NEOPLASIA
www.neoplasia.com

Volume 14 Number 11 November 2012 pp. 1087–1096 1087

Long-range Transcriptome Sequencing Reveals Cancer Cell Growth Regulatory Chimeric mRNA[1,2]

Roberto Plebani[*,3], Gavin R. Oliver[†,3], Marco Trerotola[*,4], Emanuela Guerra[*], Pamela Cantanelli[*], Luana Apicella[*], Andrew Emerson[‡], Alessandro Albiero[§], Paul D. Harkin[†,¶], Richard D. Kennedy[†,¶] and Saverio Alberti[*,#]

*Unit of Cancer Pathology, Centre of Excellence for Research on Aging, "G. D'Annunzio" University Foundation, Chieti, Italy; [†]Almac Diagnostics, Craigavon, United Kingdom; [‡]CINECA, Bologna, Italy; [§]BMR Genomics, Padova, Italy; [¶]Centre for Cancer Research & Cell Biology, Queen's University, Belfast, United Kingdom; [#]Department of Neuroscience and Imaging, "G. d'Annunzio" University of Chieti-Pescara, Chieti, Italy

Abstract

mRNA chimeras from chromosomal translocations often play a role as transforming oncogenes. However, cancer transcriptomes also contain mRNA chimeras that may play a role in tumor development, which arise as transcriptional or post-transcriptional events. To identify such chimeras, we developed a deterministic screening strategy for long-range sequence analysis. High-throughput, long-read sequencing was then performed on cDNA libraries from major tumor histotypes and corresponding normal tissues. These analyses led to the identification of 378 chimeras, with an unexpectedly high frequency of expression ($\approx 2 \times 10^{-5}$ of all mRNA). Functional assays in breast and ovarian cancer cell lines showed that a large fraction of mRNA chimeras regulates cell replication. Strikingly, chimeras were shown to include both positive and negative regulators of cell growth, which functioned as such in a cell-type–specific manner. Replication-controlling chimeras were found to be expressed by most cancers from breast, ovary, colon, uterus, kidney, lung, and stomach, suggesting a widespread role in tumor development.

Neoplasia (2012) 14, 1087–1096

Introduction

Several chimeric transcripts have been discovered in human solid tumors, which derive from chromosomal translocations. These often encode structurally and functionally altered signaling molecules or transcription factors [1] or may also function as non-coding RNA [2]. More than half of prostate cancers harbor fusion sequences, mostly *TMPRSS-ERG* [3]. The *SLC45A3-ELK4* (ETS family) fusion transcript can be generated both by chromosomal rearrangement and by trans-splicing, and it was found to be expressed in both normal prostate tissue and in prostate cancer. High levels of *SLC45A3-ELK4* mRNA are restricted to a subset of prostate cancer samples [4]. A small inversion within chromosome 2p leads to the formation of a fusion gene comprising *EML4* and *ALK* in non–small cell lung cancer [5]. The fusion of *MAML2* with *CRTC1* or *CRTC3* has a role in the development of mucoepidermoid carcinomas [6]. Rearrangements of *RAF* pathway members occur in prostate and gastric cancers [7], and a paracentric inversion of chromosome 7q results in an in-frame fusion

Abbreviations: FP, fusion point; NGS, next-generation sequencing; SD, standard deviation

Address all correspondence to: Prof. Saverio Alberti, MD, PhD, Unit of Cancer Pathology, Ce.S.I., University "G. d'Annunzio", Via Colle dell' Ara, 66100 Chieti Scalo (Chieti), Italy. E-mail: s.alberti@unich.it

[1]This work was supported by Fondazione Cassa di Risparmio della Provincia di Chieti, Italian Ministry of Health (RicOncol grant RF-EMR-2006-361866), Fondazione Compagnia di San Paolo (grant 2489IT), Ministero dello Sviluppo–Made in Italy (contract N° MI01_00424), and the Italian Foundation for Cancer Research (fellowship to M.T.).

[2]This article refers to supplementary materials, which are designated by Tables S1 to S11 and Figures S1 to S4 and are available online at www.neoplasia.com. The FusionMiner software is freely available at FusionMiner.sourceforge.net.

[3]These authors contributed equally to this work.

[4]Current address: Kimmel Cancer Center, Department of Cancer Biology, Thomas Jefferson University, Philadelphia, PA 19107.

Received 16 August 2012; Revised 16 August 2012; Accepted 30 September 2012

 1522-8002/12/$25.00
DOI 10.1593/neo.121342

between exons 1 and 8 of the *AKAP9* gene and between exons 9 and 18 of *BRAF* in radiation-induced papillary carcinomas [8]. Other thyroid carcinoma–specific events include fusion of the *RET* oncogene to various partners [9]. Further oncogenic fusions have been detected in other solid tumors [10,11].

Cancer transcriptomes also contain mRNA chimeras that arise as transcriptional (long intergenic transcription) or post-transcriptional (trans-splicing [12]) events that may play a role in tumor development. Previous findings showed that oncogenic transcripts can indeed be generated post-transcriptionally [13–15]. The fusion of *CYCLIN D1* mRNA to *TROP2* transcripts generates oncogenic *CYCLIN D1-TROP2* chimeras, whose tumor-promoting function is induced with a dramatically increased mRNA stability [13]. The oncogenic *JAZF1-JJAZ1* chimeric mRNA can be originated by trans-splicing as well as by a chromosomal translocation [14]. Similarly, the *SLC45A3-ELK4* chimeric transcript can be generated in the absence of chromosomal rearrangements [4,16]. Intergenic splicing generates a ubiquitous chimeric mRNA between the *P2Y11* and *SSF1* transcripts [17]. The generation of these chimeras appears as a regulated event [13,14] and was shown to also occur in normal tissues [4,13,14,17–20]. Several of these chimeric transcripts have been used as diagnostic or prognostic [21] markers and as targets for anti-neoplastic therapy [10,13,22,23].

Screening strategies were previously developed for *in silico* identification of mRNA chimeras in cancer cells [24]. Next-generation sequencing (NGS) approaches now provide much larger sequence information for chimera discovery [7,19,20,25–27]. However, most second-generation NGS approaches generate highly multiplexed, short-tag sequence reads, which are then condensed in strings of base-call probabilities, through a probabilistic fitting of massively parallel data sets. This makes contig assemblies and target alignments correspondingly more difficult [19,25,27,28]. Alignments to complex genomes are even more hampered, because of higher sequence complexity [29] and homology within closely related gene families and pseudogenes.

These problems have led to significant efforts for achieving longer sequence reads and higher sequencing accuracy. In 2005, 454 launched the first NGS apparatus, which was able to generate 100-bp reads. Sequence reads extended to 200 bp in 2007 [30] and are close to 900 bp at present [31]. SOLID sequencing generated 35-bp reads in 2007 [30], and this extended to 75 bp in 2011 [32]. Illumina generated 36-bp sequence reads in 2006 to 2008 [30]. These extended to 100 bp in 2010 [31] and to 300 bp in 2012 (www.illumina.com). Ion Torrent introduced its first sequencer at the end of 2010, and this was capable of 100-bp-long reads. As of 2012, reads of 525-bp average length have been obtained (www.iontorrent.com/lib/images/PDFs/pe_appnote_v12b.pdf). Pacific Biosciences (www.pacificbiosciences.com) succeeded in obtaining even longer reads, which currently are up to 1500 bp.

To take advantage of these technical advances, we have developed an analytical strategy for high-accuracy identification of mRNA chimeras in long-read DNA sequence data sets (Figure 1). This strategy was shown to work efficiently for chimera recognition (Tables S1–S7 and Figure S1). High-throughput, long-read sequencing was then performed on cDNA libraries from major tumor histotypes and corresponding normal tissues. This led to the identification of 378 chimeras, from both normal and transformed cells, indicating an unexpectedly high frequency of expression ($\approx 2 \times 10^{-5}$ of all mRNA). Functional assays in breast and ovarian cancer cell lines showed that a large fraction of mRNA chimeras regulate cell replication. Strikingly, chimeras were shown to include both positive and negative regulators of cell growth, which functioned as such in a cell-type–specific manner. Replication-controlling chimeras were found to be expressed by most cancers from breast, ovary, colon, uterus, kidney, lung, and stomach, suggesting selective pressure for a role in tumor development.

Materials and Methods

Cells

Human MCF-7, MCF-7/Almac, HBL-100, SK-BR-3, MDA-MB-231, MDA-MB-361, MDA-MB-415, MDA-MB-453, MDA-MB-468, HS578, and ZR751 breast cell lines and SKOV-3, IGROV-1, OVCAR-3, and OVCA-432 ovarian cancer cell lines were grown in RPMI 1640 medium supplemented with 10% FBS, 100 IU/ml penicillin, and 100 μg/ml streptomycin (Euroclone, Milan, Italy). All cell lines were obtained from ATCC (LGC Standards, Teddington Middlesex, United Kingdom) where they were authenticated by standardized procedures (www.atcc.org).

Cell Growth Assays

MCF-7, HBL-100, SK-OV-3, IGROV-1, and OVCAR-3 cells were seeded at 1×10^3 to 10×10^3 cells/well in 96-well plates (five replicates per data point). Cell numbers were quantified by staining with crystal violet [33]. Standard growth curves for each cell line were generated by seeding two-fold serial dilutions of defined cell numbers. Crystal violet standard curves showed good linear responses ($R^2 > 0.998$, in all cases) (Figure S2). To support the crystal violet readings, quantification was also performed by image analysis (ImageJ). Digital pictures were taken from 96-well plates after fixation. Picture noise was removed with GIMP software, after random sampling of cell-free pixels. ImageJ analysis was then performed by quantifying black areas in each culture well after image conversion to a gray scale (manuscript in preparation).

DNA Transfection

Cells were transfected with DNA in Lipofectamine 2000 (HBL-100, SKOV-3, IGROV-1, and OVCAR-3 cells) or LTX (Invitrogen, San Diego, CA), which was found to be optimal for MCF-7 cells (Figure 5*C*) [34], following the manufacturer's instructions. pEYFP transfection was used to quantify transfection efficiency [35] (EYFP expression, as measured by flow cytometry).

Flow Cytometry Immunofluorescence

Flow cytometry analysis was performed as described previously [36,37], on fluorescence-activated cell sorters (FACSCalibur, Becton-Dickinson, Sunnyvale, CA). To improve the detection of EYFP transfectants, we performed subtraction of cell autofluorescence and displacement of true transfectants in the red channel as described [35,38].

Human Samples for Tumor Transcriptome Sequencing

Non–small cell lung cancer. Non–small cell lung cancer libraries were generated from a set of frozen tissue samples, comprising 65 tumor samples (30 adenocarcinomas, 20 squamous cell carcinomas, and 15 other morphologies) from the Roy Castle Lung Cancer Research Institute (University of Liverpool) and Queens University Belfast. To maximize chances of mRNA chimera discovery, we proceeded to generate libraries from both tumor and normal tissues. Normal lung RNA was obtained from multiple commercial suppliers (Clontech, Palo Alto, CA; Ambion, Austin, TX; BioCat, Heidelberg, Germany; Stratagene, La Jolla, CA; Cybridi, Rockville, MD; and OriGene, Rockville, MD), overall from 16 donors of different ethnicity.

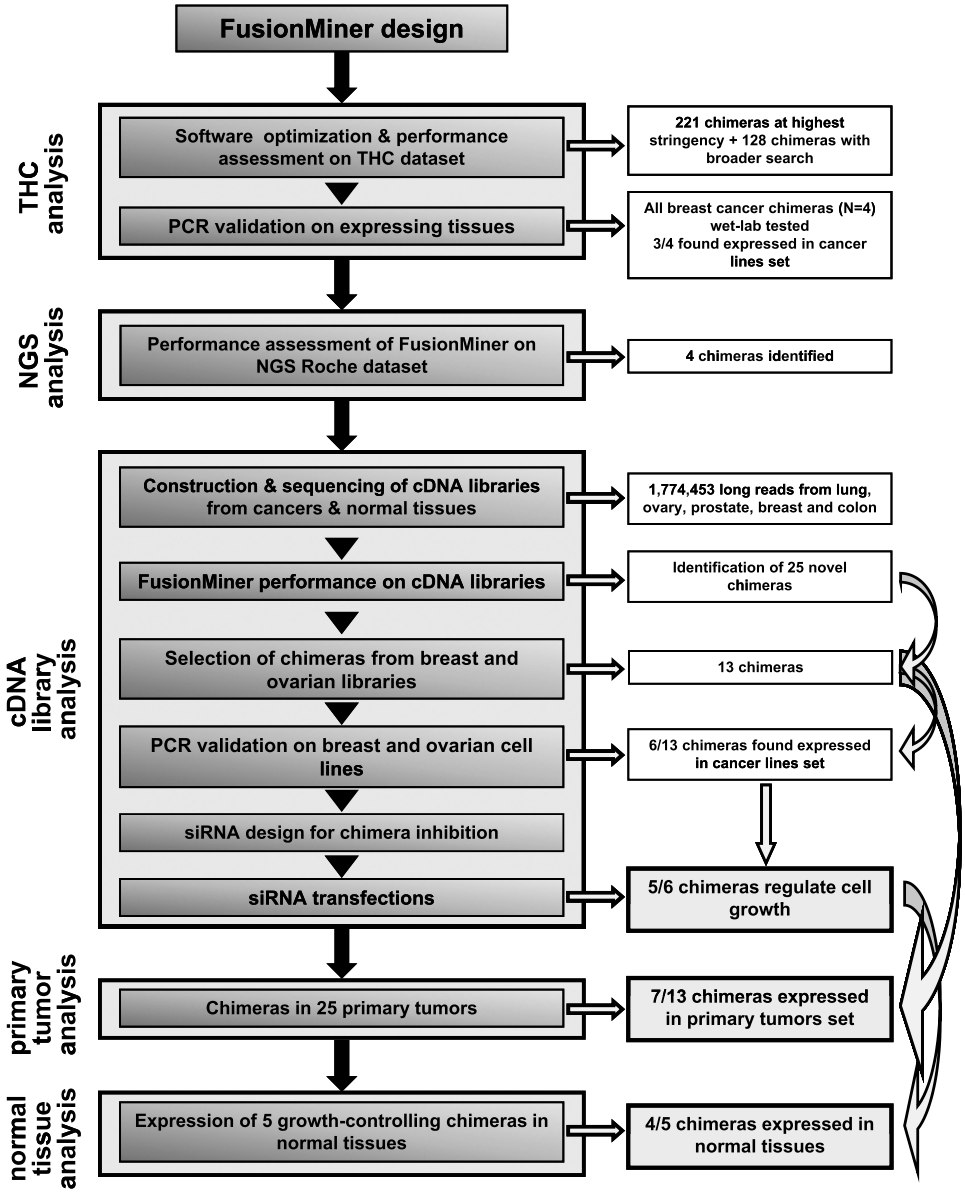


Figure 1. Flow diagram of chimera identification and validation steps.

Ovarian cancer. Ovarian library A comprised 64 ovarian tumors (31 serous, 14 endometroid, 6 mucinous, 5 clear cell, and 8 undifferentiated cancers; 52 were stage III/IV and 12 were stage I/II). For ovarian library B, RNA from normal ovarian tissue was obtained from commercial suppliers (Ambion and AMS Biotechnology [Bioggio, Switzerland]). The library was generated with equal quantities of RNA from different ethnicities (Asian, Caucasian, and African-American), with 23 donors overall. For ovarian library C, ovarian tumor total RNA was obtained from various commercial suppliers (Ambion, Clontech, Cytomyx [Lexington, MA], Biocat, and Asterand [Detroit, MI]). The library was composed of equal quantities of RNA of different ethnicity (Asian, African-American, and Caucasian), with 37 donors overall.

Prostate cancer. The prostate cancer library was constructed from 30 tumors (74% Caucasian and 26% African-American), 8 normal prostate RNA supplied by Clontech, AMS Biotechnology, and Cybridi, and 56 normal tissues adjacent to tumors obtained from St. Vincents Hospital (Dublin, Ireland).

Breast cancer. The breast cancer library was composed of 90 tumors and 18 normal samples [39–41].

Colorectal cancer. The colorectal library comprised 40 tumor samples and 40 normal tissues.

Tumor Validation Sample Set

cDNA was synthesized from 25 human primary tumors (10 breast, 6 colon, 3 stomach, 2 ovary, 2 kidney, and 2 uterus), which were independent from those used to construct the cDNA libraries. These 25 samples were used as a test set to validate chimera expression by both conventional polymerase chain reaction (PCR)/sequencing and real-time reverse transcription (RT)–PCR.

Normal Tissues

Normal breast, colon, uterus, prostate, placenta, lung, kidney, pancreas, and stomach RNA were obtained from Clontech.

cDNA Library Construction

All of the frozen tumor tissues were homogenized in RNA STAT-60 (Tel-Test, Friendswood, TX), and the RNA was extracted according to the manufacturer's instructions. Equal amounts of good quality total RNA were pooled, and the mRNA was isolated using μMACS mRNA isolation kits (Miltenyi Biotec, Bergisch Gladbach, Germany), as described by the manufacturer. Lung cDNA libraries were constructed from 3 μg of mRNA using the CloneMiner cDNA library construction kit (Invitrogen), according to the manufacturer's instructions. cDNA were inserted in the pDONR 222 vector from Invitrogen. Titer and average insert size in each cDNA library were determined according to the manufacturer's instructions. Plasmid preparations of individual clones were carried out using a modified Montáge alkaline lysis method (Millipore, Billerica, MA) that incorporates MultiScreen Plasmid 384 Miniprep clearing plates for centrifugal lysate clearing.

Sequencing of cDNA Libraries

Colony sequencing automation was implemented (QPix colony picker Biomek liquid handlers). Cycle sequencing reactions were performed in 10-μl volumes using a 1/16 dilution of Big Dye Terminator v3.1 ready reaction mix in Big Dye sequencing buffer (Applied Biosystems, Foster City, CA), 5 μM M13 primer, and 100 ng of template DNA. Cycle sequencing was performed for 40 cycles at 95°C for 10 seconds, 50°C for 5 seconds, and 60°C for 2.5 minutes. Excess dye terminators were removed using CleanSEQ (Agencourt Biosciences Corporation, Beverly, MA). Sequencing plates were analyzed on Applied Biosystems 3730/3730 × l DNA Analyzers using Applied Biosystems Sequence Analysis software. M13 forward primers were used for 5′ end sequencing of the colorectal and breast libraries; M13 reverse primers were used for 3′ end sequencing of the normal lung and prostate libraries; both M13 forward and reverse primers were used for 5′ and 3′ end sequencing of the lung tumor and ovarian cancer libraries.

Plasmids

The pEYFP expression vector (Clontech) was used to express YFP. The pSUPER vector [42] was used for RNA interference.

Small Inhibitory RNA (siRNA)

siRNA design followed four complementary strategies, i.e., Tuschl criteria (position in the mRNA, guanine-cytosine [GC] content, base composition, and flanking sequences) [43], Invitrogen algorithms (rnaidesigner.invitrogen.com/rnaiexpress/; sequence composition, nucleotide content, thermodynamic properties, and experimental validation), Whitehead Institute screening procedures (jura.wi.mit.edu/bioc/siRNAext/; Tuschl criteria, predictions of binding energies and BLAST filtering of cross-hybridizing sequences) [44], and Sonnhammer searches (www.sirnawizard.com/design_advanced.php; data mining on validated siRNA databanks, using motif rules and energy parameters) [45].

Annealed siRNA oligos were subcloned into the pSUPER vector. siRNA expression constructs were transiently transfected in MCF-7 and HBL-100 breast cancer cells and in SK-OV-3, IGROV-1, and OVCAR-3 ovarian cancer cells. siRNA-targeted transcript levels were quantified by real-time PCR. Negative-control siRNA directed toward irrelevant targets were used; these were chosen after extensive testing for lack of off-target influence on cell growth.

Quantitative RT-PCR

Hybrid sequences in cancer cell lines and tumor samples were amplified by quantitative RT-PCR. One microgram of total RNA was reverse transcribed with the M-MLV Reverse Transcriptase (Promega, Madison, WI) according to standard protocols. cDNA was quantified by ethidium bromide fluorescence in solution [46]. Quantitative RT-PCR was performed with an ABI-PRISM 7900HT Sequence Detection System (PE Applied Biosystems, Foster City, CA), using Sybr Green as the probe (Applied Biosystems). Samples were assayed as replicates (two or three independent samples), and the $1.83^{-\Delta\Delta CT}$ method was used to calculate the relative changes in gene expression [13]. The glyceraldehyde 3-phosphate dehydrogenase (GAPDH) housekeeping gene was used as an internal control. For setup curves, ΔC_T (C_T, target gene – C_T, GAPDH) was calculated for each cDNA dilution. The data were fit using least-squares linear regression analysis. As amplification efficiency was linear over the range of RNA amounts used, amplification curves were used to calculate crossover point values for siRNA-treated samples. To check for the correctness of amplified bands, amplification products were run on 3% agarose gels. Amplified products were purified and extensively sequenced (BMR Genomics, Padova, Italy). Quantitative RT-PCR was also performed with PrimeTime IDT (Integrated DNA Technologies, Bologna, Italy; www.idtdna.com) to reliably detect with higher sensitivity the interchromosomal CHD2-CHMP1A fusion in normal tissues.

Diagnostic PCR

Interchromosomal *CHD2-CHMP1A* and *ADK-DHX8* and intrachromosomal *PRKAA1-TTC33*, *SAMM50-PARVB* and *P2RX5-TAX1BP3*, *URB1-C21orf45*, *CTBS-GNG5*, *THC2538403 ZNF498-CUX1*, *THC2523555 C9orf47-S1PR3*, and *THC2668182 KLH22-SCARF* were amplified in 10 breast and 4 ovarian cancer cell lines and in 25 tumor samples by nested PCR. Chimeric mRNA were amplified by 35 amplification cycles (30 seconds at 94°C for denaturation, 30 seconds at 60°C for annealing, and 30 seconds at 72°C for extension). Hot Master Taq-polymerase 0.7 units (Eppendorf) and 12.8 pmol of forward and reverse primers were used for the amplification reaction. All of the amplified products were purified and sequenced (BMR Genomics).

Statistical Analysis

Two-way analysis of variance and *post-hoc* Bonferroni *t* tests were used for growth curve comparisons. Data were analyzed using Sigma Stat (SPSS Science Software UK Ltd, Birmingham, United Kingdom) and GraphPad Prism (GraphPad Software Inc, La Jolla, CA).

Results

Chimeric mRNA Detection Procedure

A procedure (FusionMiner) was designed to process BLAST analyses of query sequences against genomic databanks, through sequential stages of analysis and exclusion and pass-or-fail tests, as described in the Supplemental Online Material (Figure 1 and Tables S1–S7). FusionMiner performance was assessed by screening the Dana Farber Cancer Institute Gene Index Project tentative human consensus (THC) collection (Figure S1) and long-sequence-read 454 Titanium data sets (Supplemental Online Material). Samples of the identified chimeras were then validated by diagnostic PCR and by real-time quantitative PCR analysis of cancer cell lines.

Transcriptome Sequencing for Growth Regulatory Chimera Discovery

To discover growth regulatory chimeras, we then performed a large-scale sequencing and analysis of tumor and normal tissue transcriptomes. To maximize chances of discovery of growth regulatory chimeras, both major tumor histotypes, i.e., non–small cell lung, breast, prostate, ovary, and colorectal cancers, and the corresponding normal tissues were analyzed. Long-sequence-read (900 bp on average) cDNA library data sets were obtained: 481,765 from ovary, 485,049 from prostate, 157,259 from breast, 46,445 from colon, and 603,935 from lung.

These sequences were run through FusionMiner. Twenty-five mRNA chimeras were identified (15 intrachromosomal and 10 interchromosomal; Table S8, Supplemental Sequence Data). All sequences were shown to possess the structural characteristics of *bona fide* chimeric mRNA [24] (Supplemental Sequence Data). Breast and ovarian chimeras were validated by RT-PCR and functional assays (see below).

These findings led to estimate absolute chimera frequencies as 1.4×10^{-5} of all mRNA. This was in remarkable agreement with NGS sequencing data ($\approx 2 \times 10^{-5}$) (Supplemental Online Material), indicating an unexpectedly high frequency of expression of chimeric mRNA.

Chimeric Transcript Expression in Cancer Cells

Expression of the nine chimeras from the breast library and of the four chimeras from the ovarian library was analyzed in breast (MCF-7, HBL-100, SK-BR-3, MDA-MB-231, MDA-MB-361, MDA-MB-415, MDA-MB-453, MDA-MB-468, HS578, and ZR751) and ovarian (SKOV-3, IGROV-1, OVCAR-3, and OVCA-432) cancer cell lines (Figure 2 and Table S8). Six of the nine chimeras were successfully amplified by RT-PCR (Figure 2*A* and Table S9). Amplification from breast cancer cells was obtained for *PRKAA-TTC33* (10/10 lines), *SAMM50-PARVB* (5/10 lines), *P2RX5-TAX1BP3* (3/10 lines), and *CHD2-CHMP1A* (9/10 lines) (Figure 2*A*, *left*). All individual amplicons were sequence verified (Figure 2*C*). Three of these chimeras were also detected in ovarian cancer cells: *PRKAA-TTC33* (4/4 lines); *SAMM50-PARVB* (3/4 lines), and *CHD2-CHMP1A* (4/4 lines) (Figure S3).

The *URB1-C21ORF45* and *CTBS-GNG5* chimeras from the ovarian library were identified in all four ovarian cancer cell lines (Figure 2*A*, *right*). They were also detected in all breast cancer lines. Notably, different cancer cells expressed different steady-state levels of the chimeric mRNA, e.g., *CTBS-GNGS* was approximately 20 times

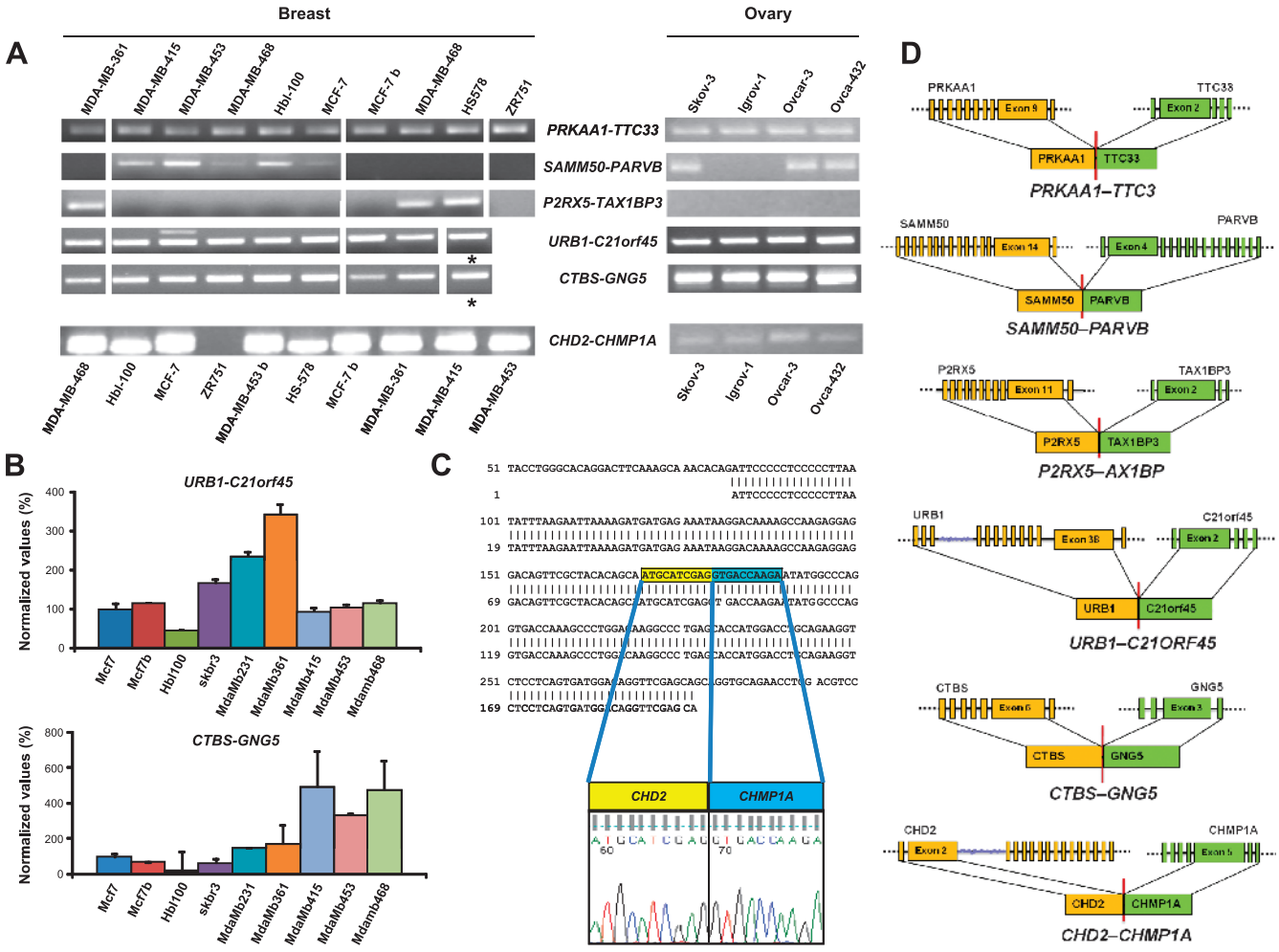


Figure 2. Chimeric mRNA expression in cancer cell lines. (A) Expression of chimeras discovered from tumor and normal tissue library sequencing; agarose electrophoresis of nested or real-time PCR products. Breast and ovarian cancer cell lines are indicated; *SK-BR-3. (B) URB1-C21ORF45 (top) and CTBS-GNG5 (bottom) expression in breast cancer cell lines by quantitative RT-PCR; results are expressed as percent values (MCF-7 = 100); three replica samples were analyzed per data point. Bars, SD. (C) *CHD2-CHMP1A*. Sequence of the PCR amplicon *versus* that of the chimera isolated from the breast library. (D) Structure of validated chimeric mRNA; 5′ partners (orange) and 3′ partners (green) are shown; exon junctions are indicated.

