGQPHHPEQLEMSWNEQMLPKSQSVNGYNEIQSLYAGEKCG
NVWEENRSKTEDRHPVYHPEEGGDEGGSGYSPPSKCEQTDLYHLPENFRPYTNGQKQE
FNNQATNVIKFSDPQWNHFQGPSCQGLEPYNKVTYKPYQSSAQNNGSPAQEITGSDTF
EGLQQQFLGANENSAENMQIIQLQVLNKAKERQLENLIEKLNESERQIRYLNHQLVII
KDEKDGLTLSLRESQKLFQNGKEREIQLEAQIKALETQIQALKVNEEQMIKKSRTTEM
ALESLKQQLVDLHHSESLQRAREQHESIVMGLTKKYEEQVLSLQKNLDATVTALKEQE
DICSRLKDHVKQLERNQEAIKLEKTEIINKLTRSLEESQKQCAHLLQSGSVQEVAQLQ
FQLQQAQKAHAMSANMNKALQEELTELKDEISLYESAAKLGIHPSDSEGELNIELTES
YVDLGIKKVNWKKSKVTSIVQEEDPNEELSKDEFILKLKAEVQRLLGSNSMKRHLVSQ
LQNDLKDCHKKIEDLHQVKKDEKSIEVETKTDTSEKPKNQLWPESSTSDVVRDDILLL
KNEIQVLQQQNQELKETEGKLRNTNQDLCNQMRQMVQDFDHDKQEAVDRCERTYQQHH
EAMKTQIRESLLAKHALEKQQLFEAYERTHLQLRSELDKLNKEVTAVQECYLEVCREK
DNLELTLRKTTEKEQQTQEKIKEKLIQQLEKEWQSKLDQTIKAMKKKTLDCGSQTDQV
TTSDVISKKEMAIMIEEQKCTIQQNLEQEKDIAIKGAMKKLEIELELKHCENITKQVE
IAVQNAHQRWLGELPELAEYQALVKAEQKKWEEQHEVSVNKRISFAVSEAKEKWKSEL
ENMRKNILPGKELEEKIHSLQKELELKNEEVPVVIRAELAKARSEWNKEKQEEIHRIQ
EQNEQDYRQFLDDHRNKINEVLAAAKEDFMKQKTELLLQKETELQTCLDQSRREWTMQ
EAKRIQLEIYQYEEDILTVLGVLLSDTQKEHISDSEDKQLLEIMSTCSSKWMSVQYFE
KLKGCIQKAFQDTLPLLVENADPEWKKRNMAELSKDSASQGTGQGDPGPAAGHHAQPL
ALQATEAEAEENNKVVEELIEENNDMKNKLEELQTLCKTPPRSLSAGAIENACLPCSG
GALEELRGQYIKAVKKIKCDMLRYIQESKERAAEMVKAEVLRERQETARKMRKYYLIC
LQQILQDDGKEGAEKKIMNAASKLATMAKLLETPISSKSQSKTTQSALPLTSEMLIAV
KKSKRNDVNQKIPCCIESKSNSVNTITRTLCEQAPKRRAACNLQRLLENSEHQSIKHV
GSKETHLEFQFGDGSCKHLNSLPRNVSPEFVPCEGEGGFGLHKKKDLLSDNGSESLPH
SAAYPFLGTLGNKPSPRCTPGPSESGCMHITFRDSNERLGLKVYKCNPLMESENAASE
KSQGLDVQEPPVKDGGDLSDCLGWPSSSATLSFDSREASFVHGRPQGTLEIPSESVKS
KQFSPSGYLSDTEESNMICQTMKCQRYQTPYLSEETTYLEPGKISVNCGHPSRHKADR
LKSDFKKLSSTLPSSVCQQPSRKLIVPLSSQQDSGFDSPFVNLD (SEQ ID NO: 12)

CFL1	1072	NM_005507	Homo sapiens cofilin 1 (non-muscle) (CFL1), mRNA

mRNA Sequence

GGCCGGCGGGAAGACTCCGTTACCCAGCGAGCGAGGCGGCGGCGCAGGGCCAGCGGACTCCATTTCCCGT
CGGCTCGCGGTGGGAGCGCCGGAAGCCCGCCCCACCCCTCATTGTGCGGCTCCTACTAAACGGAAGGGGC
CGGGAGAGGCCGCGTTCAGTCGGGTCCCGGCAGCGGCTGCAGCGCTCTCGTCTTCTGCGGCTCTCGGTGC
CCTCTCCTTTTCGTTTCCGGAAACATGGCCTCCGGTGTGGCTGTCTCTGATGGTGTCATCAAGGTGTTCA
ACGACATGAAGGTGCGTAAGTCTTCAACGCCAGAGGAGGTGAAGAAGCGCAAGAAGGCGGTGCTCTTCTG
CCTGAGTGAGGACAAGAAGAACATCATCCTGGAGGAGGGCAAGGAGATCCTGGTGGGCGATGTGGGCCAG
ACTGTCGACGACCCCTACGCCACCTTTGTCAAGATGCTGCCAGATAAGGACTGCCGCTATGCCCTCTATG
ATGCAACCTATGAGACCAAGGAGAGCAAGAAGGAGGATCTGGTGTTTATCTTCTGGGCCCCCGAGTCTGC
GCCCCTTAAGAGCAAAATGATTTATGCCAGCTCCAAGGACGCCATCAAGAAGAAGCTGACAGGGATCAAG
CATGAATTGCAAGCAAACTGCTACGAGGAGGTCAAGGACCGCTGCACCCTGGCAGAGAAGCTGGGGGGCA
GTGCCGTCATCTCCCTGGAGGGCAAGCCTTTGTGAGCCCCTTCTGGCCCCCTGCCTGGAGCATCTGGCAG
CCCCACACCTGCCCTTGGGGGTTGCAGGCTGCCCCCTTCCTGCCAGACCGGAGGGGCTGGGGGGATCCCA
GCAGGGGGAGGGCAATCCCTTCACCCCAGTTGCCAAACAGACCCCCCACCCCCTGGATTTTCCTTCTCCC
TCCATCCCTTGACGGTTCTGGCCTTCCCAAACTGCTTTTGATCTTTTGATTCCTCTTGGGCTGAAGCAGA
CCAAGTTCCCCCCAGGCACCCCAGTTGTGGGGGAGCCTGTATTTTTTTTAACAACATCCCCATTCCCCAC
CTGGTCCTCCCCCTTCCCATGCTGCCAACTTCTAACCGCAATAGTGACTCTGTGCTTGTCTGTTTAGTTC
TGTGTATAAATGGAATGTTGTGGAGATGACCCCTCCCTGTGCCGGCTGGTTCCTCTCCCTTTTCCCCTGG
TCACGGCTACTCATGGAAGCAGGACCAGTAAGGGACCTTCGATTAAAAAAAAAAAAAGACAATAATAAAAA (SEQ ID NO: 13)

Translated protein sequence

MASGVAVSDGVIKVFNDMKVRKSSTPEEVKKRKKAVLFCLSEDK
KNIILEEGKEILVGDVGQTVDDPYATFVKMLPDKDCRYALYDATYETKESKKEDLVFI
FWAPESAPLKSKMIYASSKDAIKKKLTGIKHELQANCYEEVKDRCTLAEKLGGSAVIS
LEGKPL (SEQ ID NO: 14)

CKLF	51192	NM_016326	Homo sapiens chemokine-like factor (CKLF), transcript variant 3, mRNA

mRNA Sequence

ATGCGCGCAAGAGAGCGGGAAGCCGAGCTGGGCGAGAAGTAGGGGAGGGCGGTGCTCCGCCGCGGTGGCG
GTTGCTATCGCTTCGCAGAACCTACTCAGGCAGCCAGCTGAGAAGAGTTGAGGGAAAGTGCTGCTGCTGG
GTCTGCAGACGCGATGGATAACGTGCAGCCGAAAATAAAACATCGCCCCTTCTGCTTCAGTGTGAAAGGC
CACGTGAAGATGCTGCGGCTGGTGTTTGCACTTGTGACAGCAGTATGCTGTCTTGCCGACGGGGCCCTTA
TTTACCGGAAGCTTCTGTTCAATCCCAGCGGTCCTTACCAGAAAAAGCCTGTGCATGAAAAAAAAGAAGT
TTTGTAATTTTATATTACTTTTTAGTTTGATACTAAGTATTAAACATATTTCTGTATTCTTCCACATATT
TTCTGCAGTTATTTTAACTCAGTATAGGAGCTAGAGGAAGAGATTTCCGAAGTCTGCACCCCGCGCAGAG
CACTACTGTAACTTCCAAGGGAGCGCTGGGAGCAGCGGGATCGGGTTTTCCGGCACCCGGGCCTGGGTGG
CAGGGAAGAATGTGCCGGGATCCGCCTCAGGGATCTTTGAATCTCTTTACTGCCTGGCTGGCCGGCAGCT
CCG (SEQ ID NO: 15)

Translated protein sequence

MDNVQPKIKHRPFCFSVKGHVKMLRLVFALVTAVCCLADGALIY
RKLLFNPSGPYQKKPVHEKKEVL (SEQ ID NO: 16)

COL7A1	1294	NM_000094	Homo sapiens collagen, type VII, alpha 1 (COL7A1), mRNA

mRNA Sequence

GATGACGCTGCGGCTTCTGGTGGCCGCGCTCTGCGCCGGGATCCTGGCAGAGGCGCCCCGAGTGCGAGCC
CAGCACAGGGAGAGAGTGACCTGCACGCGCCTTTACGCCGCTGACATTGTGTTCTTACTGGATGGCTCCT
CATCCATTGGCCGCAGCAATTTCCGCGAGGTCCGCAGCTTTCTCGAAGGGCTGGTGCTGCCTTTCTCTGG
AGCAGCCAGTGCACAGGGTGTGCGCTTTGCCACAGTGCAGTACAGCGATGACCCACGGACAGAGTTCGGC
CTGGATGCACTTGGCTCTGGGGGTGATGTGATCCGCGCCATCCGTGAGCTTAGCTACAAGGGGGGCAACA
CTCGCACAGGGGCTGCAATTCTCCATGTGGCTGACCATGTCTTCCTGCCCCAGCTGGCCCGACCTGGTGT
CCCCAAGGTCTGCATCCTGATCACAGACGGGAAGTCCCAGGACCTGGTGGACACAGCTGCCCAAAGGCTG
AAGGGGCAGGGGGTCAAGCTATTTGCTGTGGGGATCAAGAATGCTGACCCTGAGGAGCTGAAGCGAGTTG
CCTCACAGCCCACCAGTGACTTCTTCTTCTTCGTCAATGACTTCAGCATCTTGAGGACACTACTGCCCCT
CGTTTCCCGGAGAGTGTGCACGACTGCTGGTGGCGTGCCTGTGACCCGACCTCCGGATGACTCGACCTCT
GCTCCACGAGACCTGGTGCTGTCTGAGCCAAGCAGCCAATCCTTGAGAGTACAGTGGACAGCGGCCAGTG
GCCCTGTGACTGGCTACAAGGTCCAGTACACTCCTCTGACGGGGCTGGGACAGCCACTGCCGAGTGAGCG
GCAGGAGGTGAACGTCCCAGCTGGTGAGACCAGTGTGCGGCTGCGGGGTCTCCGGCCACTGACCGAGTAC
CAAGTGACTGTGATTGCCCTCTACGCCAACAGCATCGGGGAGGCTGTGAGCGGGACAGCTCGGACCACTG
CCCTAGAAGGGCCGGAACTGACCATCCAGAATACCACAGCCCACAGCCTCCTGGTGGCCTGGCGGAGTGT
GCCAGGTGCCACTGGCTACCGTGTGACATGGCGGGTCCTCAGTGGTGGGCCCACACAGCAGCAGGAGCTG
GGCCCTGGGCAGGGTTCAGTGTTGCTGCGTGACTTGGAGCCTGGCACGGACTATGAGGTGACCGTGAGCA
CCCTATTTGGCCGCAGTGTGGGGCCCGCCACTTCCCTGATGGCTCGCACTGACGCTTCTGTTGAGCAGAC
CCTGCGCCCGGTCATCCTGGGCCCCACATCCATCCTCCTTTCCTGGAACTTGGTGCCTGAGGCCCGTGGC
TACCGGTTGGAATGGCGGCGTGAGACTGGCTTGGAGCCACCGCAGAAGGTGGTACTGCCCTCTGATGTGA
CCCGCTACCAGTTGGATGGGCTGCAGCCGGGCACTGAGTACCGCCTCACACTCTACACTCTGCTGGAGGG
CCACGAGGTGGCCACCCCTGCAACCGTGGTTCCCACTGGACCAGAGCTGCCTGTGAGCCCTGTAACAGAC
CTGCAAGCCACCGAGCTGCCCGGGCAGCGGGTGCGAGTGTCCTGGAGCCCAGTCCCTGGTGCCACCCAGT
ACCGCATCATTGTGCGCAGCACCCAGGGGGTTGAGCGGACCCTGGTGCTTCCTGGGAGTCAGACAGCATT
CGACTTGGATGACGTTCAGGCTGGGCTTAGCTACACTGTGCGGGTGTCTGCTCGAGTGGGTCCCCGTGAG
GGCAGTGCCAGTGTCCTCACTGTCCGCCGGGAGCCGGAAACTCCACTTGCTGTTCCAGGGCTGCGGGTTG
TGGTGTCAGATGCAACGCGAGTGAGGGTGGCCTGGGGACCCGTCCCTGGAGCCAGTGGATTTCGGATTAG
CTGGAGCACAGGCAGTGGTCCGGAGTCCAGCCAGACACTGCCCCCAGACTCTACTGCCACAGACATCACA
GGGCTGCAGCCTGGAACCACCTACCAGGTGGCTGTGTCGGTACTGCGAGGCAGAGAGGAGGGCCCTGCTG
CAGTCATCGTGGCTCGAACGGACCCACTGGGCCCAGTGAGGACGGTCCATGTGACTCAGGCCAGCAGCTC
ATCTGTCACCATTACCTGGACCAGGGTTCCTGGCGCCACAGGATACAGGGTTTCCTGGCACTCAGCCCAC
GGCCCAGAGAAATCCCAGTTGGTTTCTGGGGAGGCCACGGTGGCTGAGCTGGATGGACTGGAGCCAGATA

CTGAGTATACGGTGCATGTGAGGGCCCATGTGGCTGGCGTGGATGGGCCCCCTGCCTCTGTGGTTGTGAG
GACTGCCCCTGAGCCTGTGGGTCGTGTGTCGAGGCTGCAGATCCTCAATGCTTCCAGCGACGTTCTACGG
ATCACCTGGGTAGGGGTCACTGGAGCCACAGCTTACAGACTGGCCTGGGGCCGGAGTGAAGGCGGCCCCA
TGAGGCACCAGATACTCCCAGGAAACACAGACTCTGCAGAGATCCGGGGTCTCGAAGGTGGAGTCAGCTA
CTCAGTGCGAGTGACTGCACTTGTCGGGGACCGCGAGGGCACACCTGTCTCCATTGTTGTCACTACGCCG
CCTGAGGCTCCGCCAGCCCTGGGGACGCTTCACGTGGTGCAGCGCGGGGAGCACTCGCTGAGGCTGCGCT
GGGAGCCGGTGCCCAGAGCGCAGGGCTTCCTTCTGCACTGGCAACCTGAGGGTGGCCAGGAACAGTCCCG
GGTCCTGGGGCCCGAGCTCAGCAGCTATCACCTGGACGGGCTGGAGCCAGCGACACAGTACCGCGTGAGG
CTGAGTGTCCTAGGGCCAGCTGGAGAAGGGCCCTCTGCAGAGGTGACTGCGCGCACTGAGTCACCTCGTG
TTCCAAGCATTGAACTACGTGTGGTGGACACCTCGATCGACTCGGTGACTTTGGCCTGGACTCCAGTGTC
CAGGGCATCCAGCTACATCCTATCCTGGCGGCCACTCAGAGGCCCTGGCCAGGAAGTGCCTGGGTCCCCG
CAGACACTTCCAGGGATCTCAAGCTCCCAGCGGGTGACAGGGCTAGAGCCTGGCGTCTCTTACATCTTCT
CCCTGACGCCTGTCCTGGATGGTGTGCGGGGTCCTGAGGCATCTGTCACACAGACGCCAGTGTGCCCCCG
TGGCCTGGCGGATGTGGTGTTCCTACCACATGCCACTCAAGACAATGCTCACCGTGCGGAGGCTACGAGG
AGGGTCCTGGAGCGTCTGGTGTTGGCACTTGGGCCTCTTGGGCCACAGGCAGTTCAGGTTGGCCTGCTGT
CTTACAGTCATCGGCCCTCCCCACTGTTCCCACTGAATGGCTCCCATGACCTTGGCATTATCTTGCAAAG
GATCCGTGACATGCCCTACATGGACCCAAGTGGGAACAACCTGGGCACAGCCGTGGTCACAGCTCACAGA
TACATGTTGGCACCAGATGCTCCTGGGCGCCGCCAGCACGTACCAGGGGTGATGGTTCTGCTAGTGGATG
AACCCTTGAGAGGTGACATATTCAGCCCCATCCGTGAGGCCCAGGCTTCTGGGCTTAATGTGGTGATGTT
GGGAATGGCTGGAGCGGACCCAGAGCAGCTGCGTCGCTTGGCGCCGGGTATGGACTCTGTCCAGACCTTC
TTCGCCGTGGATGATGGGCCAAGCCTGGACCAGGCAGTCAGTGGTCTGGCCACAGCCCTGTGTCAGGCAT
CCTTCACTACTCAGCCCCGGCCAGAGCCCTGCCCAGTGTATTGTCCAAAGGGCCAGAAGGGGGAACCTGG
AGAGATGGGCCTGAGAGGACAAGTTGGGCCTCCTGGCGACCCTGGCCTCCCGGGCAGGACCGGTGCTCCC
GGCCCCCAGGGGCCCCCTGGAAGTGCCACTGCCAAGGGCGAGAGGGGCTTCCCTGGAGCAGATGGGCGTC
CAGGCAGCCCTGGCCGCGCCGGGAATCCTGGGACCCCTGGAGCCCCTGGCCTAAAGGGCTCTCCAGGGTT
GCCTGGCCCTCGTGGGGACCCGGGAGAGCGAGGACCTCGAGGCCCAAAGGGGGAGCCGGGGGCTCCCGGA
CAAGTCATCGGAGGTGAAGGACCTGGGCTTCCTGGGCGGAAAGGGGACCCTGGACCATCGGGCCCCCCTG
GACCTCGTGGACCACTGGGGGACCCAGGACCCCGTGGCCCCCCAGGGCTTCCTGGAACAGCCATGAAGGG
TGACAAAGGCGATCGTGGGGAGCGGGGTCCCCCTGGACCAGGTGAAGGTGGCATTGCTCCTGGGGAGCCT
GGGCTGCCGGGTCTTCCCGGAAGCCCTGGACCCCAAGGCCCCGTTGGCCCCCCTGGAAAGAAAGGAGAAA
AAGGTGACTCTGAGGATGGAGCTCCAGGCCTCCCAGGACAACCTGGGTCTCCGGGTGAGCAGGGCCCACG
GGGACCTCCTGGAGCTATTGGCCCCAAAGGTGACCGGGGCTTTCCAGGGCCCCTGGGTGAGGCTGGAGAG
AAGGGCGAACGTGGACCCCCAGGCCCAGCGGGATCCCGGGGGCTGCCAGGGGTTGCTGGACGTCCTGGAG
CCAAGGGTCCTGAAGGGCCACCAGGACCCACTGGCCGCCAAGGAGAGAAGGGGGAGCCTGGTCGCCCTGG
GGACCCTGCAGTGGTGGGACCTGCTGTTGCTGGACCCAAAGGAGAAAAGGGAGATGTGGGGCCCGCTGGG
CCCAGAGGAGCTACCGGAGTCCAAGGGGAACGGGGCCCACCCGGCTTGGTTCTTCCTGGAGACCCTGGCC
CCAAGGGAGACCCTGGAGACCGGGGTCCCATTGGCCTTACTGGCAGAGCAGGACCCCCAGGTGACTCAGG
GCCTCCTGGAGAGAAGGGAGACCCTGGGCGGCCTGGCCCCCCAGGACCTGTTGGCCCCCGAGGACGAGAT
GGTGAAGTTGGAGAGAAAGGTGACGAGGGTCCTCCGGGTGACCCGGGTTTGCCTGGAAAAGCAGGCGAGC
GTGGCCTTCGGGGGGCACCTGGAGTTCGGGGGCCTGTGGGTGAAAAGGGAGACCAGGGAGATCCTGGAGA
GGATGGACGAAATGGCAGCCCTGGATCATCTGGACCCAAGGGTGACCGTGGGGAGCCGGGTCCCCCAGGA
CCCCCGGGACGGCTGGTAGACACAGGACCTGGAGCCAGAGAGAAGGGAGAGCCTGGGGACCGCGGACAAG
AGGGTCCTCGAGGGCCCAAGGGTGATCCTGGCCTCCCTGGAGCCCCTGGGGAAAGGGGCATTGAAGGGTT
TCGGGGACCCCCAGGCCCACAGGGGGACCCAGGTGTCCGAGGCCCAGCAGGAGAAAAGGGTGACCGGGGT
CCCCCTGGGCTGGATGGCCGGAGCGGACTGGATGGGAAACCAGGAGCCGCTGGGCCCTCTGGGCCGAATG
GTGCTGCAGGCAAAGCTGGGGACCCAGGGAGAGACGGGCTTCCAGGCCTCCGTGGAGAACAGGGCCTCCC
TGGCCCCTCTGGTCCCCCTGGATTACCGGGAAAGCCAGGCGAGGATGGCAAACCTGGCCTGAATGGAAAA
AACGGAGAACCTGGGGACCCTGGAGAAGACGGGAGGAAGGGAGAGAAAGGAGATTCAGGCGCCTCTGGGA
GAGAAGGTCGTGATGGCCCCAAGGGTGAGCGTGGAGCTCCTGGTATCCTTGGACCCCAGGGGCCTCCAGG
CCTCCCAGGGCCAGTGGGCCCTCCTGGCCAGGGTTTTCCTGGTGTCCCAGGAGGCACGGGCCCCAAGGGT
GACCGTGGGGAGACTGGATCCAAAGGGGAGCAGGGCCTCCCTGGAGAGCGTGGCCTGCGAGGAGAGCCTG
GAAGTGTGCCGAATGTGGATCGGTTGCTGGAAACTGCTGGCATCAAGGCATCTGCCCTGCGGGAGATCGT
GGAGACCTGGGATGAGAGCTCTGGTAGCTTCCTGCCTGTGCCCGAACGGCGTCGAGGCCCCAAGGGGGAC
TCAGGCGAACAGGGCCCCCCAGGCAAGGAGGGCCCCATCGGCTTTCCTGGAGAACGCGGGCTGAAGGGCG
ACCGTGGAGACCCTGGCCCTCAGGGGCCACCTGGTCTGGCCCTTGGGGAGAGGGGCCCCCCCGGGCCTTC
CGGCCTTGCCGGGGAGCCTGGAAAGCCTGGTATTCCCGGGCTCCCAGGCAGGGCTGGGGGTGTGGGAGAG
GCAGGAAGGCCAGGAGAGAGGGGAGAACGGGGAGAGAAAGGAGAACGTGGAGAACAGGGCAGAGATGGCC
CTCCTGGACTCCCTGGAACCCCTGGGCCCCCCGGACCCCCTGGCCCCAAGGTGTCTGTGGATGAGCCAGG
TCCTGGACTCTCTGGAGAACAGGGACCCCCTGGACTCAAGGGTGCTAAGGGGGAGCCGGGCAGCAATGGT
GACCAAGGTCCCAAAGGAGACAGGGGTGTGCCAGGCATCAAAGGAGACCGGGGAGAGCCTGGACCGAGGG
GTCAGGACGGCAACCCGGGTCTACCAGGAGAGCGTGGTATGGCTGGGCCTGAAGGGAAGCCGGGTCTGCA

GGGTCCAAGAGGCCCCCCTGGCCCAGTGGGTGGTCATGGAGACCCTGGACCACCTGGTGCCCCGGGTCTT
GCTGGCCCTGCAGGACCCCAAGGACCTTCTGGCCTGAAGGGGGAGCCTGGAGAGACAGGACCTCCAGGAC
GGGGCCTGACTGGACCTACTGGAGCTGTGGGACTTCCTGGACCCCCCGGCCCTTCAGGCCTTGTGGGTCC
ACAGGGGTCTCCAGGTTTGCCTGGACAAGTGGGGGAGACAGGGAAGCCGGGAGCCCCAGGTCGAGATGGT
GCCAGTGGAAAAGATGGAGACAGAGGGAGCCCTGGTGTGCCAGGGTCACCAGGTCTGCCTGGCCCTGTCG
GACCTAAAGGAGAACCTGGCCCCACGGGGGCCCCTGGACAGGCTGTGGTCGGGCTCCCTGGAGCAAAGGG
AGAGAAGGGAGCCCCTGGAGGCCTTGCTGGAGACCTGGTGGGTGAGCCGGGAGCCAAAGGTGACCGAGGA
CTGCCAGGGCCGCGAGGCGAGAAGGGTGAAGCTGGCCGTGCAGGGGAGCCCGGAGACCCTGGGGAAGATG
GTCAGAAAGGGGCTCCAGGACCCAAAGGTTTCAAGGGTGACCCAGGAGTCGGGGTCCCGGGCTCCCCTGG
GCCTCCTGGCCCTCCAGGTGTGAAGGGAGATCTGGGCCTCCCTGGCCTGCCCGGTGCTCCTGGTGTTGTT
GGGTTCCCGGGTCAGACAGGCCCTCGAGGAGAGATGGGTCAGCCAGGCCCTAGTGGAGAGCGGGGTCTGG
CAGGCCCCCCAGGGAGAGAAGGAATCCCAGGACCCCTGGGGCCACCTGGACCACCGGGGTCAGTGGGACC
ACCTGGGGCCTCTGGACTCAAAGGAGACAAGGGAGACCCTGGAGTAGGGCTGCCTGGGCCCCGAGGCGAG
CGTGGGGAGCCAGGCATCCGGGGTGAAGATGGCCGCCCCGGCCAGGAGGGACCCCGAGGACTCACGGGGC
CCCCTGGCAGCAGGGGAGAGCGTGGGGAGAAGGGTGATGTTGGGAGTGCAGGACTAAAGGGTGACAAGGG
AGACTCAGCTGTGATCCTGGGGCCTCCAGGCCCACGGGGTGCCAAGGGGGACATGGGTGAACGAGGGCCT
CGGGGCTTGGATGGTGACAAAGGACCTCGGGGAGACAATGGGGACCCTGGTGACAAGGGCAGCAAGGGAG
AGCCTGGTGACAAGGGCTCAGCCGGGTTGCCAGGACTGCGTGGACTCCTGGGACCCCAGGGTCAACCTGG
TGCAGCAGGGATCCCTGGTGACCCGGGATCCCCAGGAAAGGATGGAGTGCCTGGTATCCGAGGAGAAAAA
GGAGATGTTGGCTTCATGGGTCCCCGGGGCCTCAAGGGTGAACGGGGAGTGAAGGGAGCCTGTGGCCTTG
ATGGAGAGAAGGGAGACAAGGGAGAAGCTGGTCCCCCAGGCCGCCCCGGGCTGGCAGGACACAAAGGAGA
GATGGGGGAGCCTGGTGTGCCGGGCCAGTCGGGGGCCCCTGGCAAGGAGGGCCTGATCGGTCCCAAGGGT
GACCGAGGCTTTGACGGGCAGCCAGGCCCCAAGGGTGACCAGGGCGAGAAAGGGGAGCGGGGAACCCCAG
GAATTGGGGGCTTCCCAGGCCCCAGTGGAAATGATGGCTCTGCTGGTCCCCCAGGGCCACCTGGCAGTGT
TGGTCCCAGAGGCCCCGAAGGACTTCAGGGCCAGAAGGGTGAGCGAGGTCCCCCCGGAGAGAGAGTGGTG
GGGGCTCCTGGGGTCCCTGGAGCTCCTGGCGAGAGAGGGGAGCAGGGGCGGCCAGGGCCTGCCGGTCCTC
GAGGCGAGAAGGGAGAAGCTGCACTGACGGAGGATGACATCCGGGGCTTTGTGCGCCAAGAGATGAGTCA
GCACTGTGCCTGCCAGGGCCAGTTCATCGCATCTGGATCACGACCCCTCCCTAGTTATGCTGCAGACACT
GCCGGCTCCCAGCTCCATGCTGTGCCTGTGCTCCGCGTCTCTCATGCAGAGGAGGAAGAGCGGGTACCCC
CTGAGGATGATGAGTACTCTGAATACTCCGAGTATTCTGTGGAGGAGTACCAGGACCCTGAAGCTCCTTG
GGATAGTGATGACCCCTGTTCCCTGCCACTGGATGAGGGCTCCTGCACTGCCTACACCCTGCGCTGGTAC
CATCGGGCTGTGACAGGCAGCACAGAGGCCTGTCACCCTTTTGTCTATGGTGGCTGTGGAGGGAATGCCA
ACCGTTTTGGGACCCGTGAGGCCTGCGAGCGCCGCTGCCCACCCCGGGTGGTCCAGAGCCAGGGGACAGG
TACTGCCCAGGACTGAGGCCCAGATAATGAGCTGAGATTCAGCATCCCCTGGAGGAGTCGGGGTCTCAGC
AGAACCCCACTGTCCCTCCCCTTGGTGCTAGAGGCTTGTGTGCACGTGAGCGTGCGTGTGCACGTCCGTT
ATTTCAGTGACTTGGTCCCGTGGGTCTAGCCTTCCCCCCTGTGGACAAACCCCCATTGTGGCTCCTGCCA
CCCTGGCAGATGACTCACTGTGGGGGGGTGGCTGTGGGCAGTGAGCGGATGTGACTGGCGTCTGACCCGC
CCCTTGACCCAAGCCTGTGATGACATGGTGCTGATTCTGGGGGGCATTAAAGCTGCTGTTTTAAAAGGC (SEQ ID NO: 17)

Translated protein sequence

MTLRLLVAALCAGILAEAPRVRAQHRERVTCTRLYAADIVFLLD
GSSSIGRSNFREVRSFLEGLVLPFSGAASAQGVRFATVQYSDDPRTEFGLDALGSGGD
VIRAIRELSYKGGNTRTGAAILHVADHVFLPQLARPGVPKVCILITDGKSQDLVDTAA
QRLKGQGVKLFAVGIKNADPEELKRVASQPTSDFFFFVNDFSILRTLLPLVSRRVCTT
AGGVPVTRPPDDSTSAPRDLVLSEPSSQSLRVQWTAASGPVTGYKVQYTPLTGLGQPL
PSERQEVNVPAGETSVRLRGLRPLTEYQVTVIALYANSIGEAVSGTARTTALEGPELT
IQNTTAHSLLVAWRSVPGATGYRVTWRVLSGGPTQQQELGPGQGSVLLRDLEPGTDYE
VTVSTLFGRSVGPATSLMARTDASVEQTLRPVILGPTSILLSWNLVPEARGYRLEWRR
ETGLEPPQKVVLPSDVTRYQLDGLQPGTEYRLTLYTLLEGHEVATPATVVPTGPELPV
SPVTDLQATELPGQRVRVSWSPVPGATQYRIIVRSTQGVERTLVLPGSQTAFDLDDVQ
AGLSYTVRVSARVGPREGSASVLTVRREPETPLAVPGLRVVVSDATRVRVAWGPVPGA
SGFRISWSTGSGPESSQTLPPDSTATDITGLQPGTTYQVAVSVLRGREEGPAAVIVAR
TDPLGPVRTVHVTQASSSSVTITWTRVPGATGYRVSWHSAHGPEKSQLVSGEATVAEL
DGLEPDTEYTVHVRAHVAGVDGPPASVVVRTAPEPVGRVSRLQILNASSDVLRITWVG
VTGATAYRLAWGRSEGGPMRHQILPGNTDSAEIRGLEGGVSYSVRVTALVGDREGTPV
SIVVTTPPEAPPALGTLHVVQRGEHSLRLRWEPVPRAQGFLLHWQPEGGQEQSRVLGP
ELSSYHLDGLEPATQYRVRLSVLGPAGEGPSAEVTARTESPRVPSIELRVVDTSIDSV
TLAWTPVSRASSYILSWRPLRGPGQEVPGSPQTLPGISSSQRVTGLEPGVSYIFSLTP
VLDGVRGPEASVTQTPVCPRGLADVVFLPHATQDNAHRAEATRRVLERLVLALGPLGP
QAVQVGLLSYSHRPSPLFPLNGSHDLGIILQRIRDMPYMDPSGNNLGTAVVTAHRYML

APDAPGRRQHVPGVMVLLVDEPLRGDIFSPIREAQASGLNVVMLGMAGADPEQLRRLA
PGMDSVQTFFAVDDGPSLDQAVSGLATALCQASFTTQPRPEPCPVYCPKGQKGEPGEM
GLRGQVGPPGDPGLPGRTGAPGPQGPPGSATAKGERGFPGADGRPGSPGRAGNPGTPG
APGLKGSPGLPGPRGDPGERGPRGPKGEPGAPGQVIGGEGPGLPGRKGDPGPSGPPGP
RGPLGDPGPRGPPGLPGTAMKGDKGDRGERGPPGPGEGGIAPGEPGLPGLPGSPGPQG
PVGPPGKKGEKGDSEDGAPGLPGQPGSPGEQGPRGPPGAIGPKGDRGFPGPLGEAGEK
GERGPPGPAGSRGLPGVAGRPGAKGPEGPPGPTGRQGEKGEPGRPGDPAVVGPAVAGP
KGEKGDVGPAGPRGATGVQGERGPPGLVLPGDPGPKGDPGDRGPIGLTGRAGPPGDSG
PPGEKGDPGRPGPPGPVGPRGRDGEVGEKGDEGPPGDPGLPGKAGERGLRGAPGVRGP
VGEKGDQGDPGEDGRNGSPGSSGPKGDRGEPGPPGPPGRLVDTGPGAREKGEPGDRGQ
EGPRGPKGDPGLPGAPGERGIEGFRGPPGPQGDPGVRGPAGEKGDRGPPGLDGRSGLD
GKPGAAGPSGPNGAAGKAGDPGRDGLPGLRGEQGLPGPSGPPGLPGKPGEDGKPGLNG
KNGEPGDPGEDGRKGEKGDSGASGREGRDGPKGERGAPGILGPQGPPGLPGPVGPPGQ
GFPGVPGGTGPKGDRGETGSKGEQGLPGERGLRGEPGSVPNVDRLLETAGIKASALRE
IVETWDESSGSFLPVPERRRGPKGDSGEQGPPGKEGPIGFPGERGLKGDRGDPGPQGP
PGLALGERGPPGPSGLAGEPGKPGIPGLPGRAGGVGEAGRPGERGERGEKGERGEQGR
DGPPGLPGTPGPPGPPGPKVSVDEPGPGLSGEQGPPGLKGAKGEPGSNGDQGPKGDRG
VPGIKGDRGEPGPRGQDGNPGLPGERGMAGPEGKPGLQGPRGPPGPVGGHGDPGPPGA
PGLAGPAGPQGPSGLKGEPGETGPPGRGLTGPTGAVGLPGPPGPSGLVGPQGSPGLPG
QVGETGKPGAPGRDGASGKDGDRGSPGVPGSPGLPGPVGPKGEPGPTGAPGQAVVGLP
GAKGEKGAPGGLAGDLVGEPGAKGDRGLPGPRGEKGEAGRAGEPGDPGEDGQKGAPGP
KGFKGDPGVGVPGSPGPPGPPGVKGDLGLPGLPGAPGVVGFPGQTGPRGEMGQPGPSG
ERGLAGPPGREGIPGPLGPPGPPGSVGPPGASGLKGDKGDPGVGLPGPRGERGEPGIR
GEDGRPGQEGPRGLTGPPGSRGERGEKGDVGSAGLKGDKGDSAVILGPPGPRGAKGDM
GERGPRGLDGDKGPRGDNGDPGDKGSKGEPGDKGSAGLPGLRGLLGPQGQPGAAGIPG
DPGSPGKDGVPGIRGEKGDVGFMGPRGLKGERGVKGACGLDGEKGDKGEAGPPGRPGL
AGHKGEMGEPGVPGQSGAPGKEGLIGPKGDRGFDGQPGPKGDQGEKGERGTPGIGGFP
GPSGNDGSAGPPGPPGSVGPRGPEGLQGQKGERGPPGERVVGAPGVPGAPGERGEQGR
PGPAGPRGEKGEAALTEDDIRGFVRQEMSQHCACQGQFIASGSRPLPSYAADTAGSQL
HAVPVLRVSHAEEEERVPPEDDEYSEYSEYSVEEYQDPEAPWDSDDPCSLPLDEGSCT
AYTLRWYHRAVTGSTEACHPFVYGGCGGNANRFGTREACERRCPPRVVQSQGTGTAQD (SEQ ID NO: 18)