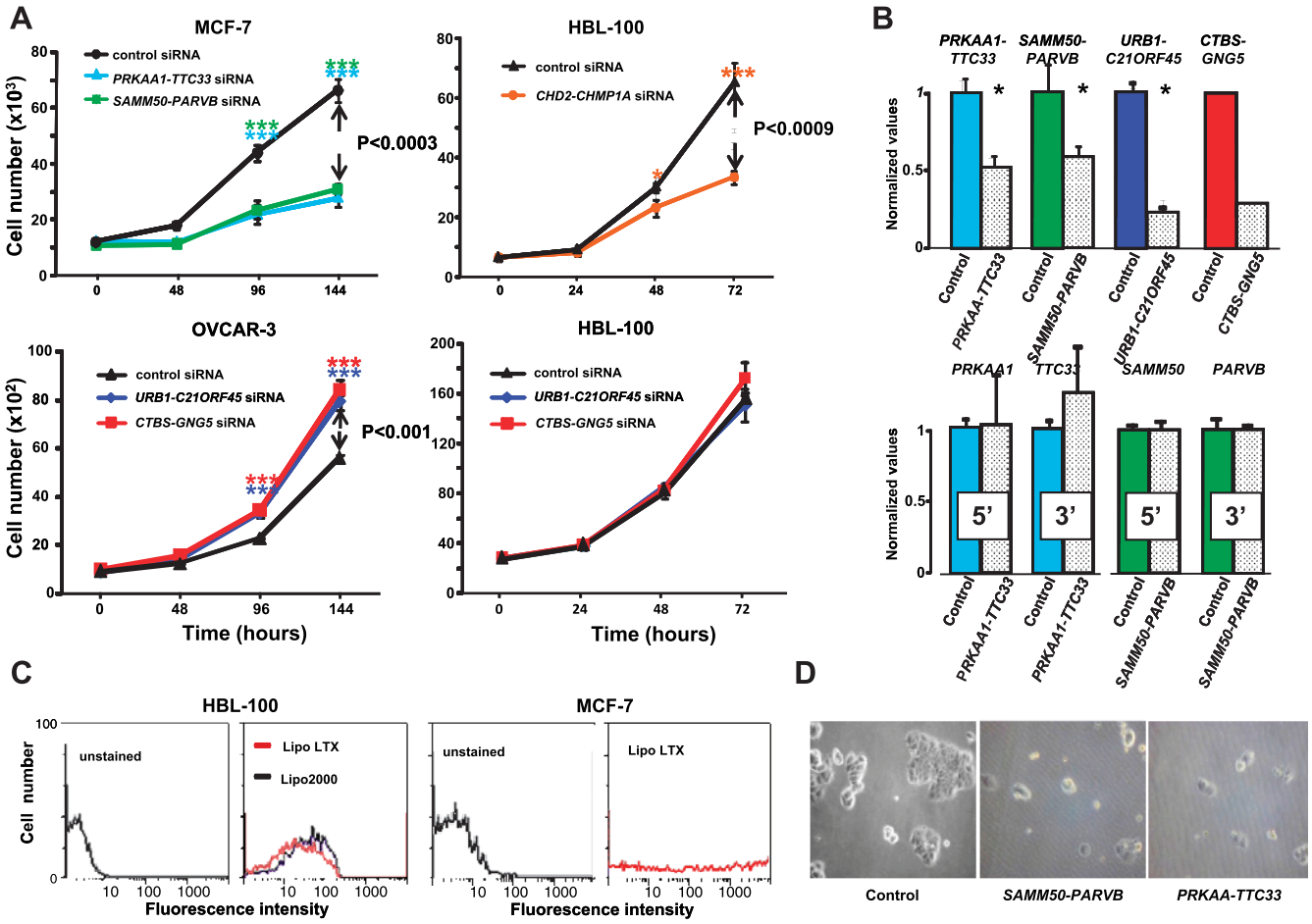


Figure 3. Cell growth modulation by chimeras. (A) Control (Table S10) and chimera-targeting siRNA are color coded. (Top, left) MCF-7; *PRKAA-TTC33* (cyan), *SAMM50-PARVB* (green), control (black). (Top, right) HBL-100; *CHD2-CHMP1A* (orange), control (black). (Bottom, left) OVCAR-3; *URB1-C21ORF45* (blue), *CTBS-GNG5* (red). (Bottom, right) HBL-100 treated with siRNA for chimeras from ovarian libraries; *URB1-C21ORF45* (blue), *CTBS-GNG5* (red). Bars, SD. Brackets, *P* value of two-way analysis of variance; Bonferroni *t* test significance: $*P \leq .05$; $***P \leq .001$. (B) Real-time PCR of siRNA-transfected cells. (Top) Chimeric RNA (left to right: *PRKAA-TTC33* and *SAMM50-PARVB* in MCF-7; *URB1-C21ORF45* and *CTBS-GNG5* in OVCAR-3). (Bottom) Single-partner RNA expression after the indicated siRNA treatment (left to right: *PRKAA-TTC33* and *SAMM50-PARVB* in MCF-7). (C) Flow cytometry analysis of transfected HBL-100 and MCF7 cells; YFP was used as a transfection efficiency benchmark. (Left) HBL-100; LTX (red) or Lipo-2000 (blue). (Right) MCF-7; LTX transfection. (D) MCF-7 cell growth blockade after *PRKAA-TTC33*–targeted or *SAMM50-PARVB*–targeted siRNA treatment (day 6 after transfection).

less expressed in HBL-100 cells, as compared with MDA-MB-415 cells (Figure 2*B*).

Overall, 75% of the THC chimeras and 54% of the chimeras from breast and ovary libraries (Tumor Transcriptome Sequencing Project) were detected in breast and ovarian cancer cell lines/primary tumors.

Fusion Proteins Encoded by the Growth Regulatory Chimeras

CHD2-CHMP1A. *CHD2* encodes the chromodomain helicase DNA-binding protein 2; *CHMP1A* encodes the chromatin-modifying protein 1A. Of interest, both of these chimera partners encode proteins with regulatory roles on chromatin/DNA structure. However, only the first 20 amino acids of helicase DNA-binding protein 2 are retained in the fusion-protein product (Table S11). This contains a casein kinase II phosphorylation site (prosite.expasy.org/). One out-of-frame C-terminal amino acid is provided by the chromatin-modifying protein 1A sequence (Table S11) and generates a hybrid N-glycosylation site, although it is not clear if this is processed *in vivo*.

CTBS-GNG5. *CTBS* encodes chitobiase; *GNG5* encodes the di-*N*-acetyl–binding and guanine-nucleotide–binding proteins. Chitobiase is a lysosomal glycosidase that is involved in degradation of asparagine-linked oligosaccharides on glycoproteins. It is also involved in the hydrolysis of *N*-acetyl-β-D-glucosamine. *GNG5* encodes the γ chain of trimeric G proteins. A fusion mRNA between chitobiase and guanine-nucleotide–binding protein was also identified by Akiva et al. [47] and by Nacu et al. [26]. The *CTBS-GNG5* is an "in-frame" fusion that preserves the first 319 amino acids from the N-terminal partner and the last 41 amino acids from the C-terminal partner (Table S11). *CTBS* provides an apparently functional chitinase catalytic domain, with a formal glycosylation site at S300. Most of Gγ5 is retained in the fusion (Supplemental Figure S4), which raises the possibility that the fusion protein can bind its Gβ partner, whether at the cell membrane or in the cytoplasm.

PRKAA1-TTC33. *PRKAA1* encodes a 5′-AMP–activated protein kinase catalytic subunit α-1; *TTC33* encodes tetratricopeptide repeat domain 33. *PRKAA1* is a Ser/Thr protein kinase that protects cells from

stress-dependent ATP depletion by switching off ATP-consuming biosynthetic pathways; *PRKAA1* also regulates fatty acid and cholesterol synthesis. The N-terminal segment retained in the chimera contains most of the protein kinase A1 protein (478/559 amino acids; Table S11). This retains the full catalytic domain (46–279). However, it loses 12 phosphorylation sites (T488, T490, T522, S496, S502, S506, S508, S516, S520, S523, S524, and S527), which suggests the loss of at least some of its physiological regulation. The fusion protein contains 32 C-terminal amino acids from the 3′ partner mRNA (*TTC33*), which do not correspond to its canonical reading frame (out-of-frame fusion; Table S11).

SAMM50-PARVB. *SAMM50* encodes the sorting and assembly machinery (SAM) component 50 homolog; *PARVB* encodes β-parvin. SAM-50 is part of the SAM complex, which has a role in integrating β-barrel proteins into the outer mitochondrial membrane. β-Parvin is an actin-binding protein that associates with focal contacts. Parvin is a key regulator of integrin-linked kinase (ILK) and of its downstream pathways. The encoded fusion protein contains an almost entire SAM-50, which only misses its last 15 C-terminal amino acids (Table S11). However, this may lead to disruption of the second major functionally relevant domain of SAM-50. Only two amino acids are contributed by the β-parvin mRNA, as an out-of-frame sequence (Table S11).

URB1-C21orf45. *URB1* encodes the pre-ribosomal–associated protein 1 (Npa1p); *C21orf45* encodes the kinetochore protein homolog A. Npa1p is a component of pre-60S ribosomal particles and associates with small nucleolar ribonucleoprotein particles (RNPs) that are required for peptidyl transferase center modification. The kinetochore protein homolog A is involved in mitosis and associates with chromatin. It also associates with centromeres in interphase cells, from late anaphase to G_1. The fusion protein keeps essentially all the exons of the N-terminal partner (38/39), including its S1385 phosphorylation site (Table S11), suggesting a largely unaltered function. Albeit the C-terminal sequence is not in its native frame, it is unusually long (121 amino acids; Table S11) and may carry novel associated functions.

Chimeras Contain Positive and Negative Regulators of Cell Growth

The six chimeras that were found to be expressed by target cell lines in culture were assayed for a role in cell growth. siRNA targeting the chimeric joint (Table S11) were used to inhibit the expression of the corresponding chimeras in breast cancer cells (Figure 3). Transfection efficiency was optimized by using a co-transfected pEYFP reporter plasmid (Figure 3*C*, *left*); the vast majority of target cells appeared successfully transfected (Figure 3*C*). Transfected siRNA downregulated mRNA chimera levels by ≈75% (Figures 3*B* and S3). To ensure absence of off-target effects, due to artifactual reduction of chimera partner transcript levels, 5′ and 3′ chimeric partners were analyzed in parallel. Levels of partner transcripts of growth regulatory chimeras remained unaffected by siRNA targeting the chimera junction regions (Figure 3*B*).

Remarkably, five of six tested chimeras appeared to regulate cell growth. The strongest growth inhibition in HBL-100 cancer cells

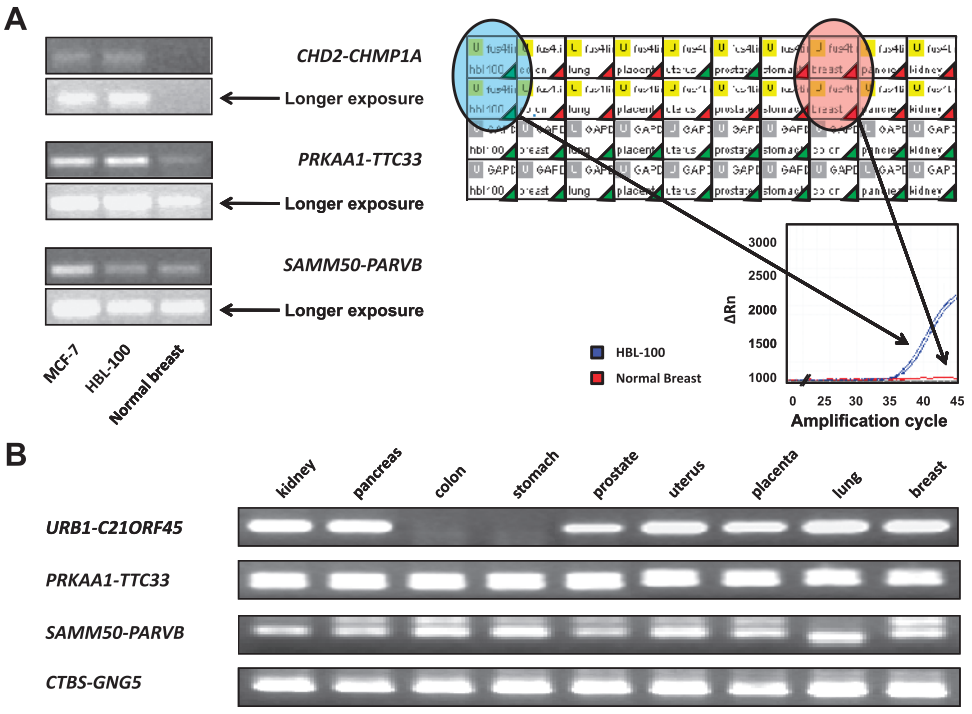


Figure 4. Expression of mRNA chimeras in normal tissues. (A) (Left) Agarose electrophoresis analysis of real-time PCR products (Sybr Green real-time PCR). (Right) Real-time PCR analysis of the *CHD2-CHMP1A* chimeric mRNA in normal tissues (PrimeTime real-time PCR). HBL-100 cell cDNA was used as a positive control. (Bottom) Amplification plots of *CHD2-CHMP1A* mRNA in the HBL-100 breast cancer cell line (blue) and normal breast tissue (red). *CHD2-CHMP1A* (top lanes) and GAPDH (bottom lanes) were amplified in duplicate. Green triangles, successful amplification; red triangles, no amplification. (B) Agarose electrophoresis of nested PCR products of *PRKAA1-TTC33*, *SAMM50-PARVB*, *CTBS-GNG5*, and *URB1-C21orf45* in normal tissues.

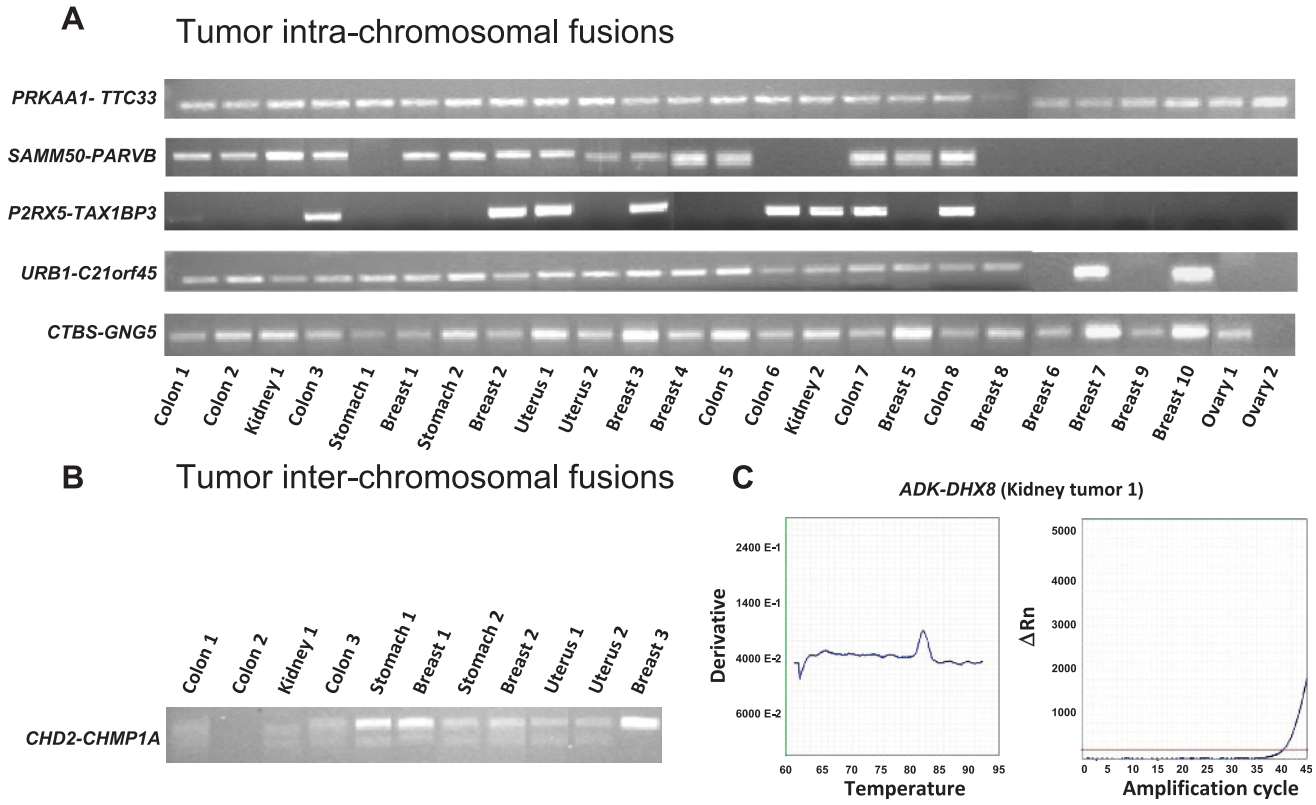


Figure 5. Expression of mRNA chimeras in primary tumors. (A, B) Agarose electrophoresis analysis of amplification products. Tumor origin and sample numbers are indicated. (A) Intrachromosomal chimeras as analyzed by real-time PCR. *PRKAA1-TTC33* and *CTBS-GNG5* were diagnosed in all 25 tumors; *SAMM50-PARVB* chimeric was found in 15 tumors, *P2RX5-TAX1BP3* in 8 tumors, and *URB1-C21orf45* in 21 tumors. (B) PCR amplicons of the interchromosomal *CHD2-CHMP1A* chimera. (C) Interchromosomal *ADK-DHX8* chimeric. Melting temperature and real-time amplification curves.

was caused by down-regulation of *CHD2-CHMP1A* (Figure 3*A*). Parallel growth blockade in MCF-7 cells was observed on shutdown of *PRKAA-TTC33* and *SAMM50-PARVB* (Figures 3*A* and S3B). Monitoring of cell growth inhibition by *PRKAA-TTC33* and *SAMM50-PARVB* siRNA through optical microscopy (Figure 3*D*) and image analysis (Figure S3, B and C) confirmed a dramatic reduction of MCF-7 cell growth. Growth inhibition by *PRKAA-TTC33* and *SAMM50-PARVB* down-regulation was also demonstrated for HBL-100 cells.

We then went on to test *URB1-C21ORF45*–targeting siRNA in ovarian cell lines. Unexpectedly, an increase in cell growth was reproducibly observed in OVCAR-3 (Figure 3*A*) and IGROV-1 cells, which indicates a growth inhibitory role of the *URB1-C21ORF45* chimera. Albeit *URB1-C21ORF45* is expressed by SKOV-3 and HBL-100 cells, the corresponding siRNA had no effects on these cells, suggesting a cell-specific function of these growth inhibitory chimera (Figure 3*A*). These tests were repeated using *CTBS-GNG5*–targeted siRNA. These assays showed that the *CTBS-GNG5* chimera also has a growth inhibitory function in OVCAR-3 and IGROV-1 cells (Figures 3 and S3B). Again, SKOV-3 and HBL-100 cancer cells were insensitive to the inhibitory function of *CTBS-GNG5*, consistent with a differential tuning of chimera-dependent growth-control circuitries in specific cell lines.

Protein-encoding reading frames of the growth regulatory chimeras were analyzed (Table S11). In all cases but one, the downstream partners did not provide in-frame sequences, generating out-of-frame, mostly short chimeric tails. This suggested altered regulation and/or dominant-negative function of a truncated molecule as a mechanism of action of these chimeric products. However, the *CTBS-GNG5* is an in-frame chimera that retains the first 319 amino acids from the

Table 1. Expression of Chimeric mRNA by Tumor Type.

Chimera	Breast* [*n*/10 (%)]	Ovary [*n*/2 (%)]	Stomach [*n*/2 (%)]	Colon [*n*/7 (%)]	Kidney [*n*/2 (%)]	Uterus [*n*/2 (%)]
PRKAA-TTC33	8 (80)	2 (100)	2 (100)	7 (100)	2 (100)	2 (100)
SAMM50-PARVB	5 (50)	–	1 (50)	5 (71)	2 (100)	1 (50)
P2RX5-TAX1BP3	2 (20)	–	–	5 (71)	1 (50)	1 (50)
URB1-C21ORF45	8 (80)	–	1 (100)	7 (100)	2 (100)	2 (100)
CTBS-GNG5	8 (80)	1 (50)	2 (100)	7 (100)	2 (100)	2 (100)
CHD2-CHMP1A	3 (30)	–	2 (100)	2 (28)	1 (50)	2 (100)
ADK-DHX8	–	–	–	1 (14)	1 (50)	–

–, Not detected.
*Tumors; total numbers are below each histotype.

N-terminal chitobiase and most of the C-terminal Gγ5 (41 amino acids), including its Gβ-binding interface (Figure S2). This suggested that the chimeric protein can bind its Gβ partner in trimeric G proteins (Supplemental Sequence Data).

Chimera Expression in Normal Tissues

We assessed the presence and expression levels of the five growth-controlling chimeras in mRNA from normal tissues (breast, lung, placenta, uterus, prostate, stomach, colon, pancreas, and kidney) by nested and real-time PCR. The four intrachromosomal chimeras (*PRKAA-TTC33*, *SAMM50-PARVB*, *URB1-C21ORF45*, and *CTBS-GNG5*) were detected in all screened normal tissues (Figure 4). This was consistent with previous findings on the expression of oncogenic mRNA chimeras in normal tissues [4,13,14,17–20]. However, we found essentially no trace of the *CHD2-CHMP1A* interchromosomal chimera in normal tissues. *CHD2-CHMP1A* was expressed by almost all cancer cell lines (13/14), thus appearing as a cancer-related event.

Expression of Growth Regulatory Chimeras in Primary Tumors

In vitro cell growth regulatory chimeras are expressed by different cancer histotypes. Total RNA was extracted from breast, ovarian, gastric, colon, kidney, and uterine tumors [13,48], was reverse transcribed, and amplified. We took advantage of chimeric-band melting-temperature specificity peaks (Figure S3E) to select for *bona fide* amplification candidates. Amplified candidates were then systematically sequenced. *PRKAA-TTC33* was detected in all 25 of these tumors, *SAMM50-PARVB* in 15 tumors, *P2RX5-TAX1BP3* in 8 tumors, and *URB1-C21ORF45* in 21 tumors; *CTBS-GNG5* was detected in almost all tumors (Figure 5*A* and Table 1); *CHD2-CHMP1A* was identified in 11 tumors (Figure 5*B*). *ADK-DHX8* was diagnosed in two tumors (Figure 5*C*). Hence, growth regulatory chimeras are broadly expressed in human tumors but in heterogeneous manners. This suggests a positive selective pressure [49] for a fusion mRNA-based growth regulatory mechanism during tumor development, which appears to operate in a chimera and tumor-type–specific manner.

Discussion

We have opened the field of the *in silico* identification of mRNA chimeras in cancer cells, through analysis of cDNA sequence databanks [24]. NGS approaches have enormously increased the amount of sequencing data of potential use for chimera discovery. However, short-read second-generation NGS analyses identify mRNA chimeras through a probabilistic fitting of highly multiplexed short-tag data sets [7,19,20,25–28,50–53], which severely affects both specificity and sensitivity of detection of mRNA chimeras. However, rapid progress is being made toward achieving longer sequence reads and higher sequencing accuracy, which allows to reduce sequence errors while improving contig assembly procedures. To permit high-throughput, high-specificity chimera discovery in long-read sequence data sets, we have developed the FusionMiner search strategy. This was shown to reach a 95.9% chimera identification specificity, with a low 4.1% false-negative classification rate. This search strategy was extensively validated by RT-PCR and cDNA sequencing (Table S1b).

Global chimera frequencies were computed for separate sequencing projects. Analysis of a human transcriptomic 454 data set of 19,527 contigs and 173,005 singletons led to the identification of four sequences as *bona fide* chimeras, for a chimera frequency of 4/192,532, i.e., 2×10^{-5} of all mRNA. High-throughput sequencing of cDNA libraries from tumors and corresponding normal tissues generated 1,774,453 long-read sequences. Twenty-five were identified by FusionMiner as *bona fide* chimeras, for a chimera frequency of 25/1,774,453, i.e., 1.4×10^{-5}, in remarkable agreement with the NGS data. Taken together, these findings suggest a chimera frequency of $\approx 2 \times 10^{-5}$ in cellular transcriptomes. Issues of data set size and of transcriptome tissue specificity suggest these to be minimal estimates. A proof of principle of this scenario was obtained, as one of the interchromosomal chimeras, which could not be detected in cell lines, and was identified in 2 of 10 primary breast cancers.

Most of the chimeras analyzed were shown to have a regulatory role in transformed cell growth [54,55]. Notably, tumor growth inhibitory mRNA chimeras, e.g., *URB1-C21ORF45* and *CTBS-GNG5*, were also discovered. Of interest, these were shown to have inhibitory capacity on the growth of a subset of ovarian cancer cells, whereas other ovarian and breast cancer cells were not affected, suggesting different regulatory contexts for chimera-driven growth control in different cell lines. Most tumors were shown to express these growth regulatory chimeras, consistent with a positive selective pressure for exploiting this growth regulatory mechanism during tumor development.

Acknowledgments

We thank M. Iacono for providing the Roche NGS data sets and C. Berrie for critical reading and editing of the manuscript.

References

[1] Mitelman F, Johansson B, and Mertens F (2007). The impact of translocations and gene fusions on cancer causation. *Nat Rev Cancer* **7**, 233–245.

[2] Mercer TR, Dinger ME, and Mattick JS (2009). Long non-coding RNAs: insights into functions. *Nat Rev Genet* **10**, 155–159.

[3] Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, et al. (2005). Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**, 644–648.

[4] Rickman DS, Pflueger D, Moss B, VanDoren VE, Chen CX, de la Taille A, Kuefer R, Tewari AK, Setlur SR, Demichelis F, et al. (2009). SLC45A3-ELK4 is a novel and frequent erythroblast transformation–specific fusion transcript in prostate cancer. *Cancer Res* **69**, 2734–2738.

[5] Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, Fujiwara S, Watanabe H, Kurashina K, Hatanaka H, et al. (2007). Identification of the transforming *EML4-ALK* fusion gene in non-small-cell lung cancer. *Nature* **448**, 561–566.

[6] Fehr A, Roser K, Heidorn K, Hallas C, Loning T, and Bullerdiek J (2008). A new type of MAML2 fusion in mucoepidermoid carcinoma. *Genes Chromosomes Cancer* **47**, 203–206.

[7] Palanisamy N, Ateeq B, Kalyana-Sundaram S, Pflueger D, Ramnarayanan K, Shankar S, Han B, Cao Q, Cao X, Suleman K, et al. (2010). Rearrangements of the RAF kinase pathway in prostate cancer, gastric cancer and melanoma. *Nat Med* **16**, 793–798.

[8] Ciampi R, Knauf JA, Kerler R, Gandhi M, Zhu Z, Nikiforova MN, Rabes HM, Fagin JA, and Nikiforov YE (2005). Oncogenic AKAP9-BRAF fusion is a novel mechanism of MAPK pathway activation in thyroid cancer. *J Clin Invest* **115**, 94–101.

[9] Santoro M, Melillo RM, and Fusco A (2006). RET/PTC activation in papillary thyroid carcinoma: European Journal of Endocrinology Prize Lecture. *Eur J Endocrinol* **155**, 645–653.

[10] Edwards PA (2010). Fusion genes and chromosome translocations in the common epithelial cancers. *J Pathol* **220**, 244–254.

[11] Skotheim RI, Thomassen GO, Eken M, Lind GE, Micci F, Ribeiro FR, Cerveira N, Teixeira MR, Heim S, Rognes T, et al. (2009). A universal assay for detection of oncogenic fusion transcripts by oligo microarray analysis. *Mol Cancer* **8**, 5.

[12] Bruzik JP and Maniatis T (1992). Spliced leader RNAs from lower eukaryotes are trans-spliced in mammalian cells. *Nature* **360**, 692–695.

[13] Guerra E, Trerotola M, Dell'Arciprete R, Bonasera V, Palombo B, El-Sewedy T, Ciccimarra T, Crescenzi C, Lorenzini F, Rossi C, et al. (2008). A bicistronic

CYCLIN D1-TROP2 mRNA chimera demonstrates a novel oncogenic mechanism in human cancer. *Cancer Res* **68**, 8113–8121.
[14] Li H, Wang J, Mor G, and Sklar J (2008). A neoplastic gene fusion mimics trans-splicing of RNAs in normal human cells. *Science* **321**, 1357–1361.
[15] Terrinoni A, Dell'Arciprete R, Fornaro M, Stella M, and Alberti S (2001). Cyclin D1 gene contains a cryptic promoter that is functional in human cancer cells. *Genes Chromosomes Cancer* **31**, 209–220.
[16] Maher CA, Kumar-Sinha C, Cao X, Kalyana-Sundaram S, Han B, Jing X, Sam L, Barrette T, Palanisamy N, and Chinnaiyan AM (2009). Transcriptome sequencing to detect gene fusions in cancer. *Nature* **458**, 97–101.
[17] Communi D, Suarez-Huerta N, Dussossoy D, Savi P, and Boeynaems J-M (2001). Cotranscription and intergenic splicing of human $P2Y_{11}$ and SSF_1 genes. *J Biol Chem* **276**, 16561–16566.
[18] Li H, Wang J, Ma X, and Sklar J (2009). Gene fusions and RNA trans-splicing in normal and neoplastic human cells. *Cell Cycle* **8**, 218–222.
[19] Pflueger D, Terry S, Sboner A, Habegger L, Esgueva R, Lin PC, Svensson MA, Kitabayashi N, Moss BJ, Macdonald TY, et al. (2010). Discovery of non-ETS gene fusions in human prostate cancer using next-generation RNA sequencing. *Genome Res* **21**, 56–67.
[20] Edgren H, Murumagi A, Kangaspeska S, Nicorici D, Hongisto V, Kleivi K, Rye IH, Nyberg S, Wolf M, Borresen-Dale AL, et al. (2011). Identification of fusion genes in breast cancer by paired-end RNA-sequencing. *Genome Biol* **12**, R6.
[21] Ambrogi F, Biganzoli E, Querzoli P, Ferretti S, Boracchi P, Alberti S, Marubini E, and Nenci I (2006). Molecular subtyping of breast cancer from traditional tumor marker profiles using parallel clustering methods. *Clin Cancer Res* **12**, 781–790.
[22] Rabbitts TH and Stocks MR (2003). Chromosomal translocation products engender new intracellular therapeutic technologies. *Nat Med* **9**, 383–386.
[23] Cimoli G, Malacarne D, Ponassi R, Valenti M, Alberti S, and Parodi S (2004). Meta-analysis of the role of p53 status in isogenic systems tested for sensitivity to cytotoxic antineoplastic drugs. *Biochim Biophys Acta* **1705**, 103–120.
[24] Romani A, Guerra M, Trerotola M, and Alberti S (2003). Detection and analysis of spliced chimeric mRNAs in sequence databanks. *Nucleic Acids Res* **31**, 1–8.
[25] Sboner A, Habegger L, Pflueger D, Terry S, Chen DZ, Rozowsky JS, Tewari AK, Kitabayashi N, Moss BJ, Chee MS, et al. (2010). FusionSeq: a modular framework for finding gene fusions by analyzing paired-end RNA-sequencing data. *Genome Biol* **11**, R104.
[26] Nacu S, Yuan W, Kan Z, Bhatt D, Rivers CS, Stinson J, Peters BA, Modrusan Z, Jung K, Seshagiri S, et al. (2011). Deep RNA sequencing analysis of readthrough gene fusions in human prostate adenocarcinoma and reference samples. *BMC Med Genomics* **4**, 11.
[27] Asmann YW, Hossain A, Necela BM, Middha S, Kalari KR, Sun Z, Chai HS, Williamson DW, Radisky D, Schroth GP, et al. (2011). A novel bioinformatics pipeline for identification and characterization of fusion transcripts in breast cancer and normal cell lines. *Nucleic Acids Res* **39**, e100.
[28] Iyer MK, Chinnaiyan AM, and Maher CA (2011). ChimeraScan: a tool for identifying chimeric transcription in sequencing data. *Bioinformatics* **27**, 2903–2904.
[29] Carletti E, Guerra E, and Alberti S (2006). The forgotten variables of DNA array hybridization. *Trends Biotechnol* **24**, 443–448.
[30] Shendure J and Ji H (2008). Next-generation DNA sequencing. *Nat Biotechnol* **26**, 1135–1145.
[31] Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, and Pallen MJ (2012). Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol* **30**, 434–439.
[32] Mardis ER (2011). A decade's perspective on DNA sequencing technology. *Nature* **470**, 198–203.
[33] Orsulic S, Li Y, Soslow RA, Vitale-Crosss LA, Gutkind JS, and Varmus HE (2002). Induction of ovarian cancer by defined multiple genetic changes in a mouse model system. *Cancer Cell* **1**, 53–62.
[34] Alberti S and Fornaro M (1990). Higher transfection efficiency of genomic DNA purified with a guanidinium thiocyanate-based procedure. *Nucleic Acids Res* **18**, 351–353.
[35] Dell'Arciprete R, Stella M, Fornaro M, Ciccocioppo R, Capri MG, Naglieri AM, and Alberti S (1996). High-efficiency expression gene cloning by flow cytometry. *J Histochem Cytochem* **44**, 629–640.
[36] Alberti S, Nutini M, and Herzenberg LA (1994). DNA methylation prevents the amplification of TROP1, a tumor-associated cell surface antigen gene. *Proc Natl Acad Sci USA* **91**, 5833–5837.
[37] Alberti S and Herzenberg LA (1988). DNA methylation prevents transfection of genes for specific surface antigens. *Proc Natl Acad Sci USA* **85**, 8391–8394.
[38] Alberti S, Parks DR, and Herzenberg LA (1987). A single laser method for subtraction of cell autofluorescence in flow cytometry. *Cytometry* **8**, 114–119.
[39] Biganzoli E, Coradini D, Ambrogi F, Garibaldi JM, Lisboa P, Soria D, Green AR, Pedriali M, Piantelli M, Querzoli P, et al. (2011). p53 status identifies two subgroups of triple-negative breast cancers with distinct biological features. *Jpn J Clin Oncol* **41**, 172–179.
[40] Querzoli P, Coradini D, Pedriali M, Boracchi P, Ambrogi F, Raimondi E, La Sorda R, Lattanzio R, Rinaldi R, Lunardi M, et al. (2010). An immunohistochemically positive E-cadherin status is not always predictive for a good prognosis in human breast cancer. *Br J Cancer* **103**, 1835–1839.
[41] Tinari N, Lattanzio R, Natoli C, Cianchetti E, Angelucci D, Ricevuto E, Ficorella C, Marchetti P, Alberti S, Piantelli M, et al. (2006). Changes of topoisomerase IIα expression in breast tumors after neoadjuvant chemotherapy predicts relapse-free survival. *Clin Cancer Res* **12**, 1501–1506.
[42] Brummelkamp TR, Bernards R, and Agami R (2002). A system for stable expression of short interfering RNAs in mammalian cells. *Science* **296**, 550–553.
[43] Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, and Tuschl T (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494–498.
[44] Semizarov D, Frost L, Sarthy A, Kroeger P, Halbert DN, and Fesik SW (2003). Specificity of short interfering RNA determined through gene expression signatures. *Proc Natl Acad Sci USA* **100**, 6347–6352.
[45] Chalk AM, Wahlestedt C, and Sonnhammer EL (2004). Improved and automated prediction of effective siRNA. *Biochem Biophys Res Commun* **319**, 264–274.
[46] Bonasera V, Alberti S, and Sacchetti A (2007). Protocol for high-sensitivity/long linear-range spectrofluorimetric DNA quantification using ethidium bromide. *Biotechniques* **43**, 173–176.
[47] Akiva P, Toporik A, Edelheit S, Peretz Y, Diber A, Shemesh R, Novik A, and Sorek R (2006). Transcription-mediated gene fusion in the human genome. *Genome Res* **16**, 30–36.
[48] Querzoli P, Pedriali M, Rinaldi R, Lombardi AR, Biganzoli E, Boracchi P, Ferretti S, Frasson C, Zanella C, Ghisellini S, et al. (2006). Axillary lymph node nanometastases are prognostic factors for disease-free survival and metastatic relapse in breast cancer patients. *Clin Cancer Res* **12**, 6696–6701.
[49] Alberti S (1997). The origin of the genetic code and protein synthesis. *J Mol Evol* **45**, 352–358.
[50] Ge H, Liu K, Juan T, Fang F, Newman M, and Hoeck W (2011). FusionMap: detecting fusion genes from next-generation sequencing data at base-pair resolution. *Bioinformatics* **27**, 1922–1928.
[51] Hu Y, Wang K, He X, Chiang DY, Prins JF, and Liu J (2010). A probabilistic framework for aligning paired-end RNA-seq data. *Bioinformatics* **26**, 1950–1957.
[52] McPherson A, Hormozdiari F, Zayed A, Giuliany R, Ha G, Sun MG, Griffith M, Heravi Moussavi A, Senz J, Melnyk N, et al. (2011). deFuse: an algorithm for gene fusion discovery in tumor RNA-seq data. *PLoS Comput Biol* **7**, e1001138.
[53] Kinsella M, Harismendy O, Nakano M, Frazer KA, and Bafna V (2011). Sensitive gene fusion detection using ambiguously mapping RNA-Seq read pairs. *Bioinformatics* **27**, 1068–1075.
[54] Guerra E, Trerotola M, Aloisi AL, Tripaldi R, Vacca G, La Sorda R, Lattanzio R, Piantelli M, and Alberti S (2012). The Trop-2 signalling network in cancer growth. *Oncogene.* DOI: 10.1038/onc.2012.151 [E-pub ahead of print].
[55] Trerotola M, Cantanelli P, Guerra E, Tripaldi R, Aloisi AL, Bonasera V, Lattanzio R, de Lange R, Weidle UH, Piantelli M, et al. (2012). Upregulation of Trop-2 quantitatively stimulates human cancer growth. *Oncogene.* DOI: 10.1038/onc.2012.36 [E-pub ahead of print].

SUPPLEMENTAL MATERIAL

Long-range transcriptome sequencing reveals cancer cell growth regulatory chimeric mRNAs

Roberto Plebani, Gavin R. Oliver, Marco Trerotola, Emanuela Guerra, Pamela Cantanelli, Luana Apicella, Andrew Emerson, Alessandro Albiero, Paul D. Harkin, Richard D. Kennedy and Saverio Alberti

Supplemental material includes:

Supplemental Material and Methods

Chimeric mRNA detection procedure. We designed the FusionMiner software workflow (Figure 1) to process BLAST analyses of query sequences against genomic databanks via sequential stages of analysis and exclusion, pass-or-fail tests. Candidate chimera were cross-validated versus experimentally verified sequences and by RT-PCR.

Alignments versus genomic assemblies were first parsed to remove spurious data on the basis of length and percent identity (=98%, over =95% of a candidate length). Individual filtered alignments were then clustered and concatenated across alignment breaks and intronic regions, based on a permissible gap criterion, to identify their genomic context. Sequences aligning to one or two chromosomes were segregated and processed separately, as intra-chromosomal (which most frequently derive from inter-genic transcription) and inter-chromosomal (which most frequently derive from chromosomal translocations and trans-splicing) chimeric candidates, respectively.

The fusion point (FP) for inter-chromosomal candidates was expected to correspond to the point in a clustered alignment where a sequence ceased to align with one chromosome and started to align with the second. Alignments on either side of a FP were assessed on the basis of length and percent identity (defaults =100 bp length, =98% identity), following clustering of the original alignments and weighted averaging of percent identities. This was to ensure that only high-quality alignment data were used in chimeras detection, while low quality or spurious alignments were dropped out. FPs were then examined and filtered, based on the degree of overlap at this position (default =10 bp). Allowing this degree of error ensured retention of true chimeras with small areas of fortuitous sequence homology across the other side of the FP. A parallel filtration procedure was applied for gaps at FP (default =10 bp), which allowed to compensate for possible sequencing errors at this position. Successful candidates were then checked for agreement of their FP with known exon boundaries, using genomic coordinate data from Ensembl, although non-canonical FP (i.e. recombinations within exons) were also identified and separately stored. Ensembl was chosen as genomic coordinate reference site as it contains confirmed gene predictions, which are integrated with external data sources, including the Sanger Institute HAVANA [1], RefSeq at NCBI [2], and the UCSC Genome Browser [3]. When candidate-chimera FPs were found to correspond to an exon-exon boundary, candidates were accepted. Intra-chromosomal candidates were treated in a corresponding manner. A permissible error threshold (default 3 bp) was applied at this stage, to compensate for sequencing errors or alignment blurring due to small areas of local homology on either side of a boundary.

Both chimera partner mRNAs were selected for occurrence in the same reading orientation versus known mRNAs (i.e. plus orientation). Joining to a 'minus' strand is, indeed, most likely generated by cDNA recombination during library construction [4]. A special case is that of a gene that transcribes both the minus and the plus strands. Of note, both mRNA classes would be available for matching in transcript datasets. FusionMiner would then operationally qualify both mRNAs as 'plus'.

Accepted candidates then entered a clustering step, whereby they were assessed and grouped if they shared the same FP. A disagreement of up to 10 bp was allowed during clustering, consistent with the errors permitted at the stage of FP definition. Chimeras were then sorted and presented in order of multiplicity of occurrence.

FusionMiner also allowed to identify and cluster candidates which were rejected during the exon-boundary checking stage. This was to permit the discovery of recurrent, non-canonical chimeras, which are expected to derive from DNA joining at recombination hot spots [5].

Prediction of reading-frame preservation at the FP was then performed.

FusionMiner search performance. The performance of the FusionMiner detection strategy (Figure S1) was assessed by screening the Dana Farber Cancer Institute (DFCI) Gene Index Project tentative human consensus (THC) collection (Figure S1). This led to the identification of 228 chimeras (105 inter-chromosomal and 123 intra-chromosomal), involving 414 genes (Tables S1-S3). Chimeras discovered by FusionMiner in the DFCI Gene Index Project encoded enzymes (16%), transcription factors/ chromatin modulators (11.5%), G proteins (5.8%), protein binding partners (5.8%), transporters (4.5%), cytoskeletal proteins (2.6%), receptors and proteases (1.9%). Curated sequence analysis [4] indicated that 221 of the 228 chimeric mRNAs candidates (96.9%) did fit all *bona fide* chimera criteria. Sixty-one of these chimeras were uniquely identified by FusionMiner (Table S3).

FusionMiner default settings were optimized to obtain maximum specificity in chimera detection. To provide differential estimates of performance (sensitivity versus specificity), FusionMiner analysis parameters were then systematically altered, and their impact on analysis outcomes was assessed.

Splicing at exon/exon borders: exon-exon boundary settings were relaxed, by extending the tolerance up to 8 bp, i.e. exon boundaries were allowed to be identified within 8 bp of BLAST alignments borders. This can be useful for specific requirements, e.g. for short-sequence-length datasets, or for alignments to poor-quality genome sequence regions. Thirty-nine additional sequences were obtained as compared with the default 3 bp tolerance (Tables S1b, S4-S6). Thirty-seven of these appeared to be

bona fide chimeras (96.6% specificity, versus 96.9% with optimal/ default parameters, -0.3% specificity; 228 + 39 = 267; +17.1% sensitivity).

Bp gap: Extending the allowed gaps in FusionMiner to 30 bp, instead of the 10 bp default, resulted in identification of only 4 additional chimeras (Tables S1b, S4-S6). Three of these (75%) were *bona fide* fusion sequences.

Percent query identity (%ID): A low %ID might be due to gaps, bad sequencing, or real mismatch regions. The default requirement for 98% ID was thus relaxed to 94%, to allow detection of these problematic sequences. Thirty-four additional chimeras were detected versus default values (Tables S1b, S4-S6). Thirty two of these (91.4%) were confirmed to be true positive *bona fide* fusion sequences (-0.4% specificity; +14.9% sensitivity).

Query length: The FusionMiner strategy/ sequential validation/ parameter combination was optimized for the recognition of small, *bona fide* chimeric sequences from NGS analysis. We validated this by analyzing chimeras with 50-base matches around a fusion joint. By shortening the minimum allowed length for matches from 100 bp to 50 bp, we identified 59 additional THC sequences versus the default searches (Tables S1b, S4-S6). Strikingly, 56 out of these 59 (94.9%) were true positives (-0.4% specificity; +25.9% sensitivity). Remarkably, 349 of 364 (95.9%) chimeric mRNAs detected by FusionMiner from THC were shown to fit chimera identification criteria.

Validation of discovered THC chimeras. THC chimeras from breast cancer (4 sequences) were searched for in breast cancer cell lines (Figure S1). cDNAs were obtained from MCF-7, MCF-7/Almac, HBL-100, SK-BR-3, MDA-MB-231, MDA-MB-361, MDA-MB-415, MDA-MB-453, and MDA-MB-468 cells, and each chimera was amplified by direct or nested PCR. All PCR amplified bands were verified by sequencing. Successful amplification was achieved for 3 out of 4 chimeras (75%), i.e. *THC2538403 ZNF498-CUX1* (4/9 lines), *THC2523555* (an additional long intergenic transcript was identified as ENST00000358157) *C9orf47-S1PR3* (8/9 lines) and *THC2668182 KLH22-SCARF2* (7/9 lines) (Figure S1). We then extended this analysis to SKOV-3, IGROV-1, OVCAR-3 and OVCA-432 ovarian cancer cells. Two of the chimeras from breast cancer, i.e. *THC2523555 C9orf47-S1PR3* and *THC2668182 KLH22-SCARF2*, were identified in all four ovarian cancer cell lines (Figure S1), suggesting broad expression across different tumor histotypes. Chimeric sequence abundance was measured by real-time quantitative PCR (Figure S1B). Chimeras were detected at considerably different levels in the tested cell lines, consistent with a regulated expression [6-11].

FusionMiner performance on long-reads NGS datasets. FusionMiner was further validated on long-sequence read from 454-generated output files. A 454 Titanium dataset of 1,241,098 reads, 355.5 bp in average length (www.bmr-genomics.it/~alex/ALBERTI/) was compiled by Newbler into 19,527 contigs and 173,005 singletons; 28,561 sequences were identified as outliers. FusionMiner identified one inter-chromosomal and three intra-chromosomal *bona fide* chimeras (Tables S1, S7, Supplemental Sequence Data). These findings reveal absolute chimeras frequencies of $2x10^{-5}$ (4 out of 192,532) in whole cell transcriptomes.submis

Supplemental Sequence Data

Sequences, BLAST alignments, and exons involved in fusions are indicated for the individual novel chimeras identified from specific sequencing datasets. The interaction networks of the proteins encoded by growth-regulatory chimera partners are shown.

Legend:
Borders of exons involved at junction points are indicated before each BLAST alignment.
Arrows indicate beginning/ end of BLAST alignments.
Chromosome location of each partner gene is indicated.
Gene descriptions are provided (Entrez gene, UniProtKB/Swiss-Prot databases).

Chimeras sequences

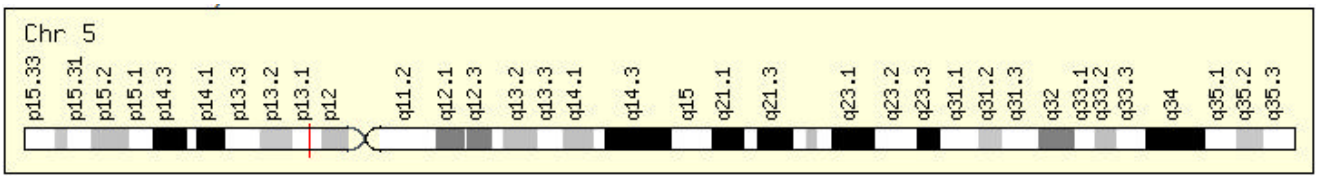
PRKAA1-TTC33

```
TTGGATCAAAGATATCAGGGAACATGAATGGTTTAAACAGGACCTTCCAAAATATCTCTTTCCTGAGGATCCATCATATAGTTCAACCATGATTGATGATGA
GCCTTAAAAGAAGTATGTGAAAAGTTTGAGTGCTCAGAAGAGGAAGTTCTCAGCTGTCTTTACAACAGAAATCACCAGGATCCTTTGGCAGTTGCCTACCAT
TCATAATAGATAACAGGAGAATAATGAATGAAGCCAAAGATTTCTATTTGGCGACAAGCCCACCTGATTCTTTTCTTGATGATCATCACCTGACTCGGCCCC
TCCTGAAAGAGTACCATTCTTGGTTGCTGAAACACCAAGGGCACGCCATACCCTTGATGAATTAAATCCACAGAAATCCAAACACCAAGGTGTAAGGAAAGC
AAATGGCATTTAGGAATTAGAAGTCAAAGTCGACCAAATGATATTATGGCAGAAGTATGTAGAGCAATCAAACAATTGGATTATGAATGGAAGGTTGTAAAC
CATATTATTTGCGTGTACGAAGGAAGAATCCTGTGACAAGCACTTACTCCAAAATGAGTCTACAGTTATACCAAGTGGATAGTAGAACTTATCTACTGGATT
CCGTAGTATTGATGAATGGCTTCCTTTGGGTGGAAGAGGAAAATTGGTGAGAAGGTCTCAAAGGTCACTTCCCAGCAGTTTGAAGCTGAAGCTGCTGATGAG
AGGATGTAGTTGACAACGATGAAAGAACTGGCTTCATGCCATTAAACGTAGAAAGAAATTCTTCTTGAACGCTGTGCTGAGAAAAGTAAACAGCTGAAGATG
AGAGCCAGTTTGGCTGAAAATAAAAGATATCGGGAGGCAATTCAGAAGTGGGATGAAGCACTACAGTTAACTCTAAATGATGCTACCCTATACGAGATGAAA
CACAGTGCTAATGTCTCTTCATGAAATGTTCCCAGCAGTACATGCAGCAGAAATGCCGTCAGCAAAATCCACATTCATGGGAGTCTGCAGACTTGGGACGTG
TCACTTGTTTAGGAGAAATATCTTGCAATTCGAGGTTTCAGTAGCCCTTCAATCTATCCATGACTGAATGGAAGGAGACCTCCTTGGCAGACGCCTTCAGGA
CGCCCGAAAAGT
```

Sanger sequencing (forward and reverse)

```
EMBOSS_001         1 TTTGCGTGTACGAAGGAAGAATCCTGTGACAAGCACTTACTCCAAAATGA     50
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
Pcr band           1 TTTGCGTGTACGAAGGAAGAATCCTGTGACAAGCACTTACTCCAAAATGA     50

EMBOSS_001        51 GTCTACAGTTATACCAAGTGGATAGTAGAACTTATCTACTGGATTTCCGT    100
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
Pcr band          51 GTCTACAGTTATACCAAGTGGATAGTAGAACTTATCTACTGGATTTCCGT    100

EMBOSS_001       101 AGTATTGATGAATGGCTTCCTTTGGGTGGAAGAGGAAAATTGGTGAGAAG    150
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
Pcr band         101 AGTATTGATGAATGGCTTCCTTTGGGTGGAAGAGGAAAATTGGTGAGAAG    150

EMBOSS_001       151 GTCTCAAAGGTCACTTCCCAGC    172
                     |||||||||||||||||||||||
Pcr band         151 GTCTCAAAGGTCACTTCCCAGC    172
```

5' partner: PRKAA1



Junction point
exon=9 1315..1441
/gene="PRKAA1"

/gene_synonym="AMPK; AMPKa1; MGC33776; MGC57364"

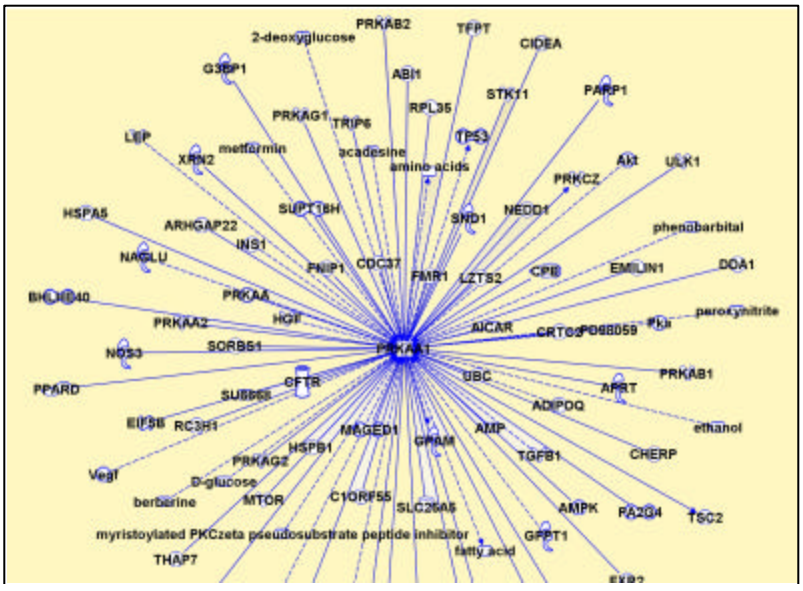
BLAST vs mRNA

>ref|NM_006251.5| UniGene info linked to NM_006251.5GEO profiles info linked to NM_006251.5Gene info linked to NM_006251.5Genome view with mapviewer linked to NM_006251.5 Homo sapiens protein kinase, AMP-activated, alpha 1 catalytic subunit (PRKAA1), transcript variant 1, mRNA
Length=5085
GENE ID: 5562 PRKAA1 | protein kinase, AMP-activated, alpha 1 catalytic subunit [Homo sapiens]
Score = 1162 bits (629), Expect = 0.0
Identities = 629/629 (100%), Gaps = 0/629 (0%)

Strand=Plus/Plus

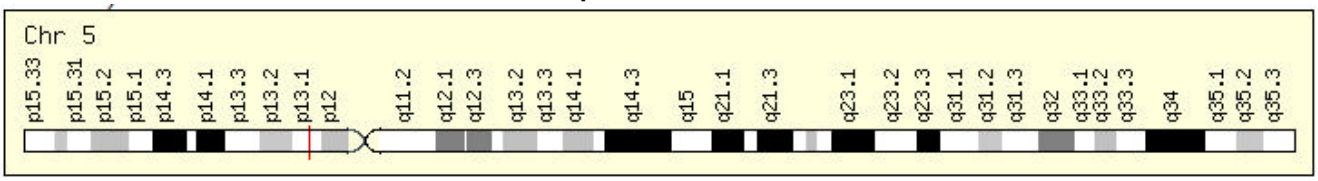
```
Query  5     ATCAAAGATATCAGGGAACATGAATGGTTTAAACAGGACCTTCCAAAATATCTCTTTCCT  64
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  814   ATCAAAGATATCAGGGAACATGAATGGTTTAAACAGGACCTTCCAAAATATCTCTTTCCT  873

Query  65    GAGGATCCATCATATAGTTCAACCATGATTGATGATGAAGCCTTAAAAGAAGTATGTGAA  124
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  874   GAGGATCCATCATATAGTTCAACCATGATTGATGATGAAGCCTTAAAAGAAGTATGTGAA  933

Query  125   AAGTTTGAGTGCTCAGAAGAGGAAGTTCTCAGCTGTCTTTACAACAGAAATCACCAGGAT  184
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  934   AAGTTTGAGTGCTCAGAAGAGGAAGTTCTCAGCTGTCTTTACAACAGAAATCACCAGGAT  993

Query  185   CCTTTGGCAGTTGCCTACCATCTCATAATAGATAACAGGAGAATAATGAATGAAGCCAAA  244
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  994   CCTTTGGCAGTTGCCTACCATCTCATAATAGATAACAGGAGAATAATGAATGAAGCCAAA  1053

Query  245   GATTTCTATTTGGCGACAAGCCCACCTGATTCTTTTCTTGATGATCATCACCTGACTCGG  304
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1054  GATTTCTATTTGGCGACAAGCCCACCTGATTCTTTTCTTGATGATCATCACCTGACTCGG  1113

Query  305   CCCCATCCTGAAAGAGTACCATTCTTGGTTGCTGAAACACCAAGGGCACGCCATACCCTT  364
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1114  CCCCATCCTGAAAGAGTACCATTCTTGGTTGCTGAAACACCAAGGGCACGCCATACCCTT  1173

Query  365   GATGAATTAAATCCACAGAAATCCAAACACCAAGGTGTAAGGAAAGCAAAATGGCATTTA  424
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1174  GATGAATTAAATCCACAGAAATCCAAACACCAAGGTGTAAGGAAAGCAAAATGGCATTTA  1233

Query  425   GGAATTAGAAGTCAAAGTCGACCAAATGATATTATGGCAGAAGTATGTAGAGCAATCAAA  484
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1234  GGAATTAGAAGTCAAAGTCGACCAAATGATATTATGGCAGAAGTATGTAGAGCAATCAAA  1293

Query  485   CAATTGGATTATGAATGGAAGGTTGTAAACCCATATTATTTGCGTGTACGAAGGAAGAAT  544
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1294  CAATTGGATTATGAATGGAAGGTTGTAAACCCATATTATTTGCGTGTACGAAGGAAGAAT  1353

Query  545   CCTGTGACAAGCACTTACTCCAAAATGAGTCTACAGTTATACCAAGTGGATAGTAGAACT  604
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1354  CCTGTGACAAGCACTTACTCCAAAATGAGTCTACAGTTATACCAAGTGGATAGTAGAACT  1413

Query  605   TATCTACTGGATTTCCGTAGTATTGATGA  633
             |||||||||||||||||||||||||||||
Sbjct  1414  TATCTACTGGATTTCCGTAGTATTGATGA  1442
```

PRKAA1 interaction network.
Protein kinase, AMP-activated, alpha 1 catalytic sub-unit (AMPK1). Major interactions are with AMP, p53, AKT, mTOR, TGFB1 and cell cycle regulatory proteins.



Summary for PRKAA1
PKA belongs to Ser/Thr protein kinases. It is the catalytic subunit of the cAMP-activated protein kinase (AMPK). AMPK is a cellular energy sensor conserved in all eukaryotic cells. The kinase activity of AMPK is activated by the stimuli that increase cell AMP/ATP ratio. AMPK regulates key metabolic enzymes through phosphorylation. It protects cells from stress that causes ATP depletion by switching-off ATP-consuming biosynthetic pathways. PKA regulates fatty acid synthesis by phosphorylation of acetyl-CoA carboxylase. It also regulates cholesterol synthesis via phosphorylation and inactivation of hormone-sensitive lipase and hydroxymethylglutaryl-CoA reductase.

3’ partner: TTC33



Junction point
exon=2 148..369
/gene="TTC33"
/gene_synonym="OSRF"

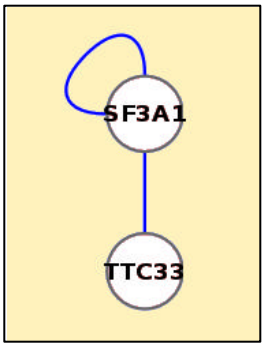
BLAST vs mRNA

>ref|NM_012382.2| UniGene info linked to NM_012382.2GEO profiles info linked to NM_012382.2Gene info linked to NM_012382.2Genome view with mapviewer linked to NM_012382.2 Homo sapiens tetratricopeptide repeat domain 33 (TTC33), mRNA
Length=5519
GENE ID: 23548 TTC33 | tetratricopeptide repeat domain 33 [Homo sapiens]
Score = 741 bits (401), Expect = 0.0
Identities = 492/530 (93%), Gaps = 30/530 (5%)

Strand=Plus/Plus

```
Query  632   GAATGGCTTCCTTTGGGTGGAAGAGGAAAATTGGTGAGAAGGTCTCAAAGGTCACTTCCC  691
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  147   GAATGGCTTCCTTTGGGTGGAAGAGGAAAATTGGTGAGAAGGTCTCAAAGGTCACTTCCC  206

Query  692   AGCAGTTTGAAGCTGAAGCTGCTGATGAGAAGGATGTAGTTGACAACGATGAA-AGAACT  750
             ||||||||||||||||||||||||||||||||||||||||||||||||||||  |||||
Sbjct  207   AGCAGTTTGAAGCTGAAGCTGCTGATGAGAAGGATGTAGTTGACAACGATGAAGGGAACT  266

Query  751   GGCTTCATGCCATTAAACGTA-GAAAGAAATTCTTCTTGAACGCTGTGCTGAGAAAAGTA  809
             |||||||||||||||||||| |||||||||||||||||| ||||||||||||||||||||
Sbjct  267   GGCTTCATGCCATTAAACGTAGGAAAGAAATTCTTCTTGAAGGCTGTGCTGAGAAAAGTA  326

Query  810   AACAGCTGAA-GATGAA-GAGCCAGTTTGGCTGAAAATAAAAGATATCGGGAGGCAATTC  867
             |||||||||| |||||| ||||||||||||||||||||||||||||||||||||||||||
Sbjct  327   AACAGCTGAAGGATGAAGGAGCCAGTTTGGCTGAAAATAAAAGATATCGGGAGGCAATTC  386

Query  868   AGAAGTGGGATGAAGCACTACAGTTAACTCTAAATGATGCTACCCTATACGAGATGAAAT  927
             ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Sbjct  387   AGAAGTGGGATGAAGCACTACAGTTAACTCCAAATGATGCTACCCTATACGAGATGAAAT  446

Query  928   CACAG-TGCTAATGTCTCTTCATGAAATGTTCCCAGCAGTACATGCAGCAGAAATG-CCG  985
             ||||| |||||||||||||||||||||||||||||||||||||||||||||||| ||| |
Sbjct  447   CACAGGTGCTAATGTCTCTTCATGAAATGTTCCCAGCAGTACATGCAGCAGAAATGGCCG  506
```

TTC33 interaction network.



Summary for TTC33
Tetratricopeptide repeat protein 33. Size: 262 amino acids; 29411 Da Secondary accessions: B2R6G0 O95105.