CYP4F3	4051	NM_000896	Homo sapiens cytochrome P450, family 4, subfamily F, polypeptide 3 (CYP4F3), mRNA

mRNA Sequence

AGAAGAAGGGGAGAGGAGGTTGTGTGGGACAAGGTGCTCCTGACAGAAGGATGCCACAGCTGAGCCTGTC
CTCGCTGGGCCTTTGGCCAATGGCAGCATCCCCGTGGCTGCTCCTGCTGCTGGTTGGGGCCTCCTGGCTC
CTGGCCCGCATCCTGGCCTGGACCTATACCTTCTATGACAACTGCTGCCGCCTCCGGTGTTTCCCGCAAC
CCCCGAAACGGAATTGGTTCTTGGGTCACCTGGGCCTGATTCACAGCTCGGAGGAAGGTCTCCTATACAC
ACAAAGCCTGGCATGCACCTTCGGTGATATGTGCTGCTGGTGGGTGGGGCCCTGGCACGCAATCGTCCGC
ATCTTCCACCCCACCTACATCAAGCCTGTGCTCTTTGCTCCAGCTGCCATTGTACCAAAGGACAAGGTCT
TCTACAGCTTCCTGAAGCCCTGGCTGGGGGATGGGCTCCTGCTGAGTGCTGGTGAAAAGTGGAGCCGCCA
CCGTCGGATGCTGACGCCTGCCTTCCATTTCAACATCCTGAAGCCCTATATGAAGATTTTCAATGAGAGT
GTGAACATCATGCATGCCAAGTGGCAGCTCCTGGCCTCAGAGGGTAGTGCCCGTCTGGACATGTTTGAGC
ACATCAGCCTCATGACCTTGGACAGTCTGCAGAAATGTGTCTTCAGCTTTGACAGCCATTGCCAGGAGAA
GCCCAGTGAATATATTGCCGCCATCTTGGAGCTCAGTGCCCTTGTGACAAAAAGACACCAGCAGATCCTC
CTGTACATAGACTTCCTGTATTATCTCACCCCTGATGGGCAGCGTTTCCGCAGGGCCTGCCGCCTGGTGC
ACGACTTCACAGATGCCGTCATCCAGGAGCGGCGCCGCACCCTCCCTAGCCAGGGTGTTGATGACTTCCT
CCAAGCCAAGGCCAAATCCAAGACTTTGGACTTCATTGATGTACTCCTGCTGAGCAAGGATGAAGATGGG
AAGAAGTTGTCCGATGAGGACATAAGAGCAGAAGCTGACACCTTTATGTTTGAGGGCCATGACACCACAG
CCAGTGGTCTCTCCTGGGTCCTGTACCACCTTGCAAAGCACCCGGAATACCAGGAGCGCTGTCGGCAGGA
GGTGCAAGAGCTTCTGAAGGACCGTGAGCCTAAAGAGATTGAATGGGACGACCTGGCCCAGCTGCCCTTC
CTGACCATGTGCATTAAGGAGAGCCTGAGGCTGCATCCCCCAGTCCCTGCCGTCTCTCGCTGCTGCACCC
AAGACATTGTGCTCCCAGACGGCCGGGTCATCCCCAAAGGCATTATCTGCCTCATCAGTGTTTTTGGAAC
CCATCACAACCCAGCCGTGTGGCCGGACCCTGAGGTCTATGACCCCTTTCGCTTTGACCCAAAGAACATC
AAGGAGAGGTCACCTCTGGCTTTTATTCCCTTCTCAGCAGGGCCCAGGAACTGCATCGGGCAGGCGTTCG
CGATGGCGGAGATGAAGGTGGTCCTGGGGCTCACGCTGCTGCGCTTCCGCGTCCTGCCTGACCACACCGA
GCCCCGCAGGAAGCCGGAGCTGGTCCTGCGCGCAGAGGGCGGACTTTGGCTGCGGGTGGAGCCCCTGAGC
TGAGTTCTGCAGAGACCCACTCTGACCCCACTAAAATGACCCCTGATTCATCAAAAGTGAGGCCTAGAAT

TACCCTAAGACCCTGTTCCACAGTCCTGTATTCCATCCTAGATATCTACTCAAAATAATTGAGACAAGTG
TTCAAACAGAAAGACGCTTGTGCGTGAATGTTCATGGCAGCCCTATTCACAGTAGCCAAACGATGAAAAC
AACCCCAAGCTATATATTACCAGATGAAAGGATAAACAAAATATGGTCCATCCATACAATGGAGTATTAC
ACAGCCATAAAAAGGAATGAAGCAGTGATCCCCACTACACTGTGGATGAACCTTGAATGCATGATACTGA
ATGAAAGACATCAGATGCAAAAGGTCACATAGTGTACTGTCCTTTTATATGAAATTTCCAGAACAGGCCA
ATCTGAAGAGATGTATAGTGGATTGGTGGCTTTCAGCAGCTGTGGGGAGGTGGGACTGAGGAGCGACTGC
TAATCAGGATGGGGTTTCCTCCTGGGATGGTGAAAATGTTCCGGACCTAGATAGTGATGAAGGTAGCACG
ACACTGTGAGTGCACTAAATGCTATTGAATTGGACACTTTAGAATGGTTGAAATAGTGATTTTTATGTGA
ATTCTACCTAAACATGCTATTACAGCTCATATATACTTTTTCCATCTGGATTCTTCACAAAAGAATATGT
TGTGAGCATCTTTCCATGATATTAAATCATCTTAGGAAACATTATTTTGTGTTCTTCAAAATGTGCATGT
TAAGTATTCAAATCAGTCTTAAATTTTTAAAAATATGTAATTTTAGAAAATAATTTAAAAGGTTTTGTTT
CAGTTTGTAAGATTTCTTTTCTGGCACTTTAATGGCTTGAGGTATCATTATCAGTTACAAATTGAGTTAT
TCTTCATCAAATGACTTTTGGAGTAGAGATTTTATTTTTATAGCAATAGATGCACAGATATTCCTGTAAG
ATACAGGTGTGGTTAGACACTTTTCTAGAACAGGCATGCCCTGCAAACTCCACAGACACTGACTGTTTTT
GTCCTATTAAGAAGTAGACCACTGAGAAGGGAGAAGGTGACATTTTAGCTTTCCCAGGTAAAAGTGGTTT
TCATCCTCACACCAATTTTATGGACTGGACGTTAACTCTCTTGCTCAAGGTCACTCTGAGTGGAAGAGTG
GGGATAAATCTGGTTCGTTTGGCATCAGAGGCCATGACTTTTCCTACCACAGAAGTAATTTTCAAAGTAA
GTCTCTGCCCTAGGCACATCAGATCACCTGGGGACCACTCCAGAGTGAGTAGACAAGACTTTGACAGGGG
TGCCTAATTTTTTTTTTTTTTTGAGATGGAGTCTCGCTCTGTTGCCCA (SEQ ID NO: 19)

Translated protein sequence

MPQLSLSSLGLWPMAASPWLLLLLVGASWLLARILAWTYTFYDN
CCRLRCFPQPPKRNWFLGHLGLIHSSEEGLLYTQSLACTFGDMCCWWVGPWHAIVRIF
HPTYIKPVLFAPAAIVPKDKVFYSFLKPWLGDGLLLSAGEKWSRHRRMLTPAFHFNIL
KPYMKIFNESVNIMHAKWQLLASEGSARLDMFEHISLMTLDSLQKCVFSFDSHCQEKP
SEYIAAILELSALVTKRHQQILLYIDFLYYLTPDGQRFRRACRLVHDFTDAVIQERRR
TLPSQGVDDFLQAKAKSKTLDFIDVLLLSKDEDGKKLSDEDIRAEADTFMFEGHDTTA
SGLSWVLYHLAKHPEYQERCRQEVQELLKDREPKEIEWDDLAQLPFLTMCIKESLRLH
PPVPAVSRCCTQDIVLPDGRVIPKGIICLISVFGTHHNPAVWPDPEVYDPFRFDPKNI
KERSPLAFIPFSAGPRNCIGQAFAMAEMKVVLGLTLLRFRVLPDHTEPRRKPELVLRA
EGGLWLRVEPLS (SEQ ID NO: 20)

DYSF	8291	NM_003494	Homo sapiens dysferlin, limb girdle muscular dystrophy 2B (autosomal recessive) (DYSF), transcript variant 8, mRNA

mRNA Sequence

GCGGCCGCCGCCCAGCCAGGTGCAAAATGCCGTGTCATTGGGAGACTCCGCAGCCGGAGCATTAGATTAC
AGCTCGACGGAGCTCGGGAAGGGCGGCGGGGGGTGGAAGATGAGCAGAAGCCCCTGTTCTCGGAACGCCGG
CTGACAAGCGGGGTGAGCGCAGCCGGGGCGGGGACCCAGCCTAGCCCACTGGAGCAGCCGGGGGTGGCCC
GTTCCCCTTTAAGAGCAACTGCTCTAAGCCAGGAGCCAGAGATTCGAGCCGGCCTCGCCCAGCCAGCCCT
CTCCAGCGAGGGGACCCACAAGCGGCGCCTCGGCCCTCCCGACCTTTCCGAGCCCTCTTTGCGCCCTGGG
CGCACGGGGCCCTACACGCGCCAAGCATGCTGAGGGTCTTCATCCTCTATGCCGAGAACGTCCACACACC
CGACACCGACATCAGCGATGCCTACTGCTCCGCGGTGTTTGCAGGGGTGAAGAAGAGAACCAAAGTCATC
AAGAACAGCGTGAACCCTGTATGGAATGAGGGATTTGAATGGGACCTCAAGGGCATCCCCCTGGACCAGG
GCTCTGAGCTTCATGTGGTGGTCAAAGACCATGAGACGATGGGGAGGAACAGGTTCCTGGGGGAAGCCAA
GGTCCCACTCCGAGAGGTCCTCGCCACCCCTAGTCTGTCCGCCAGCTTCAATGCCCCCCTGCTGGACACC
AAGAAGCAGCCCACAGGGGCCTCGCTGGTCCTGCAGGTGTCCTACACACCGCTGCCTGGAGCTGTGCCCC
TGTTCCCGCCCCCTACTCCTCTGGAGCCCTCCCCGACTCTGCCTGACCTGGATGTAGTGGCAGACACAGG
AGGAGAGGAAGACACAGAGGACCAGGGACTCACTGGAGATGAGGCGGAGCCATTCCTGGATCAAAGCGGA
GGCCCGGGGGCTCCCACCACCCCAAGGAAACTACCTTCACGTCCTCCGCCCCACTACCCCGGGATCAAAA
GAAAGCGAAGTGCGCCTACATCTAGAAAGCTGCTGTCAGACAAACCGCAGGATTTCCAGATCAGGGTCCA
GGTGATCGAGGGGCGCCAGCTGCCGGGGGTGAACATCAAGCCTGTGGTCAAGGTTACCGCTGCAGGGCAG
ACCAAGCGGACGCGGATCCACAAGGGAAACAGCCCACTCTTCAATGAGACTCTTTTCTTCAACTTGTTTG
ACTCTCCTGGGGAGCTGTTTGATGAGCCCATCTTTATCACGGTGGTAGACTCTCGTTCTCTCAGGACAGA
TGCTCTCCTCGGGGAGTTCCGGATGGACGTGGGCACCATTTACAGAGAGCCCCGGCACGCCTATCTCAGG
AAGTGGCTGCTGCTCTCAGACCCTGATGACTTCTCTGCTGGGGCCAGAGGCTACCTGAAAACAAGCCTTT
GTGTGCTGGGGCCTGGGGACGAAGCGCCTCTGGAGAGAAAAGACCCCTCTGAAGACAAGGAGGACATTGA
AAGCAACCTGCTCCGGCCCACAGGCGTAGCCCTGCGAGGAGCCCACTTCTGCCTGAAGGTCTTCCGGGCC
GAGGACTTGCCGCAGATGGACGATGCCGTGATGGACAACGTGAAACAGATCTTTGGCTTCGAGAGTAACA

AGAAGAACTTGGTGGACCCCTTTGTGGAGGTCAGCTTTGCGGGGAAAATGCTGTGCAGCAAGATCTTGGA
GAAGACGGCCAACCCTCAGTGGAACCAGAACATCACACTGCCTGCCATGTTTCCCTCCATGTGCGAAAAA
ATGAGGATTCGTATCATAGACTGGGACCGCCTGACTCACAATGACATCGTGGCTACCACCTACCTGAGTA
TGTCGAAAATCTCTGCCCCTGGAGGAGAAATAGAAGAGGAGCCTGCAGGTGCTGTCAAGCCTTCGAAAGC
CTCAGACTTGGATGACTACCTGGGCTTCCTCCCCACTTTTGGGCCCTGCTACATCAACCTCTATGGCAGT
CCCAGAGAGTTCACAGGCTTCCCAGACCCCTACACAGAGCTCAACACAGGCAAGGGGGAAGGTGTGGCTT
ATCGTGGCCGGCTTCTGCTCTCCCTGGAGACCAAGCTGGTGGAGCACAGTGAACAGAAGGTGGAGGACCT
TCCTGCGGATGACATCCTCCGGGTGGAGAAGTACCTTAGGAGGCGCAAGTACTCCCTGTTTGCGGCCTTC
TACTCAGCCACCATGCTGCAGGATGTGGATGATGCCATCCAGTTTGAGGTCAGCATCGGGAACTACGGGA
ACAAGTTCGACATGACCTGCCTGCCGCTGGCCTCCACCACTCAGTACAGCCGTGCAGTCTTTGACGGGTG
CCACTACTACTACCTACCCTGGGGTAACGTGAAACCTGTGGTGGTGCTGTCATCCTACTGGGAGGACATC
AGCCATAGAATCGAGACTCAGAACCAGCTGCTTGGGATTGCTGACCGGCTGGAAGCTGGCCTGGAGCAGG
TCCACCTGGCCCTGAAGGCGCAGTGCTCCACGGAGGACGTGGACTCGCTGGTGGCTCAGCTGACGGATGA
GCTCATCGCAGGCTGCAGCCAGCCTCTGGGTGACATCCATGAGACACCCTCTGCCACCCACCTGGACCAG
TACCTGTACCAGCTGCGCACCCATCACCTGAGCCAAATCACTGAGGCTGCCCTGGCCCTGAAGCTCGGCC
ACAGTGAGCTCCCTGCAGCTCTGGAGCAGGCGGAGGACTGGCTCCTGCGTCTGCGTGCCCTGGCAGAGGA
GCCCCAGAACAGCCTGCCGGACATCGTCATCTGGATGCTGCAGGGAGACAAGCGTGTGGCATACCAGCGG
GTGCCCGCCCACCAAGTCCTCTTCTCCCGGCGGGGTGCCAACTACTGTGGCAAGAATTGTGGGAAGCTAC
AGACAATCTTTCTGAAATATCCGATGGAGAAGGTGCCTGGCGCCCGGATGCCAGTGCAGATACGGGTCAA
GCTGTGGTTTGGGCTCTCAGTGGATGAGAAGGAGTTCAACCAGTTTGCTGAGGGGAAGCTGTCTGTCTTT
GCTGAAACCTATGAGAACGAGACTAAGTTGGCCCTTGTTGGGAACTGGGGCACAACGGGCCTCACCTACC
CCAAGTTTTCTGACGTCACGGGCAAGATCAAGCTACCCAAGGACAGCTTCCGCCCCTCGGCCGGCTGGAC
CTGGGCTGGAGATTGGTTCGTGTGTCCGGAGAAGACTCTGCTCCATGACATGGACGCCGGTCACCTGAGC
TTCGTGGAAGAGGTGTTTGAGAACCAGACCCGGCTTCCCGGAGGCCAGTGGATCTACATGAGTGACAACT
ACACCGATGTGAACGGGGAGAAGGTGCTTCCCAAGGATGACATTGAGTGCCCACTGGGCTGGAAGTGGGA
AGATGAGGAATGGTCCACAGACCTCAACCGGGCTGTCGATGAGCAAGGCTGGGAGTATAGCATCACCATC
CCCCCGGAGCGGAAGCCGAAGCACTGGGTCCCTGCTGAGAAGATGTACTACACACACCGACGGCGGCGCT
GGGTGCGCCTGCGCAGGAGGGATCTCAGCCAAATGGAAGCACTGAAAAGGCACAGGCAGGCGGAGGCGGA
GGGCGAGGGCTGGGAGTACGCCTCTCTTTTTGGCTGGAAGTTCCACCTCGAGTACCGCAAGACAGATGCC
TTCCGCCGCCGCCGCTGGCGCCGTCGCATGGAGCCACTGGAGAAGACGGGGCCTGCAGCTGTGTTTGCCC
TTGAGGGGGCCCTGGGCGGCGTGATGGATGACAAGAGTGAAGATTCCATGTCCGTCTCCACCTTGAGCTT
CGGTGTGAACAGACCCACGATTTCCTGCATATTCGACTATGGGAACCGCTACCATCTACGCTGCTACATG
TACCAGGCCCGGGACCTGGCTGCGATGGACAAGGACTCTTTTTCTGATCCCTATGCCATCGTCTCCTTCC
TGCACCAGAGCCAGAAGACGGTGGTGGTGAAGAACACCCTTAACCCCACCTGGGACCAGACGCTCATCTT
CTACGAGATCGAGATCTTTGGCGAGCCGGCCACAGTTGCTGAGCAACCGCCCAGCATTGTGGTGGAGCTG
TACGACCATGACACTTATGGTGCAGACGAGTTTATGGGTCGCTGCATCTGTCAACCGAGTCTGGAACGGA
TGCCACGGCTGGCCTGGTTCCCACTGACGAGGGGCAGCCAGCCGTCGGGGGAGCTGCTGGCCTCTTTTGA
GCTCATCCAGAGAGAGAAGCCGGCCATCCACCATATTCCTGGTTTTGAGGTGCAGGAGACATCAAGGATC
CTGGATGAGTCTGAGGACACAGACCTGCCCTACCCACCACCCCAGAGGGAGGCCAACATCTACATGGTTC
CTCAGAACATCAAGCCAGCGCTCCAGCGTACCGCCATCGAGATCCTGGCATGGGGCCTGCGGAACATGAA
GAGTTACCAGCTGGCCAACATCTCCTCCCCCAGCCTCGTGGTAGAGTGTGGGGGCCAGACGGTGCAGTCC
TGTGTCATCAGGAACCTCCGGAAGAACCCCAACTTTGACATCTGCACCCTCTTCATGGAAGTGATGCTGC
CCAGGGAGGAGCTCTACTGCCCCCCCATCACCGTCAAGGTCATCGATAACCGCCAGTTTGGCCGCCGGCC
TGTGGTGGGCCAGTGTACCATCCGCTCCCTGGAGAGCTTCCTGTGTGACCCCTACTCGGCGGAGAGTCCA
TCCCCACAGGGTGGCCCAGACGATGTGAGCCTACTCAGTCCTGGGGAAGACGTGCTCATCGACATTGATG
ACAAGGAGCCCCTCATCCCCATCCAGGAGGAAGAGTTCATCGATTGGTGGAGCAAATTCTTTGCCTCCAT
AGGGGAGAGGGAAAAGTGCGGCTCCTACCTGGAGAAGGATTTTGACACCCTGAAGGTCTATGACACACAG
CTGGAGAATGTGGAGGCCTTTGAGGGCCTGTCTGACTTTTGTAACACCTTCAAGCTGTACCGGGGCAAGA
CGCAGGAGGAGACAGAAGATCCATCTGTGATTGGTGAATTTAAGGGCCTCTTCAAAATTTATCCCCTCCC
AGAAGACCCAGCCATCCCCATGCCCCCAAGACAGTTCCACCAGCTGGCCGCCCAGGGACCCCAGGAGTGC
TTGGTCCGTATCTACATTGTCCGAGCATTTGGCCTGCAGCCCAAGGACCCCAATGGAAAGTGTGATCCTT
ACATCAAGATCTCCATAGGGAAGAAATCAGTGAGTGACCAGGATAACTACATCCCCTGCACGCTGGAGCC
CGTATTTGGAAAGATGTTCGAGCTGACCTGCACTCTGCCTCTGGAGAAGGACCTAAAGATCACTCTCTAT
GACTATGACCTCCTCTCCAAGGACGAAAAGATCGGTGAGACGGTCGTCGACCTGGAGAACAGGCTGCTGT
CCAAGTTTGGGGCTCGCTGTGGACTCCCACAGACCTACTGTGTCTCTGGACCGAACCAGTGGCGGGACCA
GCTCCGCCCCTCCCAGCTCCTCCACCTCTTCTGCCAGCAGCATAGAGTCAAGGCACCTGTGTACCGGACA
GACCGTGTAATGTTTCAGGATAAAGAATATTCCATTGAAGAGATAGAGGCTGGCAGGATCCCAAACCCAC
ACCTGGGCCCAGTGGAGGAGCGTCTGGCTCTGCATGTGCTTCAGCAGCAGGGCCTGGTCCCGGAGCACGT
GGAGTCACGGCCCCTCTACAGCCCCCTGCAGCCAGACATCGAGCAGGGGAAGCTGCAGATGTGGGTCGAC
CTATTTCCGAAGGCCCTGGGGCGGCCTGGACCTCCCTTCAACATCACCCCACGGAGAGCCAGAAGGTTTT
TCCTGCGTTGTATTATCTGGAATACCAGAGATGTGATCCTGGATGACCTGAGCCTCACGGGGGAGAAGAT

GAGCGACATTTATGTGAAAGGTTGGATGATTGGCTTTGAAGAACACAAGCAAAAGACAGACGTGCATTAT
CGTTCCCTGGGAGGTGAAGGCAACTTCAACTGGAGGTTCATTTTCCCCTTCGACTACCTGCCAGCTGAGC
AAGTCTGTACCATTGCCAAGAAGGATGCCTTCTGGAGGCTGGACAAGACTGAGAGCAAAATCCCAGCACG
AGTGGTGTTCCAGATCTGGGACAATGACAAGTTCTCCTTTGATGATTTTCTGGGCTCCCTGCAGCTCGAT
CTCAACCGCATGCCCAAGCCAGCCAAGACAGCCAAGAAGTGCTCCTTGGACCAGCTGGATGATGCTTTCC
ACCCAGAATGGTTTGTGTCCCTTTTTGAGCAGAAAACAGTGAAGGGCTGGTGGCCCTGTGTAGCAGAAGA
GGGTGAGAAGAAAATACTGGCGGGCAAGCTGGAAATGACCTTGGAGATTGTAGCAGAGAGTGAGCATGAG
GAGCGGCCTGCTGGCCAGGGCCGGGATGAGCCCAACATGAACCCTAAGCTTGAGGACCCAAGGCGCCCCG
ACACCTCCTTCCTGTGGTTTACCTCCCCATACAAGACCATGAAGTTCATCCTGTGGCGGCGTTTCCGGTG
GGCCATCATCCTCTTCATCATCCTCTTCATCCTGCTGCTGTTCCTGGCCATCTTCATCTACGCCTTCCCG
AACTATGCTGCCATGAAGCTGGTGAAGCCCTTCAGCTGAGGACTCTCCTGCCCTGTAGAAGGGGCCGTGG
GGTCCCCTCCAGCATGGGACTGGCCTGCCTCCTCCGCCCAGCTCGGCGAGCTCCTCCAGACCTCCTAGGC
CTGATTGTCCTGCCAGGGTGGGCAGACAGACAGATGGACCGGCCCACACTCCCAGAGTTGCTAACATGGA
GCTCTGAGATCACCCCACTTCCATCATTTCCTTCTCCCCCAACCCAACGCTTTTTTGGATCAGCTCAGAC
ATATTTCAGTATAAAACAGTTGGAACCACAAAAAAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 21)

Translated protein sequence

MLRVFILYAENVHTPDTDISDAYCSAVFAGVKKRTKVIKNSVNP
VWNEGFEWDLKGIPLDQGSELHVVVKDHETMGRNRFLGEAKVPLREVLATPSLSASFN
APLLDTKKQPTGASLVLQVSYTPLPGAVPLFPPPTPLEPSPTLPDLDVVADTGGEEDT
EDQGLTGDEAEPFLDQSGGPGAPTTPRKLPSRPPPHYPGIKRKRSAPTSRKLLSDKPQ
DFQIRVQVIEGRQLPGVNIKPVVKVTAAGQTKRTRIHKGNSPLFNETLFFNLFDSPGE
LFDEPIFITVVDSRSLRTDALLGEFRMDVGTIYREPRHAYLRKWLLLSDPDDFSAGAR
GYLKTSLCVLGPGDEAPLERKDPSEDKEDIESNLLRPTGVALRGAHFCLKVFRAEDLP
QMDDAVMDNVKQIFGFESNKKNLVDPFVEVSFAGKMLCSKILEKTANPQWNQNITLPA
MFPSMCEKMRIRIIDWDRLTHNDIVATTYLSMSKISAPGGEIEEEPAGAVKPSKASDL
DDYLGFLPTFGPCYINLYGSPREFTGFPDPYTELNTGKGEGVAYRGRLLLSLETKLVE
HSEQKVEDLPADDILRVEKYLRRRKYSLFAAFYSATMLQDVDDAIQFEVSIGNYGNKF
DMTCLPLASTTQYSRAVFDGCHYYYLPWGNVKPVVVLSSYWEDISHRIETQNQLLGIA
DRLEAGLEQVHLALKAQCSTEDVDSLVAQLTDELIAGCSQPLGDIHETPSATHLDQYL
YQLRTHHLSQITEAALALKLGHSELPAALEQAEDWLLRLRALAEEPQNSLPDIVIWML
QGDKRVAYQRVPAHQVLFSRRGANYCGKNCGKLQTIFLKYPMEKVPGARMPVQIRVKL
WFGLSVDEKEFNQFAEGKLSVFAETYENETKLALVGNWGTTGLTYPKFSDVTGKIKLP
KDSFRPSAGWTWAGDWFVCPEKTLLHDMDAGHLSFVEEVFENQTRLPGGQWIYMSDNY
TDVNGEKVLPKDDIECPLGWKWEDEEWSTDLNRAVDEQGWEYSITIPPERKPKHWVPA
EKMYYTHRRRRWVRLRRRDLSQMEALKRHRQAEAEGEGWEYASLFGWKFHLEYRKTDA
FRRRRWRRRMEPLEKTGPAAVFALEGALGGVMDDKSEDSMSVSTLSFGVNRPTISCIF
DYGNRYHLRCYMYQARDLAAMDKDSFSDPYAIVSFLHQSQKTVVVKNTLNPTWDQTLI
FYEIEIFGEPATVAEQPPSIVVELYDHDTYGADEFMGRCICQPSLERMPRLAWFPLTR
GSQPSGELLASFELIQREKPAIHHIPGFEVQETSRILDESEDIDLPYPPPQREANIYM
VPQNIKPALQRTAIEILAWGLRNMKSYQLANISSPSLVVECGGQTVQSCVIRNLRKNP
NFDICTLFMEVMLPREELYCPPITVKVIDNRQFGRRPVVGQCTIRSLESFLCDPYSAE
SPSPQGGPDDVSLLSPGEDVLIDIDDKEPLIPIQEEEFIDWWSKFFASIGEREKCGSY
LEKDFDTLKVYDTQLENVEAFEGLSDFCNTFKLYRGKTQEETEDPSVIGEFKGLFKIY
PLPEDPAIPMPPRQFHQLAAQGPQECLVRIYIVRAFGLQPKDPNGKCDPYIKISIGKK
SVSDQDNYIPCTLEPVFGKMFELTCTLPLEKDLKITLYDYDLLSKDEKIGETVVDLEN
RLLSKFGARCGLPQTYCVSGPNQWRDQLRPSQLLHLFCQQHRVKAPVYRTDRVMFQDK
EYSIEEIEAGRIPNPHLGPVEERLALHVLQQQGLVPEHVESRPLYSPLQPDIEQGKLQ
MWVDLFPKALGRPGPPFNITPRRARRFFLRCIIWNTRDVILDDLSLTGEKMSDIYVKG
WMIGFEEHKQKTDVHYRSLGGEGNFNWRFIFPFDYLPAEQVCTIAKKDAFWRLDKTES
KIPARVVFQIWDNDKFSFDDFLGSLQLDLNRMPKPAKTAKKCSLDQLDDAFHPEWFVS
LFEQKTVKGWWPCVAEEGEKKILAGKLEMTLEIVAESEHEERPAGQGRDEPNMNPKLE
DPRRPDTSFLWFTSPYKTMKFILWRRFRWAIILFIILFILLLFLAIFIYAFPNYAAMK
LVKPFS (SEQ ID NO: 22)

EDIL3	10085	NM_005711	Homo sapiens EGF-like repeats and discoidin I-like domains 3 (EDIL3), mRNA

mRNA Sequence

AGAAGCCCCGCAGCCGCCGCGCGGAGAACAGCGACAGCCGAGCGCCCGGTCCGCCTGTCTGCCGGTGGGT
CTGCCTGCCCGCGCAGCAGACCCGGGGCGGCCGCGGGAGCCCGCGCCCCGCCCGCCGCGCCTCTGCCGGG
ACCCACCCGCAGCGGAGGGCTGAGCCCGCCGGCGGCTCCCCGGAGCTCACCCACCTCCGCGCGCCGGAGC
GCAGGCAAAAGGGGAGGAAAGGCTCCTCTCTTTAGTCACCACTCTCGCCCTCTCCAAGAATTTGTTTAAC
AAAGCGCTGAGGAAAGAGAACGTCTTCTTGAATTCTTTAGTAGGGGCGGAGTCTGCTGCTGCCCTGCGCT
GCCACCTCGGCTACACTGCCCTCCGCGACGACCCCTGACCAGCCGGGGTCACGTCCGGGAGACGGGATCA
TGAAGCGCTCGGTAGCCGTCTGGCTCTTGGTCGGGCTCAGCCTCGGTGTCCCCCAGTTCGGCAAAGGTGA
TATTTGTGATCCCAATCCATGTGAAAATGGAGGTATCTGTTTGCCAGGATTGGCTGATGGTTCCTTTTCC
TGTGAGTGTCCAGATGGCTTCACAGACCCCAACTGTTCTAGTGTTGTGGAGGTTGCATCAGATGAAGAAG
AACCAACTTCAGCAGGTCCCTGCACTCCTAATCCATGCCATAATGGAGGAACCTGTGAAATAAGTGAAGC
ATACCGAGGGGATACATTCATAGGCTATGTTTGTAAATGTCCCCGAGGATTTAATGGGATTCACTGTCAG
CACAACATAAATGAATGCGAAGTTGAGCCTTGCAAAAATGGTGGAATATGTACAGATCTTGTTGCTAACT
ATTCCTGTGAGTGCCCAGGCGAATTTATGGGAAGAAATTGTCAATACAAATGCTCAGGCCCACTGGGAAT
TGAAGGTGGAATTATATCAAACCAGCAAATCACAGCTTCCTCTACTCACCGAGCTCTTTTTGGACTCCAA
AAATGGTATCCCTACTATGCACGTCTTAATAAGAAGGGGCTTATAAATGCGTGGACAGCTGCAGAAAATG
ACAGATGGCCGTGGATTCAGATAAATTTGCAAAGGAAAATGAGAGTTACTGGTGTGATTACCCAAGGAGC
CAAGAGGATTGGAAGCCCAGAGTATATAAAATCCTACAAAATTGCCTACAGTAATGATGGAAAGACTTGG
GCAATGTACAAAGTGAAAGGCACCAATGAAGACATGGTGTTTCGTGGAAACATTGATAACAACACTCCAT
ATGCTAACTCTTTCACACCCCCCATAAAAGCTCAGTATGTAAGACTCTATCCCCAAGTTTGTCGAAGACA
TTGCACTTTGCGAATGGAACTTCTTGGCTGTGAACTGTCGGGTTGTTCTGAGCCTCTGGGTATGAAATCA
GGACATATACAAGACTATCAGATCACTGCCTCCAGCATCTTCAGAACGCTCAACATGGACATGTTCACTT
GGGAACCAAGGAAAGCTCGGCTGGACAAGCAAGGCAAAGTGAATGCCTGGACCTCTGGCCACAATGACCA
GTCACAATGGTTACAGGTGGATCTTCTTGTTCCAACCAAAGTGACTGGCATCATTACACAAGGAGCTAAA
GATTTTGGTCATGTACAGTTTGTTGGCTCCTACAAACTGGCTTACAGCAATGATGGAGAACACTGGACTG
TATACCAGGATGAAAAGCAAAGAAAAGATAAGGTTTTCCAGGGAAATTTTGACAATGACACTCACAGAAA
AAATGTCATCGACCCTCCCATCTATGCACGACACATAAGAATCCTTCCTTGGTCCTGGTACGGGAGGATC
ACATTGCGGTCAGAGCTGCTGGGCTGCACAGAGGAGGAATGAGGGGAGGCTACATTTCACAACCCTCTTC
CCTATTTCCCTAAAAGTATCTCCATGGAATGAACTGTGCAAAATCTGTAGGAAACTGAATGGTTTTTTTT
TTTTTTCATGAAAAAGTGCTCAAATTATGGTAGGCAACTAACGGTGTTTTTAAGGGGGTCTAAGCCTGC
CTTTTCAATGATTTAATTTGATTTTATTTTATCCGTCAAATCTCTTAAGTAACAACACATTAAGTGTGAA
TTACTTTTCTCTCATTGTTTCCTGAATTATTCGCATTGGTAGAAATATATTAGGGAAAGAAAGTAGCCTT
CTTTTTATAGCAAGAGTAAAAAAGTCTCAAAGTCATCAAATAAGAGCAAGAGTTGATAGAGCTTTTACAA
TCAATACTCACCTAATTCTGATAAAAGGAATACTGCAATGTTAGCAATAAGTTTTTTTCTTCTGTAATGA
CTCTACGTTATCCTGTTTCCCTGTGCCTACCAAACACTGTCAATGTTTATTACAAAATTTTAAAGAAGAA
TATGTAACATGCAGTACTGATATTATAATTCTCATTTTACTTTCATTATTTCTAATAAGAGATTATGTGA
CTTCTTTTTCTTTTAGTTCTATTCTACATTCTTAATATTGTATATTACCTGAATAATTCAATTTTTTTCT
AATTGAATTTCCTATTAGTTGACTAAAAGAAGTGTCATGTTTACTCATATATGTAGAACATGACTGCCTA
TCAGTAGATTGATCTGTATTTAATATTCGTTAATTAAATCTGCAGTTTTATTTTTGAAGGAAGCCATAAC
TATTTAATTTCCAAATAATTGCTTCATAAAGAATCCCATACTCTCAGTTTGCACAAAAGAACAAAAAATA
TATATGTCTCTTTAAATTTAAATCTTCATTTAGATGGTAATTACATATCCTTATATTTACTTTAAAAAAT
CGGCTTATTTGTTTATTTTATAAAAAATTTAGCAAAGAAATATTAATATAGTGCTGCATAGTTTGGCCAA
GCATACTCATCATTTCTTTGTTCAGCTCCACATTTCCTGTGAAACTAACATCTTATTGAGATTTGAAACT
GGTGGTAGTTTCCCAGGAAGGCACAGGTGGAGTT (SEQ ID NO: 23)

Translated protein sequence

MKRSVAVWLLVGLSLGVPQFGKGDICDPNPCENGGICLPGLADG
SFSCECPDGFTDPNCSSVVEVASDEEEPTSAGPCTPNPCHNGGTCEISEAYRGDTFIG
YVCKCPRGFNGIHCQHNINECEVEPCKNGGICTDLVANYSCECPGEFMGRNCQYKCSG
PLGIEGGIISNQQITASSTHRALFGLQKWYPYYARLNKKGLINAWTAAENDRWPWIQI
NLQRKMRVTGVITQGAKRIGSPEYIKSYKIAYSNDGKTWAMYKVKGTNEDMVFRGNID
NNTPYANSFTPPIKAQYVRLYPQVCRRHCTLRMELLGCELSGCSEPLGMKSGHIQDYQ
ITASSIFRTLNMDMFTWEPRKARLDKQGKVNAWTSGHNDQSQWLQVDLLVPTKVTGII
TQGAKDFGHVQFVGSYKLAYSNDGEHWTVYQDEKQRKDKVFQGNFDNDTHRKNVIDPP
IYARHIRILPWSWYGRITLRSELLGCTEEE (SEQ ID NO: 24)

ERGIC3	51614	NM_015966	Homo sapiens ERGIC and golgi 3 (ERGIC3), transcript variant 2, mRNA

mRNA Sequence

GTGGCTCCAGGCCGGAAGAGGGAGTCTGTAGGGGCGGGCCGGCTGGCGTCCCCTTTCCGGCCGGTCCCCA

TGGAGGCGCTGGGGAAGCTGAAGCAGTTCGATGCCTACCCCAAGACTTTGGAGGACTTCCGGGTCAAGAC
CTGCGGGGGCGCCACCGTGACCATTGTCAGTGGCCTTCTCATGCTGCTACTGTTCCTGTCCGAGCTGCAG
TATTACCTCACCACGGAGGTGCATCCTGAGCTCTACGTGGACAAGTCGCGGGGAGATAAACTGAAGATCA
ACATCGATGTACTTTTTCCGCACATGCCTTGTGCCTATCTGAGTATTGATGCCATGGATGTGGCCGGAGA
ACAGCAGCTGGATGTGGAACACAACCTGTTCAAGCAACGACTAGATAAAGATGGCATCCCCGTGAGCTCA
GAGGCTGAGCGGCATGAGCTTGGGAAAGTCGAGGTGACGGTGTTTGACCCTGACTCCCTGGACCCTGATC
GCTGTGAGAGCTGCTATGGTGCTGAGGCAGAAGATATCAAGTGCTGTAACACCTGTGAAGATGTGCGGGA
GGCATATCGCCGTAGAGGCTGGGCCTTCAAGAACCCAGATACTATTGAGCAGTGCCGGCGAGAGGGCTTC
AGCCAGAAGATGCAGGAGCAGAAGAATGAAGGCTGCCAGGTGTATGGCTTCTTGGAAGTCAATAAGGTGG
CCGGAAACTTCCACTTTGCCCCTGGGAAGAGCTTCCAGCAGTCCCATGTGCACGTCCATGACTTGCAGAG
CTTTGGCCTTGACAACATCAACATGACCCACTACATCCAGCACCTGTCATTTGGGGAGGACTATCCAGGC
ATTGTGAACCCCCTGGACCACACCAATGTCACTGCGCCCCAAGCCTCCATGATGTTCCAGTACTTTGTGA
AGGTGGTGCCCACTGTGTACATGAAGGTGGACGGAGAGGTACTGAGGACAAATCAGTTCTCTGTGACCAG
ACATGAGAAGGTTGCCAATGGGCTGTTGGGCGACCAAGGCCTTCCCGGAGTCTTCGTCCTCTATGAGCTC
TCGCCCATGATGGTGAAGCTGACGGAGAAGCACAGGTCCTTCACCCACTTCCTGACAGGTGTGTGCGCCA
TCATTGGGGGCATGTTCACAGTGGCTGGACTCATCGATTCGCTCATCTACCACTCAGCACGAGCCATCCA
GAAGAAAATTGATCTAGGGAAGACAACGTAGTCACCCTCGGTGCTTCCTCTGTCTCCTCTTTCTCCCTGG
CCTGTGGTTGTCCCCCAGCCTCTGCCACCCTCCACCTCCTCGGTCAGCCCCAGCCCCAGGTTGATAAATC
TATTGATTGATTGTGATAGTAAAAAAAAAAAAAAAAAA (SEQ ID NO: 25)

Translated protein sequence

MEALGKLKQFDAYPKTLEDFRVKTCGGATVTIVSGLLMLLLFLS
ELQYYLTTEVHPELYVDKSRGDKLKINIDVLFPHMPCAYLSIDAMDVAGEQQLDVEHN
LFKQRLDKDGIPVSSEAERHELGKVEVTVFDPDSLDPDRCESCYGAEAEDIKCCNTCE
DVREAYRRRGWAFKNPDTIEQCRREGFSQKMQEQKNEGCQVYGFLEVNKVAGNFHFAP
GKSFQQSHVHVHDLQSFGLDNINMTHYIQHLSFGEDYPGIVNPLDHTNVTAPQASMMF
QYFVKVVPTVYMKVDGEVLRTNQFSVTRHEKVANGLLGDQGLPGVFVLYELSPMMVKL
TEKHRSFTHFLTGVCAIIGGMFTVAGLIDSLIYHSARAIQKKIDLGKTT (SEQ ID NO: 26)

GNAT1	2779	NM_000172	Homo sapiens guanine nucleotide binding protein (G protein), alpha transducing activity polypeptide 1 (GNAT1), transcript variant 2, mRNA

mRNA Sequence

AGTTGATTGCAGGTCCTCCTGGGGCCAGAAGGGTGCCTGGGAGGCCAGGTTCTGGGGATCCCCTCCATCC
AGAAGAACCACCTGCTCACTCTGTCCCTTCGCCTGCTGCTGGGACCATGGGGGCTGGGGCCAGTGCTGAG
GAGAAGCACTCCAGGGAGCTGGAAAAGAAGCTGAAAGAGGACGCTGAGAAGGATGCTCGAACCGTGAAGC
TGCTGCTTCTGGGTGCCGGTGAGTCCGGGAAGAGCACCATCGTCAAGCAGATGAAGATTATCCACCAGGA
CGGGTACTCGCTGGAAGAGTGCCTCGAGTTTATCGCCATCATCTACGGCAACACGTTGCAGTCCATCCTG
GCCATCGTACGCGCCATGACCACACTCAACATCCAGTACGGAGACTCTGCACGCCAGGACGACGCCCGGA
AGCTGATGCACATGGCAGACACTATCGAGGAGGGCACGATGCCCAAGGAGATGTCGGACATCATCCAGCG
GCTGTGGAAGGACTCCGGTATCCAGGCCTGTTTTGAGCGCGCCTCGGAGTACCAGCTCAACGACTCGGCG
GGCTACTACCTCTCCGACCTGGAGCGCCTGGTAACCCCGGGCTACGTGCCCACCGAGCAGGACGTGCTGC
GCTCGCGAGTCAAGACCACTGGCATCATCGAGACGCAGTTCTCCTTCAAGGATCTCAACTTCCGGATGTT
CGATGTGGGCGGGCAGCGCTCGGAGCGCAAGAAGTGGATCCACTGCTTCGAGGGCGTGACCTGCATCATC
TTCATCGCGGCGCTGAGCGCCTACGACATGGTGCTAGTGGAGGACGACGAAGTGAACCGCATGCACGAGA
GCCTGCACCTGTTCAACAGCATCTGCAACCACCGCTACTTCGCCACGACGTCCATCGTGCTCTTCCTTAA
CAAGAAGGACGTCTTCTTCGAGAAGATCAAGAAGGCGCACCTCAGCATCTGTTTCCCGGACTACGATGGA
CCCAACACCTACGAGGACGCCGGCAACTACATCAAGGTGCAGTTCCTCGAGCTCAACATGCGGCGCGACG
TGAAGGAGATCTATTCCCACATGACGTGCGCCACCGACACGCAGAACGTCAAATTTGTCTTCGACGCTGT
CACCGACATCATCATCAAGGAGAACCTCAAAGACTGTGGCCTCTTCTGAGGCCAGGGCCTGTGCTGCAGT
CGGGGACAAGGAGCTTCCGTCTGGCAAGGCCGGGGCACAATTTGCACTCCCCTCAGCTAGACGCACAGAC
TCAGCAATAAACCTTTGCATCAGGCTCCAGCTGTCCTTTCTTGGTGGAGGACTTAATTATCACAAGTCAT
GGGCATTTATTAAGTGCCCAGTGCTGGGTTGGGCATGAAGTGGGAAGATGGCCCCTCCCAGGAAGAAGTA
CCTGGCCTGACAAGGTGGGGCACTCTTGGGGGTATGGGACCAACTCATGGCTTTTCACGGGAGTTGAGGA
GAGAGGAGCTGTGGAAAATATTCACTGGGACAGTCTTGGATCAAGAGGGAGTTTTGAGGTGGAGGCTCAT
TCTGGCAGGGACCGTAGTGTCTACCAGCCCCAGAAACATGGGCTTATGGCCACAGGAGTTCAGTGGAGCA
AGAGCAGGGGAGGAGAGACGTGGACAGGTGCCCAAAGCCAGTCGGAGGGCCTGGGCTTTCTCAGAAGGTG
ATGGAGAGTCTTGGAAGCCCTCGAGGCAGGAACATAATTGCAGGGCTGGGATTAGGGTGAGGGAAGTGAG

GCACACTCACCTTGGGTGCAACATTTAAGGCGATGCCAAAAAATTTAGTAACCAAGGTAAATAATATTAG
GATAATATTTTTAAAAATCAAATGAATGCAAAACCCCACAATGAATGAAATATCAAAATCCAACAGAGGA
TCAAACAGAGGCATGCTAAGATATATTGGGGCTTGAAGCAAAGGGAAAACTATTTGTTGCTATATGTTTG
TAGGGATTTTTTGCCAGTTTTAAAAATACATGTATCATAAAGTTTACTATCTCAGCCACTTGCCGGTGTA
TAGTTTGGTGGTGTTAAGTACATTCATAATGTTGTACAACCACCGCAACTGTTCATCTCCAGAACTCCTT
TCCTCTTGTAAAACTGTAACTCTGTACCCATGAAAAAATAACCCCCCATTCCTGCCTTCCCCCGGCTCCT
GGCATCCACCATTCTACTTTCCATCTCTATGAATGTGACTGCTCTAAGTGCCTCAGATGTGTGGGTCCAT
GAAGTCTTTGTCTTTTTGCAACTGGCTTATTTCACTTAGCATCATGTCTTCAAGGTTTATTCATGTGTAG
CATATGGCAGAATCTCCTTCCTTTTTAAGGTTGAATAATATTCCATTGTATATATTCCACACTTTGTTTA
TTTATTCATCTATTGATGAATGGTTACATCTGCCTTTTGGCTATTGTGAATAATGCTGCTATGAACATGG
GTGTACAAATCTCTCAAAAAAAAAAAAAAAAAA (SEQ ID NO: 27)

Translated protein sequence

MGAGASAEEKHSRELEKKLKEDAEKDARTVKLLLLGAGESGKST
IVKQMKIIHQDGYSLEECLEFIAIIYGNTLQSILAIVRAMTTLNIQYGDSARQDDARK
LMHMADTIEEGTMPKEMSDIIQRLWKDSGIQACFERASEYQLNDSAGYYLSDLERLVT
PGYVPTEQDVLRSRVKTTGIIETQFSFKDLNFRMFDVGGQRSERKKWIHCFEGVTCII
FIAALSAYDMVLVEDDEVNRMHESLHLFNSICNHRYFATTSIVLFLNKKDVFFEKIKK
AHLSICFPDYDGPNTYEDAGNYIKVQFLELNMRRDVKEIYSHMTCATDTQNVKFVFDA
VTDIIIKENLKDCGLF (SEQ ID NO: 28)

GRAMD4	23151	NM_015124	Homo sapiens GRAM domain containing 4 (GRAMD4), mRNA

mRNA Sequence

CGTCATGTTAGGGTGAAGCAGAGGACCTCAGTGCTGAACATGCTAAGGAGGTTGGACAAAATCAGGTTCA
GAGGTCACAAGAGAGATGACTTCCTCGATCTAGCGGAGTCTCCAAATGCCTCGGACACCGAATGCAGCGA
CGAAATCCCCCTGAAGGTACCGCGGACCTCGCCCCGGGACAGCGAGGAGCTGAGGGACCCTGCTGGTCCA
GGGACCCTCATCATGGCCACAGGAGTCCAGGACTTTAACCGGACAGAGTTTGATCGACTGAATGAGATCA
AAGGTCACCTGGAAATTGCCTTATTGGAAAAACATTTCTTACAGGAGGAGCTCCGGAAGCTGCGAGAAGA
AACCAACGCGGAGATGCTGCGGCAGGAGCTGGACCGCGAGCGGCAGCGGCGGATGGAGCTGGAGCAGAAG
GTGCAGGAGGTGCTGAAGGCCAGAACCGAGGAGCAGATGGCTCAGCAGCCCCCAAAAGGGCAGGCCCAGG
CCAGCAATGGAGCAGAGCGCCGGAGCCAGGGGCTGTCCTCGCGCCTGCAGAAGTGGTTCTACGAGCGGTT
CGGGGAGTACGTGGAGGACTTCCGGTTCCAGCCCGAGGAGAACACTGTGGAGACAGAGGAACCCCTGAGC
GCCCGCAGGTTAACTGAAAATATGAGACGGCTCAAGCGCGGTGCCAAGCCGGTCACTAACTTTGTGAAGA
ACCTCTCTGCCTTATCCGACTGGTACTCCGTCTACACGTCTGCCATTGCCTTCACCGTGTACATGAATGC
CGTGTGGCATGGCTGGGCCATCCCATTGTTCTTATTTCTAGCAATTCTGAGGTTATCCCTCAATTACCTC
ATCGCCAGGGGGTGGCGGATACAGTGGAGCATCGTGCCCGAAGTGTCTGAGCCCGTGGAACCTCCAAAGG
AAGACCTGACTGTGTCTGAGAAGTTCCAGCTGGTGCTGGACGTCGCCCAGAAAGCCCAGAACCTTTTCGG
GAAGATGGCTGACATCCTGGAGAAGATCAAGAACTTGTTCATGTGGGTCCAGCCGGAGATCACACAGAAG
CTGTATGTGGCGCTCTGGGCTGCCTTCCTGGCCTCCTGCTTCTTCCCCTACCGCCTGGTGGGGCTTGCCG
TGGGACTCTATGCTGGTATCAAGTTCTTCCTCATTGATTTCATCTTTAAACGCTGCCCGAGGCTGCGCGC
CAAGTACGACACGCCCTATATCATCTGGAGGAGTCTCCCCACCGACCCGCAGCTCAAGGAGCGCTCCAGC
GCCGCAGTCTCACGCAGGCTGCAGACGACCTCGTCACGGAGCTACGTACCCAGCGCACCGGCCGGCCTGG
GTAAAGAGGAGGACGCCGGTCGCTTCCACAGCACCAAGAAGGGCAATTTCCACGAGATCTTCAATCTGAC
AGAAAACGAGCGTCCGCTGGCGGTGTGCGAGAATGGCTGGCGCTGCTGCCTCATCAACAGGGACCGGAAG
ATGCCCACGGACTACATCAGGAACGGGGTGCTCTACGTCACGGAGAATTACTTGTGCTTCGAAAGCTCCA
AATCTGGGTCCTCAAAGAGGAACAAAGTCATCAAGCTAGTGGACATCACGGACATCCAGAAGTACAAGGT
CCTGTCTGTCCTCCCAGGCTCAGGCATGGGGATTGCCGTGTCGACGCCATCCACCCAGAAACCGCTCGTG
TTTGGTGCCATGGTGCACAGGGATGAGGCCTTCGAGACCATTCTCAGCCAGTACATCAAGATCACCTCAG
CGGCAGCGTCTGGCGGGGACAGCTAGTATTGACTTGCCCAGGACGTTGCTGGAATTTTCTTTTTCTTTTT
CTTTTTCTTTTTTTTTTTTTACGATTTGGTAGTGGAAACAATTGGACATCCTCATGAGCTTTTGCAATAA
TTCTCCTGGACCTGTGGTTCTATTGTGTTGACCTCTGCGTTTTATCGACCAAGAAGGGGCCAGGGCTCAC
AGGGACGGGGGTGCCCCTCTCCCACAGGGCACGTCAGGTGCCTCTGAGGGCCACCCGCAGACTGGGGGAG
GGGGCAGAGGCCCTCGGGGGCCCGTGGAGAAGACACACAGGACCCCTGGCCCTGCCCTTCTCCGTTCCAG
CCTGGACAGAGAAACCTCTCCAGCCACCCCAAGAGGTTCTCGCAACCTTGTGTCCCGCTCTCCAGAGGCC
AGAAGCTCGTCCACCACCAAAGCCATAGCTGAAGAGTGCGGGGCCCTTCCTCCTGGGGACAGAAAGATGT
CGTCAAGGAGGGACATGGGGGCCTTTCACCAACCACCGAGAAACGGGCCTGGCGGCCCTCCTTCCTCTTA
CATGAGACCCTCCTGTGGCATTTGCCCTTGGTGCCGGGCTGGGGCCGGGCGCAGTGACCCTGCCTGCGCT
CCACACTCGCTCCACGGGAACAGAGAGGGTGAGAAGGGCCCACCCCTCGCCTGCCCTCAGTGTCTTTGGT

GGCACCTTCCTTGCTGGCCTCCAGGGCGCTCAGCACCGCGTCTGTAAGGGCCTGCCTGCTGCTCTCGGCC
TGACACGCCGGCCAGGAGGTCTGTAGCTGGGGACCAGTAAGGGCACAGGATGGTGCAGGTAAAAGCACAT
CTTTCTCACACTTTGCTCTTTGGAAGGCCCAGGAGAACATCCGCGAAGGCTGTTGGAGGTGCTCCGAGCA
CTGTGGCATGTCTGGCACATGGCCCCCAGGCTGCGGTTGCCTGGGTTGGTTGGGGGAGGAAGTGGGGAGG
AGTGTTCCGGGACCATGGTGGCCCAGGCTGCAGCCGCCTTTGGGCCATCCGAGAGGCTCTGGCAGCCCCT
GTGCTTTAGGGAGCAACCGTGAGCCGAGCCCAGAGGCCTGGGCCTGCACTGCCTGCAGCCGACATGCGAC
AGCGTTCCCTCCCCCGCGTGCCTAGCCGGTGCCGGTCCGGGCACAGACCCCCCCAGCCCCCGCCCTGCCC
CAGGGAAGCCTGGGCTTCCCGGGAACAAGGTGGCATTTGTGGAGGGAGCGCCCGCAGGCCTGGTCTGCTG
GGGCCGCCTGCGCTGGGCTGAAGGGAGGGAAAGGCGGCTTGGGCCTCCTGGAAGGAGGTGGCCACCCCGC
GGGCCTGCGTGTCTGCTGGGGCGGATCCCGCAGCTCCCTCAGCTTGTCCTGAGTCCCTTGGGTGTCGTTG
AGATTGTTGTTTTTTGAAGAAACAGAAGATTCTATTTTTTACAGCGAGCAAGCTGGTTTTCTTATTTTTG
TATCCTTTTTCAGATGTAATTTTTATCTTTGCTCCGATCCTCATTTGCTGGTGTGGGTGAGGGATCCGGC
GGCATGGGCTGGTTTCACCCCCTTCACGAGGGGCCGCAGAGTCACACGCTGGTGCCGGGGGTGCTTTGGG
GGGAGCTGCGCCGATCACCAGATTAAGCACATGTCCTATCCCAGGCGGTGGAGCGGAGCCCCCGTGGCTC
TGGACTGCGCGGACGTTGGCGTCAGGATGACCACACGGCGGCCTTTCCCGAATGGGGACAGAACCCGCTC
TGAGCCGTGGGTCTGGCTCCTGTAGGGGACTGGCTCTCTTGGTGCACCAGGGGAGGGGGACATATCCCAG
TGAACCCCACCTTGGCGCCTGAGGCAACACAGGGTGGGCACTGACCCACCCCCAGGGGCGGCTGCAGAGG
CAGTGCCCGCAGACAATGGCCACACCTCTCTCCCCAGGGCCCGGCAGTGCCCAAGGATGGGTCCGGGGCC
TCGGGGCCAATGAGCGCCTCTTCCTAGGTGCTGGGATTCAGTCCCCAAACACAGCGGGAGGGGTCCCTGG
GGCAGATGGGGCTTTACCAGCGTCGGGTGGTTTAGTTCGAGTCCCTTTTGTGGAGAAAGGGAGATGAAAA
CTGACCACGTGCCAGGTGTGGCCGAAGCCCCCAGGGAGGGCCACATTCGGGGAGCGGGGGGTCGGGGGAG
GGCCACCGACTGGCTCTGCTGCCAGCACAGGCCCCTCCCTGGAAGTCCTCGGGAGCGGAGCGCGGATCGG
CACGGGCTCTGGGCTCCCCGTGGAGAGAAGCTGTAGTTTTTACCAAATTGTGTACATCTGGGCAGATGTT
TAATTTCTGTGACTAATCACTGAACTAGACGAATGTTAAATTTTTTATGTCTGAAGCCTGAGTCTATTTT
GGATCTGTAAATAATCATTGCCAGTGTGACTTTTGTTCAACAAAAGGATTGTACTGTATTAAGAACCGAT
GAAAAAAATTCTCCTGTAACATTTTTTTAAGAAAACTTTGTTTGTTTAAAGAAAAAGTATTGTATAAATT
ATAATTTTTATTTAAATAAACCTAAAATGCTTTGTGCTAAGGCTCAAAAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 29)

Translated protein sequence

MLRRLDKIRFRGHKRDDFLDLAESPNASDTECSDEIPLKVPRTS
PRDSEELRDPAGPGTLIMATGVQDFNRTEFDRLNEIKGHLEIALLEKHFLQEELRKLR
EETNAEMLRQELDRERQRRMELEQKVQEVLKARTEEQMAQQPPKGQAQASNGAERRSQ
GLSSRLQKWFYERFGEYVEDFRFQPEENIVETEEPLSARRLTENMRRLKRGAKPVTNF
VKNLSALSDWYSVYTSAIAFTVYMNAVWHGWAIPLFLFLAILRLSLNYLIARGWRIQW
SIVPEVSEPVEPPKEDLTVSEKFQLVLDVAQKAQNLFGKMADILEKIKNLFMWVQPEI
TQKLYVALWAAFLASCFFPYRLVGLAVGLYAGIKFFLIDFIFKRCPRLRAKYDTPYII
WRSLPTDPQLKERSSAAVSRRLQTTSSRSYVPSAPAGLGKEEDAGRFHSTKKGNFHEI
FNLTENERPLAVCENGWRCCLINRDRKMPTDYIRNGVLYVTENYLCFESSKSGSSKRN
KVIKLVDITDIQKYKVLSVLPGSGMGIAVSTPSTQKPLVFGAMVHRDEAFETILSQYI
KITSAAASGGDS (SEQ ID NO: 30)

HYAL2	8692	NM_003773	Homo sapiens hyaluronoglucosaminidase 2 (HYAL2), transcript variant 1, mRNA

mRNA Sequence

TTTCCTCTCAGGGGGCAGCAGGAAGTGAGGAGAAAGGGCTGGGATGGGAGGCGGGAGCGGATGGGAGGGA
ATGGGGTTTATCAAGTCCTCGGCGAGCTGCCCAACGGGCAGCAGCTGGCGCAAGTAGCCTAGCTGGAGAG
GCTCACCCCAGGAAGGAGGGAGGCCACCGACCTACTGGGCCGACGGACTCCCACACAGTTCCTGAGCTGG
TGCCAGGCAGGTGACACCTCCTGCAGCCCCCAGCATGCGGGCAGGCCCAGGCCCCACCGTTACATTGGCC
CTGGTGCTGGCGGTGTCATGGGCCATGGAGCTCAAGCCCACAGCACCACCCATCTTCACTGGCCGGCCCT
TTGTGGTAGCGTGGGACGTGCCCACACAGGACTGTGGCCCACGCCTCAAGGTGCCACTGGACCTGAATGC
CTTTGATGTGCAGGCCTCACCTAATGAGGGTTTTGTGAACCAGAATATTACCATCTTCTACCGCGACCGT
CTAGGCCTGTATCCACGCTTCGATTCTGCCGGAAGGTCTGTGCATGGTGGTGTGCCACAGAATGTCAGCC
TTTGGGCACACCGGAAGATGCTGCAGAAACGTGTGGAGCACTACATTCGGACACAGGAGTCTGCGGGGCT
GGCGGTCATCGACTGGGAGGACTGGCGACCTGTGTGGGTGCGCAACTGGCAGGACAAAGATGTGTATCGC
CGGTTATCACGCCAGCTAGTGGCCAGTCGTCACCCTGACTGGCCTCCAGACCGCATAGTCAAACAGGCAC
AATATGAGTTTGAGTTCGCAGCACAGCAGTTCATGCTGGAGACACTGCGTTATGTCAAGGCAGTGCGGCC
CCGGCACCTCTGGGGCTTCTACCTCTTTCCTGACTGCTACAATCATGATTATGTGCAGAACTGGGAGAGC
TACACAGGCCGCTGCCCTGATGTTGAGGTGGCCCGCAATGACCAGCTGGCCTGGCTGTGGGCTGAGAGCA

CGGCCCTCTTCCCGTCTGTCTACCTGGACGAGACACTTGCTTCCTCCCGCCATGGCCGCAACTTTGTGAG
CTTCCGTGTTCAGGAGGCCCTTCGTGTGGCTCGCACCCACCATGCCAACCATGCACTCCCAGTCTACGTC
TTCACACGACCCACCTACAGCCGCAGGCTCACGGGGCTTAGTGAGATGGACCTCATCTCTACCATTGGCG
AGAGTGCGGCCCTGGGCGCAGCTGGTGTCATCCTCTGGGGTGACGCGGGGTACACCACAAGCACGGAGAC
CTGCCAGTACCTCAAAGATTACCTGACACGGCTGCTGGTCCCCTACGTGGTCAATGTGTCCTGGGCCACC
CAATATTGCAGCCGGGCCCAGTGCCATGGCCATGGGCGCTGTGTGCGCCGCAACCCCAGTGCCAGTACCT
TCCTGCATCTCAGCACCAACAGTTTCCGCCTAGTGCCTGGCCATGCACCTGGTGAACCCCAGCTGCGACC
TGTGGGGGAGCTCAGTTGGGCCGACATTGACCACCTGCAGACACACTTCCGCTGCCAGTGCTACTTGGGC
TGGAGTGGTGAGCAATGCCAGTGGGACCATAGGCAGGCAGCTGGAGGTGCCAGCGAGGCCTGGGCTGGGT
CCCACCTCACCAGTCTGCTGGCTCTGGCAGCCCTGGCCTTTACCTGGACCTTGTAGGGGTCTCCTGCCTA
GCTGCCTAGCAAGCTGGCCTCTACCACAAGGGCTCTCTTAGGCATGTAGGACCCTGCAGGGGGTGGACAA
ACTGGAGTCTGGAGTGGGCAGAGCCCCCAGGAAGCCCAGGAGGGCATCCATACCAGCTCGCACCCCCCTG
TTCTAAGGGGGAGGGGAAGTCCCTGGGAGGCCCCTTCTCTCCCTGCCAGAGGGGAAGGAGGGTACAGCTG
GGCTGGGGAGGACCTGACCCTACTCCCTTGCCCTAGATAGTTTATTATTATTATTATTTTGGGGTCTCTT
TTGTAAATTAAACATAAAACAATTGCTTCTCTGCTTGGATTTTGT (SEQ ID NO: 31)

Translated protein sequence

MRAGPGPTVTLALVLAVSWAMELKPTAPPIFTGRPFVVAWDVPT
QDCGPRLKVPLDLNAFDVQASPNEGFVNQNITIFYRDRLGLYPRFDSAGRSVHGGVPQ
NVSLWAHRKMLQKRVEHYIRTQESAGLAVIDWEDWRPVWVRNWQDKDVYRRLSRQLVA
SRHPDWPPDRIVKQAQYEFEFAAQQFMLETLRYVKAVRPRHLWGFYLFPDCYNHDYVQ
NWESYTGRCPDVEVARNDQLAWLWAESTALFPSVYLDETLASSRHGRNFVSFRVQEAL
RVARTHHANHALPVYVFTRPTYSRRLTGLSEMDLISTIGESAALGAAGVILWGDAGYT
TSTETCQYLKDYLTRLLVPYVVNVSWATQYCSRAQCHGHGRCVRRNPSASTFLHLSTN
SFRLVPGHAPGEPQLRPVGELSWADIDHLQTHFRCQCYLGWSGEQCQWDHRQAAGGAS
EAWAGSHLTSLLALAALAFTWTL (SEQ ID NO: 32)

IL10	3586	NM_000572	Homo sapiens interleukin 10 (IL10), mRNA

mRNA Sequence

ACACATCAGGGGCTTGCTCTTGCAAAACCAAACCACAAGACAGACTTGCAAAAGAAGGCATGCACAGCTC
AGCACTGCTCTGTTGCCTGGTCCTCCTGACTGGGGTGAGGGCCAGCCCAGGCCAGGGCACCCAGTCTGAG
AACAGCTGCACCCACTTCCCAGGCAACCTGCCTAACATGCTTCGAGATCTCCGAGATGCCTTCAGCAGAG
TGAAGACTTTCTTTCAAATGAAGGATCAGCTGGACAACTTGTTGTTAAAGGAGTCCTTGCTGGAGGACTT
TAAGGGTTACCTGGGTTGCCAAGCCTTGTCTGAGATGATCCAGTTTTACCTGGAGGAGGTGATGCCCCAA
GCTGAGAACCAAGACCCAGACATCAAGGCGCATGTGAACTCCCTGGGGGAGAACCTGAAGACCCTCAGGC
TGAGGCTACGGCGCTGTCATCGATTTCTTCCCTGTGAAAACAAGAGCAAGGCCGTGGAGCAGGTGAAGAA
TGCCTTTAATAAGCTCCAAGAGAAAGGCATCTACAAAGCCATGAGTGAGTTTGACATCTTCATCAACTAC
ATAGAAGCCTACATGACAATGAAGATACGAAACTGAGACATCAGGGTGGCGACTCTATAGACTCTAGGAC
ATAAATTAGAGGTCTCCAAAATCGGATCTGGGGCTCTGGGATAGCTGACCCAGCCCCTTGAGAAACCTTA
TTGTACCTCTCTTATAGAATATTTATTACCTCTGATACCTCAACCCCCATTTCTATTTATTTACTGAGCT
TCTCTGTGAACGATTTAGAAAGAAGCCCAATATTATAATTTTTTTCAATATTTATTATTTTCACCTGTTT
TTAAGCTGTTTCCATAGGGTGACACACTATGGTATTTGAGTGTTTTAAGATAAATTATAAGTTACATAAG
GGAGGAAAAAAAAATGTTCTTTGGGGAGCCAACAGAAGCTTCCATTCCAAGCCTGACCACGCTTTCTAGCT
GTTGAGCTGTTTTCCCTGACCTCCCTCTAATTTATCTTGTCTCTGGGCTTGGGGCTTCCTAACTGCTACA
AATACTCTTAGGAAGAGAAACCAGGGAGCCCCTTTGATGATTAATTCACCTTCCAGTGTCTCGGAGGGAT
TCCCCTAACCTCATTCCCCAACCACTTCATTCTTGAAAGCTGTGGCCAGCTTGTTATTTATAACAACCTA
AATTTGGTTCTAGGCCGGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGCGGGTG
GATCACTTGAGGTCAGGAGTTCCTAACCAGCCTGGTCAACATGGTGAAACCCCGTCTCTACTAAAAATAC
AAAAATTAGCCGGGCATGGTGGCGCGCACCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAAGAGAATTG
CTTGAACCCAGGAGATGGAAGTTGCAGTGAGCTGATATCATGCCCCTGTACTCCAGCCTGGGTGACAGAG
CAAGACTCTGTCTCAAAAAATAAAAATAAAAATAAATTTGGTTCTAATAGAACTCAGTTTTAACTAGAAT
TTATTCAATTCCTCTGGGAATGTTACATTGTTTGTCTGTCTTCATAGCAGATTTTAATTTTGAATAAATA
AATGTATCTTATTCACATC (SEQ ID NO: 33)

Translated protein sequence

MHSSALLCCLVLLTGVRASPGQGTQSENSCTHFPGNLPNMLRDL
RDAFSRVKTFFQMKDQLDNLLLKESLLEDFKGYLGCQALSEMIQFYLEEVMPQAENQD
PDIKAHVNSLGENLKTLRLRLRRCHRFLPCENKSKAVEQVKNAFNKLQEKGIYKAMSE
FDIFINYIEAYMTMKIRN (SEQ ID NO: 34)

ITFG3	83986	NM_032039	Homo sapiens integrin alpha FG-GAP repeat containing 3 (ITFG3), mRNA

mRNA Sequence

AGTGACGCCAGGGGGCGGGGCCAGCGGCGCGGTCGGGTGAGAGGCCGCGGCGGCAGGTCCACCTGGGCTT
GCGAAGGCACAGATTCCCCGTCCACAGCTCACGACCAGATGCACCAGCAGGAGTCCACATCGAGGACGTC
CTCCGGGCACTCCCACGACCAGTGACCAGGAGTTAAACTTTGGGATGTGCCCGTGATGTTGGACCACAAG
GACTTAGAGGCCGAAATCCACCCCTTGAAAAATGAAGAAAGAAAATCGCAGGAAAATCTGGGAAATCCAT
CAAAAAATGAGGATAACGTGAAAAGCGCGCCTCCACAGTCCCGGCTCTCCCGGTGCCGAGCGGCGGCGTT
TTTTCTTTCATTGTTTCTCTGCCTTTTTGTGGTGTTCGTCGTCTCATTCGTCATCCCGTGTCCAGACCGG
CCGGCGTCACAGCGAATGTGGAGGATAGACTACAGTGCCGCTGTTATCTATGACTTTCTGGCTGTGGATG
ATATAAACGGGGACAGGATCCAAGATGTTCTTTTTCTTTATAAAAACACCAACAGCAGCAACAATTTCAG
CCGATCCTGTGTGGACGAAGGCTTTTCCTCTCCCTGCACCTTTGCAGCTGCTGTGTCGGGGGCCAACGGC
AGCACGCTCTGGGAGAGACCTGTGGCCCAAGACGTGGCCCTCGTGGAGTGTGCTGTGCCCCAGCCAAGAG
GCAGTGAGGCACCTTCTGCCTGCATCCTGGTGGGCAGACCCAGTTCTTTCATTGCAGTCAACTTGTTCAC
AGGGGAAACCCTGTGGAACCACAGCAGCAGCTTCAGCGGGAATGCGTCCATCCTGAGCCCTCTGCTGCAG
GTGCCTGATGTGGACGGCGATGGGGCCCCAGACCTGCTGGTTCTCACCCAGGAGCGGGAGGAGGTTAGTG
GCCACCTCTACTCCGGCAGCACCGGGCACCAGATTGGCCTCAGAGGCAGCCTTGGTGTGGACGGGGAAAG
TGGCTTCCTCCTTCACGTCACCAGGACAGGTGCCCACTACATCCTCTTTCCCTGCGCAAGCTCCCTCTGC
GGCTGCTCTGTGAAGGGTCTCTACGAGAAGGTGACCGGGAGCGGCGGCCCGTTCAAGAGTGACCCGCACT
GGGAGAGCATGCTCAATGCCACCACCCGCAGGATGCTTTCCCACAGCTCTGGAGCAGTGCGCTACCTGAT
GCATGTCCCAGGGAACGCCGGTGCAGATGTGCTTCTTGTGGGCTCAGAGGCCTTCGTGCTGCTGGACGGG
CAGGAGCTGACGCCTCGCTGGACACCCAAGGCAGCCCATGTCCTGAGAAAACCCATCTTCGGCCGCTACA
AACCAGACACCTTGGCTGTAGCCGTTGAAAACGGAACTGGCACCGACAGACAGATCCTGTTTCTGGACCT
TGGCACTGGAGCCGTCCTGTGTAGCCTAGCCCTCCCGAGCCTCCCTGGGGGTCCACTGTCCGCCAGCCTG
CCGACCGCAGACCACCGCTCAGCCTTCTTCTTCTGGGGCCTCCACGAGCTGGGGAGCACCAGCGAGACGG
AGACCGGGGAGGCCCGGCACAGCCTGTACATGTTCCACCCCACCCTGCCGCGCGTGCTGCTGGAGCTGGC
CAATGTCTCTACCCACATTGTCGCCTTTGACGCCGTCCTGTTTGAGCCAAGCCGCCACGCCGCCTACATC
CTTCTGACAGGCCCGGCAGACTCAGAGGCACCCGGCCTGGTCTCTGTGATCAAGCACAAGGTGCGGGACC
TTGTCCCAAGCAGCAGGGTGGTCCGCCTGGGTGAGGGTGGGCCAGACAGTGACCAAGCCATCAGGGACCG
GTTCTCCCGGCTGCGGTACCAGAGTGAGGCGTAGAGGCACGCCAGCCAGAGCCTGTGGAGAGACTCCGCC
TGCTGACACTAAACGTCCTGGGAAGTGGGCCCTTCCCTGGGTCTCTGCACTGACTCCCCCACTCCTGACC
CTGGTGATGGTCGCCACTGGGCAGCAGCAGCCTTACCAGTCCTCCATGATCACACCCAGGGACCTGCATG
GGTGAGGGGACACCCTGGGCCTCTCTCCCGCCCAGCATCCTCCCTGAGTCCCCACACAGGGCCTCACTCT
GCACCCCACCAGGGTCCCGCTCACACCAGGCAGCCTTCATAGTGGTCTCCCTGGCCACCTTGGGCAGAGC
TGGGTCATGCAGCACCCCATCCTTACCCGGTGCCCTCTCCTTGCCAGCTTCTCCCCAGGCCAGAGCGGCC
ATCGCGTAGAAAGAACCAGGGTGTCCCCGGGACAGGCCGTCCCCCACCCCATCCTGTAGAAGTCCATTCC
CCTTTTCCCTCCTGTGCTCTGTCCCCCAAGGAGTCATGGAACTCAGGGTACTGGGCCTCAACGGGAACCT
GAGACAGCTCCAGCTTCGCAGCCCTTCCCGGAGCTACAGGGGGATCCTCTAGCATGGGGGGTGTGACTTG
GTTCCTTTGACCAGGTCCTGTGAGGAAGCCTGGAGCAAGGGTCTCCCCCAGCAGGATGGGTGGGGCCTGC
TCTGGAGCTGAGCCCGTGGCCGCTCACAGGTGTCCTTAGTGGTGTTGCAGCTGTCTACTGGCTGCATGTG
CTGTGAATATCCCAAGGAACTGGCTGTGGAATGCGTGTTTGGGTCAGTCTGTGCCCTCTCAGTAGACACT
GGAGCTGCTCTGTCCCTGAAGAGGCCCCGTGCCCCAGGCATGGCAAGCGCCTGCCTCTCCCCTTCCGGTG
CTCACACGCCCACGCCGTGCCACCCGATGCAGGACTCACCTCTGTGCCTTGCTGCTCCTGAGGCCCAAGG
GCAGCCATGGTGCTCTGTACTGCTCGGGCCGCCCAGGTCACAGAGCCTGAGCTTCGTAGCCAAAGCAGCC
TGATGACCCACCCACCAAGGAAGAAAGCAGAATAAACATTTTTGCACTGCCTGAAAAACCCCGGTGGTCA
GGCGTGAGCCTAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 35)