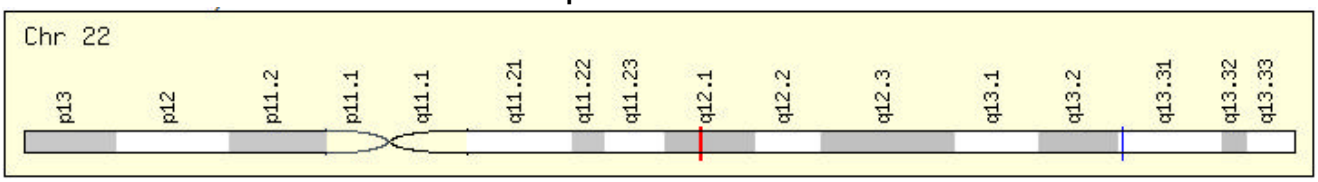
SAMM50-PARVB

TTGGGAAAGCGGACATTCACTGAAATCATCTCTTTCGCACGCCATGGTCATCGATTCTCGGAATTCTTCCATCTTACCAAGGAGAGGTGCTTTGCTGAAAGT
AACCAGGAACTGGCAGGCTACACTGGCGGGGATGTGAGCTTCATCAAAGAAGATTTTGAACTTCAGTTGAACAAGCAACTCATATTTGATTCAGTTTTTTCA
CGTCTTTCTGGGGCGGAATGTTGGTACCCATTGGTGATAAGCCGTCAAGCATTGCTGATAGGTTTTACCTCGGGGGACCCACAAGCGTCCGCGGATTCAGCA
GCACAGCATCGGGCCACAGAGCGAAGGAGACTACCTAGGTGGAGAAGCGTACTGGGCCGGCGGCCTGCACCTCTACACCCCATTACCTTTCCGGCCAGGCCA
GGTGGCTTTGGAGAACTTTTCCGAACACACTTCTTTCTCAACGCAGGAAACCTCTGCAACCTCAACTATGGGGAGGGCCCCAAAGCTCATATTCGTAAGCTG
CTGAGTGCATCCGCTGGTCGTACGGGGCCGGGATTGTCCTCAGGCTTGGCAACATCGCTCGGTTGGAACTTAATTACTGCGTCCCCATGGGAGTACAGACAG
TGACAGTGAGTGACCTGCAGGAAGAAGGCAAGAATGCCATCAACTCACCGATGTCCCCCGCCCTGGCGGATGTTCACCCTGAAGACACCCAGCTCGAGGAGA
CGAGGAGCGCACGATGATTGACCCCACTTCCACGAAGACCCCAAGTTCAAGGAACTGGTCAAGGTCCTCCTCGACTGGATTAATGACGTGCTGGTGGAGGAG
GGATCATTGTGAAGCAGCTGGAGGAAGACCTGTATGACGGCCAGGTGCTGCAGAATCTCTTGGAAAAACTGGCAAGGTGCAAGCTGAATGTGGCTGAGTGAC
CAGTTCGGAAATAGGCAGAAACAAAGCTGCTGACCGGTGCTGGAGCAGTACATGACCTGCTGCGCCCTAAGCTGGGGCTCCGTGGAGCGTGAATCAATTCCG
GAAGACTGTGCATCCTCCACTGCTGTTTCTCTGCATGACTCAGACCCCATCCGCCTCTGAACATGTACGGGCAGGTGGGGCCGGAACGGAAGCCTTGCTATA
GCACTCCGGAGAAGTGACCTACATCGAGTATATAGC

Sanger sequencing (forward and reverse)

```
RDBC0852_G11.      2 TCGCTCGGTTGGAACTTAATTACTGCGTCCCCATGGGAGTACAGACAGGT     51
                     |||||||||||||||||||||||||||||||||||||||||||||||||||.
FUSintra3-4R2      2 TCGCTCGGTTGGAACTTAATTACTGCGTCCCCATGGGAGTACAGACAGGC     51

RDBC0852_G11.    620 GACAGTGAGTGACCTGCAGGAAGAAGGCAAGAATGCCATCAACTCACCG     101
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
FUSintra3-4R2     52 GACAGTGAGTGACCTGCAGGAAGAAGGCAAGAATGCCATCAACTCACCG     101
```

5’ partner: SAMM50



Junction point
exon=14 1420..1561
/gene="SAMM50"
/gene_synonym="CGI-51; FLJ35825; FLJ42905; FLJ99036;
OMP85; SAM50; TOB55; TRG-3; YNL026W"

BLAST vs mRNA

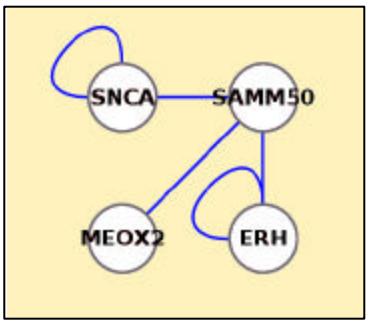
>ref|NM_015380.4| UniGene info linked to NM_015380.4GEO profiles info linked to NM_015380.4Gene info linked to NM_015380.4Genome view with mapviewer linked to NM_015380.4 Homo sapiens sorting and assembly machinery component 50 homolog (S. cerevisiae) (SAMM50), mRNA
Length=1773
GENE ID: 25813 SAMM50 | sorting and assembly machinery component 50 homolog (S. cerevisiae) [Homo sapiens]
Score = 1129 bits (611), Expect = 0.0
Identities = 617/620 (99%), Gaps = 0/620 (0%)

Strand=Plus/Plus

```
Query  5     GAAAGCGGACATTCACTGAAATCATCTCTTTCGCACGCCATGGTCATCGATTCTCGGAAT  64
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  942   GAAAGCGGACATTCACTGAAATCATCTCTTTCGCACGCCATGGTCATCGATTCTCGGAAT  1001

Query  65    TCTTCCATCTTACCAAGGAGAGGTGCTTTGCTGAAAGTTAACCAGGAACTGGCAGGCTAC  124
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1002  TCTTCCATCTTACCAAGGAGAGGTGCTTTGCTGAAAGTTAACCAGGAACTGGCAGGCTAC  1061

Query  125   ACTGGCGGGGATGTGAGCTTCATCAAAGAAGATTTTGAACTTCAGTTGAACAAGCAACTC  184
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1062  ACTGGCGGGGATGTGAGCTTCATCAAAGAAGATTTTGAACTTCAGTTGAACAAGCAACTC  1121

Query  185   ATATTTGATTCAGTTTTTTCAGCGTCTTTCTGGGGCGGAATGTTGGTACCCATTGGTGAT  244
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1122  ATATTTGATTCAGTTTTTTCAGCGTCTTTCTGGGGCGGAATGTTGGTACCCATTGGTGAT  1181

Query  245   AAGCCGTCAAGCATTGCTGATAGGTTTTACCTCGGGGGACCCACAAGCGTCCGCGGATTC  304
             |||||||||||||||||||||||||||||||| ||||||||||||||| ||||||||||||
Sbjct  1182  AAGCCGTCAAGCATTGCTGATAGGTTTTACCTTGGGGGACCCACAAGCATCCGCGGATTC  1241
```

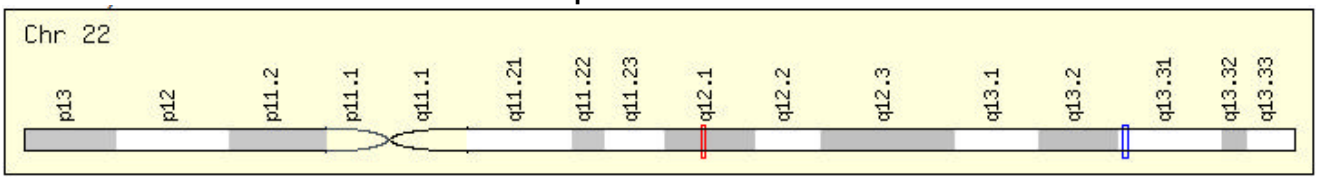
```
Query  305   AGCATGCACAGCATCGGGCCACAGAGCGAAGGAGACTACCTAGGTGGAGAAGCGTACTGG  364
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1242  AGCATGCACAGCATCGGGCCACAGAGCGAAGGAGACTACCTAGGTGGAGAAGCGTACTGG  1301

Query  365   GCCGGCGGCCTGCACCTCTACACCCCATTACCTTTCCGGCCAGGCCAGGGTGGCTTTGGA  424
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1302  GCCGGCGGCCTGCACCTCTACACCCCATTACCTTTCCGGCCAGGCCAGGGTGGCTTTGGA  1361

Query  425   GAACTTTTCCGAACACACTTCTTTCTCAACGCAGGAAACCTCTGCAACCTCAACTATGGG  484
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1362  GAACTTTTCCGAACACACTTCTTTCTCAACGCAGGAAACCTCTGCAACCTCAACTATGGG  1421

Query  485   GAGGGCCCCAAAGCTCATATTCGTAAGCTGGCTGAGTGCATCCGCTGGTCGTACGGGGCC  544
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1422  GAGGGCCCCAAAGCTCATATTCGTAAGCTGGCTGAGTGCATCCGCTGGTCGTACGGGGCC  1481

Query  545   GGGATTGTCCTCAGGCTTGGCAACATCGCTCGGTTGGAACTTAATTACTGCGTCCCCATG  604
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1482  GGGATTGTCCTCAGGCTTGGCAACATCGCTCGGTTGGAACTTAATTACTGCGTCCCCATG  1541

Query  605   GGAGTACAGACAGGTGACAG  624
              |||||||||||||| |||||
Sbjct  1542  GGAGTACAGACAGGCGACAG  1561
```

SAMM50 interaction network.



Summary for SAMM50
SAMM50 is a component of the sorting and assembly machinery (SAM) complex of the outer mitochondrial membrane. The SAM complex has a role in integrating beta-barrel proteins into the outer mitochondrial membrane.

3' partner: PARVB



Junction point
exon=4 260..349
/gene="PARVB"
/gene_synonym="CGI-56"

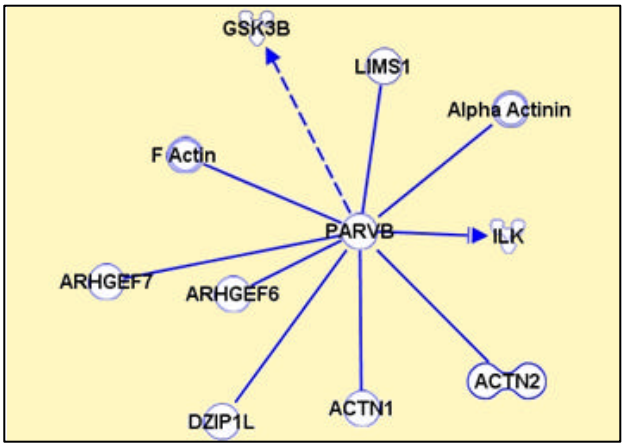
BLAST vs mRNA

```
>ref|NM_001003828.1| UniGene info linked to NM_001003828.1GEO profiles info linked to
NM_001003828.1Gene info linked to NM_001003828.1Genome view with mapviewer linked to NM_001003828.1
Homo sapiens parvin, beta (PARVB), transcript variant 1, mRNA
Length=1808
GENE ID: 29780 PARVB | parvin, beta [Homo sapiens]
Score =  684 bits (370),  Expect = 0.0
Identities = 471/514 (92%), Gaps = 30/514 (5%)

Strand=Plus/Plus
```

```
Query  623   AGTGAGTGACCTGCAGGAAGAAGGCAAGAATGCCATCAACTCACCGATGTCCCCCGCCCT  682
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  258   AGTGAGTGACCTGCAGGAAGAAGGCAAGAATGCCATCAACTCACCGATGTCCCCCGCCCT  317

Query  683   GGCGGATGTTCACCCTGAAGACACCCAGCTCGAGGAGAACGAGGAGCGCACGATGATTGA  742
              || ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Sbjct  318   GGTGGATGTTCACCCTGAAGACACCCAGCTTGAGGAGAACGAGGAGCGCACGATGATTGA  377
```

```
Query  743   CCCCACTTCC-ACGAAGACCCCAAGTTCAAGGAACTGGTCAAGGTCCTCCTCGACTGGAT  801
             |||||||||| | ||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  378   CCCCACTTCCAAGGAAGACCCCAAGTTCAAGGAACTGGTCAAGGTCCTCCTCGACTGGAT  437

Query  802   TAATGACGTGCTGGTGGAGGAGAGGATCATTGTGAAGCAGCTGGAGGAAGACCTGTATGA  861
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  438   TAATGACGTGCTGGTGGAGGAGAGGATCATTGTGAAGCAGCTGGAGGAAGACCTGTATGA  497

Query  862   CGGCCAGGTGCTGCAGAATCTCTTGGAAAAACTGGCAAGGTGCAAGCTGAATGTGGCTGA  921
             |||||||||||||||| |||||||||||||||| ||||||||||||||||||||||||||
Sbjct  498   CGGCCAGGTGCTGCAGAAGCTCTTGGAAAAACTGGCAGGGTGCAAGCTGAATGTGGCTGA  557
```

PARVB interaction network.
Beta parvin is involved in integrin linked kinase signaling. It directly interacts with ILK, ACTIN, ALPHA ACTININ and other proteins.



Summary for PARVB
Members of the parvin family, including PARVB, are actin-binding proteins associated with focal contacts. It probably has a role in the regulation of cell adhesion and cytoskeleton organization.

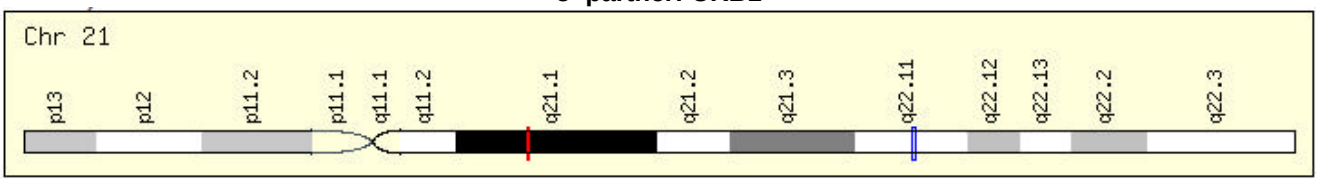
URB1-C21orf45

ATAGCCGGCTCTGTGGGGCTGAGGGGCTGGCAGGGCCTGTGCAGGAGGTGGCCTGCCTGTTCAATACGGTCATGCTGCAGCTGGTGGCTGCCCAGGGCCGGG
AGGGAGCCCTTTCCACCCGGCCATGGAAGCCCTCTCCCTGTCTTCTCTGAGTGAGAAGGATGAAGCCACACAAGGTGTTTCCTGTAATGTTTCTGTGGATAA
GAACAGAAGCTATCCAAACGTGAAAAGGAAAATGGTTGCGTCCTTGAGACTTTGTGCTGCGCGGGGTGCTCACTCAATCTTGGCTACGTGTACAGATGCACG
CCAAGAATCTTGATTACAAGAGAGACTTGTTTTGCCTCAGTGTTGAAGCCATTGAAAGTTATGTTTTAGGGTCCTCTGAAAAGCAAATTGTGTCAGAAGATA
AGAGCTTTTTAATCTTGAAAGCAGAGTTGAAATAGAAAAGTTTCTAACACAGATGGAAGATGTCTTGAAAGCATTACAAATGAAGCTGTGGGAGGCCGAATC
AAATTGTCCTTTCCCACTTGTAAAAGCTGAACTCTAGTCTGTGTCCTCCATTCTGCCCCCGCCCTTCCTCCCCTTATTTGTTAAATGAAGCAACATAGTGAG
CGTCGTCTCTAC

Sanger sequencing (forward and reverse)

```
EMBOSS_001         1 GCCCTCTCCCTGTCTTCTCTGAGTGAGAAGGATGAAGCCACACAAGGTGT     50
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
PCR band           1 GCCCTCTCCCTGTCTTCTCTGAGTGAGAAGGATGAAGCCACACAAGGTGT     50

EMBOSS_001        51 TTCCTGTAATGTTTCTGTGGATAAGGAACAGAAGCTATCCAAACGTGAAA    100
                     ||||||| ||||||||||||||||||||||||||||||||||||||||||||
PCR band          21 TTCCTGT-ATGTTTCTGTGGATAAGGAACAGAAGCTATCCAAACGTGAAA    100

EMBOSS_001       101 AGGAAAATGGTTGCGTCCTTGAG-    123
                     |||||||||||||||||||||||
PCR band         101 AGGAAAATGGTTGCGTCCTTGAGA    123
```

5’ partner: URB1



Junction point
exon=38 6128..6727
/gene="URB1"
/gene_synonym="C21orf108; KIAA0539; NPA1"

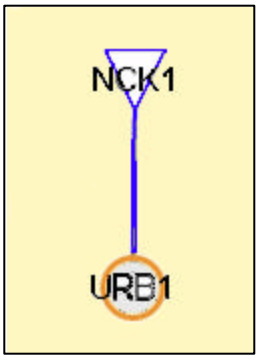
BLAST vs mRNA

>ref|NM_014825.2| UniGene info linked to NM_014825.2GEO profiles info linked to NM_014825.2Gene info linked to NM_014825.2Genome view with mapviewer linked to NM_014825.2Download subject sequence NM_014825 spanning the HSP Homo sapiens URB1 ribosome biogenesis 1 homolog (S. cerevisiae) (URB1), mRNA
Length=10808
GENE ID: 9875 URB1 | URB1 ribosome biogenesis 1 homolog (S. cerevisiae) [Homo sapiens]
Score = 322 bits (174), Expect = 7e-87
Identities = 176/177 (99%), Gaps = 0/177 (0%)

Strand=Plus/Plus

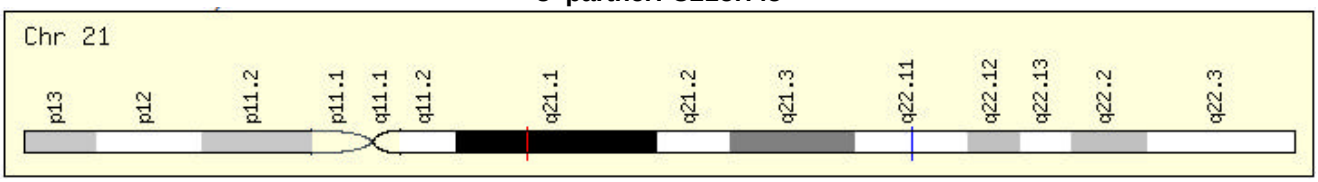
```
Query  1     ATAGCCGGCTCTGTGGGGCTGAGGGGCTGGCAGGGCCTGTGCAGGAGGTGGCCTGCCTGT  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6551  ATAGCCGGCTCTGTGGGGCTGAGGGGCTGGCAGGGCCTGTGCAGGAGGTGGCCTGCCTGT  6610

Query  61    TCAATACGGTCATGCTGCAGCTGGTGGCTGCCCAGGGCCGGGCAGGGAGCCCTTTCCACC  120
             ||||||||||||||||||||||||||||||||||||||||||||| |||||||
Sbjct  6611  TCAATACGGTCATGCTGCAGCTGGTGGCTGCCCAGGGCCGGGCAGGGAGCCCCTTCCACC  6670

Query  121   CGGCCATGGAAGCCCTCTCCCTGTCTTCTCTGAGTGAGAAGGATGAAGCCACACAAG  177
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6671  CGGCCATGGAAGCCCTCTCCCTGTCTTCTCTGAGTGAGAAGGATGAAGCCACACAAG  6727
```

URB1 interaction network.



Summary for URB1
Nucleolar pre-ribosomal-associated protein 1. Subcellular location: Nucleus, nucleolus. Secondary accessions: Q96NX1 Q9NYQ1.

3' partner: C21orf45



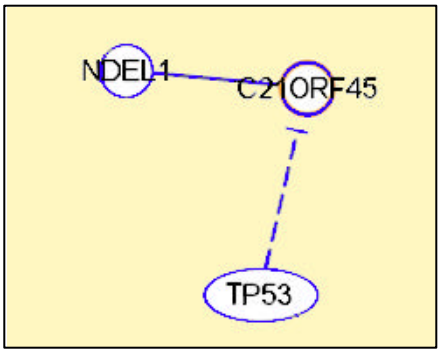
Junction point
exon=2 386..452
/gene="MIS18A"
/gene_synonym="B28; C21orf45; C21orf46; FASP1; hMis18alpha; MIS18alpha"

BLAST vs mRNA

>ref|NM_018944.2| UniGene info linked to NM_018944.2GEO profiles info linked to NM_018944.2Gene info linked to NM_018944.2Genome view with mapviewer linked to NM_018944.2 Homo sapiens MIS18 kinetochore protein homolog A (S. pombe) (MIS18A), mRNA
Length=1587
GENE ID: 54069 MIS18A | MIS18 kinetochore protein homolog A (S. pombe) [Homo sapiens]
Score = 826 bits (447), Expect = 0.0
Identities = 451/453 (99%), Gaps = 0/453 (0%)

```
Strand=Plus/Plus

Query  178  GTGTTTCCTGTAATGTTTCTGTGGATAAGGAACAGAAGCTATCCAAACGTGAAAAGGAAA  237
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  386  GTGTTTCCTGTAATGTTTCTGTGGATAAGGAACAGAAGCTATCCAAACGTGAAAAGGAAA  445

Query  238  ATGGTTGCGTCCTTGAGACTTTGTGCTGCGCGGGGTGCTCACTCAATCTTGGCTACGTGT  297
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  446  ATGGTTGCGTCCTTGAGACTTTGTGCTGCGCGGGGTGCTCACTCAATCTTGGCTACGTGT  505

Query  298  ACAGATGCACGCCCAAGAATCTTGATTACAAGAGAGACTTGTTTTGCCTCAGTGTTGAAG  357
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  506  ACAGATGCACGCCCAAGAATCTTGATTACAAGAGAGACTTGTTTTGCCTCAGTGTTGAAG  565

Query  358  CCATTGAAAGTTATGTTTTAGGGTCCTCTGAAAAGCAAATTGTGTCAGAAGATAAAGAGC  417
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  566  CCATTGAAAGTTATGTTTTAGGGTCCTCTGAAAAGCAAATTGTGTCAGAAGATAAAGAGC  625

Query  418  TTTTTAATCTTGAAAGCAGAGTTGAAATAGAAAAGTTTCTAACACAGATGGAAGATGTCT  477
            |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Sbjct  626  TTTTTAATCTTGAAAGCAGAGTTGAAATAGAAAAGTCTCTAACACAGATGGAAGATGTCT  685

Query  478  TGAAAGCATTACAAATGAAGCTGTGGGAGGCCGAATCCAAATTGTCCTTTCCCACTTGTA  537
            |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Sbjct  686  TGAAAGCATTACAAATGAAGCTGTGGGAGGCCGAATCCAAATTGTCCTTTGCCACTTGTA  745

Query  538  AAAGCTGAACTCTAGTCTGTGTCCTCCATTCTGCCCCCGCCCTTCCTCCCCTTATTTGTT  597
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  746  AAAGCTGAACTCTAGTCTGTGTCCTCCATTCTGCCCCCGCCCTTCCTCCCCTTATTTGTT  805

Query  598  AAATGAAGCAACATAGTGAGACGTCGTCTCTAC  630
            |||||||||||||||||||||||||||||||||
Sbjct  806  AAATGAAGCAACATAGTGAGACGTCGTCTCTAC  838
```

C21ORF45 interaction network.
It indirectly interacts with p53.



Summary for C21ORF45
MS18A_HUMAN, Q9NYP9. Protein Mis18-alpha. Subunit: Homodimer, and heterodimer with MIS18B. Identified in a complex containing MIS18A, MIS18B, MIS18BP1, RBBP7 and RBBP4. Subcellular location: nucleus. It associates with centromeres in interphase cells, from late anaphase to G1 phase. It is not detected on centromeres during earlier phases of mitosis and is associated with chromatin.

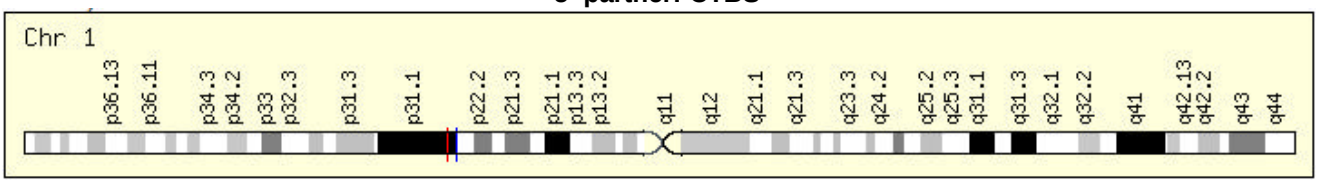
CTBS-GNG5

TGAATCTGTCTGAGGATCATGTTTGTACCATTGCAAAAGTCCCTTTCCGGGGGGCTCCTTGTAGTGACGCTGCAGGACGTCAGGTGCCCTACAAAACGATCA
GAAGCAAATAAATAGTTCTATTTCTGGAAACCTATGGGATAAAGATCAGCGGGCTCCTTATTATAACTATAAAGTTTCCCAGGCAGCTGCAGACTTGAAACAC
TTCTGTCTGCAGAATGCTCAACATGACCCTCTGCTGACTGGAGTATCTTCAAGTACAAATCCCTTCAGACCCCAGAAAGTCTGTTCCTTTTTGTAGTAAAATC
AATCTTTCAAAGGTTTCCCAAACCACTCCTTATGATCCAGTGAATATTCAAGAGAGCTACATTTGAAGCCTGTACAAAAGCTTATCCCTGTAACACATGTGC
ATAATATACAAACTTTTACTTTCGTCAGTCCTTAACATCTACCTCTCTGAATTTTCATGAATTTCTATTTCACAAGGGTAATTGTTTTATATACACTGGCAG
AGCATACAATAAAACTTAGTATGAAACTTT

Sanger sequencing (forward and reverse)

```
EMBOSS_001         1 -CCTACAAAACGATCATGAAGCAAATAAATAGTTCTATTTCTGGAAACCT     49
                       ||||||||||||||||||||||||||||||||||||||||||||||||||||
PCR band           1 TCCTACAAAACGATCATGAAGCAAATAAATAGTTCTATTTCTGGAAACCT     50

EMBOSS_001        50 ATGGGATAAAGATCAGCGGGCTCCTTATTATAACTATAAAGTTTCCCAGG     99
                       |||||||||||||||||||||||||||||||||||||||||||||||||||
PCR band          51 ATGGGATAAAGATCAGCGGGCTCCTTATTATAACTATAAAGTTTCCCAGG    100

EMBOSS_001       100 CAGCTGCAGACTTGAAACAGTTCTGTCTGCAGAATGCTCAACATGACC     147
                       |||||||||||||||||||||||||||||||||||||||||||||||||||
PCR band         101 CAGCTGCAGACTTGAAACAGTTCTGTCTGCAGAATGCTCAACATGACC     148
```

5' partner: CTBS



Junction point
exon=6 861..1022
/gene="CTBS"
/gene_synonym="CTB"

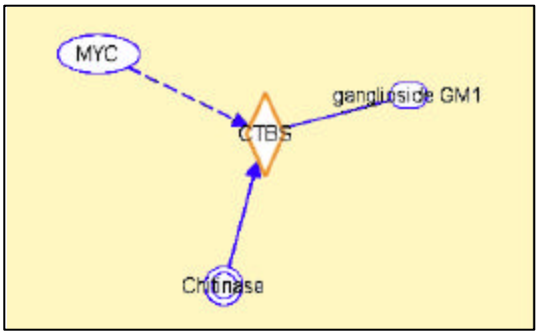
BLAST vs mRNA

>ref|NM_004388.2| UniGene info linked to NM_004388.2GEO profiles info linked to NM_004388.2Gene info linked to NM_004388.2Genome view with mapviewer linked to NM_004388.2 Homo sapiens chitobiase, di-N-acetyl- (CTBS), mRNA
Length=3152
GENE ID: 1486 CTBS | chitobiase, di-N-acetyl- [Homo sapiens]
Score = 327 bits (177), Expect = 1e-88
Identities = 177/177 (100%), Gaps = 0/177 (0%)

Strand=Plus/Plus

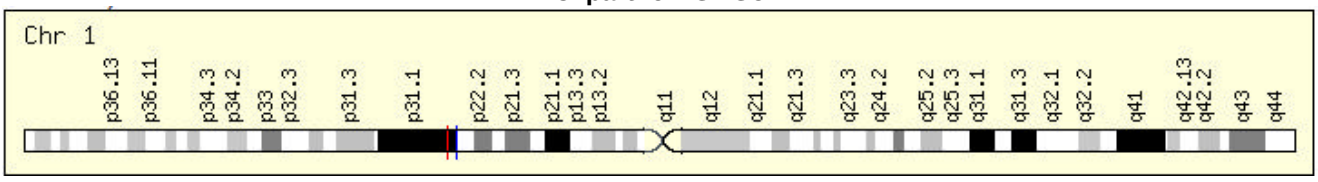
```
Query  1     TGAATCTGTCTGAGGATCATGTTTGTACCATTGCAAAAGTCCCTTTCCGGGGGGCTCCTT  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  847   TGAATCTGTCTGAGGATCATGTTTGTACCATTGCAAAAGTCCCTTTCCGGGGGGCTCCTT  906

Query  61    GTAGTGACGCTGCAGGACGTCAGGTGCCCTACAAAACGATCATGAAGCAAATAAATAGTT  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  907   GTAGTGACGCTGCAGGACGTCAGGTGCCCTACAAAACGATCATGAAGCAAATAAATAGTT  966

Query  121   CTATTTCTGGAAACCTATGGGATAAAGATCAGCGGGCTCCTTATTATAACTATAAAG  177
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  967   CTATTTCTGGAAACCTATGGGATAAAGATCAGCGGGCTCCTTATTATAACTATAAAG  1023
```

CTBS interaction network.



Summary for CTBS
Chitobiase is a lysosomal glycosidase involved in degradation of asparagine-linked oligosaccharides on glycoproteins. It is also involved on hydrolyzation of N-acetyl-beta-D-glucosamine (1-4)N-acetylglucosamine chitobiose from the reducing end of the bond. This requires prior cleavage by glycosyl-asparaginase.