Translated protein sequence

MLDHKDLEAEIHPLKNEERKSQENLGNPSKNEDNVKSAPPQSRL
SRCRAAAFFLSLFLCLFVVFVVSFVIPCPDRPASQRMWRIDYSAAVIYDFLAVDDING
DRIQDVLFLYKNTNSSNNFSRSCVDEGFSSPCTFAAAVSGANGSTLWERPVAQDVALV
ECAVPQPRGSEAPSACILVGRPSSFIAVNLFTGETLWNHSSSFSGNASILSPLLQVPD
VDGDGAPDLLVLTQEREEVSGHLYSGSTGHQIGLRGSLGVDGESGFLLHVTRTGAHYI
LFPCASSLCGCSVKGLYEKVTGSGGPFKSDPHWESMLNATTRRMLSHSSGAVRYLMHV
PGNAGADVLLVGSEAFVLLDGQELTPRWTPKAAHVLRKPIFGRYKPDTLAVAVENGTG
TDRQILFLDLGTGAVLCSLALPSLPGGPLSASLPTADHRSAFFFWGLHELGSTSETET
GEARHSLYMFHPTLPRVLLELANVSTHIVAFDAVLFEPSRHAAYILLTGPADSEAPGL
VSVIKHKVRDLVPSSRVVRLGEGGPDSDQAIRDRFSRLRYQSEA (SEQ ID NO: 36)

JAM3	83700	NM_032801	Homo sapiens junctional adhesion molecule 3 (JAM3), mRNA

mRNA Sequence

TAGACCTCAGCTTCCTCTGTCACCATGGTGCCGGCTCGGCTGGGCCCGGCGGTCGCCATGGTAACTGGGG
CGGGTCGCAGGGTCCTGGCAGGCTGGGCGCATGCGCGCGGGGACTACAAGCCGCGCCGCGCTGCCGCTGG
CCCCTCAGCAACCCTCGACATGGCGCTGAGGCGGCCACCGCGACTCCGGCTCTGCGCTCGGCTGCCTGAC
TTCTTCCTGCTGCTGCTTTTCAGGGGCTGCCTGATAGGGGCTGTAAATCTCAAATCCAGCAATCGAACCC
CAGTGGTACAGGAATTTGAAAGTGTGGAACTGTCTTGCATCATTACGGATTCGCAGACAAGTGACCCCAG
GATCGAGTGGAAGAAAATTCAAGATGAACAAACCACATATGTGTTTTTTGACAACAAAATTCAGGGAGAC
TTGGCGGGTCGTGCAGAAATACTGGGGAAGACATCCCTGAAGATCTGGAATGTGACACGGAGAGACTCAG
CCCTTTATCGCTGTGAGGTCGTTGCTCGAAATGACCGCAAGGAAATTGATGAGATTGTGATCGAGTTAAC
TGTGCAAGTGAAGCCAGTGACCCCTGTCTGTAGAGTGCCGAAGGCTGTACCAGTAGGCAAGATGGCAACA
CTGCACTGCCAGGAGAGTGAGGGCCACCCCCGGCCTCACTACAGCTGGTATCGCAATGATGTACCACTGC
CCACGGATTCCAGAGCCAATCCCAGATTTCGCAATTCTTCTTTCCACTTAAACTCTGAAACAGGCACTTT
GGTGTTCACTGCTGTTCACAAGGACGACTCTGGGCAGTACTACTGCATTGCTTCCAATGACGCAGGCTCA
GCCAGGTGTGAGGAGCAGGAGATGGAAGTCTATGACCTGAACATTGGCGGAATTATTGGGGGGGTTCTGG
TTGTCCTTGCTGTACTGGCCCTGATCACGTTGGGCATCTGCTGTGCATACAGACGTGGCTACTTCATCAA
CAATAAACAGGATGGAGAAAGTTACAAGAACCCAGGGAAACCAGATGGAGTTAACTACATCCGCACTGAC
GAGGAGGGCGACTTCAGACACAAGTCATCGTTTGTGATCTGAGACCCGCGGTGTGGCTGAGAGCGCACAG
AGCGCACGTGCACATACCTCTGCTAGAAACTCCTGTCAAGGCAGCGAGAGCTGATGCACTCGGACAGAGC
TAGACACTCATTCAGAAGCTTTTCGTTTTGGCCAAAGTTGACCACTACTCTTCTTACTCTAACAAGCCAC
ATGAATAGAAGAATTTTCCTCAAGATGGACCCGGTAAATATAACCACAAGGAAGCGAAACTGGGTGCGTT
CACTGAGTTGGGTTCCTAATCTGTTTCTGGCCTGATTCCCGCATGAGTATTAGGGTGATCTTAAAGAGTT
TGCTCACGTAAACGCCCGTGCTGGGCCCTGTGAAGCCAGCATGTTCACCACTGGTCGTTCAGCAGCCACG
ACAGCACCATGTGAGATGGCGAGGTGGCTGGACAGCACCAGCAGCGCATCCCGGCGGGAACCCAGAAAAG
GCTTCTTACACAGCAGCCTTACTTCATCGGCCCACAGACACCACCGCAGTTTCTTCTTAAAGGCTCTGCT
GATCGGTGTTGCAGTGTCCATTGTGGAGAAGCTTTTTGGATCAGCATTTTGTAAAAACAACCAAAATCAG
GAAGGTAAATTGGTTGCTGGAAGAGGGATCTTGCCTGAGGAACCCTGCTTGTCCAACAGGGTGTCAGGAT
TTAAGGAAAACCTTCGTCTTAGGCTAAGTCTGAAATGGTACTGAAATATGCTTTTCTATGGGTCTTGTTT
ATTTTATAAAATTTTACATCTAAATTTTTGCTAAGGATGTATTTTGATTATTGAAAAGAAAATTTCTATT
TAAACTGTAAATATATTGTCATACAATGTTAAATAACCTATTTTTTTAAAAAAGTTCAACTTAAGGTAGA
AGTTCCAAGCTACTAGTGTTAAATTGGAAAATATCAATAATTAAGAGTATTTTACCCAAGGAATCCTCTC
ATGGAAGTTTACTGTGATGTTCCTTTTCTCACACAAGTTTTAGCCTTTTTCACAAGGGAACTCATACTGT
CTACACATCAGACCATAGTTGCTTAGGAAACCTTTAAAAATTCCAGTTAAGCAATGTTGAAATCAGTTTG
CATCTCTTCAAAAGAAACCTCTCAGGTTAGCTTTGAACTGCCTCTTCCTGAGATGACTAGGACAGTCTGT
ACCCAGAGGCCACCCAGAAGCCCTCAGATGTACATACACAGATGCCAGTCAGCTCCTGGGGTTGCGCCAG
GCGCCCCCGCTCTAGCTCACTGTTGCCTCGCTGTCTGCCAGGAGGCCCTGCCATCCTTGGGCCCTGGCAG
TGGCTGTGTCCCAGTGAGCTTTACTCACGTGGCCCTTGCTTCATCCAGCACAGCTCTCAGGTGGGCACTG
CAGGGACACTGGTGTCTTCCATGTAGCGTCCCAGCTTTGGGCTCCTGTAACAGACCTCTTTTTGGTTATG
GATGGCTCACAAAATAGGGCCCCCAATGCTATTTTTTTTTTTTAAGTTTGTTTAATTATTTGTTAAGATT
GTCTAAGGCCAAAGGCAATTGCGAAATCAAGTCTGTCAAGTACAATAACATTTTTAAAAGAAAATGGATC
CCACTGTTCCTCTTTGCCACAGAGAAAGCACCCAGACGCCACAGGCTCTGTCGCATTTCAAAACAAACCA
TGATGGAGTGGCGGCCAGTCCAGCCTTTTAAAGAACGTCAGGTGGAGCAGCCAGGTGAAAGGCCTGGCGG
GGAGGAAAGTGAAACGCCTGAATCAAAAGCAGTTTTCTAATTTTGACTTTAAATTTTTCATCCGCCGGAG
ACACTGCTCCCATTTGTGGGGGGACATTAGCAACATCACTCAGAAGCCTGTGTTCTTCAAGAGCAGGTGT
TCTCAGCCTCACATGCCCTGCCGTGCTGGACTCAGGACTGAAGTGCTGTAAAGCAAGGAGCTGCTGAGAA
GGAGCACTCCACTGTGTGCCTGGAGAATGGCTCTCACTACTCACCTTGTCTTTCAGCTTCCAGTGTCTTG
GGTTTTTTATACTTTGACAGCTTTTTTTTAATTGCATACATGAGACTGTGTTGACTTTTTTTAGTTATGT
GAAACACTTTGCCGCAGGCCGCCTGGCAGAGGCAGGAAATGCTCCAGCAGTGGCTCAGTGCTCCCTGGTG
TCTGCTGCATGGCATCCTGGATGCTTAGCATGCAAGTTCCCTCCATCATTGCCACCTTGGTAGAGAGGGA
TGGCTCCCCACCCTCAGCGTTGGGGATTCACGCTCCAGCCTCCTTCTTGGTTGTCATAGTGATAGGGTAG
CCTTATTGCCCCCTCTTCTTATACCCTAAAACCTTCTACACTAGTGCCATGGGAACCAGGTCTGAAAAAG
TAGAGAGAAGTGAAAGTAGAGTCTGGGAAGTAGCTGCCTATAACTGAGACTAGACGGAAAAGGAATACTC
GTGTATTTTAAGATATGAATGTGACTCAAGACTCGAGGCCGATACGAGGCTGTGATTCTGCCTTTGGATG
GATGTTGCTGTACACAGATGCTACAGACTTGTACTAACACACCGTAATTTGGCATTTGTTTAACCTCATT
TATAAAAGCTTCAAAAAAACCCAAAAAAACCCAAA (SEQ ID NO: 37)

Translated protein sequence

MVPARLGPAVAMVTGAGRRVLAGWAHARGDYKPRRAAAGPSATL
DMALRRPPRLRLCARLPDFFLLLLFRGCLIGAVNLKSSNRTPVVQEFESVELSCIITD
SQTSDPRIEWKKIQDEQTTYVFFDNKIQGDLAGRAEILGKTSLKIWNVTRRDSALYRC
EVVARNDRKEIDEIVIELTVQVKPVTPVCRVPKAVPVGKMATLHCQESEGHPRPHYSW
YRNDVPLPTDSRANPRFRNSSFHLNSETGTLVFTAVHKDDSGQYYCIASNDAGSARCE
EQEMEVYDLNIGGIIGGVLVVLAVLALITLGICCAYRRGYFINNKQDGESYKNPGKPD
GVNYIRTDEEGDFRHKSSFVI (SEQ ID NO: 38)

KLHL17	339451	NM_198317	Homo sapiens kelch-like 17 (Drosophila) (KLHL17), mRNA

mRNA Sequence

AGTGAGCGACACAGAGCGGGCCGCCACCGCCGAGCAGCCCTCCGGCAGTCTCCGCGTCCGTTAAGCCCGC
GGGTCCTCCGCGAATCGGCGGTGGGTCCGGCAGCCGAATGCAGCCCCGCAGCGAGCGCCCGGCCGGCAGG
ACGCAGAGCCCGGAGCACGGCAGCCCGGGGCCCGGGCCCGAGGCGCCGCCGCCTCCACCGCCGCAGCCGC
CGGCCCCCGAGGCAGAGCGCACGCGGCCCCGGCAGGCTCGGCCCGCAGCCCCCATGGAGGGAGCCGTGCA
GCTGCTGAGCCGCGAGGGCCACAGCGTGGCCCACAACTCCAAGCGGCACTACCACGATGCCTTCGTGGCC
ATGAGCCGCATGCGCCAGCGCGGCCTCCTGTGCGACATCGTCCTGCACGTGGCTGCCAAGGAGATCCGTG
CGCACAAAGTGGTGCTGGCCTCCTGCAGCCCCTACTTCCACGCCATGTTCACAAATGAGATGAGCGAGAG
CCGCCAGACCCACGTGACGCTGCACGACATCGACCCTCAGGCCTTGGACCAGCTGGTGCAGTTTGCCTAC
ACGGCTGAGATTGTGGTGGGCGAGGGCAATGTGCAGACTCTGCTCCCAGCCGCCAGTCTCCTGCAGCTGA
ATGGCGTCCGAGACGCTTGCTGCAAGTTTCTACTGAGTCAGCTCGACCCCTCCAACTGCCTGGGTATCCG
GGGCTTTGCCGATGCGCACTCCTGCAGCGACCTGCTCAAGGCCGCCCACAGGTACGTGCTGCAGCACTTC
GTGGACGTGGCCAAGACCGAGGAGTTTATGCTGCTGCCCCTGAAACAGGTTCTGGAACTGGTCTCTAGCG
ACAGCCTGAACGTGCCTTCAGAGGAGGAGGTCTACCGAGCCGTCCTGAGCTGGGTGAAACACGACGTGGA
CGCCCGCAGGCAGCATGTCCCACGGCTCATGAAGTGTGTGCGGCTGCCCTTGCTGAGCCGCGACTTCCTG
CTGGGCCACGTGGATGCCGAGAGCCTGGTGAGGCACCACCCTGACTGCAAGGACCTCCTCATCGAGGCCC
TGAAGTTCCACCTGCTGCCTGAGCAGAGGGGCGTCCTAGGCACCAGCCGCACACGTCCCCGGCGCTGCGA
GGGGGCCGGGCCTGTGCTTTTTGCTGTGGGCGGCGGGAGCCTGTTTGCCATCCACGGAGACTGTGAGGCC
TACGACACGCGCACCGACCGCTGGCACGTGGTGGCCTCCATGTCCACGCGCCGGGCCCGGGTGGGAGTGG
CTGCGGTGGGGAACCGGCTCTATGCTGTGGGCGGCTATGATGGGACCTCAGACCTGGCTACCGTGGAGTC
CTACGACCCCGTGACTAACACGTGGCAGCCGGAGGTGTCCATGGGCACAAGGCGAAGCTGCCTGGGTGTG
GCCGCCTTGCATGGACTCCTGTACTCGGCCGGCGGCTATGACGGGGCCTCCTGCCTGAACAGTGCTGAAC
GCTACGACCCCCTGACCGGAACGTGGACGTCCGTCGCTGCCATGAGCACCCGGAGGCGCTATGTGCGAGT
GGCCACGCTTGATGGGAACCTGTATGCTGTGGGCGGCTACGACAGCTCCTCACACCTGGCCACTGTGGAG
AAGTATGAGCCCCAGGTGAACGTGTGGTCGCCCGTGGCGTCCATGCTGAGCCGACGCAGCTCAGCGGGCG
TGGCCGTGCTGGAGGGTGCCCTGTACGTGGCAGGGGGCAACGACGGCACCAGCTGCCTCAACTCGGTAGA
GAGATACAGTCCAAAGGCTGGAGCCTGGGAAAGCGTGGCGCCCATGAATATCCGCAGGAGCACGCATGAC
CTGGTGGCCATGGACGGATGGTTGTACGCCGTGGGGGGTAACGACGGTAGCTCCAGCCTCAACTCCATCG
AGAAGTACAACCCGAGGACCAACAAGTGGGTGGCCGCATCCTGCATGTTCACCCGGCGCAGCAGTGTGGG
TGTGGCGGTGCTGGAGCTGCTCAATTTCCCGCCGCCATCCTCCCCGACGCTGTCCGTGTCCTCCACCAGC
CTCTGACCCACCTACCACCAGAGGCCTGCAGCCTCCCACATGCCTTAAGGGGACCGTGGCCCCCACCAGG
GACGTCCTGCGCCATCCGTTCACGTCTCTGCATCCATTCCTTCATGTCTTTATTTAGTTGTTTATTTATT
TAGTTATTTATCTTATTTATTGAGGGGTGAGGAGTGCCACGGCTGCCCGTTTACACCTTTAGCGTCTGGT
CCTCCTGCGTGTCCTCCCCTCCACTGCCTGCATGGGGGGCGCGGGGAGTGACCAGGCGGGGGCCTCACCG
CCCCAGGGCCGTTGCCTGCTCAGACCTTGCAGGCTGTGGAGCAAGAGGCCCTGGGTCTCTCCAAGCAGCT
GCAGACCCCAGCTCGAATTTTGCACATGGCGGGGTCCCGGGAAGGGTGGGGAGCAGTTGTCCTTCCTGTC
GTCGTCTGCCGTGTGCCATCTTTCCTGGATCTTGTAGTGGGTGCACACGCGTGCACTGGGACCCCACACA
GCAATACGAGTCCAACTTAATAAACACATTTCTGGGGTTCCTCAAAAAAAAAAAAAAAAAA (SEQ ID NO: 39)

Translated protein sequence

MQPRSERPAGRTQSPEHGSPGPGPEAPPPPPPQPPAPEAERTRP
RQARPAAPMEGAVQLLSREGHSVAHNSKRHYHDAFVAMSRMRQRGLLCDIVLHVAAKE
IRAHKVVLASCSPYFHAMFTNEMSESRQTHVTLHDIDPQALDQLVQFAYTAEIVVGEG
NVQTLLPAASLLQLNGVRDACCKFLLSQLDPSNCLGIRGFADAHSCSDLLKAAHRYVL
QHFVDVAKTEEFMLLPLKQVLELVSSDSLNVPSEEEVYRAVLSWVKHDVDARRQHVPR
LMKCVRLPLLSRDFLLGHVDAESLVRHHPDCKDLLIEALKFHLLPEQRGVLGTSRTRP
RRCEGAGPVLFAVGGGSLFAIHGDCEAYDTRTDRWHVVASMSTRRARVGVAAVGNRLY
AVGGYDGTSDLATVESYDPVTNTWQPEVSMGTRRSCLGVAALHGLLYSAGGYDGASCL

NSAERYDPLTGTWTSVAAMSTRRRYVRVATLDGNLYAVGGYDSSSHLATVEKYEPQVN
VWSPVASMLSRRSSAGVAVLEGALYVAGGNDGTSCLNSVERYSPKAGAWESVAPMNIR
RSTHDLVAMDGWLYAVGGNDGSSSLNSIEKYNPRTNKWVAASCMFTRRSSVGVAVLEL
LNFPPPSSPTLSVSSTSL (SEQ ID NO: 40)

KRT24	192666	NM_019016	Homo sapiens keratin 24 (KRT24), mRNA

mRNA Sequence

ACTCTTCGTCATCACCTCTCCTATTCGCCTGGACAAGCTCATGTTTGCAGGAGCACCATGTCTTGCTCGT
CTCGCGCCTCCTCCTCCAGGGCTGGAGGCAGCAGCTCAGCCAGGGTGTCTGCTGGTGGAAGCAGCTTCAG
CAGTGGAAGCAGATGTGGTCTGGGGGGCAGCTCGGCCCAGGGCTTCCGAGGAGGAGCCAGCAGCTGCAGC
CTGAGTGGGGGGTCTAGCGGTGCTTTTGGGGGCAGCTTTGGAGGGGGCTTTGGTAGCTGCTCAGTAGGGG
GTGGTTTTGGGGGAGCTTCAGGCTCTGGGACAGGATTTGGTGGGGGTTCTAGCTTTGGCGGGGTCTCTGG
ATTTGGCAGGGGTTCTGGATTCTGTGGGAGTTCTAGATTCAGCAGTGGTGCTACTGGAGGCTTCTACAGC
TATGGTGGTGGTATGGGAGGTGGTGTTGGCGATGGGGGGCTTTTCTCTGGAGGGGAAAAGCAAACCATGC
AGAACCTCAATGACCGCTTGGCCAATTACCTAGACAAGGTCAGAGCCCTGGAGGAGGCTAACACTGATCT
GGAGAACAAAATCAAGGAGTGGTATGACAAATATGGGCCTGGGTCTGGAGACGGTGGATCGGGAAGAGAT
TATAGCAAATACTATTCAATAATTGAAGATCTCAGAAACCAGATCATTGCTGCCACTGTTGAAAATGCTG
GGATCATTTTGCACATTGACAATGCCAGATTGGCTGCTGATGACTTCAGACTGAAGTATGAGAACGAGCT
GTGTCTCCGGCAGAGCGTGGAGGCTGACATCAATGGCCTGCGGAAAGTCCTGGATGACCTGACTATGACC
CGCTCTGACCTGGAGATGCAGATTGAGAGTTTCACCGAGGAGCTAGCCTACCTGAGGAAGAACCACGAGG
AGGAAATGAAGAATATGCAAGGAAGCTCTGGAGGGGAGGTGACCGTAGAAATGAATGCTGCGCCAGGGAC
CGACCTGACCAAATTACTGAATGACATGAGGGCGCAGTACGAGGAGCTGGCTGAGCAAAACCGCCGAGAG
GCTGAGGAGCGGTTCAACAAGCAGAGCGCATCACTACAAGCACAAATCTCCACTGATGCTGGGGCAGCCA
CTTCTGCCAAGAATGAGATAACAGAACTAAAACGTACCCTGCAAGCCCTGGAAATTGAGCTTCAGTCCCA
ACTGGCCATGAAAAGCTCCCTGGAGGGAACCCTGGCTGACACAGAAGCTGGCTACGTGGCTCAGCTGTCA
GAAATTCAAACGCAGATCAGTGCCCTGGAGGAGGAGATCTGCCAGATCTGGGGTGAGACTAAATGCCAGA
ACGCAGAGTACAAGCAATTGCTGGACATCAAGACACGCCTGGAGGTGGAGATCGAGACCTACCGCCGCCT
GCTCGATGGAGAGGGAGGTGGTTCTAGTTTTGCAGAATTTGGTGGTAGAAACTCAGGATCTGTAAACATG
GGATCCAGGGATCTGGTATCTGGTGACTCAAGATCTGGAAGCTGTTCTGGTCAAGGACGAGATTCAAGCA
AGACTAGAGTGACTAAGACTATCGTAGAGGAGTTGGTGGATGGCAAGGTTGTCTCGTCTCAAGTCAGCAG
TATTTCTGAGGTGAAAGTTAAATAAGGAACTTCCAGATCAACAAAAGTGTCTTTCAAAGAAAAAAAAATC
AAGAAGGACACAAGCGAAGAAATGGCATCAATCTAGGCATCTTTCTGGATAATTTCAGGAAAAGCTTCAG
TCCAGAAATGGATGACTAGCCAACTTTTCTGCATCTTCTTATTTCCTCATTAGAATGCTCTTGAAATAGC
TGAATTAACAACTTTGCTTTAATTGTTTCTATGCTTCAATAAATTTACTTTTGCAAGTTAAAAAAAAAAA
AAAAAAA (SEQ ID NO: 41)

Translated protein sequence

MSCSSRASSSRAGGSSSARVSAGGSSFSSGSRCGLGGSSAQGFR
GGASSCSLSGGSSGAFGGSFGGGFGSCSVGGGFGGASGSGTGFGGGSSFGGVSGFGRG
SGFCGSSRFSSGATGGFYSYGGGMGGGVGDGGLFSGGEKQTMQNLNDRLANYLDKVRA
LEEANTDLENKIKEWYDKYGPGSGDGGSGRDYSKYYSIIEDLRNQIIAATVENAGIIL
HIDNARLAADDFRLKYENELCLRQSVEADINGLRKVLDDLTMTRSDLEMQIESFTEEL
AYLRKNHEEEMKNMQGSSGGEVTVEMNAAPGTDLTKLLNDMRAQYEELAEQNRREAEE
RFNKQSASLQAQISTDAGAATSAKNEITELKRTLQALEIELQSQLAMKSSLEGTLADT
EAGYVAQLSEIQTQISALEEEICQIWGETKCQNAEYKQLLDIKTRLEVEIETYRRLLD
GEGGGSSFAEFGGRNSGSVNMGSRDLVSGDSRSGSCSGQGRDSSKTRVTKTIVEELVD
GKVVSSQVSSISEVKVK (SEQ ID NO: 42)

MAD2L1BP	9587	NM_014628	Homo sapiens MAD2L1 binding protein (MAD2L1BP), transcript variant 2, mRNA

mRNA Sequence

ATTCTAACCGCAAGGAGTAGCGGAGGGGAGGTCGTGATGGCGGCGCCGGAGGCGGAGGTTCTGTCCTCAG
CCGCAGTCCCTGATTTGGAGTGGTATGAGAAGTCCGAAGAAACTCACGCCTCCCAGATAGAACTACTTGA
GACAAGCTCTACGCAGGAACCTCTCAACGCTTCGGAGGCCTTTTGCCCAAGAGACTGCATGGTACCAGTG
GTGTTTCCTGGGCCTGTGAGCCAGGAAGGCTGCTGTCAGTTTACTTGTGAACTTCTAAAGCATATCATGT
ATCAACGCCAGCAGCTCCCTCTGCCCTATGAACAGCTTAAGCACTTTTACCGAAAACCTTCTCCCCAGGC
AGAGGAGATGCTGAAGAAGAAACCTCGGGCCACCACTGAGGTGAGCAGCAGGAAATGCCAACAAGCCCTG
GCAGAACTGGAGAGTGTCCTCAGCCACCTGGAGGACTTCTTTGCACGGACACTAGTACCGCGAGTGCTGA

TTCTCCTTGGGGGCAATGCCCTAAGCCCCAAGGAGTTCTATGAACTCGACTTGTCTCTGCTGGCCCCCTA
CAGCGTGGACCAGAGCCTGAGCACAGCAGCTTGTTTGCGCCGTCTCTTCCGAGCCATATTCATGGCTGAT
GCCTTTAGCGAGCTTCAGGCTCCTCCACTCATGGGCACCGTCGTCATGGCACAGGGACACCGCAACTGTG
GAGAAGATTGGTTTCGACCCAAGCTCAACTATCGAGTGCCCAGCCGGGGCCATAAACTGACTGTGACCCT
GTCATGTGGCAGACCTTCCATCCGAACCACGGCTTGGGAAGACTACATTTGGTTCCAGGCACCAGTGACA
TTTAAAGGCTTCCGCGAGTGAATGAGTGCTTCTTAATCCTAAAAACACAATGGCTGAATTATCTTTCTCC
ATGTGGCGCTGAATCACCCATCTGGTTTGGAGCTAGAGTTGCTTCCTGGTGAGAGAGGAAGCAACTCTCC
TTCTGGTTGTCTGCCTCCCCTCAGATTTCCTGATAGGCTGATGGCATGTGGCTGTGACTGTGACTGTAAT
CATTGCTGAACAACATCTCTTTGAATCAAAGGTTGATTTTCCCAGAGGGTGCTGGGTCAGGCATTTCTAT
TAGGAGTTGGAAAGCAAAAATGGGTCCATAGACACTCTATGGAGGTGTCCCTTTCTGCTCTTTGCTGTGT
CCTTTCAGAATTTTTACCAGGAACATAATGTGGATGTGACTTATGAACTTAAATATAAAATAAATAGATT
CTTATTATATTTTCCTGAAAAAA (SEQ ID NO: 43)

Translated protein sequence

MAAPEAEVLSSAAVPDLEWYEKSEETHASQIELLETSSTQEPLN
ASEAFCPRDCMVPVVFPGPVSQEGCCQFTCELLKHIMYQRQQLPLPYEQLKHFYRKPS
PQAEEMLKKKPRATTEVSSRKCQQALAELESVLSHLEDFFARTLVPRVLILLGGNALS
PKEFYELDLSLLAPYSVDQSLSTAACLRRLFRAIFMADAFSELQAPPLMGTVVMAQGH
RNCGEDWFRPKLNYRVPSRGHKLTVTLSCGRPSIRTTAWEDYIWFQAPVTFKGFRE (SEQ ID NO: 44)

MANSC1	54682	NM_018050	Homo sapiens MANSC domain containing 1 (MANSC1), mRNA

mRNA Sequence

AACCACAAAACCCGCCAGGCCGGTGCGGGAGCTGCGGAGCATCCGCTGCGGTCCTCGCCGAGACCCCCGC
GCGGATTCGCCGGTCCTTCCCGCGGGCGCGACAGAGCTGTCCTCGCACCTGGATGGCAGCAGGGGCGCCG
GGGTCCTCTCGACGCCAGAGAGAAATCTCATCATCTGTGCAGCCTTCTTAAAGCAAACTAAGACCAGAGG
GAGGATTATCCTTGACCTTTGAAGACCAAAACTAAACTGAAATTTAAAATGTTCTTCGGGGGAGAAGGGA
GCTTGACTTACACTTTGGTAATAATTTGCTTCCTGACACTAAGGCTGTCTGCTAGTCAGAATTGCCTCAA
AAAGAGTCTAGAAGATGTTGTCATTGACATCCAGTCATCTCTTTCTAAGGGAATCAGAGGCAATGAGCCC
GTATATACTTCAACTCAAGAAGACTGCATTAATTCTTGCTGTTCAACAAAAAACATATCAGGGGACAAAG
CATGTAACTTGATGATCTTCGACACTCGAAAAACAGCTAGACAACCCAACTGCTACCTATTTTTCTGTCC
CAACGAGGAAGCCTGTCCATTGAAACCAGCAAAAGGACTTATGAGTTACAGGATAATTACAGATTTTCCA
TCTTTGACCAGAAATTTGCCAAGCCAAGAGTTACCCCAGGAAGATTCTCTCTTACATGGCCAATTTTCAC
AAGCAGTCACTCCCCTAGCCCATCATCACACAGATTATTCAAAGCCCACCGATATCTCATGGAGAGACAC
ACTTTCTCAGAAGTTTGGATCCTCAGATCACTTGGAGAAACTATTTAAGATGGATGAAGCAAGTGCCCAG
CTCCTTGCTTATAAGGAAAAAGGCCATTCTCAGAGTTCACAATTTTCCTCTGATCAAGAAATAGCTCATC
TGCTGCCTGAAAATGTGAGTGCGCTCCCAGCTACGGTGGCAGTTGCTTCTCCACATACCACCTCGGCTAC
TCCAAAGCCCGCCACCCTTCTACCCACCAATGCTTCAGTGACACCTTCTGGGACTTCCCAGCCACAGCTG
GCCACCACAGCTCCACCTGTAACCACTGTCACTTCTCAGCCTCCCACGACCCTCATTTCTACAGTTTTTA
CACGGGCTGCGGCTACACTCCAAGCAATGGCTACAACAGCAGTTCTGACTACCACCTTTCAGGCACCTAC
GGACTCGAAAGGCAGCTTAGAAACCATACCGTTTACAGAAATCTCCAACCTAACTTTGAACACAGGGAAT
GTGTATAACCCTACTGCACTTTCTATGTCAAATGTGGAGTCTTCCACTATGAATAAAACTGCTTCCTGGG
AAGGTAGGGAGGCCAGTCCAGGCAGTTCCTCCCAGGGCAGTGTTCCAGAAAATCAGTACGGCCTTCCATT
TGAAAAATGGCTTCTTATCGGGTCCCTGCTCTTTGGTGTCCTGTTCCTGGTGATAGGCCTCGTCCTCCTG
GGTAGAATCCTCTCGGAATCACTCCGCAGGAAACGTTACTCAAGACTGGATTATTTGATCAATGGGATCT
ATGTGGACATCTAAGGATGGAACTCGGTGTCTCTTAATTCATTTAGTAACCAGAAGCCCAAATGCAATGA
GTTTCTGCTGACTTGCTAGTCTTAGCAGGAGGTTGTATTTTGAAGACAGGAAAATGCCCCCTTCTGCTTT
CCTTTTTTTTTTTTGGAGACAGAGTCTTGCTTTGTTGCCCAGGCTGGAGTGCAGTAGCACGATCTCGGCT
CTCACCGCAACCTCCGTCTCCTGGGTTCAAGCGATTCTCCTGCCTCAGCCTCCTAAGTATCTGGGATTAC
AGGCATGTGCCACCACACCTGGGTGATTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATGTTGGTCAG
GCTGGTCTCAAACTCCTGACCTAGTGATCCACCCTCCTCGGCCTCCCAAAGTGCTGGGATTACAGGCATG
AGCCACCACAGCTGGCCCCCTTCTGTTTTATGTTTGGTTTTTGAGAAGGAATGAAGTGGGAACCAAATTA
GGTAATTTTGGGTAATCTGTCTCTAAAATATTAGCTAAAAACAAAGCTCTATGTAAAGTAATAAAGTATA
ATTGCCATATAAATTTCAAAATTCAACTGGCTTTTATGCAAAGAAACAGGTTAGGACATCTAGGTTCCAA
TTCATTCACATTCTTGGTTCCAGATAAAATCAACTGTTTATATCAATTTCTAATGGATTTGCTTTTCTTT
TTATATGGATTCCTTTAAAACTTATTCCAGATGTAGTTCCTTCCAATTAAATATTTG (SEQ ID NO: 45)

Translated protein sequence

MFFGGEGSLTYTLVIICFLTLRLSASQNCLKKSLEDVVIDIQSS

LSKGIRGNEPVYTSTQEDCINSCCSTKNISGDKACNLMIFDTRKTARQPNCYLFFCPN
EEACPLKPAKGLMSYRIITDFPSLTRNLPSQELPQEDSLLHGQFSQAVTPLAHHHTDY
SKPTDISWRDTLSQKFGSSDHLEKLFKMDEASAQLLAYKEKGHSQSSQFSSDQEIAHL
LPENVSALPATVAVASPHTTSATPKPATLLPTNASVTPSGTSQPQLATTAPPVTTVTS
QPPTTLISTVFTRAAATLQAMATTAVLTTTFQAPTDSKGSLETIPFTEISNLTLNTGN
VYNPTALSMSNVESSIMNKTASWEGREASPGSSSQGSVPENQYGLPFEKWLLIGSLLF
GVLFLVIGLVLLGRILSESLRRKRYSRLDYLINGIYVDI (SEQ ID NO: 46)

MOBKL1B	55233	NM_018221	Homo sapiens MOB1, Mps One Binder kinase activator-like 1B (yeast) (MOBKL1B), mRNA

mRNA Sequence

GCGAGGTGGGGTAGGCGGGCAAGGCGGGCGCCGAGGTTTGCAAAGGCTCGCAGCGGCCAGAAACCCGGCT
CCGAGCGGCGGCGGCCCGGCTTCCGCTGCCCGTGAGCTAAGGACGGTCCGCTCCCTCTAGCCAGCTCCGA
ATCCTGATCCAGGCGGGGGCCAGGGGCCCCTCGCCTCCCCTCTGAGGACCGAAGATGAGCTTCCTCTTCA
GCAGCCGCTCTTCTAAAACATTCAAACCAAAGAAGAATATCCCTGAAGGATCTCATCAGTATGAACTCTT
AAAACATGCAGAAGCAACTCTAGGAAGTGGGAATCTGAGACAAGCTGTTATGTTGCCTGAGGGAGAGGAT
CTCAATGAATGGATTGCTGTGAACACTGTGGATTTCTTTAACCAGATCAACATGTTATATGGAACTATTA
CAGAATTCTGCACTGAAGCAAGCTGTCCAGTCATGTCTGCAGGTCCGAGATATGAATATCACTGGGCAGA
TGGTACTAATATTAAAAAGCCAATCAAATGTTCTGCACCAAAATACATTGACTATTTGATGACTTGGGTT
CAAGATCAGCTTGATGATGAAACTCTTTTTCCTTCTAAGATTGGTGTCCCATTTCCCAAAAACTTTATGT
CTGTGGCAAAGACTATTCTAAAGCGTCTGTTCAGGGTTTATGCCCATATTTATCACCAGCACTTTGATTC
TGTGATGCAGCTGCAAGAGGAGGCCCACCTCAACACCTCCTTTAAGCACTTTATTTTCTTTGTTCAGGAG
TTTAATCTGATTGATAGGCGTGAGCTGGCACCTCTTCAAGAATTAATAGAGAAACTTGGATCAAAAGACA
GATAAATGTTTCTTCTAGAACACAGTTACCCCCTTGCTTCATCTATTGCTAGAACTATCTCATTGCTATC
TGTTATAGACTAGTGATACAAACTTTAAGAAAACAGGATAAAAAGATACCCATTGCCTGTGTCTACTGAT
AAAATTATCCCAAAGGTAGGTTGGTGTGATAGTTTCCGAGTAAGACCTTAAGGACACAGCCAAATCTTAA
GTACTGTGTGACCACTCTTGTTGTTATCACATAGTCATACTTGGTTGTAATATGTGATGGTTAACCTGTA
GCTTATAAATTTACTTATTATTCTTTTACTCATTTACTCAGTCATTTCTTTACAAGAAAATGATTGAATC
TGTTTTAGGTGACAGCACAATGGACATTAAGAATTTCCATCAATAATTTATGAATAAGTTTCCAGAACAA
ATTTCCTAATAACACAATCAGATTGGTTTTATTCTTTTATTTTACGAATAAAAAATGTATTTTTCAGTAA
AAAAAA (SEQ ID NO: 47)

Translated protein sequence

MSFLFSSRSSKTFKPKKNIPEGSHQYELLKHAEATLGSGNLRQA
VMLPEGEDLNEWIAVNTVDFFNQINMLYGTITEFCTEASCPVMSAGPRYEYHWADGTN
IKKPIKCSAPKYIDYLMTWVQDQLDDETLFPSKIGVPFPKNFMSVAKTILKRLFRVYA
HIYHQHFDSVMQLQEEAHLNTSFKHFIFFVQEFNLIDRRELAPLQELIEKLGSKDR (SEQ ID NO: 48)

NCBP1	4686	NM_002486	Homo sapiens nuclear cap binding protein subunit 1, 80kDa (NCBP1), mRNA

mRNA Sequence

ATTGAGAGGCCACCGGGAAACCATTGAGAAGCCCCGGAGGACCGGCCTGAGCGGAGGCGGAGACTAGAGC
GGCCGCCGGCACGACCCGCCTTCAGGCGTACGACGACCGCGGCCCGGGGGCTCTGAGTGGCCAAAGCGGC
GGCACTTTCTGCGTGGCCCCGGAAGGACATAGAGCGGAAGGCGGGAGAAAGAAGTAGCCGGCAGGCGGAG
GCAGCCCGAGGGGGCGGTTGCATGTGTGCCAGACGTTCGTAGCCCACTGAGCTTCCTCACGCCGGCTGTC
GCAGCGCCTAGCCCCACCCGGCGGCCTCTCCTGCGCTTCCGGGGCCGTGGCGAGCTAGTGCGCCTGCGTG
CCGGCCCATCCGCGCGCCTTGCAGCTGTCCTTGCGTCGGCCAGCGGCCAGACAGTTCCTGCAGCGCTTAC
CGCCTGGCCTCTCGGTTCCGCGGCGCACCGGAGGGCAGCATGTCGCGGCGGCGGCACAGCGACGAGAACG
ACGGTGGGCAGCCTCACAAAAGGAGAAAGACCTCTGATGCAAATGAAACTGAAGATCATTTGGAATCTTT
AATATGTAAAGTAGGAGAAAAGAGTGCCTGCTCTTTGGAGAGCAACCTAGAAGGCTTGGCTGGTGTTTTG
GAAGCTGATCTTCCTAACTACAAGAGCAAGATCTTAAGGCTTCTTTGTACAGTTGCACGCCTATTACCTG
AGAAGCTGACAATTTATACAACATTAGTTGGACTACTGAATGCCAGGAATTACAATTTTGGTGGAGAATT
TGTAGAAGCCATGATTCGTCAACTTAAAGAATCATTGAAAGCAAACAATTATAATGAAGCCGTGTATTTG
GTCCGTTTTTTATCTGATCTTGTGAATTGTCATGTGATTGCCGCCCCATCAATGGTTGCTATGTTTGAAA
ATTTTGTAAGCGTAACTCAGGAAGAAGATGTACCTCAGGTGCGACGAGATTGGTATGTGTATGCATTTCT
GTCATCTTTGCCCTGGGTTGGAAAGGAGTTGTACGAAAAGAAAGATGCAGAGATGGACCGCATCTTTGCC
AACACTGAAAGCTATCTTAAAAGACGCCAAAAGACTCATGTACCCATGTTACAGGTATGGACTGCTGATA
AACCACATCCACAAGAAGAGTATTTAGATTGCCTGTGGGCCCAGATTCAGAAATTGAAAAAGGATCGCTG

GCAGGAACGGCACATCCTAAGACCTTATCTTGCCTTTGACAGCATCCTGTGTGAAGCACTGCAGCACAAT
CTGCCTCCTTTTACACCACCTCCTCACACTGAAGATTCAGTGTACCCAATGCCAAGGGTCATCTTCAGAA
TGTTTGATTACACAGATGATCCCGAGGGTCCTGTCATGCCAGGGAGTCATTCAGTGGAAAGATTTGTAAT
AGAAGAGAATCTTCACTGCATCATTAAGTCCCACTGGAAGGAAAGGAAGACTTGTGCTGCACAGTTAGTG
AGCTATCCAGGGAAGAACAAGATCCCCTTGAACTACCACATAGTTGAGGTGATCTTTGCAGAGCTGTTTC
AACTTCCAGCACCCCCTCACATTGATGTGATGTACACAACACTCCTCATTGAACTGTGCAAACTTCAACC
TGGCTCTCTACCCCAAGTTCTTGCACAGGCAACTGAAATGCTATACATGCGTTTGGACACAATGAACACT
ACCTGTGTAGACAGGTTTATTAATTGGTTTTCTCATCATCTAAGTAACTTCCAGTTCCGTTGGAGCTGGG
AAGATTGGTCAGATTGTCTTAGTCAAGATCCTGAAAGTCCCAAACCGAAGTTTGTAAGAGAAGTTCTAGA
AAAATGTATGAGGTTGTCTTACCATCAGCGTATATTAGATATTGTTCCTCCTACCTTCTCAGCTCTGTGT
CCTGCAAACCCAACCTGCATTTACAAGTATGGAGATGAAAGTAGCAATTCTCTTCCTGGACATTCTGTTG
CCCTCTGTTTAGCTGTTGCCTTTAAAAGTAAGGCAACCAATGATGAAATCTTCAGCATTCTGAAAGATGT
ACCAAATCCTAACCAGGATGATGACGACGATGAAGGATTCAGTTTTAACCCATTGAAAATAGAAGTCTTT
GTACAGACTCTGCTACACTTGGCAGCCAAATCATTCAGCCACTCCTTCAGTGCTCTTGCAAAGTTTCATG
AAGTCTTCAAAACCCTAGCTGAAAGTGATGAAGGAAAGTTACATGTGCTAAGAGTTATGTTTGAGGTCTG
GAGGAACCATCCACAGATGATTGCTGTACTAGTGGATAAGATGATTCGTACACAAATAGTTGATTGTGCT
GCCGTAGCAAATTGGATCTTCTCTTCAGAACTATCTCGTGACTTTACCAGATTGTTTGTTTGGGAAATTT
TGCACTCTACAATTCGTAAGATGAACAAACATGTCCTGAAGATCCAGAAAGAGCTGGAAGAAGCTAAAGA
GAAACTTGCTAGGCAACACAAACGGCGAAGTGATGATGACGACAGAAGCAGTGACAGGAAAGACGGGGTT
CTTGAGGAACAAATAGAACGACTTCAGGAAAAAGTGGAATCTGCTCAGAGTGAACAAAAGAATCTTTTCC
TCGTTATATTTCAGCGGTTTATCATGATCTTGACCGAGCACCTAGTACGATGCGAAACTGATGGGACCAG
TGTATTAACACCATGGTATAAGAACTGTATAGAGAGGCTGCAGCAGATCTTCCTACAGCATCACCAAATA
ATCCAGCAGTACATGGTGACCCTGGAGAACCTTCTCTTCACTGCTGAATTAGACCCTCATATCTTGGCCG
TGTTCCAGCAGTTCTGTGCCCTGCAGGCCTAAGGGTCATTTTTTCCTCATGTCAAGGTTTTTTTTGATAT
CTTAAAATAATTTGTCTTATTTTTTGATGGTTTGAATGCTTGCTTTCTTGTAGTATCCTTTCACTTCTTA
AAGGAAACAAAGGGGAAGAGGACAGTGAATGAACATGGCATTACTTTTAATTGCCCTGAAAAGCAAATAC
TTCCTAACGGCAGTAATGTGACTATGACCATGATATATTATATATGTGACAGATACAAATTCTCTGTGAT
CAGTTTGTTATTTTTTTTCTCCTTAAGGCACAAAATAATTGGTTTGAGGTATGTGAAACACTAGAGGTCA
ACCTTACATAGTATATAGAACTGATGGGTTTACCCAGCTACCCAGTAGCATAACTTTTCACAGCTCGGGG
ATGAATTAACATGGCTGAAATAAAACTAAAAGTATGGTTTTTAAACTTTGGCATTTCATGATTTATCATC
TCACTCTACTCTAAAACTGGTGGTTTCTTACTGAAGGTGTTCTCCATTTGAAATTTTATCTTCAAAGTAT
TTTTAAGTAGTATCTTTAAGACATGACTTGTTAGTAATAAAAGTGTTACTAGTTGGAAGAGTAGCTCTCA
AATTTGTCTTAATGTAAATCACCTGGGAATCTTTCAAGTTATTTTGAAATTTAAACCACCGTCTGGGGGT
GGAACGCAGACATCCTCAGTAATCCTTAAAGTTTCCCCAGGTGATTCCAGGTTTGGTCACCATTATCTTA
GAGCATCTACTCACTTCCTCTAGCCTTGGGGTTATTTGTCCAAGGTCTTGTAGTGAGTTACAGAATACTA
AAGTGGATGTAGAAGTGGTCAGATTGACTGAAACTATACCCTGAATTAGATGTGAGTTTAGATTTTGTTT
ATATGGAACCTGATCCAAAAAACTACGAAGTCCTGAGCTTGTTTCCTGTATAGTACTGATGCTGAAATAA
GATGACAGCAGTTTGTAAAATAATACACAAATATGAGGAATTGTCTGACATTCCAAATTTCGAGGATTTT
TAGACTTTTTTCATTAAACCTTAGAAAAAAATTACCAGTAATCCTACAACTACTGGTAGTGTTGTTGTGC
ATTTGCACAAAATAGGTATAATTTTTTCTTATTACATCCCAAGTTTATGATGCATTAAGCGTTTTGCATA
TTTTGATATATTTTTGCTTTGGTTTACCATACATTTTAGTGGCTACAGAATGTAGTCTGCTTAATAAATG
GGAATTCCTAGAATGTTTAAATACCATACTATTTAAGACAAAATACAAAATATCCAGAAAAATCCAGGTT
GCGTGGCTGGTTAGTAAAGGACTAAAACCCAGGTTCTTGGCTAAATGTTTTCGTTTATACTGTTTATCTT
TCCCATTGCTTAAGCACAGCACAAACTATGTAATTATATATAATTACAGTTGACCCTTGAACAACATGGG
TTTGAACTGTGTGAGTCTCCTTACACACAGGTTTTCTTCCACCCCTGAGATGGCAAGACCAGCCCCTTGT
CTTCCTCAGCCTGCTCAACGTGAAGATGATGAGGATGAAGACCTTTATGATGATCCACTTCTACTTATTA
AATAGTAAATATATTTTTTTCTTATGATTTTATTTTCTTTTCTCTAGCTTCATAAGAATATAGCATATGG
GCTGGGCGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGCGGGCGGATCACAAGGTCA
GGAGATTGAGACCATCCTGGCTAACACAGTGAAACCCTGTCTCTACTAAAAACACAAAAACTTAGCCAGG
CGTGGTGGTACATGGCTGTAGTCCCAGCTACTTGGGAGGCTGAGACAGGAGAATCGCTTGAACCTGGGAG
GTGGAGGTTTCAGTGAGCCAAGATTGTGCCACTGCACTCCAGCCTGGGTGATGGAGCGAGGCTCTGTCTC
AAAAAAGAAAAAAAATATATAGCATATAACATACAAAATGAGTTTATCAACTGTTTGTTATTGGTAAGTC
AGCAGTGGGCTATTGGTGGTTAAGTTTTGGGGGAGTCAAAAGTTACATGCAAATTTTTTACTGTGCGGGG
TGTCAGCATCCCTAACCCCATGTTGTTCAAGGGTCAACTGTAGTTTAAAATGACTCCTGTCTCAAAAAAC
CAAAGGATAACCTTTAAGGGATTGGTAACTTTGACTCAAAACTGCTTTGTAATCTTTTCACAATGTACTG
AAAAGTGTGGCTAGTTATGTTTGATCCACATTCTAGAGAAATTTGTAGGTTTTAATTTCTTTTCTCTTGG
TCCTCTCTTCATGTATAATGGTTGCTTTTAACAGCTGTTCGCTGATGTGGTCCTGCTCTGTCCCAGTCTA
GCAGCTTTAGTGTATGGAAAAATTGAACTAGGAATTGAGTTTTGAAGAAATAAAGGTGTAAGAGCAAACA
TTCAACAGTTGCTGTCCCCAGTAATGAAGTTCATACAGACAAAAGATGGCATGTCACTGTACATCATACC
TTGCAATAAATATTCTGTTAAATTGTGCTGGTGCAATTTAACATGCTTTTGTCAAAGTAAA (SEQ ID NO:
49)

Translated protein sequence

MSRRRHSDENDGGQPHKRRKTSDANETEDHLESLICKVGEKSAC
SLESNLEGLAGVLEADLPNYKSKILRLLCTVARLLPEKLTIYTTLVGLLNARNYNFGG
EFVEAMIRQLKESLKANNYNEAVYLVRFLSDLVNCHVIAAPSMVAMFENFVSVTQEED
VPQVRRDWYVYAFLSSLPWVGKELYEKKDAEMDRIFANTESYLKRRQKTHVPMLQVWT
ADKPHPQEEYLDCLWAQIQKLKKDRWQERHILRPYLAFDSILCEALQHNLPPFTPPPH
TEDSVYPMPRVIFRMFDYTDDPEGPVMPGSHSVERFVIEENLHCIIKSHWKERKTCAA
QLVSYPGKNKIPLNYHIVEVIFAELFQLPAPPHIDVMYTTLLIELCKLQPGSLPQVLA
QATEMLYMRLDTMNTTCVDRFINWFSHHLSNFQFRWSWEDWSDCLSQDPESPKPKFVR
EVLEKCMRLSYHQRILDIVPPTFSALCPANPTCIYKYGDESSNSLPGHSVALCLAVAF
KSKATNDEIFSILKDVPNPNQDDDDDEGFSFNPLKIEVFVQTLLHLAAKSFSHSFSAL
AKFHEVFKTLAESDEGKLHVLRVMFEVWRNHPQMIAVLVDKMIRTQIVDCAAVANWIF
SSELSRDFTRLFVWEILHSTIRKMNKHVLKIQKELEEAKEKLARQHKRRSDDDDRSSD
RKDGVLEEQIERLQEKVESAQSEQKNLFLVIFQRFIMILTEHLVRCETDGTSVLTPWY
KNCIERLQQIFLQHHQIIQQYMVTLENLLFTAELDPHILAVFQQFCALQA (SEQ ID NO: 50)

NMU	10874	NM_006681	Homo sapiens neuromedin U (NMU), mRNA

mRNA Sequence

AGTCCTGTGTCCGGGCCCCGAGGCACAGCCAGGGCACCAGGTGGAGCACCAGCTACGCGTGGCGCAGCGC
AGCGTCCCTAGCACCGAGCCTCCCGCAGCCGCCGAGATGCTGCGAACAGAGAGCTGCCGCCCCAGGTCGC
CCGCCGGACAGGTGGCCGCGGCGTCCCCGCTCCTGCTGCTGCTGCTGCTGCTCGCCTGGTGCGCGGGCGC
CTGCCGAGGTGCTCCAATATTACCTCAAGGATTACAGCCTGAACAACAGCTACAGTTGTGGAATGAGATA
GATGATACTTGTTCGTCTTTTCTGTCCATTGATTCTCAGCCTCAGGCATCCAACGCACTGGAGGAGCTTT
GCTTTATGATTATGGGAATGCTACCAAAGCCTCAGGAACAAGATGAAAAAGATAATACTAAAAGGTTCTT
ATTTCATTATTCGAAGACACAGAAGTTGGGCAAGTCAAATGTTGTGTCGTCAGTTGTGCATCCGTTGCTG
CAGCTCGTTCCTCACCTGCATGAGAGAAGAATGAAGAGATTCAGAGTGGACGAAGAATTCCAAAGTCCCT
TTGCAAGTCAAAGTCGAGGATATTTTTTATTCAGGCCACGGAATGGAAGAAGGTCAGCAGGGTTCATTTA
AAATGGATGCCAGCTAATTTTCCACAGAGCAATGCTATGGAATACAAAATGTACTGACATTTTGTTTTCT
TCTGAAAAAAATCCTTGCTAAATGTACTCTGTTGAAAATCCCTGTGTTGTCAATGTTCTCAGTTGTAACA
ATGTTGTAAATGTTCAATTTGTTGAAAATTAAAAAATCTAAAAATAAA (SEQ ID NO: 51)

Translated protein sequence

MLRTESCRPRSPAGQVAAASPLLLLLLLLAWCAGACRGAPILPQ
GLQPEQQLQLWNEIDDTCSSFLSIDSQPQASNALEELCFMIMGMLPKPQEQDEKDNTK
RFLFHYSKTQKLGKSNVVSSVVHPLLQLVPHLHERRMKRFRVDEEFQSPFASQSRGYF
LFRPRNGRRSAGFI (SEQ ID NO: 52)

PAPLN	89932	NM_173462	Homo sapiens papilin, proteoglycan-like sulfated glycoprotein (PAPLN), mRNA

mRNA Sequence

GCGTTCCTTGCGGCCCGGCCGACCTCGCGGGCTTGGGCCTGGGCGGGCACCGACGGAGCGGCCCTGGCTG
CAGCCTCCCGGCGCCAGCGAAGACAGGCTGAGATGCGGCTGCTCCTGCTCGTGCCGCTGCTGCTGGCTCC
AGCGCCCGGGTCCTCGGCTCCCAAGGTGAGGCGGCAGAGTGACACCTGGGGACCCTGGAGCCAGTGGAGC
CCCTGCAGCCGGACCTGTGGAGGGGGTGTCAGCTTCCGGGAGCGCCCCTGCTACTCCCAGAGGAGAGATG
GAGGCTCCAGCTGCGTGGGCCCCGCCCGGAGCCACCGCTCTTGTCGCACGGAGAGCTGCCCCGACGGCGC
CCGGGACTTCCGGGCCGAGCAGTGCGCGGAGTTCGACGGAGCGGAGTTCCAGGGGCGGCGGTATCGGTGG
CTGCCCTACTACAGCGCCCCAAACAAGTGTGAACTGAACTGCATTCCCAAGGGGGAGAACTTCTACTACA
AGCACAGGGAGGCTGTGGTTGATGGGACGCCCTGCGAGCCTGGCAAGAGGGATGTCTGTGTGGATGGCAG
CTGCCGGGTTGTCGGCTGTGATCACGAGCTGGACTCGTCCAAGCAGGAGGACAAGTGTCTGCGGTGTGGG
GGTGACGGCACGACCTGCTACCCCGTCGCAGGCACCTTTGACGCTAATGACCTCAGCCGAGCTGTGAAGA
ATGTTCGTGGGGAATACTACCTCAATGGGCACTGGACCATCGAGGCGGCCCGGGCCCTGCCAGCAGCCAG
CACCATCCTGCATTACGAGCGGGGTGCTGAGGGGGACCTGGCCCCTGAGCGACTCCATGCCCGGGGCCCC
ACCTCGGAGCCCCTGGTCATCGAGCTCATCAGCCAGGAGCCCAACCCCGGTGTGCACTATGAGTACCACC
TGCCCCTGCGCCGCCCCAGCCCCGGCTTCAGCTGGAGCCACGGCTCATGGAGTGACTGCAGCGCGGAGTG
TGGCGGAGGTCACCAGTCCCGCCTGGTGTTCTGCACCATCGACCATGAGGCCTACCCCGACCACATGTGC
CAGCGCCAGCCACGGCCAGCTGACCGGCGTTCCTGCAATCTTCACCCTTGCCCGGAGACCAAGCGCTGGA

AGGCAGGGCCATGGGCACCCTGCTCAGCCTCCTGTGGAGGAGGCTCCCAGTCCCGCTCCGTGTACTGCAT
CTCGTCTGACGGGGCCGGCATCCAGGAGGCCGTGGAGGAGGCTGAGTGTGCCGGGCTGCCTGGGAAGCCC
CCTGCCATTCAGGCCTGTAACCTGCAGCGCTGTGCAGCCTGGAGCCCGGAGCCCTGGGGAGAGTGTTCTG
TCAGTTGTGGCGTTGGCGTCCGGAAGCGGAGCGTTACTTGCCGGGGTGAAAGGGGTTCTTTGCTCCATAC
CGCAGCGTGCTCCTTGGAAGACCGGCCACCTCTGACTGAGCCCTGTGTGCATGAGGACTGCCCCCTCCTC
AGTGACCAGGCCTGGCATGTTGGCACCTGGGGTCTATGCTCCAAGAGCTGCAGCTCGGGCACTCGGAGGC
GACAGGTCATCTGTGCCATTGGGCCGCCCAGCCACTGCGGGAGCCTGCAGCACTCCAAGCCTGTGGATGT
GGAGCCTTGTAACACGCAGCCCTGTCATCTCCCCCAGGAGGTCCCCAGCATGCAGGATGTGCACACCCCT
GCCAGCAACCCCTGGATGCCGTTGGGCCCTCAGGAGTCCCCTGCCTCAGACTCCAGAGGCCAGTGGTGGG
CAGCCCAGGAACACCCCTCAGCCAGGGGTGACCACAGGGGAGAACGAGGTGACCCCAGGGGCGACCAAGG
CACCCACCTGTCAGCCCTGGGCCCCGCTCCCTCTCTGCAGCAGCCCCCATACCAGCAACCCCTGCGGTCG
GGCTCAGGGCCCCACGACTGCAGACACAGTCCTCACGGGTGCTGCCCCGATGGCCACACGGCATCTCTCG
GGCCTCAGTGGCAAGGCTGCCCTGGGGCCCCCTGTCAGCAGAGCAGGTACGGGTGCTGCCCTGACAGGGT
ATCTGTCGCTGAGGGGCCCCATCACGCTGGCTGCACAAAGTCGTATGGTGGTGACAGCACCGGGGGCATG
CCCAGGTCAAGGGCAGTGGCTTCTACAGTCCACAACACCCACCAGCCCCAGGCCCAGCAGAATGAGCCCA
GTGAGTGCCGGGGCTCCCAGTTTGGCTGTTGCTATGACAACGTGGCCACTGCAGCCGGTCCTCTTGGGGA
AGGCTGTGTGGGCCAGCCCAGCCATGCCTACCCCGTGCGGTGCCTGCTGCCCAGTGCCCATGGCTCTTGC
GCAGACTGGGCTGCCCGCTGGTACTTCGTTGCCTCTGTGGGCCAATGTAACCGCTTCTGGTATGGCGGCT
GCCATGGCAATGCCAATAACTTTGCCTCGGAGCAAGAGTGCATGAGCAGCTGCCAGGGATCTCTCCATGG
GCCCCGTCGTCCCCAGCCTGGGGCTTCTGGAAGGAGCACCCACACGGATGGTGGCGGCAGCAGTCCTGCA
GGCGAGCAGGAACCCAGCCAGCACAGGACAGGGGCCGCGGTGCAGAGAAAGCCCTGGCCTTCTGGTGGTC
TCTGGCGGCAAGACCAACAGCCTGGGCCAGGGGAGGCCCCCCACACCCAGGCCTTTGGAGAATGGCCATG
GGGGCAGGAGCTTGGGTCCAGGGCCCCTGGACTGGGTGGAGATGCCGGATCACCAGCGCCACCCTTCCAC
AGCTCCTCCTACAGGATTAGCTTGGCAGGTGTGGAGCCCTCGTTGGTGCAGGCAGCCCTGGGGCAGTTGG
TGCGGCTCTCCTGCTCAGACGACACTGCCCCGGAATCCCAGGCTGCCTGGCAGAAAGATGGCCAGCCCAT
CTCCTCTGACAGGCACAGGCTGCAGTTCGACGGATCCCTGATCATCCACCCCCTGCAGGCAGAGGACGCG
GGCACCTACAGCTGTGGCAGCACCCGGCCAGGCCGCGACTCCCAGAAGATCCAACTTCGCATCATAGGGG
GTGACATGGCCGTGCTGTCTGAGGCTGAGCTGAGCCGCTTCCCTCAGCCCAGGGACCCAGCTCAGGACTT
TGGCCAAGCGGGGGCTGCTGGGCCCCTGGGGGCCATCCCCTCTTCACACCCACAGCCTGCAAACAGGCTG
CGTTTGGACCAGAACCAGCCCCGGGTGGTGGATGCCAGTCCAGGCCAGCGGATCCGGATGACCTGCCGTG
CCGAAGGCTTCCCGCCCCCAGCCATCGAGTGGCAGAGAGATGGGCAGCCTGTCTCTTCTCCCAGACACCA
GCTGCAGCCTGATGGCTCCCTGGTCATTAGCCGAGTGGCTGTAGAAGATGGCGGCTTCTACACCTGTGTC
GCTTTCAATGGGCAGGACCGAGACCAGCGATGGGTCCAGCTCAGAGTTCTGGGGGAGCTGACAATCTCAG
GACTGCCCCCTACTGTGACAGTGCCAGAGGGTGATACGGCCAGGCTATTGTGTGTGGTAGCAGGAGAAAG
TGTGAACATCAGGTGGTCCAGGAACGGGCTACCTGTGCAGGCTGATGGCCACCGTGTCCACCAGTCCCCA
GATGGCACGCTGCTCATTTACAACTTGCGGGCCAGGGATGAGGGCTCCTACACGTGCAGTGCCTACCAGG
GGAGCCAGGCAGTCAGCCGCAGCACCGAGGTGAAGGTGGTCTCACCAGCACCCACCGCCCAGCCCAGGGA
CCCTGGCAGGGACTGCGTCGACCAGCCAGAGCTGGCCAACTGTGATTTGATCCTGCAGGCCCAGCTTTGT
GGCAATGAGTATTACTCCAGCTTCTGCTGTGCCAGCTGTTCACGTTTCCAGCCTCACGCTCAGCCCATCT
GGCAGTAGGGATGAAGGCTAGTTCCAGCCCCAGTCCAAAATAGTTCATAGGGCTAGGGAGAAAGGAAGAT
GGACTCTTGGCTTCCTCTCTCTGGCTGGCAAAGGGAGTTATCTTCTGGAATACATTAGCTCTTTCAAAAA
CCCACCCAGTGTTTAGCCTCAACGGCAGCCAGTTACCAGCTTCTCTCTGTAGCCTTCAGCAGTGTTTGCA
TCTCTGACATAACCACAGGCTGCTGTTTTCAAGAAGAGCAATCTGTTTGGATAAGAAAAACCTTTACTTT
ACAGCTTCCCTTTATAATTTGTTACACAGGAATAGTTAAATGCATTTGTTTGTTTGTTTTTTGAGACAGA
GTTTCACTCTTGTTGCCCAGGCTGGAGGGCAATGGCGCGATCTCAGCTCACTGCAACCTCCGTCTCCTGG
GTTCTTGATTCTCCTGTGTCAGCCTTCTGAGTAGCTAGGATTACAGATGCCTATCACCATGCCTGGGTAA
TTTTTGTATTTTTAGTTGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCGAACTTCTGACCTCAGA
TGATCTGCCCGCCTCAGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCACGCCCAGCCATCAATG
CATTTTTTTTATTTTTTTTTTGAGACAGAGTTTCGCACTTCTTGCCCAGGCTGGAGTACAATGGTGCGAT
CTTGGCTCACTGCAACCTCCACCTCCTGGGTTCAAGCGCTTCTCCAGCCTCAGCCTCCTGAGTAGCTGGG
ATTACAGGTATGTGCCACCATGCCTGGCTAATTTTGTATTTTTAGTAGAGACGGGGTTTCTCCATGTTGG
TCAGACTGGTCTTGAACTCCCGACCTCAGGTAATCCGCCCGCCTCGGCCTCCCAAAATGCTGGGATTAGA
GGTGTGAGCCACTGTGCCCAGCCCATCAATGTGTTTTAAAGCTAGCTGTCAGGGTTCCACTTAATTTAAA
GCTGGGCAGGGAGATGTGTAATGATTTCAAAGTTAACACCTGTTTGTTTTCTAAAGGGCATGCCAAGTCC
TGCTGTATCAGGGAAGTATTCTGTGCTAAAATCAGCGATGGTTCATTGCTCTAGTCTCTCTCACCCTTCT
AGGCAGTGCATCAGTCAGCTCTAAATCTGGTGCAGAGGGTTAACAGCATAACCCTTGTTGGCAAAATGGA
ATAGATGTTAAGACCTCAAATAGGGATTTGGGATGAAACAGCTGCAGTTAGCACTGTTATCTGAGCATGA
AAGAACTGGAAACGCTCCTTACGTCGAGATGTTGGACCTTGAAGCCCTCCTGAGGCCAACATGCAAATCT
GGCTGTGACGGTTCATCTGACACCTGTGTAAAGCTGACCAGCCTGCTCTGTACAGTGACAATGAGGAGCC
CCTCTCTTCCTTAAGTAGGAATCTGTGAAGCAAAATGTTTGCTGCCAAAGACAAATCAGACTGTCAGTCA
TTAAAAACAGCATTAGCAGGATGAGGATAGCAATGGGGAAGGGTTGTGGGCAATGCAGTAACAGGGAAAT

GGCTTCAGAAATGGTTTGAGTTGGAAGACAACATTCTTCATCTCTCAGGACTTCTAATTCCTTGATGCTA
AAAGAAGAGGCATGGATTCTATGAGCTTCCAAGTCCCTTTCCACTTTAACCTTCTACAAATCTTTCAGAG
GACTGCCTAGTAGCAAAGGTTATTCCTGGACACAGGAAAGACGGGCATTACAGGGACCAAAGCTCTGAAA
GGTGACTTTTATTACCAACACACTGGCTGGAAAAGGGACAAACCACATCACGGGTGAGTGATACTTCTCA
GTCTTCTCTACTCATTCAACAAAGGAAATGTGGGCTGGGGCAGAGGTCTTTTTTCATTTAATACTGGAAA
AATATTGAAGAGCATCCATGTTCACTTATGGCTGGTTTTGCTATAGAAATTGGAAAATAAAGGCCACTTT
TTTGAAATCCCCAGTTTAATTAAAAAAAAAAAAAAAAAA (SEQ ID NO: 53)

Translated protein sequence

MRLLLLVPLLLAPAPGSSAPKVRRQSDTWGPWSQWSPCSRTCGG
GVSFRERPCYSQRRDGGSSCVGPARSHRSCRTESCPDGARDFRAEQCAEFDGAEFQGR
RYRWLPYYSAPNKCELNCIPKGENFYYKHREAVVDGTPCEPGKRDVCVDGSCRVVGCD
HELDSSKQEDKCLRCGGDGTTCYPVAGTFDANDLSRAVKNVRGEYYLNGHWTIEAARA
LPAASTILHYERGAEGDLAPERLHARGPTSEPLVIELISQEPNPGVHYEYHLPLRRPS
PGFSWSHGSWSDCSAECGGGHQSRLVFCTIDHEAYPDHMCQRQPRPADRRSCNLHPCP
ETKRWKAGPWAPCSASCGGGSQSRSVYCISSDGAGIQEAVEEAECAGLPGKPPAIQAC
NLQRCAAWSPEPWGECSVSCGVGVRKRSVTCRGERGSLLHTAACSLEDRPPLTEPCVH
EDCPLLSDQAWHVGTWGLCSKSCSSGTRRRQVICAIGPPSHCGSLQHSKPVDVEPCNT
QPCHLPQEVPSMQDVHTPASNPWMPLGPQESPASDSRGQWWAAQEHPSARGDHRGERG
DPRGDQGTHLSALGPAPSLQQPPYQQPLRSGSGPHDCRHSPHGCCPDGHTASLGPQWQ
GCPGAPCQQSRYGCCPDRVSVAEGPHHAGCTKSYGGDSTGGMPRSRAVASTVHNTHQP
QAQQNEPSECRGSQFGCCYDNVATAAGPLGEGCVGQPSHAYPVRCLLPSAHGSCADWA
ARWYFVASVGQCNRFWYGGCHGNANNFASEQECMSSCQGSLHGPRRPQPGASGRSTHT
DGGGSSPAGEQEPSQHRTGAAVQRKPWPSGGLWRQDQQPGPGEAPHTQAFGEWPWGQE
LGSRAPGLGGDAGSPAPPFHSSSYRISLAGVEPSLVQAALGQLVRLSCSDDTAPESQA
AWQKDGQPISSDRHRLQFDGSLIIHPLQAEDAGTYSCGSTRPGRDSQKIQLRIIGGDM
AVLSEAELSRFPQPRDPAQDFGQAGAAGPLGAIPSSHPQPANRLRLDQNQPRVVDASP
GQRIRMTCRAEGFPPPAIEWQRDGQPVSSPRHQLQPDGSLVISRVAVEDGGFYTCVAF
NGQDRDQRWVQLRVLGELTISGLPPTVTVPEGDTARLLCVVAGESVNIRWSRNGLPVQ
ADGHRVHQSPDGTLLIYNLRARDEGSYTCSAYQGSQAVSRSTEVKVVSPAPTAQPRDP
GRDCVDQPELANCDLILQAQLCGNEYYSSFCCASCSRFQPHAQPIWQ (SEQ ID NO: 54)

PCDH1	5097	NM_002587	Homo sapiens protocadherin 1 (PCDH1), transcript variant 1, mRNA

mRNA Sequence

CGCAAAGCCGCCGGGCTGCTGCGCCCAGAGCCAGCCGGAGCCGGAGCCGGAGCCCGAACTGCAGCTCCAG
CCCCAGCCGTGCGGAGCCGCAGCCCAGGCCGGGGCCGGCGGCGGCTCATGGACAGCGGGGCGGGCGGCCG
GCGCTGCCCGGAGGCGGCCCTCCTGATTCTGGGGCCTCCCAGGATGGAGCACCTGAGGCACAGCCCAGGC
CCTGGGGGGCAACGGCTACTGCTGCCCTCCATGCTGCTAGCACTGCTGCTCCTGCTGGCTCCATCCCCAG
GCCACGCCACTCGGGTAGTGTACAAGGTGCCGGAGGAACAGCCACCCAACACCCTCATTGGGAGCCTCGC
AGCCGACTATGGTTTTCCAGATGTGGGGCACCTGTACAAGCTAGAGGTGGGTGCCCCGTACCTTCGCGTG
GATGGCAAGACAGGTGACATTTTCACCACCGAGACCTCCATCGACCGTGAGGGGCTCCGTGAATGCCAGA
ACCAGCTCCCTGGTGATCCCTGCATCCTGGAGTTTGAGGTATCTATCACAGACCTCGTGCAGAATGGCAG
CCCCCGGCTGCTAGAGGGCCAGATAGAAGTACAAGACATCAATGACAACACACCCAACTTCGCCTCACCA
GTCATCACTCTGGCCATCCCTGAGAACACCAACATCGGCTCACTCTTCCCCATCCCGCTGGCTTCAGACC
GTGATGCTGGTCCCAACGGTGTGGCATCCTATGAGCTGCAGGCTGGGCCTGAGGCCCAGGAGCTATTTGG
GCTGCAGGTGGCAGAGGACCAGGAGGAGAAGCAACCACAGCTCATTGTGATGGGCAACCTGGACCGTGAG
CGCTGGGACTCCTATGACCTCACCATCAAGGTGCAGGATGGCGGCAGCCCCCCACGCGCCAGCAGTGCCC
TGCTGCGTGTCACCGTGCTTGACACCAATGACAACGCCCCCAAGTTTGAGCGGCCCTCCTATGAGGCCGA
ACTATCTGAGAATAGCCCCATAGGCCACTCGGTCATCCAGGTGAAGGCCAATGACTCAGACCAAGGTGCC
AATGCAGAAATCGAATACACATTCCACCAGGCGCCCGAAGTTGTGAGGCGTCTTCTTCGACTGGACAGGA
ACACTGGACTTATCACTGTTCAGGGCCCGGTGGACCGTGAGGACCTAAGCACCCTGCGCTTCTCAGTGCT
TGCTAAGGACCGAGGCACCAACCCCAAGAGTGCCCGTGCCCAGGTGGTTGTGACCGTGAAGGACATGAAT
GACAATGCCCCCACCATTGAGATCCGGGGCATAGGGCTAGTGACTCATCAAGATGGGATGGCTAACATCT
CAGAGGATGTGGCAGAGGAGACAGCTGTGGCCCTGGTGCAGGTGTCTGACCGAGATGAGGGAGAGAATGC
AGCTGTCACCTGTGTGGTGGCAGGTGATGTGCCCTTCCAGCTGCGCCAGGCCAGTGAGACAGGCAGTGAC
AGCAAGAAGAAGTATTTCCTGCAGACTACCACCCCGCTAGACTACGAGAAGGTCAAAGACTACACCATTG
AGATTGTGGCTGTGGACTCTGGCAACCCCCCACTCTCCAGCACTAACTCCCTCAAGGTGCAGGTGGTGGA