3' partner: GNG5



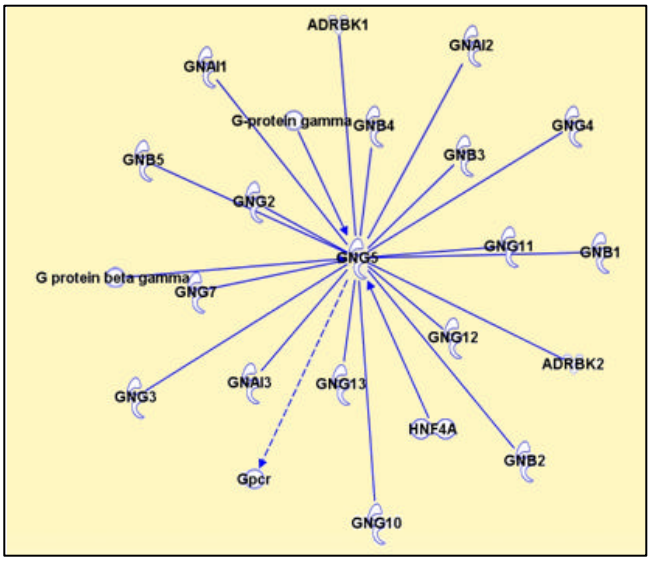
Junction point
exon=3 436..580
/gene="GNG5"
/gene_synonym="FLJ92393"

BLAST vs mRNA

>ref|NM_005274.2| UniGene info linked to NM_005274.2GEO profiles info linked to NM_005274.2Gene info linked to NM_005274.2Genome view with mapviewer linked to NM_005274.2 Homo sapiens guanine nucleotide binding protein (G protein), gamma 5 (GNG5), mRNA
Length=823

```
GENE ID: 2787 GNG5 | guanine nucleotide binding protein (G protein), gamma 5 [Homo sapiens]
Score =  682 bits (369),  Expect = 0.0
Identities = 371/372 (99%), Gaps = 0/372 (0%)

Strand=Plus/Plus

Query  174  AAAGTTTCCCAGGCAGCTGCAGACTTGAAACAGTTCTGTCTGCAGAATGCTCAACATGAC  233
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  433  AAAGTTTCCCAGGCAGCTGCAGACTTGAAACAGTTCTGTCTGCAGAATGCTCAACATGAC  492

Query  234  CCTCTGCTGACTGGAGTATCTTCAAGTACAAATCCCTTCAGACCCCAGAAAGTCTGTTCC  293
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  493  CCTCTGCTGACTGGAGTATCTTCAAGTACAAATCCCTTCAGACCCCAGAAAGTCTGTTCC  552

Query  294  TTTTTGTAGTAAAATGAATCTTTCAAAGGTTTCCCAAACCACTCCTTATGATCCAGTGAA  353
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  553  TTTTTGTAGTAAAATGAATCTTTCAAAGGTTTCCCAAACCACTCCTTATGATCCAGTGAA  612

Query  354  TATTCAAGAGAGCTACATTTGAAGCCTGTACAAAAGCTTATCCCTGTAACACATGTGCCA  413
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  613  TATTCAAGAGAGCTACATTTGAAGCCTGTACAAAAGCTTATCCCTGTAACACATGTGCCA  672

Query  414  TAATATACAAACTTTTACTTTCGTCAGTCCTTAACATCTACCTCTCTGAATTTTCATGAA  473
            |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Sbjct  673  TAATATACAAACTTCTACTTTCGTCAGTCCTTAACATCTACCTCTCTGAATTTTCATGAA  732

Query  474  TTTCTATTTCACAAGGGTAATTGTTTTATATACACTGGCAGCAGCATACAATAAAACTTA  533
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  733  TTTCTATTTCACAAGGGTAATTGTTTTATATACACTGGCAGCAGCATACAATAAAACTTA  792

Query  534  GTATGAAACTTT  545
            ||||||||||||
Sbjct  793  GTATGAAACTTT  804
```

GNG5 interaction network.



Summary for GNG5

G proteins are trimeric proteins that regulate flow of information from cell surface receptors to internal metabolic effectors. Interactions of a G protein with its activated receptor promotes exchange of GTP for GDP (alpha subunit). The alpha-GTP complex dissociates from the beta-gamma heterodimer so as to interact with and regulate effector molecules. G proteins are involved as modulators or transducers in transmembrane signaling systems. The beta and gamma chains are required for the GTPase activity, for replacement of GDP by GTP, and for G protein-effector interactions.

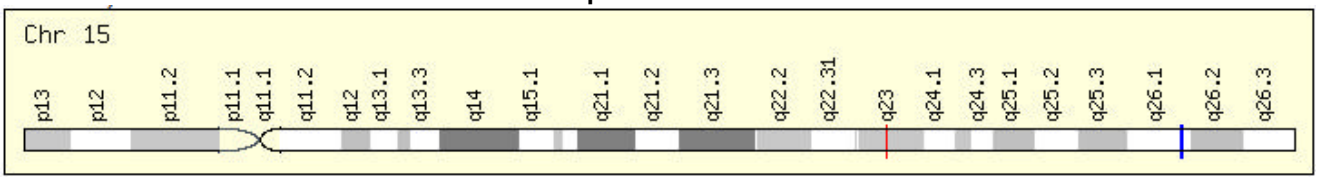
CHD2- CHMP1A

CTACTAGTAGATAGGACTCTTGGTTTGGACATACTACATGGATCAGTAAATACCTGGGCACAGGACTTCAAAGCAAACACAGATTCCCCCTCCCCCTTAATA
TTAAGAATTAAAAGATGATGAGAAATAAGGACAAAAGCCAAGAGGAGGACAGTTCGCTACACAGCAATGCATCGAGGTGACCAAGAATATGGCCCAGGTGAC
AAAGCCCTGGACAAGGCCCTGAGCACCATGGACCTGCAGAAGGTCTCCTCAGTGATGGACAGGTTCGAGCAGCAGGTGCAGAACCTGGACGTCCATACATCG
TGATGGAGGACTCCATGAGCTCGGCCACCACCCTGACCACGCCGCAGGAGCAGGTGGACAGCCTCATCATGCAGATCGCCGAGGAGAATGGCCTGGAGGTGC
GGACCAGCTCAGCCAGCTGCCCGAGGGCGCCTCTGCCGTGGGCGAGAGCTCTGTGXXXXXXXXXXXXXXXXXXXXXXXXXXXXXTTGGCCGCCTTGAGGAA
TAGCCGTGCCCCGCCGGTGTGCACCGCCTCTGCCCCGTGATGTGCTGGAAGGCTCCTGTCCTCTCCCCACCGCGTCTTGCCTTTGTGCTGACCCCGCGGGGC
GCGGCCGGCAGCCACTCTGCGTCTCTCACCTGCCAGGCCTGCGTGGCCTTAGGGTTGTTCCTGTTCTTTTAGGTTGGGCGGTGGGTCTGTGTCCTGGTGTTG
GTTTCTGCAAATTTCTGGGGGTGATTTCTGTGACTCTGGGCCCACAGCGGGGAGGCCAAGAAGGGCCCTGTGGACTTTCACCCAGCACTGTGGGGGCCTTCA
ACTCTGGGGCAGCAGACATGCTGCTTCCCATCAGCAGAGGGGGTCAGGCCTGCCCTGTTGCCAAACAACTCCTTGAGGCCTCTCCGCACCAACTCATCGGGC
GGAGGTCTCACCCATGTTGGACGACATAGCCCTAGGAGGACACCACAGGTCTAGTGTGGCTTGGGGGATGTCAGGTCACTAAGCGTTCTGGGTGTATTCAAT
ACATCCTTCCTTCTTCAAATTACTTTCAAGAG

Sanger sequencing (forward and reverse)

```
EMBOSS_001        51 TACCTGGGCACAGGACTTCAAAGCAAACACAGATTCCCCCTCCCCCTTAA    100
                                                 ||||||||||||||||||||
PCR band           1                             ATTCCCCCTCCCCCTTAA     18

EMBOSS_001       101 TATTTAAGAATTAAAAGATGATGAGAAATAAGGACAAAAGCCAAGAGGAG    150
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
PCR band          19 TATTTAAGAATTAAAAGATGATGAGAAATAAGGACAAAAGCCAAGAGGAG     68

EMBOSS_001       151 GACAGTTCGCTACACAGCAATGCATCGAGGTGACCAAGAATATGGCCCAG    200
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
PCR band          69 GACAGTTCGCTACACAGCAATGCATCGAGGTGACCAAGAATATGGCCCAG    118

EMBOSS_001       201 GTGACCAAAGCCCTGGACAAGGCCCTGAGCACCATGGACCTGCAGAAGGT    250
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
PCR band         119 GTGACCAAAGCCCTGGACAAGGCCCTGAGCACCATGGACCTGCAGAAGGT    168

EMBOSS_001       251 CTCCTCAGTGATGGACAGGTTCGAGCAGCAGGTGCAGAACCTGGACGTCC    300
                     ||||||||||||||||||||||||||||
PCR band         169 CTCCTCAGTGATGGACAGGTTCGAGCA                         195
```

5’ partner: CHD2



Junction point
exon=2 505..637
/gene="CHD2"
/gene_synonym="DKFZp547I1315; DKFZp686E01200;
DKFZp781D1727; FLJ38614"

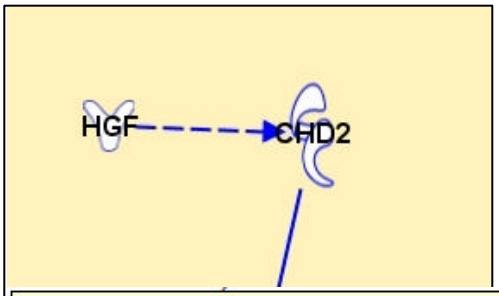
BLAST vs mRNA

>ref|NM_001271.3| Homo sapiens chromodomain helicase DNA binding protein 2 (CHD2), transcript variant 1, mRNA
Length=9374
GENE ID: 1106 CHD2| chromodomain helicase DNA binding protein 2 [Homo sapiens]
Score = 324 bits (358), Expect = 3e-91
Identities = 179/179 (100%), Gaps = 0/179 (0%)

Strand=Plus/Plus

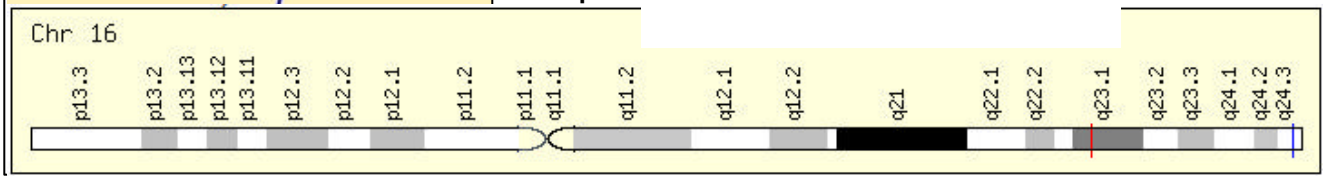
```
Query  1    CTACTAGTAGATAGGACTCTTGGTTTGGACATACTACATGGATCAGTAAATACCTGGGCA  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  459  CTACTAGTAGATAGGACTCTTGGTTTGGACATACTACATGGATCAGTAAATACCTGGGCA  518

Query  61   CAGGACTTCAAAGCAAACACAGATTCCCCCTCCCCCTTAATATTTAAGAATTAAAAGATG  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  519  CAGGACTTCAAAGCAAACACAGATTCCCCCTCCCCCTTAATATTTAAGAATTAAAAGATG  578

Query  121  ATGAGAAATAAGGACAAAAGCCAAGAGGAGGACAGTTCGCTACACAGCAATGCATCGAG  179
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  579  ATGAGAAATAAGGACAAAAGCCAAGAGGAGGACAGTTCGCTACACAGCAATGCATCGAG  637
```

CHD2 interaction network.



Summary for CHD2
The CHD family of proteins is characterized by the presence of chromo (chromatin organization modifier) and SNF2-related helicase/ATPase domains. CHD domains modify gene expression by altering chromatin structure and access of the transcriptional apparatus to target DNA sites. HGF regulates cell growth, and cell motility and modulates CHD2 activity.

3' partn



Junction point
exon=5 366..494
/gene="CHMP1A"
/gene_synonym="CHMP1; KIAA0047; PCOLN3; PRSM1"

BLAST vs mRNA

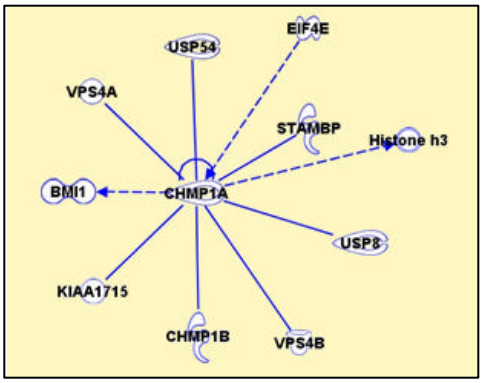
>ref|NM_001083314.1| Homo sapiens chromatin modifying protein 1A (CHMP1A), transcript variant 1, mRNA
Length=2353
GENE ID: 5119 CHMP1A| chromatin modifying protein 1A [Homo sapiens]
Score = 1292 bits (1432), Expect = 0.0
Identities = 814/878 (92%), Gaps = 14/878 (1%)

Strand=Plus/Plus

```
Query  179   GGTGACCAAGAATATGGCCCAGGTGACCAAAGCCCTGGACAAGGCCCTGAGCACCATGGA  238
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  365   GGTGACCAAGAATATGGCCCAGGTGACCAAAGCCCTGGACAAGGCCCTGAGCACCATGGA  424

Query  239   CCTGCAGAAGGTCTCCTCAGTGATGGACAGGTTCGAGCAGCAGGTGCAGAACCTGGACGT  298
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  425   CCTGCAGAAGGTCTCCTCAGTGATGGACAGGTTCGAGCAGCAGGTGCAGAACCTGGACGT  484

Query  299   CCATACATCGGTGATGGAGGACTCCATGAGCTCGGCCACCACCCTGACCACGCCGCAGGA  358
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  485   CCATACATCGGTGATGGAGGACTCCATGAGCTCGGCCACCACCCTGACCACGCCGCAGGA  544

Query  359   GCAGGTGGACAGCCTCATCATGCAGATCGCCGAGGAGAATGGCCTGGAGGTGCTGGACCA  418
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  545   GCAGGTGGACAGCCTCATCATGCAGATCGCCGAGGAGAATGGCCTGGAGGTGCTGGACCA  604

Query  419   GCTCAGCCAGCTGCCCGAGGGCGCCTCTGCCGTGGGCGAGAGCTCTGTnnnnnnnnnnn  478
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  605   GCTCAGCCAGCTGCCCGAGGGCGCCTCTGCCGTGGGCGAGAGCTCTGTGCGCAGCCAGGA  664

Query  479   nnnnnnnnnnnnnnnnnnnTGGCCGCCTTGAGGAACTAGCCGTGCCCCGCCGGTGTGCA  538
                               ||||||||||||||||||||||||||||||||||||||||||
Sbjct  665   GGACCAGCTGTCACGGAGGTTGGCCGCCTTGAGGAACTAGCCGTGCCCCGCCGGTGTGCA  724

Query  539   CCGCCTCTGCCCCGTGATGTGCTGGAAGGCTCCTGTCCTCTCCCCACCGCGTCTTGCCTT  598
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  725   CCGCCTCTGCCCCGTGATGTGCTGGAAGGCTCCTGTCCTCTCCCCACCGCGTCTTGCCTT  784

Query  599   TGTGCTGACCCCGCGGGGCTGCGGCCGGCAGCCACTCTGCGTCTCTCACCTGCCAGGCCT  658
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  785   TGTGCTGACCCCGCGGGGCTGCGGCCGGCAGCCACTCTGCGTCTCTCACCTGCCAGGCCT  844

Query  659   GCGTGGCCTTAGGGTTGTTCCTGTTCTTTTAGGTTGGGCGGTGGGTCTGTGTCCTGGTGT  718
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  845   GCGTGGCCTTAGGGTTGTTCCTGTTCTTTTAGGTTGGGCGGTGGGTCTGTGTCCTGGTGT  904

Query  719   TGAGTTTCTGCAAATTTCTGGGGGTGATTTCTGTGACTCTGGGCCCACAGCGGGGAGGCC  778
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  905   TGAGTTTCTGCAAATTTCTGGGGGTGATTTCTGTGACTCTGGGCCCACAGCGGGGAGGCC  964

Query  779   AAGAAGGGCCCTGTGGACTTTCACCCAGCACTGTGGGGGCCTTCAGACTCTGGGGCAGCA  838
             |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  965   AAGAGGGGCCCTGTGGACTTTCACCCAGCACTGTGGGGGCCTTCAGACTCTGGGGCAGCA  1024

Query  839   GACATGCTGCTTCCCATCAG-CAGAGGGGGTCAGGCCTGCCCTGTTGCCAAACAACTCCT  897
             |||||||||||||||||||| ||||||||||||||| ||||||||||||||||||||||
```

```
Sbjct  1025  GACATGCTGCTTCCCATCAGCCAGAGGGGGTCAGGGCTGCCCTGTTGCCAAACAACTCCC  1084

Query  898   TGAGGCCTCTCCGCACCAACTCATCGGGCAGGAGGTCTCACCCATGTTGGAC-GACATAG  956
             |||||||||||||||||| |||| |||||||||||||| || ||||| ||||| |||||||
Sbjct  1085  TGAGGCCTCTCCGCACCACCTCAGCGGGCAGGAGGTCCCA-CCATG-TGGACAGACATAG  1142

Query  957   CCCTAGGAGGACACCACAGGTCTAGTGTGGCTTGGGGGATGTCAGGT--CACTAAGC---  1011
             ||| |||||| |||||||||||| ||||   ||||||||||||||   |||  | |
Sbjct  1143  CCCAAGGAGG-CACCACAGGTCTA-TGTGTGCTGGGGGATGTCAGGTGCCACCCAACGCT  1200

Query  1012  GTTCTGGGTGTATT--CAATGACATCCTTCCTTCTTCA  1047
              || ||||  |||||  |||||||||||| |||| || ||
Sbjct  1201  GTCCTGGTGGTATTTACAATGACATCC-TCCTCCTCCA  1237
```

CHMP1A interaction network.



Summary for CHMP1A

CHMP1A a member of the CHMP/Chmp family. CHMP1A is a component of endosomal sorting transport complex III (ESCRT-III) which is involved in multivesicular bodies (MVBs) formation and sorting of endosomal cargo proteins into MVBs. The MVB pathway appears to require the sequential function of ESCRT-O, -I,-II and -III complexes. ESCRT-III proteins mostly dissociate from the invaginating membrane before the ILV is released. The ESCRT machinery also functions in topologically equivalent membrane fission events, such as the terminal stages of cytokinesis and the budding of enveloped viruses (HIV-1). Involved in recruiting VPS4A and/or VPS4B to the midbody of dividing cells. May also be involved in chromosome condensation. Targets the Polycomb group (PcG) protein BMI1/PCGF4 to regions of condensed chromatin.

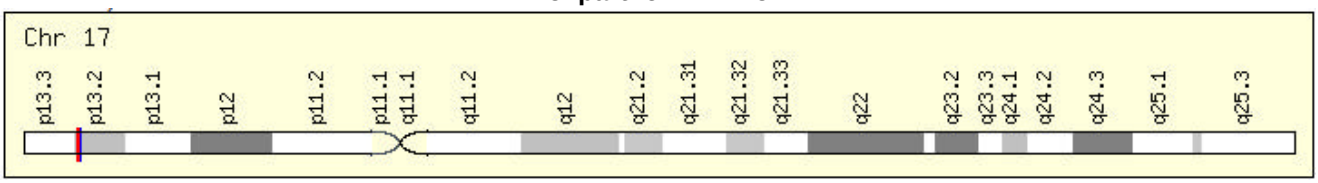
P2RX5-TAX1BP3

TTGGCTCACATCTGGGCCAGGGCTGCTGGGGATGCCGGAGCAGCAGGAGCTGCAGGAGCCACCCGAGGCGAAGCGTGGAAGCAGCAGTCAGAAGGGGAACGG
TCTGTGTGCCCACAGCTCCTGGAGCCCCACAGCAAAGAGTTGAAATTCACAAGCTGCGTCAAGGTGAGAACTTAATCCTGGGTTTCAGCATTGGAGGTGGAA
CGACCAGGATCCTTCCCAGAATCCCTTCTCTGAAGACAAGACGGACAAGGGTATTTATGTCACACGGGTGTCTGAAGGAGGCCCTGCTGAAATCGCTGGGCT
CAGATTGGAGACAAGATCATGCAGGTGAACGGCTGGGACATGACCATGGTCACACACGACCAGGCCCGCAAGCGGCTCACCAAGCGCTCGGAGGAGGTGGTG
GTCTGCTGGTGACGCGGCAGTCGCTGCAGAAGGCCGTGCAGCAGTCCATGCTGTCCTAGCAGCCACCACCATCTGCGACTCCTGCCTGCCGCCTCTCTGTAC
GTAACGCCACTTCCACACTCTGTCCCCATCTGGCTTCTGCTGACCGCTGGGCCCCAGCTCAGAAGGGCTATAGCTGGTCCCAGAGGCCTGGCCTGGCCTTCC
TCCCTTCTCCCATCCCTGGCCTGGGGCCTCTGGGACCGGCTTTCTCTCCTGGACACCGAGGATTGGAAATAAAGGCCCTGGAGCTGAGTAGTAGCCCAGTCT
CTGTGACCACAGGCTCACGTCCGACCCTGCTGCTTGGCCACAGCAGTGGCTGGGGCAAGTGGGAACCACTACCTCTTGGGGACCCCCAAAAGCCTGGGAAA

Sanger sequencing (reverse)

```
EMBOSS_001         2 GTCAGAAGGGGAACGGATCTGTGTGCCCACAGCTCCTGGAGCCCCACAGC     51
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
PCR band           2 GTCAGAAGGGGAACGGATCTGTGTGCCCACAGCTCCTGGAGCCCCACAGC     51

EMBOSS_001        52 AAAGAGTTGAAATTCACAAGCTGCGTCAAGGTGAGAACTTAATCCTGGGT    101
                     |||||||||||| ||||||||||||||||||||||||||||||||||||
PCR band          52 AAAGAGTTGAAATTCACAAGCTGCGTCAAGGTGAGAACTTAATCCTGGGT    101

EMBOSS_001       102 TTCAGCATTGGAGGTGGAAT    151
                     ||||||||||||||||||||
PCR band         102 TTCAGCATTGGAGGTGGAAT    151
```

5' partner: P2RX5



Junction point
exon=11 1392..1586
/gene="P2RX5"
/gene_synonym="LRH-1; MGC47755; P2X5; P2X5R"

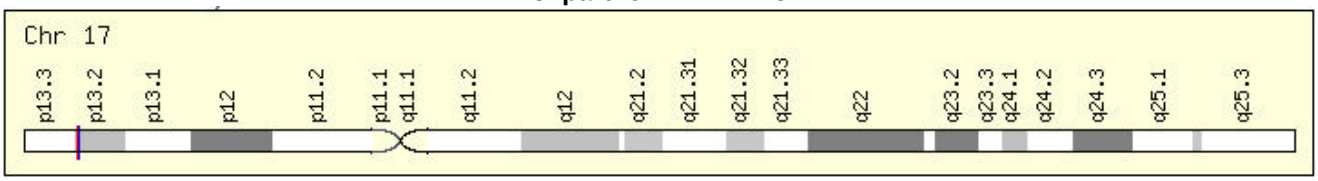
BLAST vs mRNA

>ref|NM_001204520.1| Gene info linked to NM_001204520.1 Homo sapiens purinergic receptor P2X, ligand-gated ion channel, 5 (P2RX5), transcript variant 5, mRNA
Length=2269
GENE ID: 5026 P2RX5 | purinergic receptor P2X, ligand-gated ion channel, 5 [Homo sapiens]
Score = 244 bits (132), Expect = 2e-63
Identities = 132/132 (100%), Gaps = 0/132 (0%)

Strand=Plus/Plus

```
Query  4     GCTCACATCTGGGCCAGGGCTGCTGGGGATGCCGGAGCAGCAGGAGCTGCAGGAGCCACC  63
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1455  GCTCACATCTGGGCCAGGGCTGCTGGGGATGCCGGAGCAGCAGGAGCTGCAGGAGCCACC  1514

Query  64    CGAGGCGAAGCGTGGAAGCAGCAGTCAGAAGGGGAACGGATCTGTGTGCCCACAGCTCCT  123
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1515  CGAGGCGAAGCGTGGAAGCAGCAGTCAGAAGGGGAACGGATCTGTGTGCCCACAGCTCCT  1574

Query  124   GGAGCCCCACAG  135
             ||||||||||||
Sbjct  1575  GGAGCCCCACAG  1586
```

3' partner: TAX1BP3



Junction point
exon=2 193..312
/gene="TAX1BP3"
/gene_synonym="TIP-1"

BLAST vs mRNA

>ref|NM_014604.3| Gene info linked to NM_014604.3 Homo sapiens Tax1 (human T-cell leukemia virus type I) binding protein 3 (TAX1BP3), transcript variant 1, mRNA
Length=1398
GENE ID: 30851 TAX1BP3 | Tax1 (human T-cell leukemia virus type I) binding protein 3 [Homo sapiens]
Score = 1218 bits (659), Expect = 0.0
Identities = 680/689 (99%), Gaps = 6/689 (0%)

Strand=Plus/Plus

```
Query  135   GCAAAGAGTTGAAATTCACAAGCTGCGTCAAGGTGAGAACTTAATCCTGGGTTTCAGCAT  194
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  192   GCAAAGAGTTGAAATTCACAAGCTGCGTCAAGGTGAGAACTTAATCCTGGGTTTCAGCAT  251

Query  195   TGGAGGTGGAATCGACCAGGATCCTTCCCAGAATCCCTTCTCTGAAGACAAGACGGACAA  254
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  252   TGGAGGTGGAATCGACCAGGATCCTTCCCAGAATCCCTTCTCTGAAGACAAGACGGACAA  311

Query  255   GGGTATTTATGTCACACGGGTGTCTGAAGGAGGCCCTGCTGAAATCGCTGGGCTGCAGAT  314
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  312   GGGTATTTATGTCACACGGGTGTCTGAAGGAGGCCCTGCTGAAATCGCTGGGCTGCAGAT  371

Query  315   TGGAGACAAGATCATGCAGGTGAACGGCTGGGACATGACCATGGTCACACACGACCAGGC  374
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  372   TGGAGACAAGATCATGCAGGTGAACGGCTGGGACATGACCATGGTCACACACGACCAGGC  431

Query  375   CCGCAAGCGGCTCACCAAGCGCTCGGAGGAGGTGGTGCGTCTGCTGGTGACGCGGCAGTC  434
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  432   CCGCAAGCGGCTCACCAAGCGCTCGGAGGAGGTGGTGCGTCTGCTGGTGACGCGGCAGTC  491

Query  435   GCTGCAGAAGGCCGTGCAGCAGTCCATGCTGTCCTAGCAGCCACCACCATCTGCGACTCC  494
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  492   GCTGCAGAAGGCCGTGCAGCAGTCCATGCTGTCCTAGCAGCCACCACCATCTGCGACTCC  551

Query  495   TGCCTGCCGCCTCTCTGTACAGTAACGCCACTTCCACACTCTGTCCCCATCTGGCTTCTG  554
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  552   TGCCTGCCGCCTCTCTGTACAGTAACGCCACTTCCACACTCTGTCCCCATCTGGCTTCTG  611

Query  555   CTGACCGCTGGGCCCCAGCTCAGAAGGGCTATAGCTGGTCCCAGAGGCCTGGCCTGGCCT  614
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  612   CTGACCGCTGGGCCCCAGCTCAGAAGGGCTATAGCTGGTCCCAGAGGCCTGGCCTGGCCT  671
```

NXPH1-TXNL4A

TTGGCAACTGTGCCAAGCCTTGGCTCCCGCGAACCAATCCTGAGCGCGACCCGGGCACTGGGACGGCGACTCCGCCAAAGCTGGACGAGGCAGCCGGACCCG
CTGCGCTCGAGCATGGAGACGGAGCGCCTGGGAGGGCACGTCCGGGGCGCTGGAGACGCCAGGCCCGAGTAGCTTCTCCATGGAGCCTGCCCAGAGCGGTCC
TTCTCGCAGGATTCGCCCCAAGTCCTGTGCGGCTGCTGAGAGCGCTCCTTGCTCTGTAAAGTGGATGTCAGGTGGATCTATGTTTCTGAAGGAACAAAGACT
AAAGAAGGCACCGCCAAGGAAGTTTGAGACGCGGGAGAATGCAGGCTGCGTGCTGGTACGTGCTTTTCCTCCTGCAGCCCACCGTCTACTTGGTTAAAAATT
TGCAGTTATTTATCTTGTGGATATTACAGAAGTGCCTGACTTCAACAAAATGTATGAGTTATACGATCCATGTACTGTCATGTTTTTCTTCAGGAACAAGCA
ATCATGATTGACTTGGGGACTGGCAACAACAACAAGATTAACTGGGCCATGGAGGACAAGCAGGAGATGGTGGACATCATCGAGACGGTGTACCGCGGGGCC
GCAAAGGCCGCGGCCTGGTGGTGTCCCCCAAGGACTACTCCACCAAGTACCGCTACTGAGGCGCCCTCAGTCTGCGCGGATAAATGTCGTGGAGACCTTTTT
TATAGAAACATATTTAAGCTATTTAAAGCCTTTGGAAATACAGGAAGCTCCCGGGCTGG

5’ partner: NXPH1

Junction point
exon=2 802..965
/gene="NXPH1"
/gene_synonym="Nbla00697; NPH1"

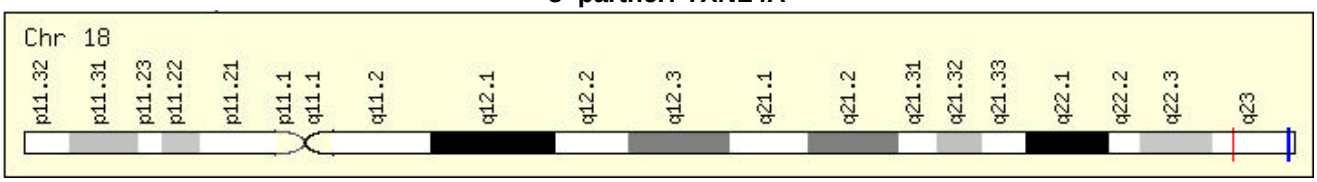
BLAST vs mRNA

>ref|NM_152745.2| UniGene infoGeoGene info Homo sapiens neurexophilin 1 (NXPH1), mRNA
Length=2931
GENE ID: 30010 NXPH1 | neurexophilin 1 [Homo sapiens]
Score = 720 bits (798), Expect = 0.0
Identities = 399/399 (100%), Gaps = 0/399 (0%)