CGTCAATGACAACGCACCTGTCTTCACTCAGAGTGTCACTGAGGTCGCCTTCCCGGAAAACAACAAGCCT
GGTGAAGTGATTGCTGAGATCACTGCCAGTGATGCTGACTCTGGCTCTAATGCTGAGCTGGTTTACTCTC
TGGAGCCTGAGCCGGCTGCTAAGGGCCTCTTCACCATCTCACCCGAGACTGGAGAGATCCAGGTGAAGAC
ATCTCTGGATCGGGAACAGCGGGAGAGCTATGAGTTGAAGGTGGTGGCAGCTGACCGGGGCAGTCCTAGC
CTCCAGGGCACAGCCACTGTCCTTGTCAATGTGCTGGACTGCAATGACAATGACCCCAAATTTATGCTGA
GTGGCTACAACTTCTCAGTGATGGAGAACATGCCAGCACTGAGTCCAGTGGGCATGGTGACTGTCATTGA
TGGAGACAAGGGGGAGAATGCCCAGGTGCAGCTCTCAGTGGAGCAGGACAACGGTGACTTTGTTATCCAG
AATGGCACAGGCACCATCCTATCCAGCCTGAGCTTTGATCGAGAGCAACAAAGCACCTACACCTTCCAGC
TGAAGGCAGTGGATGGTGGCGTCCCACCTCGCTCAGCTTACGTTGGTGTCACCATCAATGTGCTGGACGA
GAATGACAACGCACCCTATATCACTGCCCCTTCTAACACCTCTCACAAGCTGCTGACCCCCCAGACACGT
CTTGGTGAGACGGTCAGCCAGGTGGCAGCCGAGGACTTTGACTCTGGTGTCAATGCTGAGCTGATCTACA
GCATTGCAGGTGGCAACCCTTATGGACTCTTCCAGATTGGGTCACATTCAGGTGCCATCACCCTGGAGAA
GGAGATTGAGCGGCGCCACCATGGGCTACACCGCCTGGTGGTGAAGGTCAGTGACCGCGGCAAGCCCCCA
CGCTATGGCACAGCCTTGGTCCATCTTTATGTCAATGAGACTCTGGCCAACCGCACGCTGCTGGAGACCC
TCCTGGGCCACAGCCTGGACACGCCGCTGGATATTGACATTGCTGGGGATCCAGAATATGAGCGCTCCAA
GCAGCGTGGCAACATTCTCTTTGGTGTGGTGGCTGGTGTGGTGGCCGTGGCCTTGCTCATCGCCCTGGCG
GTTCTTGTGCGCTACTGCAGACAGCGGGAGGCCAAAAGTGGTTACCAGGCTGGTAAGAAGGAGACCAAGG
ACCTGTATGCCCCCAAGCCCAGTGGCAAGGCCTCCAAGGGAAACAAAAGCAAAGGCAAGAAGAGCAAGTC
CCCAAAGCCCGTGAAGCCAGTGGAGGACGAGGATGAGGCCGGGCTGCAGAAGTCCCTCAAGTTCAACCTG
ATGAGCGATGCCCCTGGGGACAGTCCCCGCATCCACCTGCCCCTCAACTACCCACCAGGCAGCCCTGACC
TGGGCCGCCACTATCGCTCTAACTCCCCACTGCCTTCCATCCAGCTGCAGCCCCAGTCACCCTCAGCCTC
CAAGAAGCACCAGGTGGTACAGGACCTGCCACCTGCAAACACATTCGTGGGCACCGGGGACACCACGTCC
ACGGGCTCTGAGCAGTACTCCGACTACAGCTACCGCACCAACCCCCCCAAATACCCCAGCAAGCAGGTAG
GCCAGCCCTTTCAGCTCAGCACACCCCAGCCCCTACCCCACCCCTACCACGGAGCCATCTGGACCGAGGT
GTGGGAGTGATGGAGCAGGTTTACTGTGCCTGCCCGTGTTGGGGGCCAGCCTGAGCCAGCAGTGGGAGGT
GGGGCCTTAGTGCCTCACCGGGCACACGGATTAGGCTGAGTGAAGATTAAGGGAGGGTGTGCTCTGTGGT
CTCCTCCCTGCCCTCTCCCCACTGGGGAGAGACCTGTGATTTGCCAAGTCCCTGGACCCTGGACCAGCTA
CTGGGCCTTATGGGTTGGGGGTGGTAGGCAGGTGAGCGTAAGTGGGGAGGGAAATGGGTAAGAAGTCTAC
TCCAAACCTAGGTCTCTATGTCAGACCAGACCTAGGTGCTTCTCTAGGAGGGAAACAGGGAGACCTGGGG
TCCTGTGGATAACTGAGTGGGGAGTCTGCCAGGGGAGGGCACCTTCCCATTGTGCCTTCTGTGTGTATTG
TGCATTAACCTCTTCCTCACCACTAGGCTTCTGGGGCTGGGTCCCACATGCCCTTGACCCTGACAATAAA
GTTCTCTATTTTTGGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
A (SEQ ID NO: 55)

Translated protein sequence

MDSGAGGRRCPEAALLILGPPRMEHLRHSPGPGGQRLLLPSMLL
ALLLLLAPSPGHATRVVYKVPEEQPPNTLIGSLAADYGFPDVGHLYKLEVGAPYLRVD
GKTGDIFTTETSIDREGLRECQNQLPGDPCILEFEVSITDLVQNGSPRLLEGQIEVQD
INDNTPNFASPVITLAIPENTNIGSLFPIPLASDRDAGPNGVASYELQAGPEAQELFG
LQVAEDQEEKQPQLIVMGNLDRERWDSYDLTIKVQDGGSPPRASSALLRVTVLDTNDN
APKFERPSYEAELSENSPIGHSVIQVKANDSDQGANAEIEYTFHQAPEVVRRLLRLDR
NTGLITVQGPVDREDLSTLRFSVLAKDRGTNPKSARAQVVVTVKDMNDNAPTIEIRGI
GLVTHQDGMANISEDVAEETAVALVQVSDRDEGENAAVTCVVAGDVPFQLRQASETGS
DSKKKYFLQTTTPLDYEKVKDYTIEIVAVDSGNPPLSSTNSLKVQVVDVNDNAPVFTQ
SVTEVAFPENNKPGEVIAEITASDADSGSNAELVYSLEPEPAAKGLFTISPETGEIQV
KTSLDREQRESYELKVVAADRGSPSLQGTATVLVNVLDCNDNDPKFMLSGYNFSVMEN
MPALSPVGMVTVIDGDKGENAQVQLSVEQDNGDFVIQNGTGTILSSLSFDREQQSTYT
FQLKAVDGGVPPRSAYVGVTINVLDENDNAPYITAPSNTSHKLLTPQTRLGETVSQVA
AEDFDSGVNAELIYSIAGGNPYGLFQIGSHSGAITLEKEIERRHHGLHRLVVKVSDRG
KPPRYGTALVHLYVNETLANRTLLETLLGHSLDTPLDIDIAGDPEYERSKQRGNILFG
VVAGVVAVALLIALAVLVRYCRQREAKSGYQAGKKETKDLYAPKPSGKASKGNKSKGK
KSKSPKPVKPVEDEDEAGLQKSLKFNLMSDAPGDSPRIHLPLNYPPGSPDLGRHYRSN
SPLPSIQLQPQSPSASKKHQVVQDLPPANTFVGTGDTTSTGSEQYSDYSYRTNPPKYP
SKQVGQPFQLSTPQPLPHPYHGAIWTEVWE (SEQ ID NO: 56)

PDGFB	5155	NM_002608	Homo sapiens platelet-derived growth factor beta polypeptide (simian sarcoma viral (v-sis) oncogene homolog) (PDGFB), transcript variant 1, mRNA

mRNA Sequence
CCTGCCTGCCTCCCTGCGCACCCGCAGCCTCCCCCGCTGCCTCCCTAGGGCTCCCCTCCGGCCGCCAGCG CCCATTTTTCATTCCCTAGATAGAGATACTTTGCGCGCACACACATACATACGCGCGCAAAAAGGAAAAA AAAAAAAAAAAGCCCACCCTCCAGCCTCGCTGCAAAGAGAAAACCGGAGCAGCCGCAGCTCGCAGCTCGC AGCTCGCAGCCCGCAGCCCGCAGAGGACGCCCAGAGCGGCGAGCGGGCGGGCAGACGGACCGACGGACTC GCGCCGCGTCCACCTGTCGGCCGGGCCCAGCCGAGCGCGCAGCGGGCACGCCGCGCGCGCGGAGCAGCCG TGCCCGCCGCCCGGGCCCCGCGCCAGGGCGCACACGCTCCCGCCCCCCTACCCGGCCCGGGCGGGAGTTT GCACCTCTCCCTGCCCGGGTGCTCGAGCTGCCGTTGCAAAGCCAACTTTGGAAAAAGTTTTTTGGGGGAG ACTTGGGCCTTGAGGTGCCCAGCTCCGCGCTTTCCGATTTTGGGGGCCTTTCCAGAAAATGTTGCAAAAA AGCTAAGCCGGCGGGCAGAGGAAAACGCCTGTAGCCGGCGAGTGAAGACGAACCATCGACTGCCGTGTTC CTTTTCCTCTTGGAGGTTGGAGTCCCCTGGGCGCCCCCACACGGCTAGACGCCTCGGCTGGTTCGCGACG CAGCCCCCCGGCCGTGGATGCTCACTCGGGCTCGGGATCCGCCCAGGTAGCGGCCTCGGACCCAGGTCCT GCGCCCAGGTCCTCCCCTGCCCCCCAGCGACGGAGCCGGGGCCGGGGGCGGCGGCGCCCGGGGGCCATGC GGGTGAGCCGCGGCTGCAGAGGCCTGAGCGCCTGATCGCCGCGGACCCGAGCCGAGCCCACCCCCCTCCC CAGCCCCCCACCCTGGCCGCGGGGGCGGCGCGCTCGATCTACGCGTCCGGGGCCCCGCGGGGCCGGGCCC GGAGTCGGCATGAATCGCTGCTGGGCGCTCTTCCTGTCTCTCTGCTGCTACCTGCGTCTGGTCAGCGCCG AGGGGGACCCCATTCCCGAGGAGCTTTATGAGATGCTGAGTGACCACTCGATCCGCTCCTTTGATGATCT CCAACGCCTGCTGCACGGAGACCCCGGAGAGGAAGATGGGGCCGAGTTGGACCTGAACATGACCCGCTCC CACTCTGGAGGCGAGCTGGAGAGCTTGGCTCGTGGAAGAAGGAGCCTGGGTTCCCTGACCATTGCTGAGC CGGCCATGATCGCCGAGTGCAAGACGCGCACCGAGGTGTTCGAGATCTCCCGGCGCCTCATAGACCGCAC CAACGCCAACTTCCTGGTGTGGCCGCCCTGTGTGGAGGTGCAGCGCTGCTCCGGCTGCTGCAACAACCGC AACGTGCAGTGCCGCCCCACCCAGGTGCAGCTGCGACCTGTCCAGGTGAGAAAGATCGAGATTGTGCGGA AGAAGCCAATCTTTAAGAAGGCCACGGTGACGCTGGAAGACCACCTGGCATGCAAGTGTGAGACAGTGGC AGCTGCACGGCCTGTGACCCGAAGCCCGGGGGGTTCCCAGGAGCAGCGAGCCAAAACGCCCCAAACTCGG GTGACCATTCGGACGGTGCGAGTCCGCCGGCCCCCCAAGGGCAAGCACCGGAAATTCAAGCACACGCATG ACAAGACGGCACTGAAGGAGACCCTTGGAGCCTAGGGGCATCGGCAGGAGAGTGTGTGGGCAGGGTTATT TAATATGGTATTTGCTGTATTGCCCCCATGGGGTCCTTGGAGTGATAATATTGTTTCCCTCGTCCGTCTG TCTCGATGCCTGATTCGGACGGCCAATGGTGCTTCCCCCACCCCTCCACGTGTCCGTCCACCCTTCCATC AGCGGGTCTCCTCCCAGCGGCCTCCGGCGTCTTGCCCAGCAGCTCAAGAAGAAAAAGAAGGACTGAACTC CATCGCCATCTTCTTCCCTTAACTCCAAGAACTTGGGATAAGAGTGTGAGAGAGACTGATGGGGTCGCTC TTTGGGGGAAACGGGCTCCTTCCCCTGCACCTGGCCTGGGCCACACCTGAGCGCTGTGGACTGTCCTGAG GAGCCCTGAGGACCTCTCAGCATAGCCTGCCTGATCCCTGAACCCCTGGCCAGCTCTGAGGGGAGGCACC TCCAGGCAGGCCAGGCTGCCTCGGACTCCATGGCTAAGACCACAGACGGGCACACAGACTGGAGAAAACC CCTCCCACGGTGCCCAAACACCAGTCACCTCGTCTCCCTGGTGCCTCTGTGCACAGTGGCTTCTTTTCGT TTTCGTTTTGAAGACGTGGACTCCTCTTGGTGGGTGTGGCCAGCACACCAAGTGGCTGGGTGCCCTCTCA GGTGGGTTAGAGATGGAGTTTGCTGTTGAGGTGGCTGTAGATGGTGACCTGGGTATCCCCTGCCTCCTGC CACCCCTTCCTCCCCACACTCCACTCTGATTCACCTCTTCCTCTGGTTCCTTTCATCTCTCTACCTCCAC CCTGCATTTTCCTCTTGTCCTGGCCCTTCAGTCTGCTCCACCAAGGGGCTCTTGAACCCCTTATTAAGGC CCCAGATGATCCCAGTCACTCCTCTCTAGGGCAGAAGACTAGAGGCCAGGGCAGCAAGGGACCTGCTCAT CATATTCCAACCCAGCCACGACTGCCATGTAAGGTTGTGCAGGGTGTGTACTGCACAAGGACATTGTATG CAGGGAGCACTGTTCACATCATAGATAAAGCTGATTTGTATATTTATTATGACAATTTCTGGCAGATGTA GGTAAAGAGGAAAAGGATCCTTTTCCTAATTCACACAAAGACTCCTTGTGGACTGGCTGTGCCCCTGATG CAGCCTGTGGCTTGGAGTGGCCAAATAGGAGGGAGACTGTGGTAGGGGCAGGGAGGCAACACTGCTGTCC ACATGACCTCCATTTCCCAAAGTCCTCTGCTCCAGCAACTGCCCTTCCAGGTGGGTGTGGGACACCTGGG AGAAGGTCTCCAAGGGAGGGTGCAGCCCTCTTGCCCGCACCCCTCCCTGCTTGCACACTTCCCCATCTTT GATCCTTCTGAGCTCCACCTCTGGTGGCTCCTCCTAGGAAACCAGCTCGTGGGCTGGGAATGGGGGAGAG AAGGGAAAAGATCCCCAAGACCCCCTGGGGTGGGATCTGAGCTCCCACCTCCCTTCCCACCTACTGCACT TTCCCCCTTCCCGCCTTCCAAAACCTGCTTCCTTCAGTTTGTAAAGTCGGTGATTATATTTTTGGGGGCT TTCCTTTTATTTTTTAAATGTAAAATTTATTTATATTCCGTATTTAAAGTTGTAAAAAAAAATAACCACA AAACAAAACCAAATGAAAAAAAAAAAAAAAAAA (SEQ ID NO: 57)
Translated protein sequence
MNRCWALFLSLCCYLRLVSAEGDPIPEELYEMLSDHSIRSFDDL QRLLHGDPGEEDGAELDLNMTRSHSGGELESLARGRRSLGSLTIAEPAMIAECKTRTE VFEISRRLIDRINANFLVWPPCVEVQRCSGCCNNRNVQCRPTQVQLRPVQVRKIEIVR KKPIFKKATVTLEDHLACKCETVAAARPVTRSPGGSQEQRAKTPQTRVTIRTVRVRRP PKGKHRKFKHTHDKTALKETLGA (SEQ ID NO: 58)

PHOSPHO2	493911	NM_001008489	Homo sapiens phosphatase, orphan 2 (PHOSPHO2),

			mRNA

mRNA Sequence

AACAAGGGAGGTGCTGCAGTTGGCGGTCGGGCTAGAGAAGAGGCGCCTGCGCTTGCGAGCTGGGCTTG
TGAGTGGGGCTGCCGAGAGGGCAGGCGTGGGGCGAGGCCAAAGGACTGAACCCGCAGGAGCGTCACGGGC
GCCGGGGCGGCTGCCGACGGCGGGACTGGGTTTTCTATCAGATGTTCCACGTAATAATGCTGGAGTTAAG
AAGTTTCCATTATTTTGCTCCAAACCAGAAGACTCTGTTCCCTGTATATAGAATAGGAGTAATATTTGAA
AACAACTGGCTGATGTTTAAAACTGAAGATTGTCATGATTGTTTATCCTAATCCCAATGCTGAAGTAAGA
TTGTCTTGGAAATACTAAGTTGGGGTAATCCAAATCTATTTCTGGAACCATGAAAATTTTGCTAGTTTTT
GACTTTGACAATACAATCATAGATGACAATAGTGACACTTGGATTGTACAATGTGCTCCCAACAAAAAGC
TTCCTATTGAACTACGTGATTCTTATCGAAAAGGATTTTGGACAGAATTTATGGGCAGAGTCTTTAAGTA
TTTGGGAGATAAGGGTGTAAGAGAACATGAAATGAAAAGAGCAGTGACATCATTGCCTTTCACTCCAGGG
ATGGTGGAACTCTTCAACTTTATAAGAAAGAATAAGGATAAATTTGACTGCATTATTATTTCAGATTCAA
ATTCGGTCTTCATAGATTGGGTTTTAGAAGCTGCCAGTTTTCATGACATATTTGATAAAGTGTTTACAAA
TCCAGCAGCTTTTAATAGCAATGGTCATCTCACTGTTGAAAATTATCATACTCATTCTTGCAATAGATGC
CCAAAGAATCTTTGCAAAAAGGTAGTTTTGATAGAATTTGTAGATAAACAGTTACAACAGGGAGTGAATT
ATACACAAATTGTTTATATTGGTGATGGTGGAAATGATGTCTGTCCAGTCACCTTTTTAAAGAATGATGA
TGTTGCCATGCCACGGAAAGGATATACCTTACAGAAAACTCTTTCCAGAATGTCTCAAAATCTTGAGCCT
ATGGAATATTCTGTTGTAGTTTGGTCCTCAGGTGTTGATATAATTTCTCATTTACAATTTCTAATAAAGG
ATTAATATGTCAGCAAAAAAAAAAAAAAA (SEQ ID NO: 59)

Translated protein sequence

MKILLVFDFDNTIIDDNSDTWIVQCAPNKKLPIELRDSYRKGFW
TEFMGRVFKYLGDKGVREHEMKRAVTSLPFTPGMVELFNFIRKNKDKFDCIIISDSNS
VFIDWVLEAASFHDIFDKVFTNPAAFNSNGHLTVENYHTHSCNRCPKNLCKKVVLIEF
VDKQLQQGVNYTQIVYIGDGGNDVCPVTFLKNDDVAMPRKGYTLQKTLSRMSQNLEPM
EYSVVVWSSGVDIISHLQFLIKD (SEQ ID NO: 60)

PSENEN	55851	NM_172341	Homo sapiens presenilin enhancer 2 homolog (C. elegans) (PSENEN), mRNA

mRNA Sequence

CTCGCCCAAAGAAGACTACAATCTCCAGGGAAACCTGGGGCGTCTCGCGCAAACGTCCATAACTGAAAGT
AGCTAAGGCACCCCAGCCGGAGGAAGTGAGCTCTCCTGGGGCGTGGTTGTTCGTGATCCTTGCATCTGTT
ACTTAGGGTCAAGGCTTGGGTCTTGCCCCGCAGACCCTTGGGACGACCCGGCCCCAGCGCAGCTATGAAC
CTGGAGCGAGTGTCCAATGAGGAGAAATTGAACCTGTGCCGGAAGTACTACCTGGGGGGGTTTGCTTTCC
TGCCTTTTCTCTGGTTGGTCAACATCTTCTGGTTCTTCCGAGAGGCCTTCCTTGTCCCAGCCTACACAGA
ACAGAGCCAAATCAAAGGCTATGTCTGGCGCTCAGCTGTGGGCTTCCTCTTCTGGGTGATAGTGCTCACC
TCCTGGATCACCATCTTCCAGATCTACCGGCCCCGCTGGGGTGCCCTTGGGGACTACCTCTCCTTCACCA
TACCCCTGGGCACCCCCTGACAACTTCTGCACATACTGGGGCCCTGCTTATTCTCCCAGGACAGGCTCCT
TAAAGCAGAGGAGCCTGTCCTGGGAGCCCCTTCTCAAACTCCTAAGACTTGTTTTCATGTCCCACGTTCT
CTGCTGACATCCCCCAATAAAGGACCCTAACTTTCAAAAAAAAAAAAA (SEQ ID NO: 61)

Translated protein sequence

MNLERVSNEEKLNLCRKYYLGGFAFLPFLWLVNIFWFFREAFLV
PAYTEQSQIKGYVWRSAVGFLFWVIVLTSWITIFQIYRPRWGALGDYLSFTIPLGTP (SEQ ID NO: 62)

SATB1	6304	NM_002971	Homo sapiens SATB homeobox 1 (SATB1), transcript variant 1, mRNA

mRNA Sequence

TTTCTCGCTCGCTCCCGTTCCCCGGACGCGGCGGATGAGCCGGCCCCGCTGGGGAAGGCTCCGGGCGGCG
GCGGGCGGCCGGGAGGAGGCTGCGTGCTCGGGGCTGGGGCTGCGAGCGGGGTGATTTTGTATTAAAATGA
GGAGGAGGAAGAAGAGGCACCCACAGCGGCAGCGGCGGCGGCGGCAGCAGCAGCAGGAGCAGCGGCG
GAGAGGGCTGCAGCCCGGGCGGACGCGCGGAGCCGAGCGGGGCACGGCGGCGGCAGCGACAGCGGCCGGG
ATGAGTCAACTAATAATTTAATGGGGGCAGAGACGGCAGCGAGGGGTAGAGCTAGCGAGGGAGAGAGCGA
GAGAAGCAGCCCCGTCCGGGGACTCGCGCTCACACTCACGCACACACACAAACACACACACACCTCTCCC
TGTGCCACCCAGCAACACCCGGCCTCGTCACAACAACAACAGCCGCGGCCGCCCTCTATCCTGCCCGGGG
GCCCAGCCGAAAGCCAGGGCGACTCTAGAGGACGCTGCCCGCCCCCCTCTTTCATTTCGGGAAACTCCTG

ATCAGTTTTGTCGGGGTTTCTGGGTTTCTTTTCCCCCAAAGTCCTAGTGCCATTGTGGTGCTCGTTGTTT
ACCTCGGACTCTGGACGAGTGAGAGCTTGGCGACTTTTTGGGGGGAGGGGGCGGGGAGTTTGTCGCTGCC
TAGGCGGTGGAGGTGGCTGGGGGTGCCTTCTGATCTTCCTCCTCCTCCCCCTCCCCCCGAACCTCTTCTC
TCCTCACTTGCTGGGACCCCAGACGCTCACAGCCCCGCGTCAATGGGCAGGGAGAGGGTCCTTGCGGCTG
TTGTCAGCGAGGGCAGAATCAAAAGTGGCATTTTAGTGCCTTTCCGGGGCTTTTCTCGCGACCCCCTGCC
CCCCACCCTCGCTGTCCCCCGCTAGATGCCCTCGTTGGGGGTGCGAGGCTGTGGGGAAAAGTTTAAGGTT
TGTTAATATTAGTCGCGATTGTTGGCGAGGGGGGTGGGGGTGATTGGAAGGGAGGCGAGGTGGCCTTCCC
AATGCGCGTTATTCGGGGTTATTGAAGAATAATATTGCAAGTGACAGCCAGAAGTAGACTTTCTGTCCTC
ACACCGAAGAACCCGAGTGAGCAGGAGGGAGGGAGAGACGCGAAGAGACCTTTTTTCCTTTTTGGAGACC
TTGTCCGCAGTGATTTTTTTTTTTTTAAGAGAATCCTCAGTCACCACGTCGTTTCCCCAGCACCATCACA
GTGTACAGCTCATAACGGGTTTTGCTTTGTTTTTACGATTTCCCCCCAACGAATCACTTGTCAGATCAAT
TTTATCTTCTTCCTCCTCCCTGCTTCCCACTCTCCCCTCCTCCCCATCGCAAACCCTGTTCTCTGAGGTT
AGACATTTTACAAACCCCTATATGTTGGTTTTCGAATTGTGATTTTTTTTTTAAACCCCTTTCTCATGGC
TACTCTTCTAGACGTTTATTTCTGCCCTTCCCCCGCTTAGGGGGGCGGGGGTAGGGGAAAGGAAAATAAT
ACAATTTCAGGGGAAGTCGCCTTCAGGTCTGCTGCTTTTTTATTTTTTTTTTTTTAATTAAAAAAAAAAA
GGACATAGAAAACATCAGTCTTGAACTTCTCTTCAAGAACCCGGGCTGCAAAGGAAATCTCCTTTGTTTT
TGTTATTTATGTGCTGTCAAGTTTTGAAGTGGTGATCTTTAGACAGTGACTGAGTATGGATCATTTGAAC
GAGGCAACTCAGGGGAAAGAACATTCAGAAATGTCTAACAATGTGAGTGATCCGAAGGGTCCACCAGCCA
AGATTGCCCGCCTGGAGCAGAACGGGAGCCCGCTAGGAAGAGGAAGGCTTGGGAGTACAGGTGCAAAAAT
GCAGGGAGTGCCTTTAAAACACTCGGGCCATCTGATGAAAACCAACCTTAGGAAAGGAACCATGCTGCCA
GTTTTCTGTGTGGTGGAACATTATGAAAACGCCATTGAATATGATTGCAAGGAGGAGCATGCAGAATTTG
TGCTGGTGAGAAAGGATATGCTTTTCAACCAGCTGATCGAAATGGCATTGCTGTCTCTAGGTTATTCACA
TAGCTCTGCTGCCCAGGCCAAAGGGCTAATCCAGGTTGGAAAGTGGAATCCAGTTCCACTGTCTTACGTG
ACAGATGCCCCTGATGCTACAGTAGCAGATATGCTTCAAGATGTGTATCATGTGGTCACATTGAAAATTC
AGTTACACAGTTGCCCCAAACTAGAAGACTTGCCTCCCGAACAATGGTCGCACACCACAGTGAGGAATGC
TCTGAAGGACTTACTGAAAGATATGAATCAGAGTTCATTGGCCAAGGAGTGCCCCCTTTCACAGAGTATG
ATTTCTTCCATTGTGAACAGTACTTACTATGCAAATGTCTCAGCAGCAAAATGTCAAGAATTTGGAAGGT
GGTACAAACATTTCAAGAAGACAAAAGATATGATGGTTGAAATGGATAGTCTTTCTGAGCTATCCCAGCA
AGGCGCCAATCATGTCAATTTTGGCCAGCAACCAGTTCCAGGGAACACAGCCGAGCAGCCTCCATCCCCT
GCGCAGCTCTCCCATGGCAGCCAGCCCTCTGTCCGGACACCTCTTCCAAACCTGCACCCTGGGCTCGTAT
CAACACCTATCAGTCCTCAATTGGTCAACCAGCAGCTGGTGATGGCTCAGCTGCTGAACCAGCAGTATGC
AGTGAATAGACTTTTAGCCCAGCAGTCCTTAAACCAACAATACTTGAACCACCCTCCCCCTGTCAGTAGA
TCTATGAATAAGCCTTTGGAGCAACAGGTTTCGACCAACACAGAGGTGTCTTCCGAAATCTACCAGTGGG
TACGCGATGAACTGAAACGAGCAGGAATCTCCCAGGCGGTATTTGCACGTGTGGCTTTTAACAGAACTCA
GGGCTTGCTTTCAGAAATCCTCCGAAAGGAAGAGGACCCCAAGACTGCATCCCAGTCTTTGCTGGTAAAC
CTTCGGGCTATGCAGAATTTCTTGCAGTTACCGGAAGCTGAAAGAGACCGAATATACCAGGACGAAAGGG
AAAGGAGCTTGAATGCTGCCTCGGCCATGGGTCCTGCCCCCCTCATCAGCACACCACCCAGCCGTCCTCC
CCAGGTGAAAACAGCTACTATTGCCACTGAAAGGAATGGGAAACCAGAGAACAATACCATGAACATTAAT
GCTTCCATTTATGATGAGATTCAGCAGGAAATGAAGCGTGCTAAAGTGTCTCAAGCACTGTTTGCAAAGG
TTGCAGCAACCAAAAGCCAGGGATGGTTGTGCGAGCTGTTACGCTGGAAAGAAGATCCTTCTCCAGAAAA
CAGAACCCTGTGGGAGAACCTCTCCATGATCCGAAGGTTCCTCAGTCTTCCTCAGCCAGAACGTGATGCC
ATTTATGAACAGGAGAGCAACGCGGTGCATCACCATGGCGACAGGCCGCCCCACATTATCCATGTTCCAG
CAGAGCAGATTCAGCAACAGCAGCAGCAACAGCAACAGCAGCAGCAGCAGCAGGCACCGCCGCCTCC
ACAGCCACAGCAGCAGCCACAGACAGGCCCTCGGCTCCCCCCACGGCAACCCACGGTGGCCTCTCCAGCA
GAGTCAGATGAGGAAAACCGACAGAAGACCCGGCCACGAACAAAAATTTCAGTGGAAGCCTTGGGAATCC
TCCAGAGTTTCATACAAGACGTGGGCCTGTACCCTGACGAAGAGGCCATCCAGACTCTGTCTGCCCAGCT
CGACCTTCCCAAGTACACCATCATCAAGTTCTTTCAGAACCAGCGGTACTATCTCAAGCACCACGGCAAA
CTGAAGGACAATTCCGGTTTAGAGGTCGATGTGGCAGAATATAAAGAAGAGGAGCTGCTGAAGGATTTGG
AAGAGAGTGTCCAAGATAAAAATACTAACACCCTTTTTTCAGTGAAACTAGAAGAAGAGCTGTCAGTGGA
AGGAAACACAGACATTAATACTGATTTGAAAGACTGAGATAAAAGTATTTGTTTCGTTCAACAGTGCCAC
TGGTATTTACTAACAAAATGAAAAGTCCACCTTGTCTTCTCTCAGAAAACCTTTGTTGTTCATTGTTTGG
CCAATGAATCTTCAAAAACTTGCACAAACAGAAAAGTTGGAAAAGGATAATACAGACTGCACTAAATGTT
TTCCTCTGTTTTACAAACTGCTTGGCAGCCCCAGGTGAAGCATCAAGGATTGTTTGGTATTAAAATTTGT
GTTCACGGGATGCACCAAAGTGTGTACCCCGTAAGCATGAAACCAGTGTTTTTTGTTTTTTTTTTAGTTC
TTATTCCGGAGCCTCAAACAAGCATTATACCTTCTGTGATTATGATTTCCTCTCCTATAATTATTTCTGT
AGCACTCCACACTGATCTTTGGAAACTTGCCCCTTATTTAAAAAAAAAAAAGAAAAAAAAGAGTTTGTTA
CTCTATTGTATGTTACAAAAGAACTATAGACTGTGGAATGCAGTTTAAAGATGACATATGCCAACAAATG
CCTTGTATTATATGGCACTGCCGTAATTCAAATTTGTTTTTATTTTGGAAATAAAAGTTCACTGTACTTT
TTTTTCATTCTCATTGTTACATGATTTTTTAAAAAAAGGAAAAGAAAATGTGAAACACAATTTAGTCCTC
ATTATTTATTTGTAGATCCTGCAGCATCATGTTGTAATTAATTTTTTGGAAGTTTCCGTTAAATGTAATA
TTGCTTCTCTTGTTACCATACTGATTCTTTTCTATTTATAAATGTATTTTGATGGGCAGTAAAACAAAGT

GTCTTAAAAGTTTTAAATAGAGAAAATGTGCTTTACACAGTTGCCTATAAAAAGTGCTCTATGTTATCCA
AGCAATTCATACTATAAGCTTCACTCTTATTGTTGTATGCAATTTTTACTATCATGCAAATAAGCTTAGG
TAAATAAAACTAATAGATCACCTTAGAAAATTATGCAATTAATGTGAAAATAATTGATGTTTGCAATGTG
TCTTCCTTTGGTTTACAATCAATTTTAAAGCTACATCTGTATAAAATTTCTGTATAAAGGTGTATTTCTT
TTTTATGAGTTTATGGCTATGAAAACAGCTATTTTGTTACAGCTGGCTGTTTTTATAAGTGTATCACAAT
TTTCTTTATGCAGAAATGTTCTGACTAGGAGTGGTTATTGACTGTAACTACACAATTAAAATTGTTTGTA
TCGTATGACATGGTAGGGTTTGTCTGCTTATGTGAAGTAACTAAAGGAGTCAAAGGATGGCCCTCTCATT
TAGGTGCATGTTAATAACTTGTTATTTCACTGATTTTAAAAAGAGCAATTGACAAGTTACTTGAAACACT
GTAAATTTAAATCACAAACACATGCTCATTTTTAAATAGGTATGAAATTTCACAATGAAAATAACCTGTT
TGGTTAACATTTTGCTTAATAAGTAGAGATAGGATGGTCAAAAGACTCTCCGACAAAAACAAATCCAGTC
TCTAGCAGTTATGTTGTTAGAATGGA (SEQ ID NO: 63)

Translated protein sequence

MDHLNEATQGKEHSEMSNNVSDPKGPPAKIARLEQNGSPLGRGR
LGSTGAKMQGVPLKHSGHLMKTNLRKGTMLPVFCVVEHYENAIEYDCKEEHAEFVLVR
KDMLFNQLIEMALLSLGYSHSSAAQAKGLIQVGKWNPVPLSYVTDAPDATVADMLQDV
YHVVTLKIQLHSCPKLEDLPPEQWSHTTVRNALKDLLKDMNQSSLAKECPLSQSMISS
IVNSTYYANVSAAKCQEFGRWYKHFKKTKDMMVEMDSLSELSQQGANHVNFGQQPVPG
NTAEQPPSPAQLSHGSQPSVRTPLPNLHPGLVSTPISPQLVNQQLVMAQLLNQQYAVN
RLLAQQSLNQQYLNHPPPVSRSMNKPLEQQVSTNTEVSSEIYQWVRDELKRAGISQAV
FARVAFNRTQGLLSEILRKEEDPKTASQSLLVNLRAMQNFLQLPEAERDRIYQDERER
SLNAASAMGPAPLISTPPSRPPQVKTATIATERNGKPENNTMNINASIYDEIQQEMKR
AKVSQALFAKVAATKSQGWLCELLRWKEDPSPENRTLWENLSMIRRFLSLPQPERDAI
YEQESNAVHHHGDRPPHIIHVPAEQIQQQQQQQQQQQQQQAPPPQPQQQPQTGPRL
PPRQPTVASPAESDEENRQKTRPRTKISVEALGILQSFIQDVGLYPDEEAIQTLSAQL
DLPKYTIIKFFQNQRYYLKHHGKLKDNSGLEVDVAEYKEEELLKDLEESVQDKNTNTL
FSVKLEEELSVEGNTDINTDLKD (SEQ ID NO: 64)

SNX11	29916	NM_013323	Homo sapiens sorting nexin 11 (SNX11), transcript variant 2, mRNA

mRNA Sequence

CCGGCGTCCCAAGTGAGTGGAGGGGGGATCCCGACTCCAGTCCGGGGCCTTGGCCAGCGGAGCCGCGCTA
TTCGGAAGCGGGAATCCCACTCAGAGCCCGGGCCTGTAGGGGCGGGGCGTCCCGGGCACCCGGGATTGGG
GCGTCTCCCGTCGTGCACCGGGGCACCGGCGACTCACCCGGAAGGAGAAGCCGTGATCTGGCTATATGGT
GGGGCGCGGGCGGTGTCGCTGTGGGGAGCTGGTGCTGTTCTCAGATGTTTCCTTCCAATGGGCTTTTGGT
GTAGGATGTCGGAGAACCAAGAACAGGAGGAGGTGATTACAGTGCGTGTTCAGGACCCCCGAGTGCAGAA
TGAGGGCTCCTGGAACTCTTATGTGGATTATAAGATATTCCTCCATACCAACAGCAAAGCCTTTACTGCC
AAGACTTCCTGTGTGCGGCGCCGCTACCGTGAGTTCGTGTGGCTGAGAAAGCAGCTACAGAGAAATGCTG
GTTTGGTGCCTGTTCCTGAACTTCCTGGGAAGTCAACCTTCTTCGGCACCTCAGATGAGTTCATTGAGAA
GCGACGACAAGGTCTGCAGCACTTCCTTGAAAAGGTCCTGCAGAGTGTGGTTCTCCTGTCAGACAGCCAG
TTGCACCTATTCCTGCAAAGCCAGCTCTCGGTGCCTGAGATAGAAGCCTGTGTCCAGGGCCGAAGTACCA
TGACTGTGTCTGATGCCATTCTTCGATATGCTATGTCAAACTGTGGCTGGGCCCAGGAAGAGAGGCAGAG
CTCTTCTCACCTGGCTAAAGGAGACCAGCCTAAGAGTTGCTGCTTTCTTCCAAGATCGGGTAGGAGGAGC
TCTCCCTCACCGCCTCCCAGTGAAGAAAAGGACCATTTAGAAGTGTGGGCTCCAGTTGTTGACTCTGAGG
TTCCTTCCTTGGAAAGTCCCACTCTCCCACCCCTCTCCTCACCATTATGCTGTGATTTTGGAAGACCCAA
AGAGGGAACCTCCACTCTTCAGTCTGTGAGGAGGGCTGTGGGAGGAGATCATGCTGTGCCTTTGGACCCT
GGTCAGTTAGAAACAGTTTTGGAAAAGTGAGCTCTGGGTTCTGCTCTGAGATGGTCAGAGAAGATGCGGG
CCAGGAGACTTACTCAGGTGGGACTGGGCACAGGGCAGGTATGTGGGAGGCTGGGCTGCTTAGTGTCTTC
TAGTCACCTCTGCTTGGGCTGATTGACAGAGGTCAGTCATTACAGCCCCTTATGCCTCTTCCATGGGAAC
AAATACTGTGCAGATGTTTGTAAGTTAAACATAAGACACAGGGGCTGTTGCTTTTGAACAGAACCCTATA
TTACTCTCCTGGGATCTGAGTTTCTGCAGGTCATTTGTATGTAGGACCAGGAGTATCTCCTCAGGTGACC
AGTTTTGGGGACCCGTATGTGGCAAATTCTAAGCTGCCATATTGAACATCATCCCACTGGGAGTGGTTAT
GTTGTATCCCCATCTTGGCTGGCTTCAGTTTTTGCTGTAGCCCTAGAGCACTTTGTTTGTGGGAGGCTGG
CCTCTTGCCTACCTCCTTGCATGGACAGGGGGATGAATATTTACTTTCCCACCTCCTTGCTTTTTCTTTC
ACTGATACCACTGAATGGAACTGGTGCTGTGACTCCTGCTGCTGGGGATTTATGTCCCGAGACCTTAGCC
TGGCTGAGTGGAGCCTGAGACCTGCACAACAGCTCATGGTCATGCATGAGAGAGAAGTGGCTGGCCACAG
CCAGAGGGAACAGTAACAGCCCAGGGGCCTTTATTTTGGGAAAGGCTGTCCCGGGCTGTTACTGTCTCTT
CTGGTTATAAAGCAGACATGTGGCCATCTTTTCCGCAGGGTTAGAGTGGGCTCCTTTCTTTTTGGAATCC

TTTTCTTCTCCTTTGGTAGCAGCTCCCTGCCTCCAGGGCTTCCGCCACCAGCGTCTCTGCTGTGTTGCGC
AGTGCAGTGGGGTGCAAGGGCTTTGTTTCTGCCTGCCTGAAAGAGAGGGCTCTGGGGATGGAGATGAGAA
ACAACACGCTCTCCTTCAGACAATGAGGCATTCTGTCCTCCTGCTGCCATTCTTCATCTCCACTGAGAGC
CAGAGCTGGTAGGAGCCGAGTGCCACAGGCATTCTGCATTGCTCTACTCTTAGGTTTGTGTGTGTGATCC
TTCCCCTCCCTGTCGCCCACTCCTCCCTCCTCTGGCTATCCTACCCTGTCTGTGGGCTCTTTTACTACCA
GCCTATGCTGTGGGACTGTCATGGCATTTAGTTCAGAGTGGAGGGGCTTTGGCCTGAAATAAAATGCAAG
TATTT (SEQ ID NO: 65)

Translated protein sequence

MGFWCRMSENQEQEEVITVRVQDPRVQNEGSWNSYVDYKIFLHT
NSKAFTAKTSCVRRRYREFVWLRKQLQRNAGLVPVPELPGKSTFFGTSDEFIEKRRQG
LQHFLEKVLQSVVLLSDSQLHLFLQSQLSVPEIEACVQGRSTMTVSDAILRYAMSNCG
WAQEERQSSSHLAKGDQPKSCCFLPRSGRRSSPSPPPSEEKDHLEVWAPVVDSEVPSL
ESPTLPPLSSPLCCDFGRPKEGTSTLQSVRRAVGGDHAVPLDPGQLETVLEK (SEQ ID NO: 66)

SPTA1	6708	NM_003126	Homo sapiens spectrin, alpha, erythrocytic 1 (elliptocytosis 2) (SPTA1), mRNA

mRNA Sequence

TATGTCTTCTAAAGATAATGTCGATTGTGTATGGCTGATGGGATTCTAGGACCAAGCAAGAGGTTTTTTT
TTTTCCCCCACATACTTAACGTTTCTATATTTCTATTTGAATTCGACTGGACAGTTCCATTTGAATTATT
TCTCTCTCTCTCTCTCTCTGACACATTTTATCTTGCCAGGTTCTAAACCTTTAGGAAAAATGGAGCAATT
TCCAAAGGAAACCGTTGTGGAGAGCAGTGGGCCAAAGGTTTTGGAAACAGCAGAAGAGATCCAGGAGAGG
CGTCAGGAAGTGTTGACTCGGTATCAAAGTTTCAAGGAGCGGGTCGCTGAGAGGGGTCAGAAGCTTGAGG
ATTCCTATCACTTACAAGTTTTCAAGCGAGATGCAGATGATCTGGGGAAGTGGATCATGGAGAAAGTCAA
TATCTTAACCGATAAGAGCTATGAAGACCCAACTAATATACAGGGGAAATATCAGAAGCATCAATCCCTT
GAAGCAGAGGTGCAAACAAAATCAAGACTCATGTCTGAACTGGAAAAAACAAGGGAAGAACGATTTACCA
TGGGTCATTCTGCCCACGAAGAAACGAAGGCCCATATAGAGGAGCTACGCCACCTGTGGGACCTGCTGTT
AGAGCTGACCCTGGAGAAGGGTGACCAGTTGCTGCGGGCCCTGAAGTTCCAGCAGTATGTACAGGAGTGT
GCTGACATCTTAGAGTGGATTGGAGACAAGGAGGCTATAGCGACATCAGTGGAGCTAGGTGAAGACTGGG
AGCGCACCGAAGTTCTGCATAAGAAATTTGAAGACTTCCAAGTGGAGCTGGTAGCTAAAGAAGGGAGAGT
TGTTGAAGTGAACCAATATGCCAATGAGTGTGCCGAGGAAAACCATCCTGACCTACCCTTAATTCAGTCT
AAGCAAAATGAGGTGAATGCTGCCTGGGAGCGCCTTCGTGGTTTGGCTCTCCAGAGACAGAAAGCTCTGT
CCAATGCTGCAAACTTACAACGATTCAAAAGGGATGTGACTGAAGCCATCCAGTGGATCAAGGAGAAGGA
ACCTGTACTCACCTCTGAGGACTATGGCAAAGACCTTGTTGCCTCTGAAGGACTGTTTCACAGTCACAAG
GGACTTGAGAGAAATCTTGCTGTCATGAGTGACAAGGTGAAGGAGTTATGTGCTAAAGCAGAGAAGCTGA
CACTTTCCCATCCTTCAGATGCACCTCAGATCCAGGAGATGAAAGAAGATCTGGTCTCCAGCTGGGAGCA
TATTCGTGCCCTGGCCACCAGCAGATATGAAAAACTGCAGGCTACTTATTGGTACCATCGATTTTCATCT
GACTTTGATGAACTCTCAGGCTGGATGAACGAGAAGACTGCTGCGATCAATGCTGATGAGCTGCCAACAG
ATGTGGCTGGTGGAGAAGTTCTGCTGGACAGGCATCAGCAGCATAAGCATGAGATTGACTCTTACGATGA
CCGATTTCAATCTGCTGATGAGACTGGTCAAGACCTCGTGAATGCCAATCATGAAGCCTCTGATGAAGTT
CGGGAAAAGATGGAAATACTTGACAACAACTGGACTGCCCTGCTGGAACTGTGGGACGAGCGTCATCGTC
AGTATGAGCAGTGCTTGGACTTTCATCTCTTCTACAGAGACAGTGAGCAAGTGGACAGTTGGATGAGTAG
ACAAGAGGCCTTCCTGGAAAACGAGGATCTGGGAAACTCACTGGGCAGTGCAGAAGCCCTTCTTCAGAAG
CATGAAGACTTTGAGGAAGCCTTTACTGCCCAGGAAGAGAAGATCATAACTGTAGACAAGACTGCAACCA
AATTGATTGGTGATGACCATTATGATTCAGAGAACATCAAGGCTATCCGTGACGGGCTGTTAGCCCGGCG
GGATGCCCTACGTGAAAAGGCTGCCACTAGACGTAGATTGCTGAAGGAGTCATTGCTTCTGCAAAAACTG
TATGAGGACTCAGATGACCTAAAGAACTGGATCAACAAGAAGAAAAAGTTGGCAGATGATGAAGATTACA
AGGACATACAGAACTTGAAGAGCAGGGTTCAAAAGCAGCAAGTCTTTGAAAAGGAGTTGGCAGTTAATAA
GACCCAGCTGGAAAACATACAGAAAACTGGCCAAGAGATGATTGAGGGTGGTCACTATGCCTCTGACAAT
GTGACCACTCGTCTGAGTGAAGTTGCCAGCCTCTGGGAGGAGTTGCTGGAGGCTACAAAACAGAAAGGGA
CCCAGTTGCATGAGGCCAACCAGCAGCTGCAATTTGAAAATAATGCAGAAGATTTGCAGCGCTGGCTGGA
GGATGTTGAGTGGCAAGTCACCTCTGAGGATTATGGGAAAGGCCTGGCCGAGGTACAGAATCGACTCAGG
AAACACGGCCTCCTGGAGTCGGCTGTGGCTGCTCGTCAGGATCAGGTGGATATCCTTACAGACCTGGCTG
CATATTTTGAAGAAATAGGCCATCCTGATTCTAAGGATATAAGGGCAAGGCAAGAGTCCTTGGTATGCCG
ATTTGAAGCTCTGAAAGAGCCACTGGCCACCCGAAAGAAGAAGCTCTTAGACCTTCTCCATCTGCAGCTG
ATTTGTAGAGACACAGAGGATGAGGAGGCCTGGATCCAAGAGACTGAACCCTCAGCTACTTCCACCTACC
TTGGAAAGGACCTGATTGCTTCCAAAAAGCTTCTGAATAGGCATAGAGTCATCCTGGAGAACATTGCCAG
CCATGAACCACGCATTCAAGAGATAACAGAAAGGGGAAACAAAATGGTAGAGGAAGGACACTTTGCTGCA
GAAGATGTGGCCTCTAGGGTCAAGAGTTTGAACCAGAATATGGAGTCTCTCCGTGCTCGAGCTGCTAGGC

GACAAAATGATCTTGAAGCCAATGTCCAGTTCCAGCAGTACCTGGCTGACCTGCATGAAGCAGAAACATG
GATCAGAGAGAAGGAACCTATTGTAGATAATACTAACTATGGTGCTGATGAAGAAGCAGCTGGGGCTCTT
CTAAAGAAGCATGAGGCCTTTCTATTAGATCTCAATTCATTTGGAGACAGTATGAAAGCTCTGCGGAATC
AGGCAAACGCCTGCCAGCAACAACAGGCTGCACCAGTGGAGGGAGTTGCTGGAGAACAAAGGGTCATGGC
TTTATATGACTTCCAGGCCCGCAGCCCCCGAGAAGTCACCATGAAGAAAGGTGATGTCTTAACGCTGCTC
AGTTCCATCAATAAGGACTGGTGGAAGGTGGAAGCTGCTGATCATCAGGGCATTGTCCCAGCTGTCTATG
TCAGAAGACTGGCCCACGATGAGTTCCCGATGCTCCCACAGCGGCGACGAGAAGAGCCAGGAAACATCAC
CCAGCGCCAGGAGCAGATTGAGAACCAATACCGCTCCCTCTTGGATCGGGCAGAAGAACGCAGACGTCGT
CTATTGCAACGTTATAATGAATTTTTATTGGCCTATGAGGCAGGAGACATGCTGGAATGGATTCAAGAGA
AAAAGGCAGAAAACACTGGAGTGGAACTAGATGATGTTTGGGAGCTGCAGAAAAAGTTTGATGAGTTCCA
AAAGGATTTGAATACCAATGAGCCTCGGCTAAGGGATATCAACAAGGTAGCTGATGATCTACTATTTGAA
GGACTTCTAACACCAGAAGGAGCTCAAATCCGGCAGGAATTGAATTCCCGCTGGGGTTCTTTGCAGAGGC
TTGCAGATGAACAGCGGCAGCTGCTGGGCAGTGCCCATGCTGTTGAAGTGTTTCACAGAGAAGCAGATGA
CACGAAGGAGCAGATTGAGAAGAAATGCCAGGCCCTCAGTGCTGCAGACCCTGGCTCAGATCTGTTCAGT
GTTCAGGCTCTTCAGCGACGGCATGAGGGCTTTGAAAGGGACCTCGTACCCCTGGGAGATAAGGTGACCA
TACTGGGGGAGACAGCAGAGCGGCTCAGTGAGTCCCATCCAGATGCCACTGAGGACCTGCAGAGACAGAA
AATGGAGCTGAATGAGGCCTGGGAAGACCTGCAGGGGCGTACAAAGGATCGTAAGGAGAGCCTAAATGAG
GCCCAGAAATTCTACCTGTTCCTCAGCAAGGCCAGGGATCTGCAGAACTGGATCAGTAGCATTGGTGGCA
TGGTATCATCACAGGAGCTGGCCGAAGACTTAACTGGCATAGAGATCTTGCTGGAGAGACATCAGGAGCA
CCGTGCTGACATGGAGGCAGAGGCTCCCACCTTCCAGGCCTTAGAGGACTTCAGTGCAGAACTTATCGAC
AGTGGGCACCATGCTAGCCCTGAAATTGAAAAAAAGCTTCAAGCTGTCAAGCTAGAGAGAGATGATTTGG
AGAAGGCTTGGGAAAAACGCAAGAAGATCCTAGACCAGTGCCTGGAGTTGCAGATGTTCCAGGGGAACTG
TGATCAAGTTGAGAGCTGGATGGTGGCACGTGAGAATTCCCTGAGGTCAGATGACAAAAGTTCCTTAGAC
AGTCTGGAGGCTTTGATGAAGAAACGGGACGATTTGGACAAAGCAATCACTGCCCAGGAAGGGAAGATCA
CTGACCTAGAACATTTTGCTGAGAGCCTCATTGCTGATGAACACTATGCCAAAGAAGAGATTGCTACGCG
GCTCCAACGTGTACTAGACAGGTGGAAGGCTCTCAAAGCACAACTGATTGATGAGCGGACAAAGCTTGGA
GACTATGCCAACCTAAAACAATTCTACCGAGACCTTGAGGAGCTGGAAGAATGGATCAGTGAGATGCTGC
CCACAGCCTGTGATGAATCCTACAAAGACGCCACTAACATTCAGAGGAAATACCTGAAACACCAGACCTT
TGCACATGAAGTCGATGGCCGATCTGAGCAGGTGCATGGCGTCATCAACCTGGGGAACTCCCTGATTGAG
TGTAGCGCTTGTGATGGCAATGAAGAGGCCATGAAGGAGCAACTGGAACAGCTGAAGGAACATTGGGATC
ATCTGCTTGAGAGAACAAATGACAAAGGGAAGAAGCTCAATGAGGCCAGTCGTCAACAGAGGTTCAACAC
AAGCATCCGGGACTTTGAGTTCTGGCTCTCAGAGGCAGAGACATTGCTGGCCATGAAAGATCAGGCCAGG
GACTTGGCTTCAGCAGGAAACCTACTCAAGAAGCATCAGCTATTGGAGAGAGAGATGTTGGCTCGAGAGG
ATGCACTCAAGGACCTGAATACATTGGCTGAAGATTTGCTCTCCAGCGGGACTTTCAACGTTGATCAGAT
TGTGAAGAAAAAAGATAATGTCAACAAGCGTTTCCTGAATGTCCAAGAATTGGCAGCTGCACACCACGAA
AAATTGAAAGAGGCCTATGCCTTGTTCCAGTTCTTCCAGGATCTAGATGATGAGGAATCCTGGATAGAGG
AGAAGTTGATACGAGTGAGCTCCCAGGACTATGGGAGAGATCTTCAGGGGGTTCAGAACTTGCTGAAGAA
GCACAAACGCCTAGAGGGGGAGCTGGTGGCCCATGAGCCTGCCATCCAGAATGTGCTGGATATGGCAGAG
AAGCTGAAAGACAAGGCTGCTGTGGGGCAAGAGGAGATCCAGTTGCGGCTGGCTCAGTTTGTTGAACACT
GGGAGAAGCTCAAAGAGTTGGCCAAGGCCCGAGGACTTAAGTTGGAAGAATCCCTAGAATACTTGCAATT
CATGCAGAATGCTGAGGAAGAGGAAGCTTGGATCAATGAAAAGAATGCTTTGGCTGTCCGAGGAGATTGT
GGAGATACATTAGCTGCTACTCAGAGCTTGCTAATGAAGCATGAAGCTTTGGAAAATGACTTTGCTGTCC
ATGAGACCCGAGTACAAAATGTGTGTGCACAAGGAGAAGACATCCTAAATAAGGTGTTGCAGGAGGAAAG
TCAGAACAAAGAGATTTCTTCCAAGATAGAGGCTCTGAATGAAAAGACCCCTTCTCTGGCTAAGGCAATA
GCTGCTTGGAAGTTGCAATTGGAAGACGATTATGCCTTTCAGGAATTCAACTGGAAGGCTGATGTGGTAG
AGGCTTGGATAGCTGATAAGGAAACAAGCCTAAAGACCAATGGCAATGGTGCAGACCTTGGTGACTTCCT
CACTCTTCTGGCAAAACAGGACACTCTGGATGCCAGTCTGCAGAGTTTCCAGCAAGAGAGACTTCCCGAG
ATCACTGACCTGAAGGACAAACTGATTTCTGCTCAACACAACCAGTCTAAAGCCATTGAAGAGCGTTATG
CCGCTCTGCTGAAGCGCTGGGAACAGTTGCTGGAAGCCTCGGCAGTCCACAGACAGAAATTGCTGGAGAA
ACAGCTGCCTCTACAGAAGGCTGAGGACCTGTTCGTGGAATTTGCACATAAGGCTTCAGCTTTGAACAAC
TGGTGTGAAAAGATGGAAGAAAACTTGTCAGAGCCTGTGCACTGTGTCTCCCTGAATGAAATTCGGCAGC
TGCAGAAAGACCATGAGGACTTCTTGGCCTCCCTGGCTAGGGCTCAAGCAGACTTTAAATGTTTGCTGGA
GCTAGACCAGCAGATTAAGGCCTTAGGTGTGCCTTCCAGCCCTTATACCTGGTTAACAGTGGAGGTGCTG
GAAAGGACCTGGAAGCACCTATCTGACATCATTGAGGAACGGGAGCAGGAGCTGCAAAAGGAAGAGGCAA
GACAGGTCAAGAACTTTGAGATGTGTCAGGAGTTTGAACAGAATGCCAGTACCTTCCTTCAATGGATCCT
GGAAACCAGGGCTTACTTTCTGGATGGATCATTGCTCAAAGAAACAGGAACTCTGGAATCTCAGCTGGAA
GCAAATAAAAGAAAACAGAAGGAGATCCAGGCGATGAAGCGTCAACTAACCAAGATTGTGGACCTGGGGG
ACAACTTGGAAGACGCTCTGATCCTTGATATCAAATACAGCACCATTGGATTGGCTCAGCAGTGGGACCA
GCTCTACCAGCTTGGGTTGCGGATGCAACACAACCTGGAGCAACAGATCCAGGCCAAGGACATCAAAGGT
GTGAGTGAAGAGACTCTAAAGGAATTTAGCACAATCTATAAACACTTTGATGAGAATTTGACAGGGCGCC
TGACTCACAAAGAGTTCCGGTCCTGCCTGAGAGGACTCAATTACTACTTGCCCATGGTGGAGGAGGATGA

ACATGAGCCCAAGTTTGAGAAGTTCCTGGATGCTGTGGATCCAGGGAGGAAGGGCTATGTCTCACTGGAG
GACTATACTGCTTTCCTGATTGACAAGGAGTCAGAAAACATCAAGTCCAGTGATGAAATAGAGAATGCCT
TCCAAGCCCTGGCAGAGGGCAAGTCATATATTACCAAAGAAGACATGAAGCAGGCCCTTACCCCAGAGCA
AGTGTCATTCTGTGCCACACATATGCAGCAATATATGGACCCACGGGGTCGAAGCCATCTCTCTGGCTAT
GACTACGTTGGCTTCACCAATTCCTACTTTGGCAACTAATAAGCAGCTCCTCGTGGATCGTAGAAAATCT
TAGTGTCGTGGGAAATTTACTGGGGGGCAAAGAGTACAGGCAAATGTGGAAGATAAAGATGGCCTCGTGT
GTGTGTGTGTGTGTGTGTGTGTGTGTGTGCTTGTGTTTGTGTGCATATTACATTTATTGTAGGATCTT
AAAAAATCTCAAGGGTGGGAGATAGAAAGGTTAATAGAGTTGGAGGAGTGGAAGCTATTTTGTATGCAAC
TAGTCACTGCTGAGGGGTGTCAAAGTTTCTATTTTTATTTGTTCTGTTTTGCACGTCTTTATCATTTTGC
TTTATTCCGATTATAGAATAAAGTAATTCTTTTTAAAAATATTTTTTGGGGCAAAGTTAAGTAAAATGTT
GAGCTTCTATATTTCTGGGAACTGTACTCATATAAGAGTGGGCAGCTAATTTTACTGTAAAGAAGGGCCA
TGGTATAGTAGATAAATAAAATCCAAGGCAATTTTCAAACAATTTTTTTAAACTTTGGAATGTGTTTAAA
TTTAAATTTGAAAATAAAGATATTTGATTTTCTGGGG (SEQ ID NO: 67)

Translated protein sequence

MEQFPKETVVESSGPKVLETAEEIQERRQEVLTRYQSFKERVAE
RGQKLEDSYHLQVFKRDADDLGKWIMEKVNILTDKSYEDPTNIQGKYQKHQSLEAEVQ
TKSRLMSELEKTREERFTMGHSAHEETKAHIEELRHLWDLLLELTLEKGDQLLRALKF
QQYVQECADILEWIGDKEAIATSVELGEDWERTEVLHKKFEDFQVELVAKEGRVVEVN
QYANECAEENHPDLPLIQSKQNEVNAAWERLRGLALQRQKALSNAANLQRFKRDVTEA
IQWIKEKEPVLTSEDYGKDLVASEGLFHSHKGLERNLAVMSDKVKELCAKAEKLTLSH
PSDAPQIQEMKEDLVSSWEHIRALATSRYEKLQATYWYHRFSSDFDELSGWMNEKTAA
INADELPTDVAGGEVLLDRHQQHKHEIDSYDDRFQSADETGQDLVNANHEASDEVREK
MEILDNNWTALLELWDERHRQYEQCLDFHLFYRDSEQVDSWMSRQEAFLENEDLGNSL
GSAEALLQKHEDFEEAFTAQEEKIITVDKTATKLIGDDHYDSENIKAIRDGLLARRDA
LREKAATRRRLLKESLLLQKLYEDSDDLKNWINKKKKLADDEDYKDIQNLKSRVQKQQ
VFEKELAVNKTQLENIQKTGQEMIEGGHYASDNVTTRLSEVASLWEELLEATKQKGTQ
LHEANQQLQFENNAEDLQRWLEDVEWQVTSEDYGKGLAEVQNRLRKHGLLESAVAARQ
DQVDILTDLAAYFEEIGHPDSKDIRARQESLVCRFEALKEPLATRKKKLLDLLHLQLI
CRDTEDEEAWIQETEPSATSTYLGKDLIASKKLLNRHRVILENIASHEPRIQEITERG
NKMVEEGHFAAEDVASRVKSLNQNMESLRARAARRQNDLEANVQFQQYLADLHEAETW
IREKEPIVDNTNYGADEEAAGALLKKHEAFLLDLNSFGDSMKALRNQANACQQQQAAP
VEGVAGEQRVMALYDFQARSPREVTMKKGDVLTLLSSINKDWWKVEAADHQGIVPAVY
VRRLAHDEFPMLPQRRREEPGNITQRQEQIENQYRSLLDRAEERRRRLLQRYNEFLLA
YEAGDMLEWIQEKKAENTGVELDDVWELQKKFDEFQKDLNTNEPRLRDINKVADDLLF
EGLLTPEGAQIRQELNSRWGSLQRLADEQRQLLGSAHAVEVFHREADDTKEQIEKKCQ
ALSAADPGSDLFSVQALQRRHEGFERDLVPLGDKVTILGETAERLSESHPDATEDLQR
QKMELNEAWEDLQGRTKDRKESLNEAQKFYLFLSKARDLQNWISSIGGMVSSQELAED
LTGIEILLERHQEHRADMEAEAPTFQALEDFSAELIDSGHHASPEIEKKLQAVKLERD
DLEKAWEKRKKILDQCLELQMFQGNCDQVESWMVARENSLRSDDKSSLDSLEALMKKR
DDLDKAITAQEGKITDLEHFAESLIADEHYAKEEIATRLQRVLDRWKALKAQLIDERT
KLGDYANLKQFYRDLEELEEWISEMLPTACDESYKDATNIQRKYLKHQTFAHEVDGRS
EQVHGVINLGNSLIECSACDGNEEAMKEQLEQLKEHWDHLLERTNDKGKKLNEASRQQ
RFNTSIRDFEFWLSEAETLLAMKDQARDLASAGNLLKKHQLLEREMLAREDALKDLNT
LAEDLLSSGTFNVDQIVKKKDNVNKRFLNVQELAAAHHEKLKEAYALFQFFQDLDDEE
SWIEEKLIRVSSQDYGRDLQGVQNLLKKHKRLEGELVAHEPAIQNVLDMAEKLKDKAA
VGQEEIQLRLAQFVEHWEKLKELAKARGLKLEESLEYLQFMQNAEEEEAWINEKNALA
VRGDCGDTLAATQSLLMKHEALENDFAVHETRVQNVCAQGEDILNKVLQEESQNKEIS
SKIEALNEKTPSLAKAIAAWKLQLEDDYAFQEFNWKADVVEAWIADKETSLKTNGNGA
DLGDFLTLLAKQDTLDASLQSFQQERLPEITDLKDKLISAQHNQSKAIEERYAALLKR
WEQLLEASAVHRQKLLEKQLPLQKAEDLFVEFAHKASALNNWCEKMEENLSEPVHCVS
LNEIRQLQKDHEDFLASLARAQADFKCLLELDQQIKALGVPSSPYIWLTVEVLERTWK
HLSDIIEEREQELQKEEARQVKNFEMCQEFEQNASTFLQWILETRAYFLDGSLLKETG
TLESQLEANKRKQKEIQAMKRQLTKIVDLGDNLEDALILDIKYSTIGLAQQWDQLYQL
GLRMQHNLEQQIQAKDIKGVSEETLKEFSTIYKHFDENLTGRLTHKEFRSCLRGLNYY
LPMVEEDEHEPKFEKFLDAVDPGRKGYVSLEDYTAFLIDKESENIKSSDEIENAFQAL
AEGKSYITKEDMKQALTPEQVSFCATHMQQYMDPRGRSHLSGYDYVGFTNSYFGN (SEQ ID NO: 68)

TMEM79	84283	NM_032323	Homo sapiens transmembrane protein 79

			(TMEM79), transcript variant 1, mRNA

mRNA Sequence

AGGTTTTGAGACACAGGTAAAGGGAGGGAGACAGAGAGAAATACTTGCAGAGCCAGCAGGTAGCTGGGCA
GCTCCTTCCCGGACGGACGGATGGACAGACGCTGGGGACCCTCCACTCCATATGGAAAGATGACATGACC
TTGTGGTAGATCCCAGAACTGAGGCCCCAGGATGACAGAACAGGAGACCCTGGCCCTACTGGAAGTGAAG
AGGTCTGATTCCCCAGAGAAGAGCTCACCCCAGGCCTTGGTTCCCAATGGCCGGCAGCCAGAAGGGGAAG
GTGGGGCCGAATCCCCGGGAGCTGAGTCCCTCAGAGTGGGGTCTTCAGCTGGATCTCCCACAGCCATAGA
GGGGGCTGAGGATGGTCTAGACAGCACAGTAAGTGAGGCTGCCACCTTGCCCTGGGGGACTGGCCCTCAG
CCCAGTGCTCCGTTCCCGGATCCCCCTGGCTGGCGGGACATTGAACCAGAGCCCCCTGAGTCAGAACCAC
TTACCAAGCTAGAGGAGCTGCCCGAAGACGATGCCAACCTGCTGCCTGAGAAAGCGGCCCGTGCCTTCGT
GCCTATTGACCTACAGTGCATTGAGCGGCAGCCCCAAGAAGACCTTATCGTGCGCTGTGAGGCAGGCGAG
GGCGAGTGCCGAACCTTCATGCCCCCCCGGGTCACCCACCCCGACCCCACTGAGCGCAAGTGGGCTGAGG
CAGTGGTGAGGCCGCCTGGCTGTTCCTGTGGGGGCTGCGGGAGCTGTGGAGACCGTGAGTGGCTAAGGGC
TGTGGCCTCCGTGGGAGCCGCACTCATTCTCTTCCCTTGCCTACTATACGGGGCATATGCCTTCCTGCCG
TTTGATGTCCCACGGCTGCCCACCATGAGTTCCCGCCTGATCTACACACTGCGCTGCGGGGTCTTTGCCA
CCTTCCCCATTGTGCTGGGGATCCTGGTGTACGGGCTGAGCCTGTTATGCTTTTCTGCCCTTCGGCCCTT
TGGGGAGCCACGGCGGGAGGTGGAGATCCACCGGCGATATGTGGCCCAGTCGGTCCAGCTCTTTATTCTC
TACTTCTTCAACCTGGCCGTGCTTTCCACTTACCTGCCCCAGGATACCCTCAAACTGCTCCCTCTGCTCA
CTGGTCTCTTTGCCGTCTCCCGGCTGATCTACTGGCTGACCTTTGCCGTGGGCCGCTCCTTCCGAGGCTT
CGGCTACGGCCTGACGTTTCTGCCACTGCTGTCGATGCTGATGTGGAACCTCTACTACATGTTCGTGGTG
GAGCCGGAGCGCATGCTCACTGCCACCGAGAGCCGCCTGGACTACCCGGACCACGCCCGCTCGGCCTCCG
ACTACAGGCCCCGCCCCTGGGGCTGAGCCTCTCCGCCCTCGCCCTCGGAGTAGGGGGTAGCGGCTTGGGT
CTGACACATCTTTGAACCTTGTGGCCAGGCCTGGACTTCGCCCCCAGGCCTAGGACCGCGGTGGGTGGAA
CCCTGCTACTGCCCCAACAGGGACTCCAATCAATCGGAGTTCTCCCCTTGCCGGAGCTGCCCTTCACCTT
TGGGGCCCGAGACAGTCATAAGGGATGGACTTAGTTTTCTTGCAGGGAAAAAGGTGGACAGCCGTGTTTC
TTAAGGATGCTGAGGGCATGGGGCCAGGACCAGGGGAGAGGCACAGCTCCTTCCTGAGCAGCCTCTCACC
ACTGCCACAAGGCTCCCTAATGCTGGTCTCTGCTCCACTCCCCGGCTTCCCGTGAGGCAGGAGGCAGAGC
CACAGCCAAGGCCCTGACCACTTCTGTGCCAGTTGTCTAAGCAGAGCGCCTCAGGGACGCTGGAAATGCC
TTAAGGATAGAGGCTGGGCATCACATCAAATGGGACTGTGGTGTTTGGTGAAAACCTTCCTGAGGATCTG
GATTCAGGACCCTCCATGACTGGCCTATTTACTGTTTACAGCTGGCCAGTGCAGAGCTGCTGCTCTTTTA
CCTTTTTAGGCCCCTGTAACTTCCCACCTTTAAACTGCCCAGAAGGCATGCCTCTCCCACAGGAAGAGGG
GAGCAGACAGGGAAATCTGCCTACCAAGAGGGGTGTGTGTGTCTTTGTGCCCACACGTGGTGGCTGGGGA
GTGCCTGGATGGTGCGGTGGTTGATGTTAACCTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
GTGTGTGTGTAACAATAAATTACTACCAGTCAAAAAAAAAAAAAA (SEQ ID NO: 69)

Translated protein sequence

MTEQETLALLEVKRSDSPEKSSPQALVPNGRQPEGEGGAESPGA
ESLRVGSSAGSPTAIEGAEDGLDSTVSEAATLPWGTGPQPSAPFPDPPGWRDIEPEPP
ESEPLTKLEELPEDDANLLPEKAARAFVPIDLQCIERQPQEDLIVRCEAGEGECRTFM
PPRVTHPDPTERKWAEAVVRPPGCSCGGCGSCGDREWLRAVASVGAALILFPCLLYGA
YAFLPFDVPRLPTMSSRLIYTLRCGVFATFPIVLGILVYGLSLLCFSALRPFGEPRRE
VEIHRRYVAQSVQLFILYFFNLAVLSTYLPQDTLKLLPLLTGLFAVSRLIYWLTFAVG
RSFRGFGYGLTFLPLLSMLMWNLYYMFVVEPERMLTATESRLDYPDHARSASDYRPRP
WG (SEQ ID NO: 70)

TMIGD2	126259	NM_144615	Homo sapiens transmembrane and immunoglobulin domain containing 2 (TMIGD2), mRNA

mRNA Sequence

GGAAGTCTGTCAACTGGGAGGGGGAGAGGGGGGTGATGGGCCAGGAATGGGGTCCCCGGGCATGGTGCTG
GGCCTCCTGGTGCAGATCTGGGCCCTGCAAGAAGCCTCAAGCCTGAGCGTGCAGCAGGGGCCCAACTTGC
TGCAGGTGAGGCAGGGCAGTCAGGCGACCCTGGTCTGCCAGGTGGACCAGGCCACAGCCTGGGAACGGCT
CCGTGTTAAGTGGACAAAGGATGGGGCCATCCTGTGTCAACCGTACATCACCAACGGCAGCCTCAGCCTG
GGGGTCTGCGGGCCCCAGGGACGGCTCTCCTGGCAGGCACCCAGCCATCTCACCCTGCAGCTGGACCCTG
TGAGCCTCAACCACAGCGGGGCGTACGTGTGCTGGGCGGCCGTAGAGATTCCTGAGTTGGAGGAGGCTGA
GGGCAACATAACAAGGCTCTTTGTGGACCCAGATGACCCCACACAGAACAGAAACCGGATCGCAAGCTTC
CCAGGATTCCTCTTCGTGCTGCTGGGGGTGGGAAGCATGGGTGTGGCTGCGATCGTGTGGGGTGCCTGGT
TCTGGGGCCGCCGCAGCTGCCAGCAAAGGGACTCAGGTAACAGCCCAGGAAATGCATTCTACAGCAACGT

CCTATACCGGCCCCGGGGGGCCCCAAAGAAGAGTGAGGACTGCTCTGGAGAGGGGAAGGACCAGAGGGGC
CAGAGCATTTATTCAACCTCCTTCCCGCAACCGGCCCCCCGCCAGCCGCACCTGGCGTCAAGACCCTGCC
CCAGCCCGAGACCCTGCCCCAGCCCCAGGCCCGGCCACCCCGTCTCTATGGTCAGGGTCTCTCCTAGACC
AAGCCCCACCCAGCAGCCGAGGCCAAAAGGGTTCCCCAAAGTGGGAGAGGAGTGAGAGATCCCAGGAGAC
CTCAACAGGACCCCACCCATAGGTACACACAAAAAAGGGGGGATCGAGGCCAGACACGGTGGCTCACGCC
TGTAATCCCAGCAGTTTGGGAAGCCGAGGCGGGTGGAACACTTGAGGTCAGGGGTTTGAGACCAGCCTGG
CTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCCGAGATTGCGCCACTGCACTCCAGCCTGGGCGACAGAG
TGAGACTCCGTCTCAAAAAAAACAAAAAGCAGGAGGATTGGGAGCCTGTCAGCCCCATCCTGAGACCCCG
TCCTCATTTCTGTAATGATGGATCTCGCTCCCACTTTCCCCCAAGAACCTAATAAAGGCTTGTGAAGAAA
AAGCAAAAAAAAAAAAAAAAAA (SEQ ID NO: 71)