Strand=Plus/Plus

```
Query  5    CAACTGTGCCAAGCCTTGGCTCCCGCGAACCAATCCTGAGCGCGACCCGGGCACTGGGAC  64
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  569  CAACTGTGCCAAGCCTTGGCTCCCGCGAACCAATCCTGAGCGCGACCCGGGCACTGGGAC  628

Query  65   GGCGACTCCGCCAAAGCTGGACGAGGCAGCCGGACCCGTCTGCGCTCGAGCATGGAGACG  124
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  629  GGCGACTCCGCCAAAGCTGGACGAGGCAGCCGGACCCGTCTGCGCTCGAGCATGGAGACG  688

Query  125  GAGCGCCTGGGAGGGCACGTCCGGGGCGCTGGAGACGCCAGGCCCGAGTAGCTTCTCCAT  184
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  689  GAGCGCCTGGGAGGGCACGTCCGGGGCGCTGGAGACGCCAGGCCCGAGTAGCTTCTCCAT  748

Query  185  GGAGCCTGCCCAGAGCGGTCCCTTCTCGCAGGATTCGCCCCAAGTCCTGTGCGGCTGCTG  244
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  749  GGAGCCTGCCCAGAGCGGTCCCTTCTCGCAGGATTCGCCCCAAGTCCTGTGCGGCTGCTG  808

Query  245  AGAGCGCTCCTTGCTCTGTAAAGTGGATGTCAGGTGGATCTATGTTTCTGAAGGAACAAA  304
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  809  AGAGCGCTCCTTGCTCTGTAAAGTGGATGTCAGGTGGATCTATGTTTCTGAAGGAACAAA  868

Query  305  GACTCAAAGAAGGCACCGCCAAGGAAGTTTGAGACGCGGGAGAATGCAGGCTGCGTGCTG  364
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  869  GACTCAAAGAAGGCACCGCCAAGGAAGTTTGAGACGCGGGAGAATGCAGGCTGCGTGCTG  928

Query  365  GTACGTGCTTTTCCTCCTGCAGCCCACCGTCTACTTGGT  403
            |||||||||||||||||||||||||||||||||||||||
Sbjct  929  GTACGTGCTTTTCCTCCTGCAGCCCACCGTCTACTTGGT  967
```

3’ partner: TXNL4A



Junction point
exon=2 294..397

/gene="TXNL4A"
/gene_synonym="DIB1; DIM1; HsT161; TXNL4; U5-15kD"

BLAST vs mRNA

>ref|NM_006701.2| Homo sapiens thioredoxin-like 4A (TXNL4A), mRNA
Length=1415
Score = 654 bits (724), Expect = 0.0
Identities = 375/381 (98%), Gaps = 2/381 (0%)

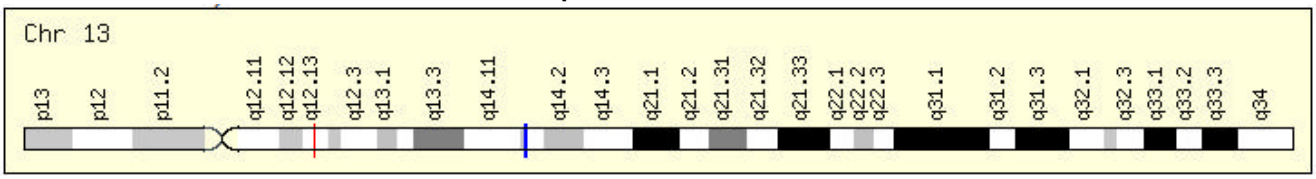
Strand=Plus/Plus

```
Query  401  GGTTAAAAATTTTGCAGTTATTTATCTTGTGGATATTACAGAAGTGCCTGACTTCAACAA  460
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  293  GGTTAAAAATTTTGCAGTTATTTATCTTGTGGATATTACAGAAGTGCCTGACTTCAACAA  352

Query  461  AATGTATGAGTTATACGATCCATGTACTGTCATGTTTTTCTTCAGGAACAAGCACATCAT  520
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  353  AATGTATGAGTTATACGATCCATGTACTGTCATGTTTTTCTTCAGGAACAAGCACATCAT  412

Query  521  GATTGACTTGGGGACTGGCAACAACAACAAGATTAACTGGGCCATGGAGGACAAGCAGGA  580
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  413  GATTGACTTGGGGACTGGCAACAACAACAAGATTAACTGGGCCATGGAGGACAAGCAGGA  472

Query  581  GATGGTGGACATCATCGAGACGGTGTACCGCGGGGCCCGCAAAGGCCGCGGCCTGGTGGT  640
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  473  GATGGTGGACATCATCGAGACGGTGTACCGCGGGGCCCGCAAAGGCCGCGGCCTGGTGGT  532

Query  641  GTCCCCCAAGGACTACTCCACCAAGTACCGCTACTGAGGCGCCCTCAGTCTGCGCGGATA  700
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  533  GTCCCCCAAGGACTACTCCACCAAGTACCGCTACTGAGGCGCCCTCAGTCTGCGCGGATA  592

Query  701  AATGTCGTGGAGACCTTTTTGTATAGAAACATATTTAAGCTATTTAAAGCCTTTGG-AAA  759
            |||||||||||| |||||||||||| ||||| | ||||||||||||||||||||||| |||
Sbjct  593  AATGTCGTGGAGCCCTTTTTGTATGGAAACGT-TTTAAGCTATTTAAAGCCTTTGGAAAA  651

Query  760  TACAGGAAGCTCCCGGGCTGG  780
            |||||||||||||| |||||||
Sbjct  652  TACAGGAAGCTCCAGGGCTGG  672
```

KIAA1704- ITGB3BP

TTGGCAGTAAGAAAGATGAAGAACATATATTATCAGGAAGAGATAAGAGACTGGCTGAGCAGGTATCTTCATACAATGAATCAAAAAGATCAGAATCTCTTA
GGACATACATCATAAAAAGTTAAAGAGTAAGGCTGCTGAAGACAAAAATAAGCCTCAAGAGAGAATACCATTTGACCGTGATAAAGATCTCAAGGTTAATCG
TTTGATGAAGCTCAGAAAAAAGCCCTAATAAAAAAATCTAGAGAACTAAACACCAGATTTTCACACGGCAAAGGCAATATGTTTTTATAAATTCATGATGTT
CTATCAAAAGTTGAGAAATTGTCAGAAGAAATCATGGAGATAATGCAAAATTTAAGTAGTATACAGGCTTTGGAGGGCAGTAGAGAGCTTGAAAATCTCATT
GAATCTCCTGTGCATCXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXGTGAATAAACAAAAACTGTTTGAAAAGAGTACAGGAC
TCCTCACAAAGGTAAGTAGTTGTGTTTTAGTTTATTACATAAAGATGCTTTCATCCCATTTTGCCAATTAAAACTAATCATCATGAATCCTTTTTAGTGGTT
TCTTTTTTCCTCAGAGATCCCTGTTCTGAAGTATGTTCTTTTAGAAGATCACAAATCTCATTTGAAAAGCTTTATATTGCATTTAGCAGAAGATGATATCTG
TGAGATCAACTTCTCTGATGATTGAACTTTTTAAATGTCTTTTGAAATGCAATCTTCAACATAAGATTTGAATTTTGTCTTTTTATTAAAATTTCTTTTTGA
AATACAAATAGGAGCAACAATGCATTAATATACTGCAGTTTTTAGTGTTGCACTAGAAGCATTGAAAGTTACTTTCTAACTAGTATTCTGGGAAACTTTCAG
TCATTAATATGCTTTGTTCTACTATGACTCTTATGTCTGAGCATTTCACAGTATGGAAAACTAAATAACAGTTTTAGTTTCCATAGAAGTACACATCACACC
TCACACCTCATAGCAAGTATCATGACTGTAAATACCAAACCTTAAACTAATAAGGAGGATAAAGTACTACGATTGCCATAGCATCTGCATGAATAGGCTTAA
GGGTGGGGCGGTTCAGCGCTAACACTTTTAT

5' partner: KIAA1704



Junction point
CDS 104..1126
/gene="KIAA1704"
/gene_synonym="AD029; bA245H20.2; LSR7; RP11-245H20.2"

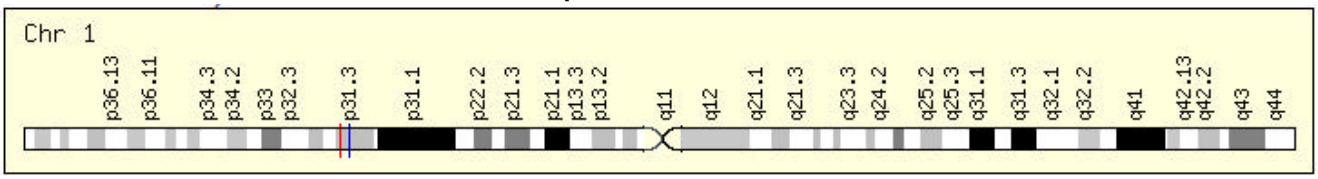
BLAST vs mRNA

>ref|NM_018559.2| Homo sapiens KIAA1704 (KIAA1704), mRNA
Length=1431
Score = 527 bits (584), Expect = 2e-153
Identities = 292/292 (100%), Gaps = 0/292 (0%)

Strand=Plus/Plus

```
Query  5     CAGTAAGAAAGATGAAGAACATATATTATCAGGAAGAGATAAGAGACTGGCTGAGCAGGT  64
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  835   CAGTAAGAAAGATGAAGAACATATATTATCAGGAAGAGATAAGAGACTGGCTGAGCAGGT  894

Query  65    ATCTTCATACAATGAATCAAAAAGATCAGAATCTCTTATGGACATACATCATAAAAAGTT  124
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  895   ATCTTCATACAATGAATCAAAAAGATCAGAATCTCTTATGGACATACATCATAAAAAGTT  954

Query  125   AAAGAGTAAGGCTGCTGAAGACAAAAATAAGCCTCAAGAGAGAATACCATTTGACCGTGA  184
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  955   AAAGAGTAAGGCTGCTGAAGACAAAAATAAGCCTCAAGAGAGAATACCATTTGACCGTGA  1014

Query  185   TAAAGATCTCAAGGTTAATCGGTTTGATGAAGCTCAGAAAAAAGCCCTAATaaaaaaaTC  244
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1015  TAAAGATCTCAAGGTTAATCGGTTTGATGAAGCTCAGAAAAAAGCCCTAATAAAAAAATC  1074

Query  245   TAGAGAACTAAACACCAGATTTTCACACGGCAAAGGCAATATGTTTTTATAA  296
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1075  TAGAGAACTAAACACCAGATTTTCACACGGCAAAGGCAATATGTTTTTATAA  1126
```

3’ partner: ITGB3BP



Junction point
exon=5 286..364
/gene="ITGB3BP"
/gene_synonym="CENP-R; CENPR; HSU37139; NRIF3; TAP20"

BLAST vs mRNA

>ref|NM_014288.3| Homo sapiens integrin beta 3 binding protein (beta3-endonexin) (ITGB3BP), mRNA
Length=892
Score = 241 bits (266), Expect = 3e-67
Identities = 135/136 (99%), Gaps = 0/136 (0%)

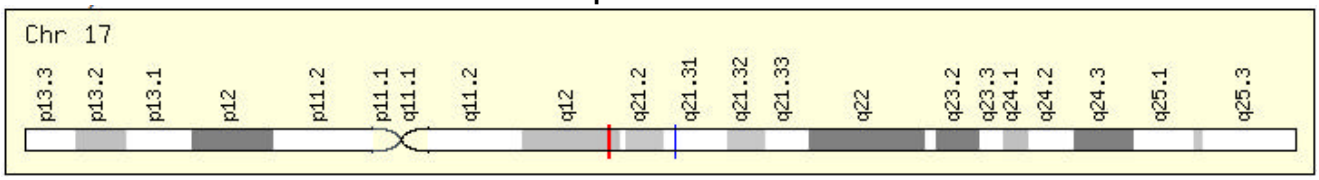
Strand=Plus/Plus

```
Query  293   ATAAATTCATGATGTTGCTATCAAAAGTTGAGAAATTGTCAGAAGAAATCATGGAGATAA  352
             || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  282   ATGAATTCATGATGTTGCTATCAAAAGTTGAGAAATTGTCAGAAGAAATCATGGAGATAA  341

Query  353   TGCAAAATTTAAGTAGTATACAGGCTTTGGAGGGCAGTAGAGAGCTTGAAAATCTCATTG  412
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  342   TGCAAAATTTAAGTAGTATACAGGCTTTGGAGGGCAGTAGAGAGCTTGAAAATCTCATTG  401

Query  413   GAATCTCCTGTGCATC  428
             ||||||||||||||||
Sbjct  402   GAATCTCCTGTGCATC  417
```

DHX8 -ADK

TTGGCTGTAGCCATGGCGGGAGCCTTAATCGGGTCGGAGCCAGGCCCCGCGGAAGAACTTGCCAAACTCGAGTACCTGTCTTTGGTGTCAAAGGTTTGCACTC
AGCTGGACAATCACTTGGGGATCAACGACAAGGACCTTGCTGAATTTGTGATCAGTCTTGCTGAGAAAAATACCACCTTTGATACTTTTAAGGCTTCTCTCGT
CAAAAATGGTGCAGAATTTACGGATTCTCTTATTAGTAACTTGCTGCGTCTCATACAAACCATGCGGCCTCCAGCGAAGCCTTCCACTAGCAAAGGTTTTCTC
TCTCAACTGGTCTCTGACAAGCCTCTGACTGAATGTATCCGTGCTGGCCACTATGCAGCAAGCATCATAATTAGACGGACTGGCTGCACCTTTCCTGAGAAGC
CAGACTTCCACTGATGGAAGAGCTGAAAACACAAGCCCAGGAGTGCAGACACTGCCCTAATTGCTTCCTGAGAATTCCCATATTAATAAAGAAGAAAATTATC
TGCCATTTTTTCCTACTATAATAATGCTGAATCTTAATTTAGAGGGTACAAGGGTATGGTAATGCTTGTAGAATCTTTATTATCTCAACAATCTAAAAAATGA
TGTTTATTTCCATAGTTTGATAGTGCCACTTAAATGCCAATTAAACAAGAATATAACATTTCAATAGAAATTTTTATTTCATTTTCAATTACTTTGTATACAT
TTCTGCTTTGAATGCAGATGCAAATTTTAATATAATAGATTTTTTAATGAATTAATCTTAACATAGTAATCTTTAGCTTTTTATACAAATATATTTAATTTAC
GAGTATATGTGTGTCTATACACACACATACATAAATATACCACATATACACTGATAGTCAAATAGGGTACAGAAATTTTATCTTGTCAATATGCCAATTATCT
CTTTAATGTGCACTCCAACATGTAATAACTTTGGATATTTAAAAAACACCATCACGCCAATAG

5’ partner: DHX8



Junction point
exon=3 308..380
/gene="DHX8"
/gene_synonym="DDX8; HRH1; PRP22; PRPF22"

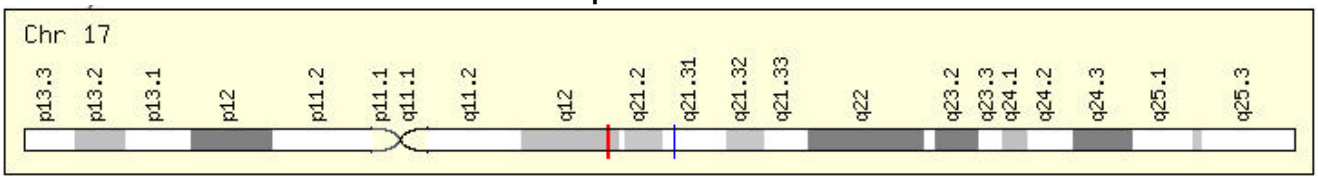
BLAST vs mRNA

>ref|NM_004941.1| Homo sapiens DEAH (Asp-Glu-Ala-His) box polypeptide 8 (DHX8), mRNA
Length=4201
Gene id: 4826690 ATP-dependent RNA helicase DHX8
Score = 542 bits (600), Expect = 2e-157
Identities = 300/300 (100%), Gaps = 0/300 (0%)

Strand=Plus/Plus

```
Query  2    TGGCTGTAGCCATGGCGGGAGCCTTAATCGGGTCGGAGCCAGGCCCCGCGGAAGAACTTG  61
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  81   TGGCTGTAGCCATGGCGGGAGCCTTAATCGGGTCGGAGCCAGGCCCCGCGGAAGAACTTG  140

Query  62   CCAAACTCGAGTACCTGTCTTTGGTGTCAAAGGTTTGCACTGAGCTGGACAATCACTTGG  121
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  141  CCAAACTCGAGTACCTGTCTTTGGTGTCAAAGGTTTGCACTGAGCTGGACAATCACTTGG  200

Query  122  GGATCAACGACAAGGACCTTGCTGAATTTGTGATCAGTCTTGCTGAGAAAAATACCACCT  181
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  201  GGATCAACGACAAGGACCTTGCTGAATTTGTGATCAGTCTTGCTGAGAAAAATACCACCT  260

Query  182  TTGATACTTTTAAGGCTTCTCTCGTCAAAAATGGTGCAGAATTTACGGATTCTCTTATTA  241
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  261  TTGATACTTTTAAGGCTTCTCTCGTCAAAAATGGTGCAGAATTTACGGATTCTCTTATTA  320

Query  242  GTAACTTGCTGCGTCTCATACAAACCATGCGGCCTCCAGCGAAGCCTTCCACTAGCAAAG  301
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  321  GTAACTTGCTGCGTCTCATACAAACCATGCGGCCTCCAGCGAAGCCTTCCACTAGCAAAG  380
```

3’ partner: ADK



Junction point
exon=12 1037..2018
/gene="ADK"
/gene_synonym="AK"

BLAST vs mRNA

>ref|NM_006721.2| Homo sapiens adenosine kinase (ADK), transcript variant ADK-long, mRNA
Length=2018
GENE ID: 132 ADK| adenosine kinase [Homo sapiens]
Score = 1088 bits (1206), Expect = 0.0
Identities = 667/713 (93%), Gaps = 41/713 (5%)

Strand=Plus/Plus

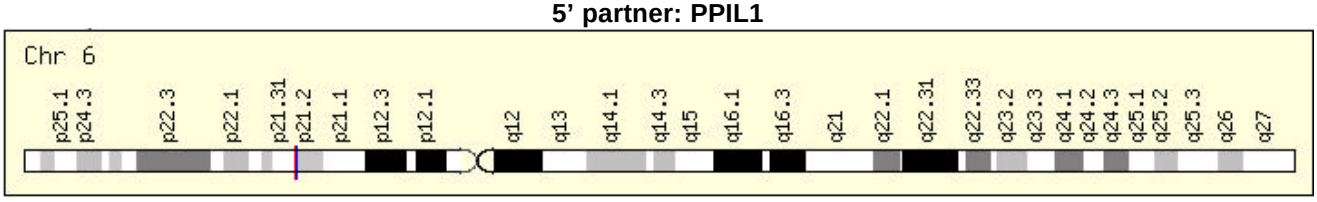
```
Query  300   AGGTTTTCTGTCTCAACTGGTCTCTGACAAGCCTCTGACTGAATGTATCCGTGCTGGCCA  359
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1035  AGGTTTTCTGTCTCAACTGGTCTCTGACAAGCCTCTGACTGAATGTATCCGTGCTGGCCA  1094

Query  360   CTATGCAGCAAGCATCATAATTAGACGGACTGGCTGCACCTTTCCTGAGAAGCCAGACTT  419
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1095  CTATGCAGCAAGCATCATAATTAGACGGACTGGCTGCACCTTTCCTGAGAAGCCAGACTT  1154
```

```
Query  420   CCACTGATGGAAGAGCTGAAAACACAAGCCCAGGAGTGCAGACACTGCCCTAATTGCTTC  479
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1155  CCACTGATGGAAGAGCTGAAAACACAAGCCCAGGAGTGCAGACACTGCCCTAATTGCTTC  1214

Query  480   CTGAGAATTCCCATATTAATAAAGAAGAAAATTATCTGCCATTTTTTCCTACTATAATAA  539
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1215  CTGAGAATTCCCATATTAATAAAGAAGAAAATTATCTGCCATTTTTTCCTACTATAATAA  1274

Query  540   TGCTGAATCTTAATTTAGAGGGTACAAGGGTATGGTAATGCTTGTAGAATCTTTATTATC  599
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1275  TGCTGAATCTTAATTTAGAGGGTACAAGGGTATGGTAATGCTTGTAGAATCTTTATTATC  1334

Query  600   TCAACAATCTAAAAAATGATGTTTATTTCCATAGTTTGATAGTGCCACTTAAATGCCAAT  659
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1335  TCAACAATCTAAAAAATGATGTTTATTTCCATAGTTTGATAGTGCCACTTAAATGCCAAT  1394

Query  660   TAAACAAGAATATAACATTTCAATAGAAATTTTTATTTCATTTTCAATTACTTTGTA---  716
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1395  TAAACAAGAATATAACATTTCAATAGAAATTTTTATTTCATTTTCAATTACTTTGTAAAT  1454

Query  717   --------------------------------TACATTTCTGCTTTGAATGCAGATGCAA  744
                                             ||||||||||||||||||||||||||||
Sbjct  1455  TCGTGTGTATTTAGTACACTGATTTGTTTTTTTACATTTCTGCTTTGAATGCAGATGCAA  1514

Query  745   ATTTTAATATAATAGATTTTTTAATGAATTAATCTTAACATAGTAATCTTTAGCTTTTTA  804
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1515  --TTTAATATAATAGATTTTTTAATGAATTAATCTTAACATAGTAATCTTTAGCTTTTTA  1572

Query  805   TACAAATATATTTAATTTAGGAGTATATGTGTGTCTATACACACACATACATAAATATAC  864
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1573  TACAAATATATTTAATTTAGGAGTATATGTGTGTCTATACACACACATACATAAATATAC  1632

Query  865   CACATATACA-CTGATAGTCAAATAGGGTACAGAAATTTTATCTTGTCAA-TATGCC-AA  921
             |||||||||| ||||||||||||| |||||||||||||||||||||||| |||||| ||
Sbjct  1633  CACATATACACCTGATAGTCAAATAAGGTACAGAAATTTTATCTTGTCAATTATGCCAAA  1692

Query  922   TTATCTCTTTAATGTGCACTCCAACATGTAAT-AACTTTGGATATTTAAAAAA  973
             | ||||||||||||||||| |||||||||| |||||||||| ||||| ||
Sbjct  1693  TAATCTCTTTAATGTGCACTCAAACATGTAATAAACTTTGGATAATTAAATAA  1745
```

PPIL1-KIAA1614

TTGGCGCTTGCCTAGACAAGCATTCCGCCGCCGGCTTCGCTATGGCGGCAATTCCCCCAGATTCCTGGCAGCCACCCAACGTTTACTTGGAGACCAGCATGG
AATCATTGTGCTGGAGCTGTACTGGAAGCATGCTCCAAAGACCTGTAAGAACTTTGCTGAGTTGGCTCGTCGAGGTTACTACAATGGCACAAAATTCCACAG
ATTATCAAAGACTTCATGATCCAAGGAGGTGACCCAACAGGGACAGGAGGACCCCAGGGCTTTCTTGGCTCAGCAGATGTTGCCACCATCAACTCCACGGGC
TCACCCTCTCCCTGTCCTCAGAGGAGTCAGAGTCCAGCAAGGAATCAGAGGGAAGCCTGCAGAGGACAGGGTCAGGATCTGGAGGACATGTGCTGTCAAGAG
ATCAGCAGGAGCTGGCACAGGACCCGGCTCCCCCTCGGCTGCCCCTTTGGACCAGAACAAGAAAAGGAGCAGCAGCATAGCCTCCACCCTGGGGCTGAAAAA
CTCTTCTCAGCCCTGGGCCAGAGTTCCCGGCCCAAGCTGGGCAAGTCCCGCAGCTACAGTGTGGAGCAGTTGCAGCCCGCCCCGCCTGGCCTGACGTCACAG
CCAGGGCCCCATCGTTACAATCCCTGCACCCGGTGTCACCCTCTCACCAGCGTCGGAAAGCTGCCTCTTTTCAGAACCTCCATTCTCTGCTGAGCAGCAAGG
GGACCGGTCCAGCCTCTACCTGGTAGCAGGGCCAGGGGACCACAGTGCAGCTGGCAGCCGGCCAAGACTTCACCACGGCGTGCCCTCAGTGTGGAGGACGTG
GTGCTCCCAGCCTGTCTCGCACCGTGGGCCGCCTGGTGGAGGTGTTCCCAGACAGCACCAGCCAGCTGCAGCTGCAGCGCTCTCCAAGGGCACTTTCGGCTT
TGCGTGCCTCTGGGAATGGCGCCCAGACTCAGGGACGCACTCTCTGAGATATTTCGCCTGGGCGTTTCGATCCATCCATGGGGTCCCTGATTATTTCAGTGG
TATGTGACTCGCACCTATCACCACATATGACATGTCACTCAGCGAGGATCACGCCATGAATGCGAAAACTCTGCTGATCCTGGAAGTACCTCACCTGCGAAA
CCTGGCACCCAAAGTTTTCCAAGTGA

5’ partner: PPIL1



Junction point
exon=2 309..463
/gene="PPIL1"
/gene_synonym="CGI-124; CYPL1; hCyPX; MGC678; PPIase"

BLAST vs mRNA

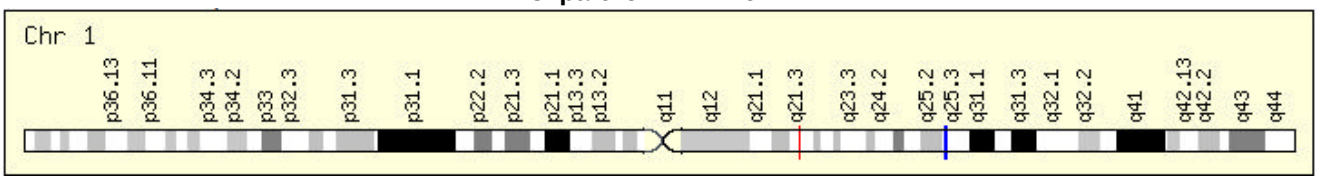
>ref|NM_016059.4| Homo sapiens peptidylprolyl isomerase (cyclophilin)-like 1 (PPIL1), mRNA
Length=1750
GENE ID: 51645 PPIL1| peptidylprolyl isomerase (cyclophilin)-like 1 [Homo sapiens]
Score = 437 bits (484), Expect = 3e-126
Identities = 242/242 (100%), Gaps = 0/242 (0%)

Strand=Plus/Plus

```
Query  12   CTAGACAAGCATTCCGCCGCCGGCTTCGCTATGGCGGCAATTCCCCCAGATTCCTGGCAG  71
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  223  CTAGACAAGCATTCCGCCGCCGGCTTCGCTATGGCGGCAATTCCCCCAGATTCCTGGCAG  282
```

```
Query  72   CCACCCAACGTTTACTTGGAGACCAGCATGGGAATCATTGTGCTGGAGCTGTACTGGAAG  131
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  283  CCACCCAACGTTTACTTGGAGACCAGCATGGGAATCATTGTGCTGGAGCTGTACTGGAAG  342

Query  132  CATGCTCCAAAGACCTGTAAGAACTTTGCTGAGTTGGCTCGTCGAGGTTACTACAATGGC  191
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  343  CATGCTCCAAAGACCTGTAAGAACTTTGCTGAGTTGGCTCGTCGAGGTTACTACAATGGC  402

Query  192  ACAAAATTCCACAGAATTATCAAAGACTTCATGATCCAAGGAGGTGACCCAACAGGGACA  251
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  403  ACAAAATTCCACAGAATTATCAAAGACTTCATGATCCAAGGAGGTGACCCAACAGGGACA  462

Query  252  GG  253
            ||
Sbjct  463  GG  464
```


3’ partner: KIAA1614



Junction point
exon=6 2829..2985
/gene="KIAA1614"
/gene_synonym="RP11-46A10.3"

BLAST vs mRNA

>ref|NM_020950.1| Homo sapiens KIAA1614 (KIAA1614), mRNA
Length=4155
GENE ID: 57710 KIAA1614| KIAA1614 [Homo sapiens]
Score = 1238 bits (1372), Expect = 0.0
Identities = 712/724 (98%), Gaps = 4/724 (0%)

Strand=Plus/Plus

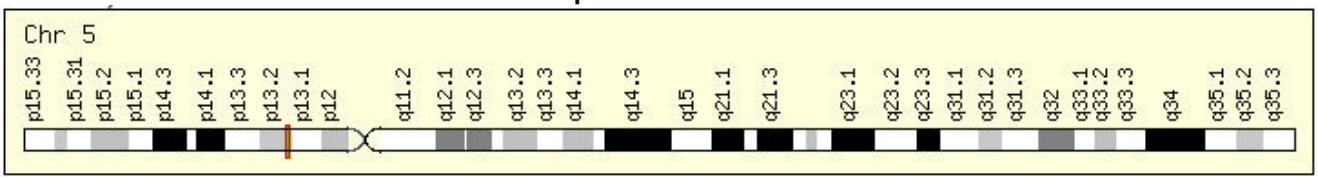

```
Query  252   GGAGGACCCCAGGGCTTTCTTGGCTCAGCAGATGTTGCCACCATCAACTCCACGGGCATC  311
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2828  GGAGGACCCCAGGGCTTTCTTGGCTCAGCAGATGTTGCCACCATCAACTCCACGGGCATC  2887

Query  312   ACCCTCTCCCTGTCCTCAGAGGAGTCAGAGTCCAGCAAGGAATCAGAGGGAAGCCTGCAG  371
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2888  ACCCTCTCCCTGTCCTCAGAGGAGTCAGAGTCCAGCAAGGAATCAGAGGGAAGCCTGCAG  2947

Query  372   AGGACAGGGTCAGGATCTGGAGGACATGTGCTGTCAAGAGCATCAGCAGGAGCTGGCACA  431
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2948  AGGACAGGGTCAGGATCTGGAGGACATGTGCTGTCAAGAGCATCAGCAGGAGCTGGCACA  3007

Query  432   GGACCCGGCTCCCCCTCGGCTGCCCCTTTGGACCAGAACAAGAAAAGGAGCAGCAGCATA  491
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3008  GGACCCGGCTCCCCCTCGGCTGCCCCTTTGGACCAGAACAAGAAAAGGAGCAGCAGCATA  3067

Query  492   GCCTCCACCCTGGGGCTGAAAAAGCTCTTCTCAGCCCTGGGCCAGAGTTCCCGGCCCAAG  551
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3068  GCCTCCACCCTGGGGCTGAAAAAGCTCTTCTCAGCCCTGGGCCAGAGTTCCCGGCCCAAG  3127

Query  552   CTGGGCAAGTCCCGCAGCTACAGTGTGGAGCAGTTGCAGCCCGCCCCGCCTGGCCTGACG  611
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3128  CTGGGCAAGTCCCGCAGCTACAGTGTGGAGCAGTTGCAGCCCGCCCCGCCTGGCCTGACG  3187

Query  612   TCACAGTCCAGGGCCCCATCGTTACAATCCCTGCACCCGGTGTCACCCTCTCACCAGCGT  671
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3188  TCACAGTCCAGGGCCCCATCGTTACAATCCCTGCACCCGGTGTCACCCTCTCACCAGCGT  3247

Query  672   CGGAAAGCTGCCTCTTTTCAGAACCTCCATTCTCTGCTGAGCAGCAAGGGGGACCGGTCC  731
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Sbjct  3248  CGGAAAGCTGCCTCTTTTCAGAACCTCCATTCTCTGCTGAGCAGCAAGGGGAACCGGTCC  3307

Query  732   AGCCTCTACCTGGTAGCAGGGCCAGGGGACCACAGTGCAGCTGGCA-GCCGGCCAAGACT  790
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Sbjct  3308  AGCCTCTACCTGGTAGCAGGGCCAGGGGACCACAGTGCAGCTGGCAGGCCGGCCAAGACT  3367

Query  791   TCACCACGGCGTGCCCTCAGTGTGGAGGACGTGGGTGCTCCCAGCCTGTCTCGCACCGTG  850
              ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Sbjct  3368  TCACCACGGCGTGCCCTCAGTGTGGAGGACGTGGGTGCTCCCAGCCTGGCTCGCACCGTG  3427

Query  851   GGCCGCCTGGTGGAGGTGTTCCCAGACAGCACCAGCCAGCTGCAGCTGCAGCGCTCTCCA  910
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||| |||
Sbjct  3428  GGCCGCCTGGTGGAGGTGTTCCCAGACGGCACCAGCCAGCTGCAGCTGCAGCGCTCCCCA  3487
```

```
Query  911   -AGGGCACTTTCGGCTTCTGCGT-GCCTCTGGGAAT-GGCGCCCAGACTCAGGGACGCAC  967
              ||||||||||||||||||||| |||||||||||| |||||||||||||||| | || |
Sbjct  3488  GGGGGCACTTTCGGCTTCTGCGTGGCCTCTGGGAATGGGCGCCCAGACTCAGGTATGCCC  3547

Query  968   TCTC  971
              ||||
Sbjct  3548  TCTC  3551
```