Translated protein sequence

MGSPGMVLGLLVQIWALQEASSLSVQQGPNLLQVRQGSQATLVC
QVDQATAWERLRVKWTKDGAILCQPYITNGSLSLGVCGPQGRLSWQAPSHLTLQLDPV
SLNHSGAYVCWAAVEIPELEEAEGNITRLFVDPDDPTQNRNRIASFPGFLFVLLGVGS
MGVAAIVWGAWFWGRRSCQQRDSGNSPGNAFYSNVLYRPRGAPKKSEDCSGEGKDQRG
QSIYSTSFPQPAPRQPHLASRPCPSPRPCPSPRPGHPVSMVRVSPRPSPTQQPRPKGF
PKVGEE (SEQ ID NO: 72)

TUBB6	84617	NM_032525	Homo sapiens tubulin, beta 6 (TUBB6), mRNA

mRNA Sequence

GGGCACGAGGGCAGAGCCAGTTCCTAGCGCAGAGCCGCGCCCGCCATGAGGGAGATCGTGCACATCCAGG
CGGGCCAGTGCGGGAACCAGATCGGCACCAAGTTTTGGGAAGTGATCAGCGATGAGCACGGCATCGACCC
GGCCGGAGGCTACGTGGGAGACTCGGCGCTGCAGCTGGAGAGAATCAACGTCTACTACAATGAGTCATCG
TCTCAGAAATATGTGCCCAGGGCCGCCCTGGTGGACTTAGAGCCAGGCACCATGGACAGCGTGCGGTCTG
GGCCTTTTGGGCAGCTTTTCCGGCCTGACAACTTCATCTTTGGCCAGACGGGTGCAGGGAACAACTGGGC
GAAAGGGCACTACACGGAGGGCGCGGAGCTGGTGGACGCAGTGCTGGACGTGGTGCGGAAGGAGTGCGAG
CACTGCGACTGCCTGCAGGGCTTCCAGCTCACGCACTCGCTGGGCGGCGGCACGGGCTCAGGCATGGGCA
CGCTGCTCATCAGCAAGATCCGTGAGGAGTTCCCGGACCGCATCATGAACACCTTCAGCGTCATGCCCTC
GCCCAAGGTGTCGGACACGGTGGTGGAGCCCTACAATGCCACACTGTCGGTGCACCAGCTGGTGGAGAAT
ACAGACGAGACCTACTGCATCGACAACGAGGCGCTCTATGACATCTGCTTCCGCACTCTGAAGCTGACAA
CGCCCACCTACGGGGACCTCAACCACCTGGTGTCCGCCACCATGAGTGGGGTCACCACCTCGCTGCGCTT
CCCGGGCCAGCTCAATGCTGACCTGCGCAAGCTGGCGGTGAACATGGTGCCCTTCCCGCGCCTGCACTTC
TTCATGCCTGGCTTCGCGCCGCTCACCAGCCGCGGCAGCCAGCAGTACCGGGCCCTGACCGTGCCCGAGC
TCACCCAGCAGATGTTCGACGCCAGGAACATGATGGCCGCCTGCGATCCGCGCCATGGCCGCTACCTGAC
CGTGGCCACCGTGTTCCGCGGGCCCATGTCCATGAAGGAGGTGGACGAGCAGATGCTGGCCATCCAGAGT
AAGAACAGCAGCTACTTCGTGGAGTGGATTCCCAACAACGTGAAGGTGGCCGTGTGCGACATCCCGCCCC
GCGGCCTGAAGATGGCCTCCACCTTCATCGGCAACAGCACGGCCATCCAGGAGCTGTTCAAGCGCATCTC
CGAGCAGTTCTCAGCCATGTTCCGGCGCAAGGCCTTCCTGCACTGGTTCACGGGTGAGGGCATGGATGAA
ATGGAGTTCACCGAGGCGGAGAGCAACATGAACGACCTGGTATCCGAGTACCAGCAGTACCAGGATGCCA
CCGCCAATGACGGGGAGGAAGCTTTTGAGGATGAGGAAGAGGAGATCGATGGATAGTCGGAATAGAGCCG
CCCCAACTCAGATCCTACAACACGCAAGTTCCTTCTTGAACCCTGGTGCCTCCTACCCTATGGCCCTGAA
TGGTGCACTGGTTTAATTGTGTTGGTGTCGGCCCCTCACAAATGCAGCCAAGTCATGTAATTAGTCATCT
GGAACAAAGACTAAAAACAGCAGAGAATTGCGGGTTCTACCCAGTCAGAAGATCACACCATGGAGACTTT
CTACTAGAGGACTTGAAAGAGAACTGAGGGGCCACAAAATAAACTTCACCTTCCATTAAGTGTTCAAGCA
TGTCTGCAAATTAGGAGGGAGTTAGAAACAGTCTTTTTCATCCTTTGTGATGAAGCCTGAAATTGTGCCG
TGTTGCCTTATATGAATATGCAGTATGGGACTTTGAAATAATGATTCATAATAAAATACTAAACGTGTGT
CTTCAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 73)

Translated protein sequence

MREIVHIQAGQCGNQIGTKFWEVISDEHGIDPAGGYVGDSALQL
ERINVYYNESSSQKYVPRAALVDLEPGTMDSVRSGPFGQLFRPDNFIFGQTGAGNNWA
KGHYTEGAELVDAVLDVVRKECEHCDCLQGFQLTHSLGGGTGSGMGTLLISKIREEFP
DRIMNTFSVMPSPKVSDTVVEPYNATLSVHQLVENTDETYCIDNEALYDICFRTLKLT
TPTYGDLNHLVSATMSGVTTSLRFPGQLNADLRKLAVNMVPFPRLHFFMPGFAPLTSR
GSQQYRALTVPELTQQMFDARNMMAACDPRHGRYLTVATVFRGPMSMKEVDEQMLAIQ
SKNSSYFVEWIPNNVKVAVCDIPPRGLKMASTFIGNSTAIQELFKRISEQFSAMFRRK
AFLHWFTGEGMDEMEFTEAESNMNDLVSEYQQYQDATANDGEEAFEDEEEEIDG (SEQ ID NO: 74)

TYROBP	7305	NM_198125	Homo sapiens TYRO protein tyrosine kinase

			binding protein (TYROBP), transcript variant 2, mRNA

mRNA Sequence

AGACTTCCTCCTTCACTTGCCTGGACGCTGCGCCACATCCCACCGGCCCTTACACTGTGGTGTCCAGCAG
CATCCGGCTTCATGGGGGGACTTGAACCCTGCAGCAGGCTCCTGCTCCTGCCTCTCCTGCTGGCTGTAAG
TGGTCTCCGTCCTGTCCAGGCCCAGGCCCAGAGCGATTGCAGTTGCTCTACGGTGAGCCCGGGCGTGCTG
GCAGGGATCGTGATGGGAGACCTGGTGCTGACAGTGCTCATTGCCCTGGCCGTGTACTTCCTGGGCCGGC
TGGTCCCTCGGGGGCGAGGGGCTGCGGAGGCGACCCGGAAACAGCGTATCACTGAGACCGAGTCGCCTTA
TCAGGAGCTCCAGGGTCAGAGGTCGGATGTCTACAGCGACCTCAACACACAGAGGCCGTATTACAAATGA
GCCCGAATCATGACAGTCAGCAACATGATACCTGGATCCAGCCATTCCTGAAGCCCACCCTGCACCTCAT
TCCAACTCCTACCGCGATACAGACCCACAGAGTGCCATCCCTGAGAGACCAGACCGCTCCCCAATACTCT
CCTAAAATAAACATGAAGCACAAAAACAAAAAAAAAAAAAAAAAA (SEQ ID NO: 75)

Translated protein sequence

MGGLEPCSRLLLLPLLLAVSGLRPVQAQAQSDCSCSTVSPGVLA
GIVMGDLVLTVLIALAVYFLGRLVPRGRGAAEATRKQRITETESPYQELQGQRSDVYS
DLNTQRPYYK (SEQ ID NO: 76)

YTHDF1	54915	NM_017798	Homo sapiens YTH domain family, member 1 (YTHDF1), mRNA

mRNA Sequence

GCGTGCACGCTGACGCCGCGCAGTCTCGTCCCCTGCCGCCGCCGTCGCCGCTGCTGTCGCCGCCGCCGCC
GCCATTGGAGTCGACGCCTCCTCAGTGCGTCCGCGTCCCGGGCTCACCGCCGCTGCCGCCTCGCCAGGGG
CCCGCGCGCCCAGCAGCCGCCGCCGCCGCCCGGCCGGCGCCCGGGGAATTGGCGGCGGGGCCCGGGGCCG
CGCGAGCTAGGGTGACAGGCCCGGCCTCTAGGGGAGGCCCGAGCCGGCGGGCGCCCCGGCCCCGCGTCTA
GTTGTTCATGAAGCATGTCGGCCACCAGCGTGGACACCCAGAGAACAAAAGGACAAGATAATAAAGTACA
AAATGGTTCGTTACATCAGAAGGATACAGTTCATGACAATGACTTTGAGCCCTACCTTACTGGACAGTCA
AATCAGAGTAACAGTTACCCCTCAATGAGCGACCCCTACCTGTCCAGCTATTACCCGCCGTCCATTGGAT
TTCCTTACTCCCTCAATGAGGCTCCGTGGTCTACTGCAGGGGACCCTCCGATTCCATACCTCACCACCTA
CGGACAGCTCAGTAACGGAGACCATCATTTTATGCACGATGCTGTTTTTGGGCAGCCTGGGGGCCTGGGG
AACAACATCTATCAGCACAGGTTCAATTTTTTCCCTGAAAACCCTGCGTTCTCAGCATGGGGGACAAGTG
GGTCTCAAGGTCAGCAGACCCAGAGCTCCGCGTATGGGAGCAGCTACACCTACCCCCCGAGCTCCCTGGG
TGGCACGGTGGTTGATGGGCAGCCAGGCTTTCACAGCGACACCCTCAGCAAGGCCCCCGGGATGAACAGC
CTGGAGCAGGGCATGGTTGGCCTGAAGATTGGGGACGTCAGCTCCTCCGCCGTCAAGACGGTGGGCTCTG
TCGTCAGCAGCGTGGCACTGACTGGTGTCCTTTCTGGCAACGGTGGGACAAATGTGAACATGCCAGTTTC
AAAGCCGACCTCGTGGGCTGCCATTGCCAGCAAGCCTGCAAAACCACAGCCTAAAATGAAAACAAAGAGC
GGGCCTGTCATGGGGGGTGGGCTGCCCCCTCCACCCATAAAGCATAACATGGACATTGGCACCTGGGATA
ACAAGGGGCCTGTGCCGAAGGCCCCAGTCCCCCAGCAGGCACCCTCTCCACAGGCTGCCCCACAGCCCCA
GCAGGTGGCTCAGCCTCTCCCAGCACAGCCCCCAGCTTTGGCTCAACCGCAGTATCAGAGCCCTCAGCAG
CCACCCCAGACCCGCTGGGTTGCCCCACGCAACAGAAACGCGGCGTTTGGGCAGAGCGGAGGGGCTGGCA
GCGATAGCAACTCTCCTGGAAACGTCCAGCCTAATTCTGCCCCCAGCGTCGAATCCCACCCCGTCCTTGA
AAAACTGAAGGCTGCTCACAGCTACAACCCGAAAGAGTTTGAGTGGAATCTGAAAAGCGGGCGTGTGTTC
ATCATCAAGAGCTACTCTGAGGACGACATCCACCGCTCCATTAAGTACTCCATCTGGTGTAGCACAGAGC
ACGGCAACAAGCGCCTGGACAGCGCCTTCCGCTGCATGAGCAGCAAGGGGCCCGTCTACCTGCTCTTCAG
CGTCAATGGGAGTGGGCATTTTTGTGGGGTGGCCGAGATGAAGTCCCCCGTGGACTACGGCACCAGTGCC
GGGGTCTGGTCTCAGGACAAGTGGAAGGGGAAGTTTGATGTCCAGTGGATTTTTGTTAAGGATGTACCCA
ATAACCAGCTCCGGCACATCAGGCTGGAGAATAACGACAACAAACCGGTCACAAACTCCCGGGACACCCA
GGAGGTGCCCTTAGAAAAAGCCAAGCAAGTGCTGAAAATTATCAGTTCCTACAAGCACACAACCTCCATC
TTCGACGACTTTGCTCACTACGAGAAGCGCCAGGAGGAGGAGGAGGTGGTGCGCAAGGAACGGCAGAGTC
GAAACAAACAATGAGGGCGAACCAGTTTCTTACATGTTCTAACGTTTGACTTTGAAAACAGTTTAAAACA
CGTGTGCTTGGTCAGCTCCAGTGTGTCGTCCCGTGCGGGGGTTGAGTGTTGCATCTTTGCCTTTCTTGTC
GTTGATTTTTGCCCAGATGGATCTGCATTTATTTGTACTTTTTCTATGTATTATAATCCTGTAGAAGTCA
CTAATAAAGGAGTATTTTTTTTGTCAGCTTATCAATCAGACTGATCTAATGTGAAATGTAAGTATCCTTA
AAAACAAAGCATCTATTTTGGCAGAAATTGTGTTCTTAAATTCAGTCATTTGATATTCTGTGAGACTTCA
TATTTCTCATCCCTTTATTGCTTTTTAGCAAACATAAGAAACCATGAGTCATTTTGTCATTTAGAGTATT
CTGATAAAATCTCTTGAAAATACTGAAATCAAAAGGTTAATGATTTTTTGTTCATTCTGATTTGTCATTT
TATTATCTGTTATCGGTCTAAAGTGCTAATTTACCCATTTGATTTTTCTGCTAGACAGATAACTTTTAAT

TTTTCAAATTTGGCAGACACTTTTTTTTTTTTTTTGAAAATCTTTCCTTCCAGATCTGTTGCCCACTGAA
CAGCCACCCGTCCCTCACTGTCCTGGTGTCCGATTGGGCTGGATGGTGTTGGGGCATGATGTGTGGAGGA
ACTGGAAGGTGCTTTAGGTCTGGTTCAGGGTCGGGCATTCTTTGTTGTTTGCACATCTTTTTAAATTTTA
CACCTTTTCTTAAGAATTCTAATGCCGTCTTAAGTTTTTATACCAATAATGCTGAGCTTTAAGTGTAGGA
TCTGGTAGTACAGACAGTGTGATGGATGATGCTGCTGGTTGTAAATTTCATCGTGTGTGTCTAATTTTTT
TTCCTGTTGAATGGGTAAAAACAAAACAAAACTTTTTTTAGAAGATGAATTTGCTGTCATGTTTTGTGGA
ATGAGGGACCGTTGAGCTCACTACCACCTGGAGTTTGAGTTGAAGCATGAAAATGGTGCCCATGCCTGAC
GCTCCAGCGCCTGGATCTGCACGTGCCCTTGTAGAGGATCCTTACCGTCCTAGAGAGCAGACGCTTTCTG
AAAACTACTTGCTCCAAAAGACCCTCTGAGTTAACGTTTCAGCTGTATCATTAGACTTGTATTTAGAGCG
TGTCACTTCCTCTGAACTGTTACTGCCTGAATGGAGTCCTGGACGACATTGGGTTTTTCCTCTAGGAGAA
TACAAGCCTTAATAAACAATACTATTTAGCAAAAAAAAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 77)

Translated protein sequence

MSATSVDTQRTKGQDNKVQNGSLHQKDTVHDNDFEPYLTGQSNQ
SNSYPSMSDPYLSSYYPPSIGFPYSLNEAPWSTAGDPPIPYLTTYGQLSNGDHHFMHD
AVFGQPGGLGNNIYQHRFNFFPENPAFSAWGTSGSQGQQTQSSAYGSSYTYPPSSLGG
TVVDGQPGFHSDTLSKAPGMNSLEQGMVGLKIGDVSSSAVKTVGSVVSSVALTGVLSG
NGGTNVNMPVSKPTSWAAIASKPAKPQPKMKTKSGPVMGGGLPPPPIKHNMDIGTWDN
KGPVPKAPVPQQAPSPQAAPQPQQVAQPLPAQPPALAQPQYQSPQQPPQTRWVAPRNR
NAAFGQSGGAGSDSNSPGNVQPNSAPSVESHPVLEKLKAAHSYNPKEFEWNLKSGRVF
IIKSYSEDDIHRSIKYSIWCSTEHGNKRLDSAFRCMSSKGPVYLLFSVNGSGHFCGVA
EMKSPVDYGTSAGVWSQDKWKGKFDVQWIFVKDVPNNQLRHIRLENNDNKPVTNSRDT
QEVPLEKAKQVLKIISSYKHTTSIFDDFAHYEKRQEEEEVVRKERQSRNKQ (SEQ ID NO: 78)

ZIC5	85416	NM_033132	Homo sapiens Zic family member 5 (odd-paired homolog, Drosophila) (ZIC5), mRNA

mRNA Sequence

GCGGCCGCAAGCACGGGGGCGAATCCCCGCTGGGTCGAGGGCCTGAACGGGAGCCAATCGAGCAGCCGAG
GCTACTGCCAATCACGCGGCTCCCTCCAATCCCACCCGTGCCATTTCCAAAATCTCGGTCCCACTGTGCA
GCTCAAATGTGGTGTTCACTCTGCCAATCGCTGGAGGATAGAGTGGGAACAGGAATAAGCAGAGTTAAGA
GGCCAGGACAAAAGAAGTTAAAGAGCGCCCAATACATACATGTTTTTGAAGGCGGGCAGAGGGAATAAAG
TCCCCCCAGTGAGGGTCTATGGGCCTGATTGTGTAGTTCTGATGGAGCCCCCTTTGAGCAAGAGGAACCC
GCCAGCGCTGAGATTAGCGGATTTGGCAACGGCTCAGGTCCAGCCGCTTCAGAATATGACAGGCTTCCCG
GCGCTGGCCGGCCCGCCCGCCCACTCCCAACTCCGGGCCGCCGTCGCGCACCTCCGCCTGCGGGACCTGG
GCGCTGACCCCGGCGTGGCCACCACTCCGCTCGGACCCGAGCACATGGCCCAGGCGAGCACGCTGGGCCT
CAGCCCTCCCTCCCAGGCGTTCCCGGCACACCCGGAGGCTCCGGCAGCCGCCGCCCGTGCTGCAGCCTTG
GTCGCGCACCCCGGCGCGGGCAGCTACCCCTGCGGCGGGGGCAGCAGTGGCGCGCAGCCCTCCGCGCCCC
CGCCCCCAGCCCCTCCTCTTCCTCCCACCCCTTCACCCCCTCCCCCTCCCCCGCCTCCTCCTCCTCCTGC
CCTCTCGGGCTACACCACCACCAACAGTGGCGGCGGCGGCAGCAGCGGCAAAGGCCACAGCAGGGACTTC
GTCCTCCGGAGGGACCTTTCCGCCACGGCCCCCGCGGCGGCCATGCACGGGGCCCCGCTCGGAGGGGAGC
AGCGGTCCGGCACCGGCTCCCCCCAGCACCCGGCCCCGCCTCCCCACTCGGCCGGCATGTTCATCTCCGC
CAGCGGCACCTACGCGGGCCCGGACGGCAGCGGCGGCCCCGGCGCTCTTCCCCGCGCTGCACGACACGCCG
GGGGCCCCAGGCGGCCACCCGCACCCGCTCAACGGCCAGATGCGCCTGGGGCTGGCGGCGGCAGCGGCAG
CCGCGGCGGCTGAGCTGTACGGCCGCGCCGAACCGCCCTTCGCGCCGCGCTCTGGGGACGCGCACTACGG
GGCGGTTGCGGCCGCAGCGGCGGCCGCCCTGCACGGCTACGGAGCCGTGAACTTAAACCTGAACCTGGCG
GCTGCGGCGGCCGCAGCAGCGGCCGGGCCCGGGCCCCACCTGCAGCACCACGCGCCGCCCCCGGCGCCGC
CGCCGCCGCCGGCGCCCGCGCAGCACCCGCACCAGCACCACCCCCACCTCCCAGGGGCGGCTGGGGCCTT
CCTGCGCTACATGCGGCAGCCAATCAAGCAGGAGCTCATCTGCAAGTGGATCGACCCCGACGAGCTGGCC
GGGCTGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCACCGCCCCCGGCCGGCGGCGCCAAGCCCT
GCTCCAAAACTTTCGGCACCATGCACGAGCTGGTGAATCACGTCACGGTGGAGCACGTGGGAGGCCCCGA
GCAGAGCAGCCACGTCTGCTTCTGGGAGGACTGTCCGCGCGAGGGCAAGCCCTTCAAGGCCAAATACAAG
CTCATCAACCACATCCGCGTGCACACCGGCGAGAAGCCCTTTCCCTGCCCTTTCCCCGGCTGCGGCAAGG
TCTTCGCGCGCTCCGAGAACCTCAAGATCCACAAGCGTACTCATACAGGGGAAAAGCCTTTCAAATGTGA
ATTTGATGGCTGTGACAGGAAGTTTGCCAATAGCAGTGATCGGAAGAAACATTCCCATGTCCACACCAGT
GACAAGCCCTACTACTGCAAGATTCGAGGCTGTGACAAATCCTACACTCACCCAAGCTCCCTGAGGAAGC
ACATGAAGATTCACTGCAAGTCCCCGCCACCTTCTCCAGGACCCCTTGGTTACTCATCAGTGGGGACTCC
AGTGGGCGCCCCCTTGTCCCCTGTGCTGGACCCAGCCAGGAGTCACTCCAGCACTCTGTCCCCTCAGGTG
ACCAACCTCAATGAGTGGTACGTTTGCCAGGCCAGTGGGGCCCCCAGCCACCTCCACACCCCTTCCAGCA
ACGGAACCACCTCTGAGACTGAAGATGAGGAAATTTACGGGAACCCTGAAGTTGTGCGGACGATACATTA

GAATTTATTATTAATAATAATAAGTGAAATAATAAGTGGGAGTCCTTGGACCACATCCTAACCTGAGACA
ATGCCGAGCCTGAGACAAACCCGTGACTCAGACTTGCCACCGGGTCTAATTAGCCCTATTTATTCAGTAT
GAAACCCTATGGTGTTTGTACATTTAATTAATTTAATTAAGATATTTGGGCTTTTTTTTTTTTTTTCTT
AAAAAACAAACAAAAAACAACCAAGCTGGACTTGTACATTGCAGGAGGATGGGGCTGGGGGCAAATTGTA
CCAAGGAAAATGAATGGAGAGATTAGTTAATGGCGATACACACTGCCGATGCAATATATATATATATATA
TATATACATATATATATATATTATTTTTTTTAAAAGGGGGAGAAAAAGAGCATTAAGTCAGAACTTAACA
CAGCACCAAGGCCCTCTGCATTTCCCAGAGTGCCTCTCAAATGCCTTTGACACCATACCATGGGCTGCTT
TTGAGCCTCCTTGTTGGACCCTAATTCTGCCAAGGCCTCTTGATTGTAAACCACACACCTGCTGCATTGC
CAACAGATCCTGTTCCGTACCTGTGTCCAAAAACATTTGTAAAAACCCTTTGAGTTTAATATTTGTAATT
TTTAATTTCCACTCTTTTATTACTGATCTTAGCTTAATACAATATTTTTATACAGGATTATTTCTTCAGT
ATCCTACTGTGTGATTTTAAAAAAAGATGCAGCAACCTTAATATATCTCCATATCTTGTGCTACTGTGAT
TGTTCAAGCAAAAGTGGAGAGAAGAAAAGCTGCTGCAAAAGACAACTGTGAAACTGTGATATTTTATAAA
ATAGAAGAAATTCAAGTGCTTTCTTTTTCCTATATGTTTTTTTTTTTTTATCTGAATTCTCAGATACTGC
CTCCTAACTGTGTCCAAACTTCTTGTGTAATAAAGAGATTCTGTTTTCGATCCTAAGTTCTTTGGGATGC
CAACATTCACAGTCAAGTCTTGAGGAGGTGTGATGATGGCATCATGCCTATTTTTTTGGAAAGCTGTTGT
TTTTAAAACAGGCCAACACCTCTTTTATACTGTTGTATCAGCCTTTTAAAAAGTCTATTTTTCAATGCCT
GAAACTGCATTTTAATGCATTTTCTTCCACCTGAGCACTGAGCACACCAAACTGGAATCCATTTGAAAAT
GACAGTGTGTGAAGTGTATGATTTACATTAAAAGAGGGGAGGGAGTTGCCATACATATTAAAAATTTTTA
AAAGGTTTATAGTTACCACCAAACACTGATGAATGTGTGACCTTTGCCAGAGCTGTCAAGCTAGGATAAA
AAAGGTCAAGGACCTAGGACAATAACTCTTAGTCGATTTATTTTCGGTTGGTACAACACATCTCCTGTGC
AAAATGTAGTCCATCAGAAACATCCTACAGATACACTAAAGAGCACTAATTTATCCTTAGAGACCCCGAA
GACACCCCCTCCCCAGGGTTTGTAGAAATTTGTTTTGTGTGCTGTGAGTGGTTGATGTAGTCTTGTCATT
GTTAATAACTTGTATGTGAACACTATTATTTGTACAGTTGAATTAATTTATTTTCAGACATCATCCTTTT
TTTTTTCTTTCCTGGAAGAGTTCAAAGCACACCAAAGAATTATATTATACATTTTGGTGAAAGATTGTC
ATTTATGATCCATGGTTTATTTAAAAAAAAAAGGAAAGAAAATGGAAAAATATATTTTTAAGCTTACTTG
AATGAACAACGTAATGTGAAAACCAAGACTCTTCCTGCATGTCTTTTTTGCATTGTGTTGATAAGATTAT
ATATAGTTTATAGATATATTATATTACTAGTACAGTGCATGGTGCTGTCACTTGGAAAGCCTTTCAATGT
TGTCTTCAGATTGTTGTGATGAATATGAAACATGCAGACCCTCCTTTATAAAGAAAAAGACCTTAAAACT
TGAATATGAGATAATTTTACATTTTAAAAGTTTATTTGATTTTCATATTATTCACTTTCAAAGCCCTTTC
AAATAGAAAAGGTATGAACTTTTGGGGGGGATAATTTATGTATCGTAAACTTATTAGAACAAAATATTCCT
GATGTATAATGAGTTGTTTTATTTATACAACTTTTTCAATGGTAGTTTGCACTATTCTTTATTATGCTAC
AGGTTTATTTATTATGAAACAAAGGAATATGTATTTTATGTATTTTACCATGCATAGGTTAACTCTTTGC
CACAGATTTATTGGTTCTTGATACACCTAAAATAAAAAAAAAATGTGTACCTCCAATAGAGAGCAAGCAAG
AATGATTATGAAGTAACAAATTTAATAAAGGTATTCTTGTTATTATTAAAAAAAAAA (SEQ ID NO: 79)

Translated protein sequence

MFLKAGRGNKVPPVRVYGPDCVVLMEPPLSKRNPPALRLADLAT
AQVQPLQNMTGFPALAGPPAHSQLRAAVAHLRLRDLGADPGVATTPLGPEHMAQASTL
GLSPPSQAFPAHPEAPAAAARAAALVAHPGAGSYPCGGGSSGAQPSAPPPPAPPLPPT
PSPPPPPPPPPPALSGYTTTNSGGGGSSGKGHSRDFVLRRDLSATAPAAAMHGAPLG
GEQRSGTGSPQHPAPPPHSAGMFISASGTYAGPDGSGGPALFPALHDTPGAPGGHPHP
LNGQMRLGLAAAAAAAAAELYGRAEPPFAPRSGDAHYGAVAAAAAAALHGYGAVNLNL
NLAAAAAAAAAGPGPHLQHHAPPPAPPPPPAPAQHPHQHHPHLPGAAGAFLRYMRQPI
KQELICKWIDPDELAGLPPPPPPPPPPPPPPAGGAKPCSKTFGTMHELVNHVTVEHV
GGPEQSSHVCFWEDCPREGKPFKAKYKLINHIRVHTGEKPFPCPFPGCGKVFARSENL
KIHKRTHTGEKPFKCEFDGCDRKFANSSDRKKHSHVHTSDKPYYCKIRGCDKSYTHPS
SLRKHMKIHCKSPPPSPGPLGYSSVGTPVGAPLSPVLDPARSHSSTLSPQVTNLNEWY
VCQASGAPSHLHTPSSNGTTSETEDEEIYGNPEVVRTIH (SEQ ID NO: 80)

CLAIMS

What is claimed is:

1. A method for determining whether eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof, may be used to treat a subject having breast cancer, the method comprising

determining the level of expression of a biomarker selected from the group of biomarkers listed in Table 1 in a sample derived from said subject, wherein a low level of expression of the biomarker is indicative that eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof, will be effective in treating said subject.

2. A method for determining whether eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof, may be used to treat a subject having breast cancer, the method comprising

assaying a sample derived from said subject to determine the level of expression in said sample of a biomarker selected from the group of biomarkers listed in Table 1, wherein a low level of expression of the biomarker is indicative that eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof, will be effective in treating said subject.

3. A method for predicting whether eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof, may be used to treat a subject having breast cancer, the method comprising

determining the level of expression of a biomarker selected from the group of biomarkers listed in Table 1 in a sample derived from said subject, and

predicting that eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof, will be effective in treating a subject having breast cancer when said sample is determined to have a low level of expression of said biomarker.

4. A method for determining the sensitivity of a breast tumor to treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof, the method comprising

determining the level of expression of a biomarker selected from the group of biomarkers listed in Table 1 in said tumor, wherein a low level of expression of the biomarker in said tumor indicates that said tumor is sensitive to treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof.

5. A method for determining the sensitivity of a breast tumor to treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof, the method comprising

determining the level of expression of a biomarker selected from the group of biomarkers listed in Table 1 in said tumor, and

identifying said tumor as being sensitive to treatment with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof, when there is determined to be a low level of expression of said biomarker in said tumor.

6. The method of claim 4 or 5, wherein a sample of the breast tumor is obtained from a subject having breast cancer.

7. A method of treating a subject having breast cancer, the method comprising

identifying a subject having breast cancer in which a biomarker selected from the group of biomarkers listed in Table 1 has a low level of expression, and

administering a therapeutically effective amount of eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof, to said subject.

8. A method of treating a subject having breast cancer, the method comprising

assaying a sample derived from said subject to determine the level of expression in said sample of a biomarker selected from the group of biomarkers listed in Table 1, and

administering a therapeutically effective amount of eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof, to said subject when a low level of expression of the biomarker is detected in said sample.

9. The method of any one of claims 1-8, wherein the pharmaceutically acceptable salt of eribulin is eribulin mesylate.

10. The method of any one of claims 1-3 or 6-8, wherein said subject has not been previously treated with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof.

11. The method of any one of claims 1-3 or 6-8, wherein said subject has been previously treated with eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof.

12. The method of any one of claims 1-3 or 6-8, wherein said breast cancer is an Estrogen Receptor (ER) negative breast cancer.

13. The method of any one of claims 1-3 or 6-8, wherein said breast cancer is a Progesterone Receptor (PR) negative breast cancer.

14. The method of any one of claims 1-3 or 6-8, wherein said breast cancer is a HER-2 negative breast cancer.

15. The method of any one of claims 1-3 or 6-8, wherein said breast cancer is an Estrogen Receptor (ER) negative and Progesterone Receptor (PR) negative breast cancer.

16. The method of any one of claims 1-3 or 6-8, wherein said breast cancer is an Estrogen Receptor (ER) negative and HER-2 negative breast cancer.

17. The method of any one of claims 1-3 or 6-8, wherein said breast cancer is a Progesterone Receptor (PR) negative and HER-2 negative breast cancer.

18. The method of any one of claims 1-3 or 6-8, wherein said breast cancer is an Estrogen Receptor (ER) negative, Progesterone Receptor (PR) negative and HER-2 negative breast cancer.

19. The method of any one of claims 1-8, wherein at least 2 biomarkers selected from the group of biomarkers listed in Table 1 have a low level of expression

20. The method of any one of claims 1-8, wherein the level of expression of at least 3 biomarkers selected from the group of biomarkers listed in Table 1 have a low level of expression.

21. The method of any one of claims 1-8, wherein the level of expression of at least 4 biomarkers selected from the group of biomarkers listed in Table 1 have a low level of expression.

22. The method of any one of claims 1-8, wherein the level of expression of at least 5 biomarkers selected from the group of biomarkers listed in Table 1 have a low level of expression.

23. The method of any one of claims 1-8, wherein there is a low level of expression of said biomarker as compared to a control.

24. The method of claim 23, where said biomarker is not expressed at a detectable level.

25. The method of any one of claims 1-8, wherein the level of expression of said biomarker is determined at the nucleic acid level.

26. The method of claim 25, the level of expression of said biomarker is determined by detecting cDNA.

27. The method of claim 25, the level of expression of said biomarker is determined by detecting mRNA or miRNA.

28. The method of claim 25, the level of expression of said biomarker is determined by detecting DNA.

29. The method claim 25, wherein the level of expression of said biomarker is determined by using a technique selected from the group consisting of polymerase chain reaction (PCR) amplification reaction, reverse-transcriptase PCR analysis, quantitative reverse-transcriptase PCR analysis, Northern blot analysis, RNAase protection assay, digital RNA detection/ quantitation, and combinations or sub-combinations thereof.

30. The method of any one of claims 1-8, wherein the level of expression of said biomarker is determined at the protein level.

31. The method of claim 30, wherein the presence of the protein is detected using an antibody or antigen binding fragment thereof, which specifically binds to the protein.

32. The method of claim 31, wherein the antibody or antigen binding fragment thereof is selected from the group consisting of a murine antibody, a human antibody, a humanized antibody, a bispecific antibody, a chimeric antibody, a Fab, Fab', $F(ab')_2$, ScFv, SMIP, affibody, avimer, versabody, nanobody, a domain antibody, and an antigen binding fragment of any of the foregoing.

33. The method of claim 31, wherein the antibody or antigen binding fragment thereof is labeled.

34. The method of claim 33, wherein the antibody or antigen binding fragment thereof is labeled with a label selected from the group consisting of a radio-label, a biotin-label, a chromophore-label, a fluorophore-label, and an enzyme-label.

35. The method of claim 28, wherein the level of expression of said biomarker is determined by using a technique selected from the group consisting of an immunoassay, a western blot analysis, a radioimmunoassay, immunofluorimetry, immunoprecipitation, equilibrium dialysis, immunodiffusion, electrochemiluminescence immunoassay (ECLIA), ELISA assay, immunopolymerase chain reaction and combinations or sub-combinations thereof.

36. The method of claim 35, wherein the immunoassay is a solution-based immunoassay selected from the group consisting of electrochemiluminescence, chemiluminescence, fluorogenic chemiluminescence, fluorescence polarization, and time-resolved fluorescence.

37. The method of claim 35, wherein the immunoassay is a sandwich immunoassay selected from the group consisting of electrochemiluminescence, chemiluminescence, and fluorogenic chemiluminescence.

38. The method of any one of claims 1-3, 6 or 8, wherein said sample comprises a fluid, or component thereof, obtained from said subject.

39. The method of claim 38, wherein the fluid is selected from the group consisting of blood, lymph, serum, plasma, cystic fluid, nipple aspirates, urine, sputum, and fluid collected from a biopsy.

40. The method of any one of claims 1-3, 6 or 8, wherein the sample comprises a tissue, or component thereof, obtained from said subject.

41. The method of claim 40, wherein said tissue is selected from the group consisting of breast tissue, connective tissue, lymphatic tissue, tissue obtained from a biopsy and tissue obtained from a lump biopsy.

42. The method of claim 40, wherein the tissue is breast tissue or a component thereof.

43. The method of claim 42, wherein said component of said breast tissue comprises breast tissue cells.

44. The method of claim 43, wherein said breast tissue cells are circulating breast tumor cells.

45. The method of any one of claims 1-3 or 6-8, wherein said subject is a human subject.

46. A kit for predicting whether eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof, may be used to treat a subject having breast cancer, the kit comprising

reagents for determining the level of expression of a biomarker selected from the group of biomarkers listed in Table 1; and

instructions for use of the kit to predict whether eribulin, an analog thereof, or a pharmaceutically acceptable salt thereof, may be used to treat a subject having breast cancer.

47. The kit of claim 46, wherein the pharmaceutically acceptable salt of eribulin is eribulin mesylate.

48. The kit of claim 46, wherein the reagent for determining the level of expression of the biomarker is a probe for identifying a null mutation in the biomarker.

49. The kit of claim 46, wherein the reagent for determining the level of expression of the biomarker is a probe for amplifying and/or detecting the biomarker.

50. The kit of claim 46, wherein the reagent for determining the level of expression of the biomarker is an antibody.

51. The kit of claim 46, further comprising reagents for obtaining a biological sample from a subject.

52. The kit of claim 46, further comprising a control sample.

5