WDR70- ANKLE2

TTGGGGTTCTGCGGCGTGCGGCCAGCCATGGAGCGCTCTGGGCCCAGCGAAGTGACAGGCTCAGACGCATCGGGACCGGACCCGCAGCTTGCGGTCACCATG
GCTTCACGGGGTTCGGTAAAAAAGCTCGCACATTTGACTTGGAAGCAATGTTTGAACAAACTCGAAGAACAGCTGTGGAAAGAAGTTGCAAAACACTGGGTGA
AATGACAATGGATGCTCTGTTGGCTCGATTGAAACTTCTGAATCCAGATGACCTTAGAGAAGAAATCGTCAAAGCCGGATTGAAATGTGGACCCATTACATCA
ACTACAAGGTTCATTTTTGAGAAAAAATTGGCTCAGGCTTTACTGGAGCAAGGAGGAAGGCTGTCTTCTTTCTACCACCATGAGGCAGGTGTCACAGCTCTCA
GCCAGGACCCACAAAGGATTTTGAAGCCAGCTGAAGGGAACCCAACTGATCAGGCTGGTTTTTCTGAAGACAGAGATTTTGGTTACAGTGTGGGCCTGAATCC
TCCAGAGGAGGAAGCTGTGACATCCAAGACCTGCTCGGTGCCTCCTAGTGACACCGACACCTACAGAGCTGGAGCGACTGCGTCTAAGGAGCCGCCCCTGTAC
TATGGGGTGTGTCCAGTGTATGAGGACGTCCCAGCGAGAAATGAAAGGATCTATGTTTATGAAAATAAAAAGGAAGCATTGCAAGCTGTCAAGATGATCAAAC
GGTCCCGATTTAAAGCTTTTTCTACCAGAGAAGACGCTGAGAAATTTGCTAGAGGAATTTGTGATTATTTCCCTTCTCCAAGCAAAACGTCCTTACCACTGTC
TCCTGTGAAAACAGCTCCACTCTTTAGCAATGACAGGTTGAAAGATGGTTTGTGCTTGTCGGATCAGAACAGTCAACAAAGAGCGAGCGAACAGTTACAAAA
TCCCCGCACGCAGACCTCACCGCCAGCTTCGGAAAGCTGTGAGAGAGACGAGACACCATTTCTGACCTTATCTGAGCACCCCGGTATCTGATAGCTCAGGAG
CACCCACTATCGTGCAGAGATGCAGTACACGTGATGCATGTGCTGCAAGAGAACAGCTTCATCTGCAGCTGACTCTGACGTCTGAAACCTGACTTCTGAGGC
GAGTACTGATGACACAAGCATGCGTGCGAGCGATTCCGTTACTGGGGGGGACCTGTTATTCTCTACACCCCCGACA

5’ partner: WDR70



Junction point
exon=3 150..233
/gene="WDR70"
/gene_synonym="FLJ10233"

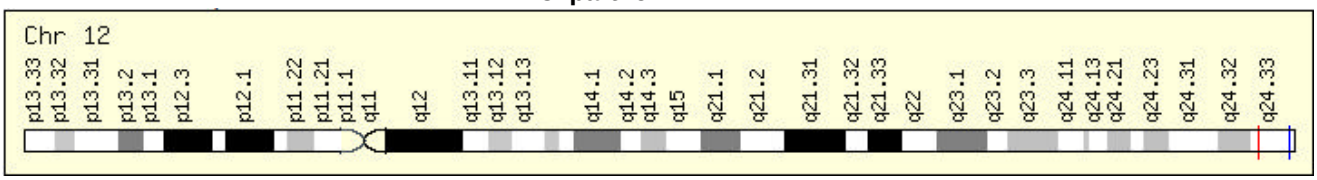
BLAST vs mRNA

>ref|NM_018034.2| Homo sapiens WD repeat domain 70 (WDR70), mRNA
Length=2247
Score = 334 bits (370), Expect = 4e-95
Identities = 193/198 (97%), Gaps = 0/198 (0%)

Strand=Plus/Plus

```
Query  5    GGTTCTGCGGCGTGCGGCCAGCCATGGAGCGCTCTGGGCCCAGCGAAGTGACAGGCTCAG  64
            ||||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  36   GGTTCTGGGGTGTGCGGCCAGCCATGGAGCGCTCTGGGCCCAGCGAAGTGACAGGCTCAG  95

Query  65   ACGCATCGGGACCGGACCCGCAGCTTGCGGTCACCATGGGCTTCACGGGGTTCGGTAAAA  124
            |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  96   ACGCGTCGGGACCGGACCCGCAGCTTGCGGTCACCATGGGCTTCACGGGGTTCGGTAAAA  155

Query  125  AAGCTCGCACATTTGACTTGGAAGCAATGTTTGAACAAACTCGAAGAACAGCTGTGGAAA  184
            |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Sbjct  156  AAGCTCGCACATTTGACTTGGAAGCAATGTTTGAACAAACTCGAAGGACAGCTGTGGAAA  215

Query  185  GAAGTTGCAAAACACTGG  202
            ||||| ||||||||||||
Sbjct  216  GAAGTCGCAAAACACTGG  233
```

3’ partner: ANKLE2



Junction point
exon=2 249..707
/gene="ANKLE2"

/gene_synonym="FLJ22280; FLJ36132; KIAA0692; LEMD7"

BLAST vs mRNA

>ref|NM_015114.1| Homo sapiens ankyrin repeat and LEM domain containing 2 (ANKLE2), mRNA
Length=4491
GENE ID: 23141 ANKLE2 | ankyrin repeat and LEM domain containing 2 [Homo sapiens]
Score = 1517 bits (1682), Expect = 0.0
Identities = 986/1049 (93%), Gaps = 45/1049 (4%)

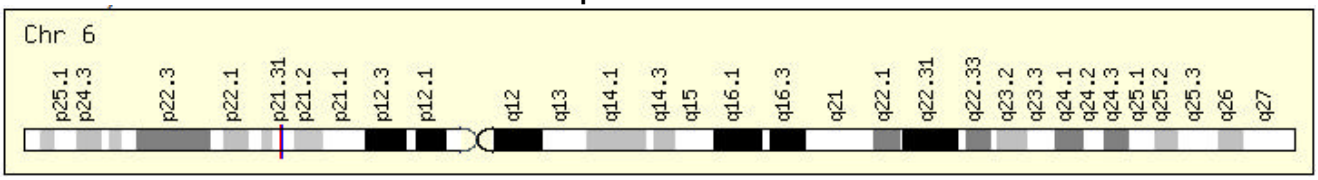
Strand=Plus/Plus

```
Query  202   GGTGAAATGACAATGGATGCTCTGTTGGCTCGATTGAAACTTCTGAATCCAGATGACCTT  261
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  248   GGTGAAATGACAATGGATGCTCTGTTGGCTCGATTGAAACTTCTGAATCCAGATGACCTT  307

Query  262   AGAGAAGAAATCGTCAAAGCCGGATTGAAATGTGGACCCATTACATCAACTACAAGGTTC  321
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  308   AGAGAAGAAATCGTCAAAGCCGGATTGAAATGTGGACCCATTACATCAACTACAAGGTTC  367

Query  322   ATTTTTGAGAAAAAATTGGCTCAGGCTTTACTGGAGCAAGGAGGAAGGCTGTCTTCTTTC  381
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  368   ATTTTTGAGAAAAAATTGGCTCAGGCTTTACTGGAGCAAGGAGGAAGGCTGTCTTCTTTC  427

Query  382   TACCACCATGAGGCAGGTGTCACAGCTCTCAGCCAGGACCCACAAAGGATTTTGAAGCCA  441
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  428   TACCACCATGAGGCAGGTGTCACAGCTCTCAGCCAGGACCCACAAAGGATTTTGAAGCCA  487

Query  442   GCTGAAGGGAACCCAACTGATCAGGCTGGTTTTTCTGAAGACAGAGATTTTGGTTACAGT  501
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  488   GCTGAAGGGAACCCAACTGATCAGGCTGGTTTTTCTGAAGACAGAGATTTTGGTTACAGT  547

Query  502   GTGGGCCTGAATCCTCCAGAGGAGGAAGCTGTGACATCCAAGACCTGCTCGGTGCCTCCT  561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Sbjct  548   GTGGGCCTGAATCCTCCAGAGGAGGAAGCTGTGACATCCAAGACCTGCTCGGTGCCCCCT  607

Query  562   AGTGACACCGACACCTACAGAGCTGGAGCGACTGCGTCTAAGGAGCCGCCCCTGTACTAT  621
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  608   AGTGACACCGACACCTACAGAGCTGGAGCGACTGCGTCTAAGGAGCCGCCCCTGTACTAT  667

Query  622   GGGGTGTGTCCAGTGTATGAGGACGTCCCAGCGAGAAATGAAAGGATCTATGTTTATGAA  681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  668   GGGGTGTGTCCAGTGTATGAGGACGTCCCAGCGAGAAATGAAAGGATCTATGTTTATGAA  727

Query  682   AATAAAAAGGAAGCATTGCAAGCTGTCAAGATGATCAAAGGGTCCCGATTTAAAGCTTTT  741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  728   AATAAAAAGGAAGCATTGCAAGCTGTCAAGATGATCAAAGGGTCCCGATTTAAAGCTTTT  787

Query  742   TCTACCAGAGAAGACGCTGAGAAATTTGCTAGAGGAATTTGTGATTATTTCCCTTCTCCA  801
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  788   TCTACCAGAGAAGACGCTGAGAAATTTGCTAGAGGAATTTGTGATTATTTCCCTTCTCCA  847

Query  802   AGCAAAACGTCCTTACCACTGTCTCCTGTGAAAACAGCTCCACTCTTTAGCAATGACAGG  861
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  848   AGCAAAACGTCCTTACCACTGTCTCCTGTGAAAACAGCTCCACTCTTTAGCAATGACAGG  907

Query  862   TTGAAAGATGGTTTGTGCTTGTCGG-ATCAG-AACAGTCAACAAAGAGCGAGCGAACAGT  919
             |||||||||||||||||||||||| ||||| ||||||||||||||||||||||||||||
Sbjct  908   TTGAAAGATGGTTTGTGCTTGTCGGAATCAGAAACAGTCAACAAAGAGCGAGCGAACAGT  967
```

TAF11- FAM62B

```
TTGGCTTCTGCCTCAGGCATCTCCGCGATCTCCTCTCCCCTCCAATCCTATCCGTGATGGACGATGCCCACGAGTCGCCCTCCGACAAAGGTGGAGAGACAG
GGAGTCGGATGAGACGGCCGCTGTGCCCGGGGACCCGGGGGCTACCGACACCGATGGAATCCCAGAGGAAACTGACGGAGACGCAGATGTGGACTTGAAAGA
GCTGCAGCGGAGGAAGGCGAGATTCGATACAAAACCAATGAACCTGTGTGGGAGGAAAACTTCACTTTCTTCATTCACAATCCCAAGCGCCAGGACCTTGAA
TTGAGGTCAGAGACGAGCAGCACCAGTGTTCCCTGGGGAACCTGAAGGTCCCCCTCAGCCAGCTGCTCACCAGTGAGGACATGACTGTGAGCCAGCGCTTCC
GCTCAGTAACTCGGGTCCAAACAGCACCATCAAGATGAAGATTGCCCTGCGGGTGCTCCATCTCGAAAAGCGAGAAAGGCCTCCAGACCACCAACACTCAGC
CAAGTCAAACGTCCCTCTGTGTCCAAAGAGGGGAGGAAAACATCCATCAAATCTCATATGTCTGGGTCTCCAGGCCCTGGTGGCAGCAACACAGCTCCATCC
CACCAGTCATTGGGGGCAGTGATAAGCCTGGTATGGAAGAAAAGGCCCAGCCCCCTGAGGCCGGCCCTCAAGGGCTGCACGACCTGGCAGAAGCTCCTCCAG
CTCCTGGCCTCCCCAGGCCACATCTCAGTCAAGGAGCCGACCCCCAGCATCGCCTCGGACATCTCGCTGCCCATCGCCACCCAGGAGCTGCGGCAAAACTGA
CAGCTGGAAAACGGGACGACCCTGGGACAGTCTCCACTGGGGCAGATCCAGCTGACCATCCGGCACAGCTCGCAGAGAACAAGCTTATCGTGGTCGTGCATG
CTGCAGAAACCTCATTGCTTCTCTGAGACGCTCTGACCCTATGTCGCATGTATTTATTACGGACAGAGCGGTCAGGAGAGAAACCCCCGGTCAAGAAAATTA
TCAATGTTGATCAGCTGATCAGTTCGTACAAAATGCAAGAACCTGACGTGCTGAGACTGCGCTCGTCAGACAGGTCCTGGCAGATTGTGTCTGCACCTGAAA
CTTGCCA
```

5’ partner: TAF11



Junction point
exon=1 1..256
/gene="TAF11"
/gene_synonym="MGC:15243; PRO2134; TAF2I; TAFII28"

BLAST vs mRNA

>ref|NM_005643.2| Homo sapiens TAF11 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 28kDa (TAF11), mRNA
Length=1599
GENE ID: 6882 TAF11 | TAF11 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 28kDa [Homo sapiens]
Score = 405 bits (448), Expect = 2e-116
Identities = 224/224 (100%), Gaps = 0/224 (0%)

Strand=Plus/Plus

```
Query  4    GCTTCTGCCTCAGGCATCTCCGCGATCTCCTCTCCCCTCCAATCCTATCCGTGATGGACG  63
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  33   GCTTCTGCCTCAGGCATCTCCGCGATCTCCTCTCCCCTCCAATCCTATCCGTGATGGACG  92

Query  64   ATGCCCACGAGTCGCCCTCCGACAAAGGTGGAGAGACAGGGGAGTCGGATGAGACGGCCG  123
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  93   ATGCCCACGAGTCGCCCTCCGACAAAGGTGGAGAGACAGGGGAGTCGGATGAGACGGCCG  152

Query  124  CTGTGCCCGGGGACCCGGGGGCTACCGACACCGATGGAATCCCAGAGGAAACTGACGGAG  183
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  153  CTGTGCCCGGGGACCCGGGGGCTACCGACACCGATGGAATCCCAGAGGAAACTGACGGAG  212

Query  184  ACGCAGATGTGGACTTGAAAGAAGCTGCAGCGGAGGAAGGCGAG  227
            |||||||||||||||||||||||||||||||||||||||||||
Sbjct  213  ACGCAGATGTGGACTTGAAAGAAGCTGCAGCGGAGGAAGGCGAG  256
```

3’ partner: FAM62B

Junction point
exon=15 1705..1791
/gene="FAM62B"
/gene_synonym="CHR2SYT; ESYT2; KIAA1228"

BLAST vs mRNA

>ref|NM_020728.2| Homo sapiens family with sequence similarity 62 (C2 domain containing) member B (FAM62B), mRNA
Length=5957
GENE ID: 57488 FAM62B | family with sequence similarity 62 (C2 domain containing) member B [Homo sapiens]
Score = 1321 bits (1464), Expect = 0.0
Identities = 802/828 (96%), Gaps = 18/828 (2%)

Strand=Plus/Plus

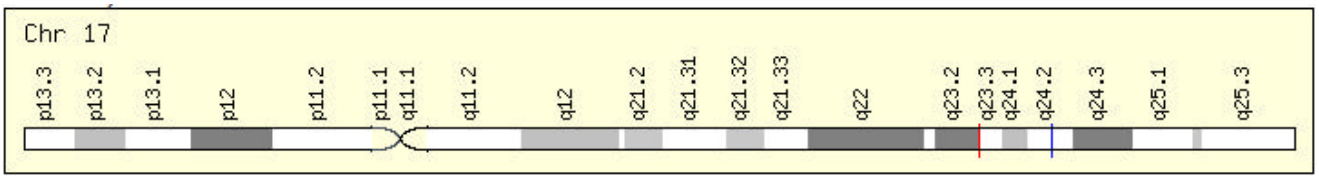
```
Query  217   AGGAAGGCGAGATTCGATACAAAACCAATGAACCTGTGTGGGAGGAAAACTTCACTTTCT  276
             ||||  || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1694  AGGAGAGCAAGATTCGATACAAAACCAATGAACCTGTGTGGGAGGAAAACTTCACTTTCT  1753
```

```
Query  277   TCATTCACAATCCCAAGCGCCAGGACCTTGAAGTTGAGGTCAGAGACGAGCAGCACCAGT  336
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1754  TCATTCACAATCCCAAGCGCCAGGACCTTGAAGTTGAGGTCAGAGACGAGCAGCACCAGT  1813

Query  337   GTTCCCTGGGGAACCTGAAGGTCCCCCTCAGCCAGCTGCTCACCAGTGAGGACATGACTG  396
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1814  GTTCCCTGGGGAACCTGAAGGTCCCCCTCAGCCAGCTGCTCACCAGTGAGGACATGACTG  1873

Query  397   TGAGCCAGCGCTTCCAGCTCAGTAACTCGGGTCCAAACAGCACCATCAAGATGAAGATTG  456
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1874  TGAGCCAGCGCTTCCAGCTCAGTAACTCGGGTCCAAACAGCACCATCAAGATGAAGATTG  1933

Query  457   CCCTGCGGGTGCTCCATCTCGAAAAGCGAGAAAGGCCTCCAGACCACCAACACTCAGCTC  516
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1934  CCCTGCGGGTGCTCCATCTCGAAAAGCGAGAAAGGCCTCCAGACCACCAACACTCAGCTC  1993

Query  517   AAGTCAAACGTCCCTCTGTGTCCAAAGAGGGGAGGAAAACATCCATCAAATCTCATATGT  576
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1994  AAGTCAAACGTCCCTCTGTGTCCAAAGAGGGGAGGAAAACATCCATCAAATCTCATATGT  2053

Query  577   CTGGGTCTCCAGGCCCTGGTGGCAGCAACACAGCTCCATCCACACCAGTCATTGGGGGCA  636
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2054  CTGGGTCTCCAGGCCCTGGTGGCAGCAACACAGCTCCATCCACACCAGTCATTGGGGGCA  2113

Query  637   GTGATAAGCCTGGTATGGAAGAAAAGGCCCAGCCCCCTGAGGCCGGCCCTCAAGGGCTGC  696
             ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Sbjct  2114  GTGATAAGCCTGGTATGGAAGAAAAGGCCCAGCCCCCTGAGGCCGGCCCTCAGGGGCTGC  2173

Query  697   ACGACCT-GGCAGAAGCTCCTCCAGCCTCCTGGCCTCCCCAGGCCACATCTCAGTCAAGG  755
             ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2174  ACGACCTGGGCAGAAGCTCCTCCAGCCTCCTGGCCTCCCCAGGCCACATCTCAGTCAAGG  2233

Query  756   AGCCGACCCCCAGCATCGCCTCGGACATCTCGCTGCCCATCGCCACCCAGGAGCTGCGGC  815
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2234  AGCCGACCCCCAGCATCGCCTCGGACATCTCGCTGCCCATCGCCACCCAGGAGCTGCGGC  2293
```

NOL11- C7orf42

TCAAATTATAACAAGTCCAGCTGTGTGCAACTTTCAAACTGGAGAGTATGTTGTTGTACACGGTTAATAAGGTTTTAAGAATATGGAATAATGAAGATGTAA
CCTGGATAAAGTATTTAAAGCTACATTGTCAGCAGAAGTATATAGGATACTTTCAGTGCAAGGGACAGAACCCTTGGTGCTCTTCAAGGAAGGTGCTGTTCG
GGTTTAGAGGCCTTGCTTGCAGACCCCCAGCAGAAAATTGAAACTGTTATCTCTGATGAAGAAGTGATTAAACAGCCACCGCACTGTGTTCCTGACACGTAC
GCAACGCCACGTTCTGGTACAAGATCTTCACAACTGCCAGGGATGCCAACACAAAATACGCCCAAGATTACAATCCTTTCTGGTGTTATAAGGGGGCCATTG
AAAAGTTTATCATGCTTTAAATCCCAAGCTTACAGTGATTGTTCCAGATGATGACCGTTCATTAATAAATTTGCATATCAAGCACACCAGTTACTTCCTAAA
GTGATGGTGTTAACAATGTTTTGAAAAGATGTAATCTTGGGCTGACGTAGCAAAAAGAGTCAGAGCAATCAAGAAAAAATCCCGAGAAGGTGGATTTGG

5’ partner: NOL11



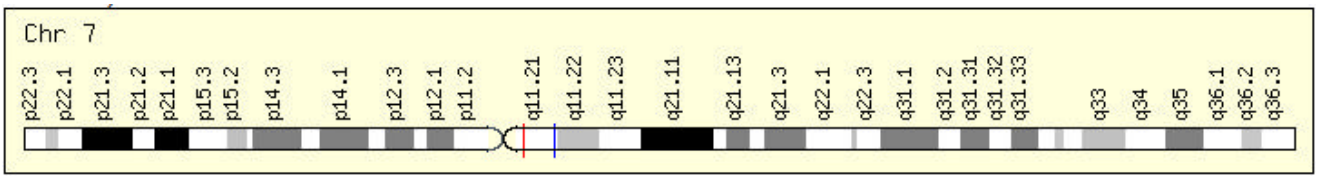
Junction point
exon=4 316..464
/gene="NOL11"
/gene_synonym="DKFZp586L0724"

BLAST vs mRNA

```
>ref|NM_015462.3| Homo sapiens nucleolar protein 11 (NOL11), mRNA
Length=2454
GENE ID: 25926 NOL11 | nucleolar protein 11 [Homo sapiens]
Score =  484 bits (536),  Expect = 2e-140
Identities = 275/278 (98%), Gaps = 1/278 (0%)

Strand=Plus/Plus

Query  5    CTTGGGGAGCTGGTCAGTGAAACAAGGTCAAATTATAACATGTCCAGCTGTGTGCAACTT  64
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  162  CTTGGGGAGCTGGTCAGTGAAACAAGGTCAAATTATAACATGTCCAGCTGTGTGCAACTT  221

Query  65   TCAAACTGGAGAGTATGTTGTTGTACACGATAATAAGGTTTTAAGAATATGGAATAATGA  124
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  222  TCAAACTGGAGAGTATGTTGTTGTACACGATAATAAGGTTTTAAGAATATGGAATAATGA  281

Query  125  AGATGTAAACCTGGATAAAGTATTTAAAGCTACATTGTCAGCAGAAGTATATAGGATACT  184
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  282  AGATGTAAACCTGGATAAAGTATTTAAAGCTACATTGTCAGCAGAAGTATATAGGATACT  341

Query  185  TTCAGTGCAAGGGACAGAACCCTTGGTGCTCTTCAAGGAAGGTGCTGTTCGTGGTTTAGA  244
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  342  TTCAGTGCAAGGGACAGAACCCTTGGTGCTCTTCAAGGAAGGTGCTGTTCGTGGTTTAGA  401
```

```
Query  245  GGCCTTGCTTGCAGACCCCCAGCAGAAAATTGAAACTGTTATCTCTGATGAAGAAGTGAT  304
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  402  GGCCTTGCTTGCAGACCCCCAGCAGAAAATTGAAACTGTTATCTCTGATGAAGAAGTGAT  461

Query  305  TAAA  308
            ||||
Sbjct  462  TAAA  465
```


3’ partner: C7orf42



Junction point
exon=5 861..1044
/gene="C7orf42"
/gene_synonym="FLJ10099; FLJ13090"

BLAST vs mRNA

>ref|NM_017994.4| Homo sapiens chromosome 7 open reading frame 42 (C7orf42), mRNA
Length=4238
GENE ID: 55069 C7orf42| chromosome 7 open reading frame 42 [Homo sapiens]
Score = 475 bits (526), Expect = 2e-137
Identities = 309/338 (91%), Gaps = 1/338 (0%)

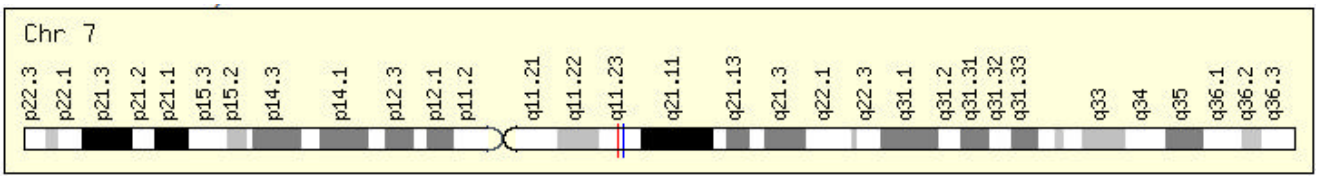
Strand=Plus/Plus

```
Query  308   ACAGCCACCGCACTGTGTTCCTGACACGTACAGCAACGCCACGCTCTGGTACAAGATCTT  367
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  861   ACAGCCACCGCACTGTGTTCCTGACACGTACAGCAACGCCACGCTCTGGTACAAGATCTT  920

Query  368   CACAACTGCCAGAGATGCCAACACAAAATACGCCCAAGATTACAATCCTTTCTGGTGTTA  427
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  921   CACAACTGCCAGAGATGCCAACACAAAATACGCCCAAGATTACAATCCTTTCTGGTGTTA  980

Query  428   TAAGGGGGCCATTGGAAAAGTCTATCATGCTTTAAATCCCAAGCTTACAGTGATTGTTCC  487
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  981   TAAGGGGGCCATTGGAAAAGTCTATCATGCTTTAAATCCCAAGCTTACAGTGATTGTTCC  1040

Query  488   AGATGATGACCGTTCATTAATAAATTTGCATCTCATGCACACCAGTTACTTCCTCTTTGT  547
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1041  AGATGATGACCGTTCATTAATAAATTTGCATCTCATGCACACCAGTTACTTCCTCTTTGT  1100

Query  548   GATGGTGATAACAATGTTTTGCTATGCTGTTATCAAGGGCAGACCTAGCAAATTGCGTCA  607
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1101  GATGGTGATAACAATGTTTTGCTATGCTGTTATCAAGGGCAGACCTAGCAAATTGCGTCA  1160

Query  608   GAGCAATCCTGAATTTTGTCCCGAGAAGGTGGCTTTGGCTGAAGCCTAATTCCACAGCTC  667
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1161  GAGCAATCCTGAATTTTGTCCCGAGAAGGTGGCTTTGGCTGAAGCCTAATTCCACAGCTC  1220

Query  668   CTTGTTTTTAGAGAGAGACTGAGAGAACCATAATCCTTGCCTGCTGAACCCAGCCTGGGC  727
             ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1221  CTTGTTTTTTGAGAGAGACTGAGAGAACCATAATCCTTGCCTGCTGAACCCAGCCTGGGC  1280

Query  728   CTGGATGCTCTGTGAATACATTATCTTGCGATGTTGGGTTATTCCAGCCAAAGACATTTC  787
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1281  CTGGATGCTCTGTGAATACATTATCTTGCGATGTTGGGTTATTCCAGCCAAAGACATTTC  1340

Query  788   AAGTGCCTGTAACTGATTTGTACATATTTAT-AAAATCTATTCAGA  832
             |||||||||||||||||||||||||||||| ||||||||||||||
Sbjct  1341  AAGTGCCTGTAACTGATTTGTACATATTTATAAAAATCTATTCAGA  1386
```

NSUN5B- CPSF3

TTGGGCCGCCAGGGCTCTATCAAGGGGCTGGTGTACTCCAGCAACTTCCAGAACGTGAAGCAGCTGTACGCGCTGGTGTGCGAAACGCAGCGCTACTCCGCC
TGCTGGATGCCGTGATCTCCAGCGCCGGCCTCCTCAGTGCGAAGAAGCTGCAGCCGCACCTGGCCAAGGGTGCAGTACAGAAGGTTTTTAAAAAATTAGAAA
GCACGTTTACAGCAAGAGGTTGGAGATCATGCTCCAGGACATATTTGGAGAAGACTGTGTAAGTGTAAAGGATGACTCTATTTTTAGCGTCACAGTGGACGG
AAAACTGCCAACCTTAACTTGGAGACACGGACTGTAGAATGTGAAGAGGGAAGTGAAGACGATGAATCCCTCCGAGAAATGGTGGAGCTGGCTGCACAGAGA
TGTACGAGGCCCTGACGCCAGTTCACTGAGACTGTGCCTGTATATGAACTTTGAAAAAATACTTGACTTTACTTTTGTTACCTAAAATAAAATGCATTCGTT
TTTTG

5’ partner: NSUN5B



Junction point
exon=2 121..241
/gene="NSUN5P1"
/gene_synonym="FLJ99347; MGC129801; NSUN5B; WBSCR20B"

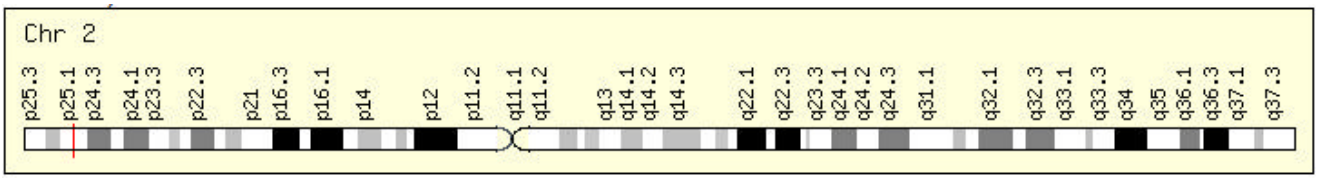
BLAST vs mRNA

>ref|NR_033322.2| Homo sapiens NOP2/Sun domain family, member 5 pseudogene 1 (NSUN5P1), non-coding RNA
Length=1751
GENE ID: 155400 NSUN5P1| NOP2/Sun domain family, member 5 pseudogene 1 [Homo sapiens]
Score = 311 bits (168), Expect = 3e-84
Identities = 168/168 (100%), Gaps = 0/168 (0%)

Strand=Plus/Plus

```
Query  5    GCCGCCAGGGCTCTATCAAGGGGCTGGTGTACTCCAGCAACTTCCAGAACGTGAAGCAGC  64
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  74   GCCGCCAGGGCTCTATCAAGGGGCTGGTGTACTCCAGCAACTTCCAGAACGTGAAGCAGC  133

Query  65   TGTACGCGCTGGTGTGCGAAACGCAGCGCTACTCCGCCGTGCTGGATGCCGTGATCTCCA  124
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  134  TGTACGCGCTGGTGTGCGAAACGCAGCGCTACTCCGCCGTGCTGGATGCCGTGATCTCCA  193

Query  125  GCGCCGGCCTCCTCAGTGCGAAGAAGCTGCAGCCGCACCTGGCCAAGG  172
            ||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  194  GCGCCGGCCTCCTCAGTGCGAAGAAGCTGCAGCCGCACCTGGCCAAGG  241
```

3’ partner: CPSF3



Junction point
exon=16 1822..1891
/gene="CPSF3"
/gene_synonym="CPSF; CPSF-73; CPSF73; YSH1"

BLAST vs mRNA

>ref|NM_016207.2| Homo sapiens cleavage and polyadenylation specific factor 3, 73kDa (CPSF3), mRNA
Length=2286
Score = 609 bits (674), Expect = 5e-178
Identities = 343/347 (98%), Gaps = 0/347 (0%)

Strand=Plus/Plus

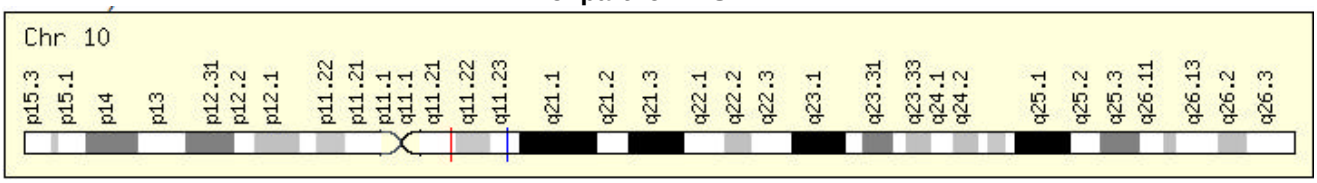
```
Query  169   AAGGGTGCAGTACAGAAGGTTTTTAAAAAATTAGAAATGCACGTTTACAGCAAGAGGTTG  228
             || |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Sbjct  1818  AAAGGTGCAGTACAGAAGGTTTCTAAAAAATTAGAAATGCACGTTTACAGCAAGAGGTTG  1877

Query  229   GAGATCATGCTCCAGGACATATTTGGAGAAGACTGTGTAAGTGTAAAGGATGACTCTATT  288
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1878  GAGATCATGCTCCAGGACATATTTGGAGAAGACTGTGTAAGTGTAAAGGATGACTCTATT  1937

Query  289   TTTAGCGTCACAGTGGACGGGAAAACTGCCAACCTTAACTTGGAGACACGGACTGTAGAA  348
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1938  CTTAGCGTCACAGTGGACGGGAAAACTGCCAACCTTAACTTGGAGACACGGACTGTAGAA  1997
```

```
Query  349   TGTGAAGAGGGAAGTGAAGACGATGAATCCCTCCGAGAAATGGTGGAGCTGGCTGCACAG  408
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1998  TGTGAAGAGGGAAGTGAAGACGATGAATCCCTCCGAGAAATGGTGGAGCTGGCTGCACAG  2057

Query  409   AGACTGTACGAGGCCCTGACGCCAGTTCACTGAGACTGTGCCTGTATATGAACTTTGAAA  468
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2058  AGACTGTACGAGGCCCTGACGCCAGTTCACTGAGACTGTGCCTGTATATGAACTTTGAAA  2117

Query  469   AAATACTTGACTTTACTTTTGTTACCTAAAATAAAATGCATTCGTTT  515
             |||||||||||| ||||||||||||||||||||||||||||||||||
Sbjct  2118  AAATACTTGACTCTACTTTTGTTACCTAAAATAAAATGCATTCGTTT  2164
```

MSMB- POLE4

AATGCCAACTTTGTACAAAAAAGTTGGAGGAGTCCTGCTTATCACAATGAATGTTCTCCTGGGCAGCGTTGTGATCTTTGCCACCTTCGTGACTTTATGCAA
GCATCATGCTATTTCATACCTAATGAGGGAGTTCCAGGAGATTCAACCAGGAAATGCATGGATCTCAAAGGAAACAAACACCCAATAAACTCGGAGTGGCAG
CTGACAACTGTGAGACATGCACTTGCTACGAAACAGAAATTTCATGTTGCACCCTATAATGCAATAGAAGCTGTGGATGAATTTGCTTTTCTGGAAGGTACT
TAGATTGATTGCCGAGCGGGGCAGTTTTGTGAGCCTTCATCTGAAGCCTTCAGTTCACCCCTCTGCACAGGCCTCAGCTTTGAAGAACGGAGTCTTTGCACT
ACACACACTCTTCCTGTTCTGCCTTCACCTATGCCGGGATAAGCAGAGATCTCATCAATTAGCTCTTCTCTGCAAGGTCTTCCACTGTTTCTGTCTGTCTTC
ATATCAAGCCTGGATGCAGCTGCTGCTGCTTAGAGCAGAGATGAAGAAAGTGTTCTGCATAAGTGGCTTCCTGAATGATGAGGACCAGAATAAAGGTTTTTG
TCAACCTC

5’ partner: MSMB



Junction point
exon=3 142..247
/gene="MSMB"
/gene_synonym="HPC13; IGBF; MSP; MSPB; PN44; PRPS; PSP;
PSP-94; PSP57; PSP94"

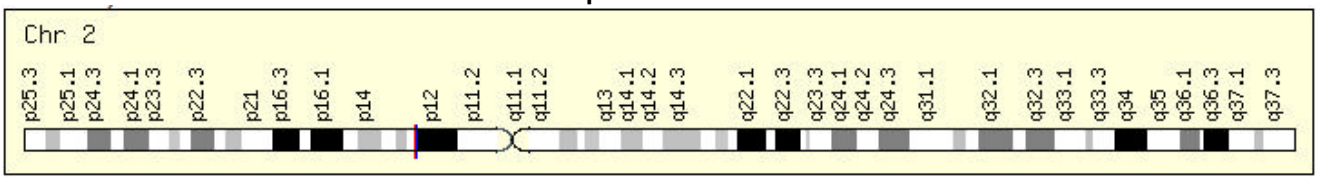
BLAST vs mRNA

>ref|NM_002443.2| Homo sapiens microseminoprotein, beta- (MSMB), transcript variant PSP94, mRNA
Length=572
GENE ID: 4477 MSMB microseminoprotein, beta- [Homo sapiens]
Score = 423 bits (468), Expect = 1e-122
Identities = 234/234 (100%), Gaps = 0/234 (0%)

Strand=Plus/Plus

```
Query  28    AGGAGTCCTGCTTATCACAATGAATGTTCTCCTGGGCAGCGTTGTGATCTTTGCCACCTT  87
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  14    AGGAGTCCTGCTTATCACAATGAATGTTCTCCTGGGCAGCGTTGTGATCTTTGCCACCTT  73

Query  88    CGTGACTTTATGCAATGCATCATGCTATTTCATACCTAATGAGGGAGTTCCAGGAGATTC  147
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  74    CGTGACTTTATGCAATGCATCATGCTATTTCATACCTAATGAGGGAGTTCCAGGAGATTC  133

Query  148   AACCAGGAAATGCATGGATCTCAAAGGAAACAAACACCCAATAAACTCGGAGTGGCAGAC  207
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  134   AACCAGGAAATGCATGGATCTCAAAGGAAACAAACACCCAATAAACTCGGAGTGGCAGAC  193

Query  208   TGACAACTGTGAGACATGCACTTGCTACGAAACAGAAATTTCATGTTGCACCCT  261
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  194   TGACAACTGTGAGACATGCACTTGCTACGAAACAGAAATTTCATGTTGCACCCT  247
```

3’ partner: POLE4



Junction point
exon=3 331..372
/gene="POLE4"
/gene_synonym="p12"

BLAST vs mRNA

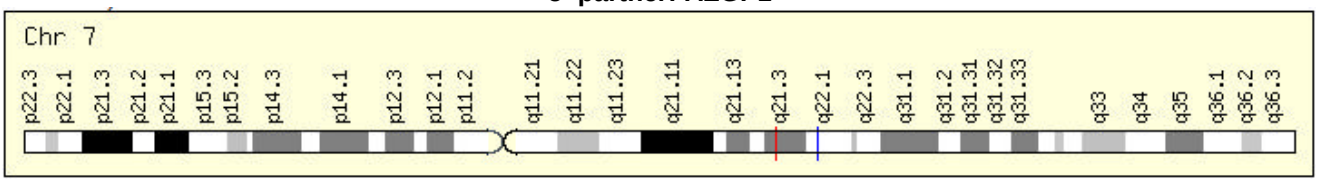
```
>ref|NM_019896.2| Homo sapiens polymerase (DNA-directed), epsilon 4 (p12 subunit) (POLE4), mRNA
Length=710
GENE ID: 56655 POLE4| polymerase (DNA-directed), epsilon 4 (p12 subunit) [Homo sapiens]
Score =  654 bits (724),  Expect = 0.0
Identities = 364/365 (99%), Gaps = 0/365 (0%)

Strand=Plus/Plus
```

```
Query  262  ATAATGCAATAGAAGCTGTGGATGAATTTGCTTTTCTGGAAGGTACTTTAGATTGATTGC  321
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  331  ATAATGCAATAGAAGCTGTGGATGAATTTGCTTTTCTGGAAGGTACTTTAGATTGATTGC  390

Query  322  CGAGCGGGGCAGTTTTGTGAGCCTTCATCTGAAGCCTTCAGTTCACCCCTCTGCACAGGC  381
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  391  CGAGCGGGGCAGTTTTGTGAGCCTTCATCTGAAGCCTTCAGTTCACCCCTCTGCACAGGC  450

Query  382  CTCAGCTTTGAAGAACGGAGTCTTTGCACTTACACACACTCTTCCTGTTCTGCCTTCACC  441
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  451  CTCAGCTTTGAAGAACGGAGTCTTTGCACTTACACACACTCTTCCTGTTCTGCCTTCACC  510

Query  442  TATGCCGGGATAAGCAGAGATCTCATCAATTAGCTCTTCTCTGCAAGGTCTTCCACTGTT  501
            |||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Sbjct  511  TATGCCGGGATAAGCAGAGATCTCATCAATTAGCTCTTCTCTGCAAGGTCTTCCACTATT  570

Query  502  TCTGTCTGTCTTCCATATCAAGCCTGGATGCAGCTGCTGCTGCTTAGAGCAGAGATGAAG  561
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  571  TCTGTCTGTCTTCCATATCAAGCCTGGATGCAGCTGCTGCTGCTTAGAGCAGAGATGAAG  630

Query  562  AAAGTGTTCTGCATAAGTGGCTTCCTGAATGATGAGGACCAGAATAAAGGTTTTTGATCA  621
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  631  AAAGTGTTCTGCATAAGTGGCTTCCTGAATGATGAGGACCAGAATAAAGGTTTTTGATCA  690

Query  622  ACCTC  626
            |||||
Sbjct  691  ACCTC  695
```

AZGP1-GJC3

TTGGAGAATGGTGCCTGTCCTGCTGTCTCTGCTGCTGCTTCTGGGTCCTGCTGTCCCCCAGGAGAACCAAGATGGTCGTTACTCTCTGACCTATATCTACAC
GGGCTGTCCAAGCATGTTGAAGACGTCCCCGCGTTTCAGGCCCTTGGCTCACTCAATGACCTCCAGTTCTTTAGATACAACAGTAAAGACAGGAAGTCTCAG
CCATGGGACTCTGGAGACAGGTGGAAGGAATGGAGGATTGGAAGCAGGACAGCCAACTTCAGAAGGCCAGGGAGGACATCTTTATGGAGACCCTGAAAGACA
CGTGGAGTATTACAACGACAGTAACGGATGTGTGGCAGGTTCCTGCGGCGGCTGCTGGCGGAGGAGAGCCGGCGCTCCACCCCCGTGGGGCGCCTCTTGCTT
CCGTGCTCCTGGGATTCCGCCTTGTGCTGCTGGCTGCCAGTGGGCCTGGAGTCTATGGTGATGAGCAGAGTGAATTCGTGTGTCACACCCAGCAGCCGGGCT
CAAGGCTGCCTGCTTCGATGCCTTCCACCCCCTCTCCCCGCTGCGTTTCTGGGTCTTCCAGGTCATCTTGGTGGCTGTACCCAGCGCCCTCTATATGGGTTTC
ACTCTGTATCACGTGATCTGGCACTGGGAATTATCAGGAAAGGGGAAGGAGGAGGAGATCTCTGATCCAGGGACGGGAGGGCAACACAGATGTCTCTAGGGGC
TGGAAGCTCTCACGCTGCTCTGGGCTTATGTGGCTCAGCTGGGGGCTCGGCTTGTCCTGTAGGGGGCAGCCCTGGGGTTGCAGTACCACCTGTATGGGTTCCA
GATGCCCAGCTCCTTTGCATGTCGCCGAGAACCTTGCCTTGGTAGTATAACCTGCAATCTGTCCCGCTCCTCTGAGAAGACCATTTTTCCTAAAGACCATGTT
TGGAGTCAGCGGTTTCTGTCTCTTGTTTACTTTTTTGGAGCTTGTGCTTCTGGGTTTGGGGAGATGGTGGAGGACTGAGCACAATCTTCCTCTTTCTAAATAC
TCTAACTCGAAAGCACCGAAGACCAAGAAGCACGATAGCCTCCCAGTGGGTGATCAAGAGCCATTCAGAGCAGTTCAGGAGAGCTAGCTAGGAAAACAAGAAC
AGTTGAACTGAGAATGTCTGGAAGTGTGGAGAATAAAACTTTTTTGC

5’ partner: AZGP1



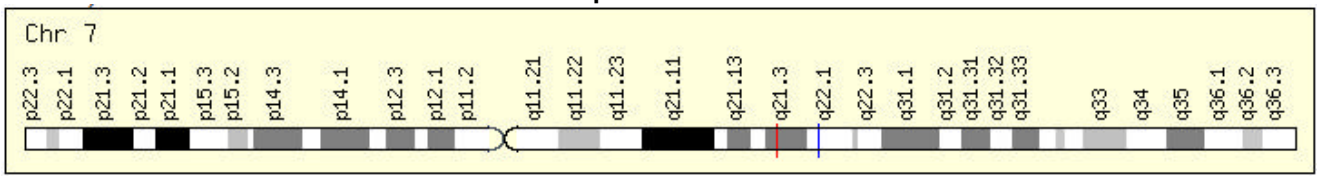
Junction point
exon=2 169..429
/gene="AZGP1"
/gene_synonym="ZA2G; ZAG"

BLAST vs mRNA

```
>ref|NM_001185.3| UniGene info linked to NM_001185.3GEO profiles info linked to NM_001185.3Gene info
linked to NM_001185.3Genome view with mapviewer linked to NM_001185.3 Homo sapiens alpha-2-glycoprotein
1, zinc-binding (AZGP1), mRNA
Length=1278
GENE ID: 563 AZGP1 | alpha-2-glycoprotein 1, zinc-binding [Homo sapiens]
Score =  614 bits (332),  Expect = 2e-174
Identities = 332/332 (100%), Gaps = 0/332 (0%)

Strand=Plus/Plus
```

```
Query  5    AGAATGGTGCCTGTCCTGCTGTCTCTGCTGCTGCTTCTGGGTCCTGCTGTCCCCCAGGAG  64
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  99   AGAATGGTGCCTGTCCTGCTGTCTCTGCTGCTGCTTCTGGGTCCTGCTGTCCCCCAGGAG  158

Query  65   AACCAAGATGGTCGTTACTCTCTGACCTATATCTACACTGGGCTGTCCAAGCATGTTGAA  124
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  159  AACCAAGATGGTCGTTACTCTCTGACCTATATCTACACTGGGCTGTCCAAGCATGTTGAA  218

Query  125  GACGTCCCCGCGTTTCAGGCCCTTGGCTCACTCAATGACCTCCAGTTCTTTAGATACAAC  184
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  219  GACGTCCCCGCGTTTCAGGCCCTTGGCTCACTCAATGACCTCCAGTTCTTTAGATACAAC  278

Query  185  AGTAAAGACAGGAAGTCTCAGCCCATGGGACTCTGGAGACAGGTGGAAGGAATGGAGGAT  244
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  279  AGTAAAGACAGGAAGTCTCAGCCCATGGGACTCTGGAGACAGGTGGAAGGAATGGAGGAT  338

Query  245  TGGAAGCAGGACAGCCAACTTCAGAAGGCCAGGGAGGACATCTTTATGGAGACCCTGAAA  304
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  339  TGGAAGCAGGACAGCCAACTTCAGAAGGCCAGGGAGGACATCTTTATGGAGACCCTGAAA  398

Query  305  GACATCGTGGAGTATTACAACGACAGTAACGG  336
            ||||||||||||||||||||||||||||||||
Sbjct  399  GACATCGTGGAGTATTACAACGACAGTAACGG  430
```

3’ partner: GJC3



Junction point
exon=1 1..781
/gene="GJC3"
/gene_synonym="CX29; CX30.2; CX31.3; GJE1"

BLAST vs mRNA

>ref|NM_181538.2| UniGene info linked to NM_181538.2GEO profiles info linked to NM_181538.2Gene info linked to NM_181538.2Genome view with mapviewer linked to NM_181538.2 Homo sapiens gap junction protein, gamma 3, 30.2kDa (GJC3), mRNA
Length=1131
GENE ID: 349149 GJC3 | gap junction protein, gamma 3, 30.2kDa [Homo sapiens]
Score = 1266 bits (685), Expect = 0.0
Identities = 806/857 (95%), Gaps = 37/857 (4%)

Strand=Plus/Plus

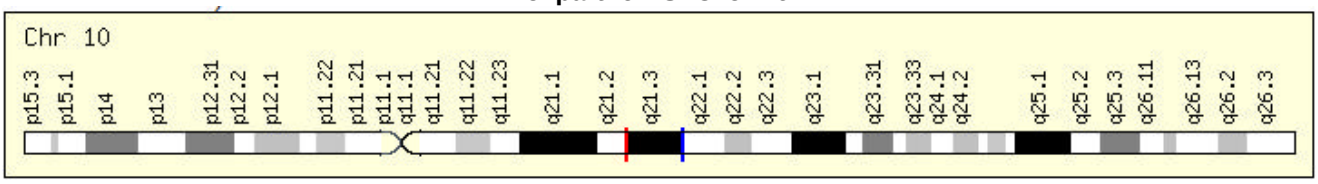
```
Query  337  ATGTGTGGCAGGTTCCTGCGGCGGCTGCTGGCGGAGGAGAGCCGGCGCTCCACCCCCGTG  396
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGTGTGGCAGGTTCCTGCGGCGGCTGCTGGCGGAGGAGAGCCGGCGCTCCACCCCCGTG  60

Query  397  GGGCGCCTCTTGCTTCCCGTGCTCCTGGGATTCCGCCTTGTGCTGCTGGCTGCCAGTGGG  456
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GGGCGCCTCTTGCTTCCCGTGCTCCTGGGATTCCGCCTTGTGCTGCTGGCTGCCAGTGGG  120

Query  457  CCTGGAGTCTATGGTGATGAGCAGAGTGAATTCGTGTGTCACACCCAGCAGCCGGGCTGC  516
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  CCTGGAGTCTATGGTGATGAGCAGAGTGAATTCGTGTGTCACACCCAGCAGCCGGGCTGC  180

Query  517  AAGGCTGCCTGCTTCGATGCCTTCCACCCCCTCTCCCCGCTGCGTTTCTGGGTCTTCCAG  576
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  AAGGCTGCCTGCTTCGATGCCTTCCACCCCCTCTCCCCGCTGCGTTTCTGGGTCTTCCAG  240

Query  577  GTCATCTTGGTGGCTGTACCCAGCGCCCTCTATATGGGTTTCACTCTGTATCACGTGATC  636
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GTCATCTTGGTGGCTGTACCCAGCGCCCTCTATATGGGTTTCACTCTGTATCACGTGATC  300

Query  637  TGGCACTGGGAATTATCAGGAAAGGGGAAGGAGGAGGAGATCTCTGATCCAGGGACGGGA  696
            |||||||||||||||||||||||||||||||||||||||| | ||||||||||||||||||
Sbjct  301  TGGCACTGGGAATTATCAGGAAAGGGGAAGGAGGAGGAGACC-CTGATCCAGGGACGGGA  359

Query  697  GGGCAACACAGATGTCTCTAGGGGCTGGAAGCTCTCACGCTGCTCTGGGCTTATGTGGCT  756
            |||||||||||||||| | |||||||||||||| |||| ||||||||||||||||||||||
Sbjct  360  GGGCAACACAGATGTCCC-AGGGGCTGGAAGC-CTCAGGCTGCTCTGGGCTTATGTGGCT  417

Query  757  CAGCTGGGGGCTCGGCTTGTCCTGTAGGGGGCAGCCCTGGGGTTGCAGTACCACCTGTAT  816
            |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Sbjct  418  CAGCTGGGGGCTCGGCTTGTCCTGGAGGGGGCAGCCCTGGGGTTGCAGTACCACCTGTAT  477
```

```
Query  817   GGGTTCCAGATGCCCAGCTCCTTTGCATGTCGCCGAGAACCTTGCCTTGGTAGTATAACC  876
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  478   GGGTTCCAGATGCCCAGCTCCTTTGCATGTCGCCGAGAACCTTGCCTTGGTAGTATAACC  537

Query  877   TGCAATCTGTCCCGCTCCTCTGAGAAGACCATTTTTCCTAAAGACCATGTTTGGAGTCAG  936
             |||||||||||||| |||||||||||||||||||| ||||||||||||||||||||||||
Sbjct  538   TGCAATCTGTCCCGCCCCCTCTGAGAAGACCATTTT-CCTAAAGACCATGTTTGGAGTCAG  596

Query  937   CGGTTTCTGTCTCTTGTTTACTTTTTTGGAGCTTGTGCTTCTGGGTTTGGGGAGATGGTG  996
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  597   CGGTTTCTGTCTCTTGTTTACTTTTTTGGAGCTTGTGCTTCTGGGTTTGGGGAGATGGTG  656
```

SLC25A16-DNA2

TTGGTACCCTTCTTGGCAGACCTTCATCAGACAATCCTAATGTCTTAGTTTTGAAAACTCATGTAAACTTACTTTGTGGTGGTGTTGCTGGAGCAATAGCGC
GACAATATCCTACCCATTTGATGTGACTCGTCGGCGAATGCAATTAGGAACTGTTCTGCCGGAATTTGAAAAGTGCCTATTTCAGAAGAAAGTGGTAGCTTC
TTTCCAAGAACAGTTCTGAGCACAGGAATGGATAACCGGTACCTGGTGTTGGCAGTCAATACTGTACAGAACAAAGAGGGAAACTGTGAAAAGCGCCTGGTC
TCACTGCTTCACAGTCACTAGAAAATAAAGAACTATGCATCCTTAGGAATGACTGGTGTTCTGTTCCAGTAGAGCCAGGAGATATCATTCATTTGGAGGGAG
CTGCACATCTGACACTTGGATAATAGATAAAGATTTTGGATATTTGATTCTGTATCCAGACATGCTGATTTCTGGCACCAGCATAGCCAGTAGTATTCGATG
ATGAGAAGAGCTGTCCTGAGTGAAACTTTTAGGAGCTCTGATCCAGCCACACGCCAAATGCTAATTGGTACGGTTCTCCATGAGGTGTTTCAAAAAGCAATA
ATAATAGCTTTGCCCCAGAAAAGCTACAAGAACTTGCTTTTCAAACAATTCAAGAATAAGACATTTGAAGGAAATGTACCGCTTAAATCTAAGTCAAGATGA
ATAAAACAAGAAGTAGAGGACTATCTTCCTTCGTTTTGTAAATGGGCAGGAGATTTCATGCATAAAAACACTTCGACTGACTTCCCTCAGATGCAGCTCTCT
TGCCAAGTGATAATAGTAAGGATAATTCAACATGTAACATTGAAGTCGTGAAACCAATGGATATTGAAGAAGCATTTGGTCCCCTAAGTTTGGATGAAAGGC
AAATAAATGTTACAGTTGGTGTGAAATACATCGAGGTATAAACAATACAGATATGCCGCTGGACTAGACTGCAAGATCAAATTCTATGAACTACGTAGTCAG
GTTCTGTAACTCTTACTAGTCAGAGAGAAAAGCTGATCAAGCCTGACTGCTCCCTACTCAGAACTGTCAGTACCTGGCCTGCACATCTAAATAAAGAATTAA
CTGAACCGAATGCATCCATTGTTTACGATTAGCAAATCGCCTCTTTCCGGGACACT

5’ partner: SLC25A16



Junction point
exon=8 922..990
/gene="SLC25A16"
/gene_synonym="D10S105E; GDA; GDC; HGT.1; hML7; MGC39851; ML7"

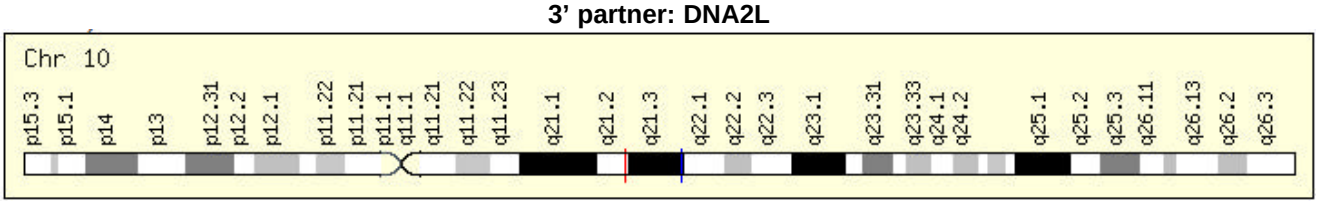
BLAST vs mRNA

>ref|NM_152707.3| Homo sapiens solute carrier family 25 (mitochondrial carrier; Graves disease autoantigen), member 16 (SLC25A16), nuclear gene encoding mitochondrial protein, mRNA
Length=2264
GENE ID: 8034 SLC25A16 solute carrier family 25 (mitochondrial carrier; Graves disease autoantigen), member 16 [Homo sapiens]
Score = 327 bits (177), Expect = 3e-88
Identities = 177/177 (100%), Gaps = 0/177 (0%)

Strand=Plus/Plus

```
Query  5    TACCCTTCTTGGCAGACCTTCATCAGACAATCCTAATGTCTTAGTTTTGAAAACTCATGT  64
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  814  TACCCTTCTTGGCAGACCTTCATCAGACAATCCTAATGTCTTAGTTTTGAAAACTCATGT  873

Query  65   AAACTTACTTTGTGGTGGTGTTGCTGGAGCAATAGCGCAGACAATATCCTACCCATTTGA  124
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  874  AAACTTACTTTGTGGTGGTGTTGCTGGAGCAATAGCGCAGACAATATCCTACCCATTTGA  933

Query  125  TGTGACTCGTCGGCGAATGCAATTAGGAACTGTTCTGCCGGAATTTGAAAAGTGCCT  181
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  934  TGTGACTCGTCGGCGAATGCAATTAGGAACTGTTCTGCCGGAATTTGAAAAGTGCCT  990
```

3’ partner: DNA2L



exon=2 184..366
/gene="DNA2"
/gene_synonym="DNA2L; FLJ10063; KIAA0083; MGC133297"

BLAST vs mRNA

>ref|NM_001080449.2| Gene info linked to NM_001080449.2 Homo sapiens DNA replication helicase 2 homolog (yeast) (DNA2), mRNA
Length=4287
GENE ID: 1763 DNA2 | DNA replication helicase 2 homolog (yeast) [Homo sapiens]
Score = 1496 bits (810), Expect = 0.0
Identities = 933/984 (95%), Gaps = 41/984 (4%)

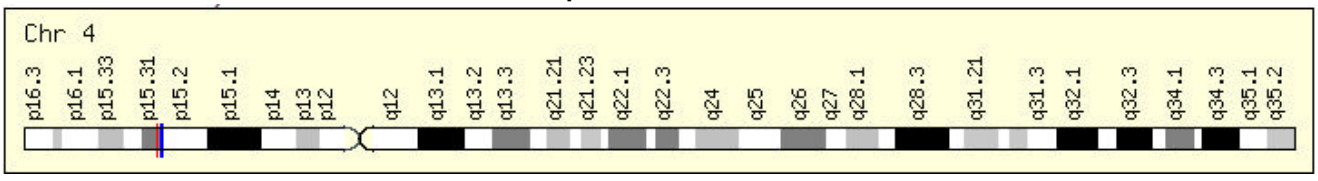
Strand=Plus/Plus

```
Query  180   CTATTTCAGAAGAAAGTGGTAGCTTCCTTTCCAAGAACAGTTCTGAGCACAGGAATGGAT  239
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  182   CTATTTCAGAAGAAAGTGGTAGCTTCCTTTCCAAGAACAGTTCTGAGCACAGGAATGGAT  241

Query  240   AACCGGTACCTGGTGTTGGCAGTCAATACTGTACAGAACAAAGAGGGAAACTGTGAAAAG  299
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  242   AACCGGTACCTGGTGTTGGCAGTCAATACTGTACAGAACAAAGAGGGAAACTGTGAAAAG  301

Query  300   CGCCTGGTCATCACTGCTTCACAGTCACTAGAAAATAAAGAACTATGCATCCTTAGGAAT  359
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  302   CGCCTGGTCATCACTGCTTCACAGTCACTAGAAAATAAAGAACTATGCATCCTTAGGAAT  361

Query  360   GACTGGTGTTCTGTTCCAGTAGAGCCAGGAGATATCATTCATTTGGAGGGAGACTGCACA  419
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  362   GACTGGTGTTCTGTTCCAGTAGAGCCAGGAGATATCATTCATTTGGAGGGAGACTGCACA  421

Query  420   TCTGACACTTGGATAATAGATAAAGATTTTGGATATTTGATTCTGTATCCAGACATGCTG  479
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  422   TCTGACACTTGGATAATAGATAAAGATTTTGGATATTTGATTCTGTATCCAGACATGCTG  481

Query  480   ATTTCTGGCACCAGCATAGCCAGTAGTATTCGATGTATGAGAAGAGCTGTCCTGAGTGAA  539
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  482   ATTTCTGGCACCAGCATAGCCAGTAGTATTCGATGTATGAGAAGAGCTGTCCTGAGTGAA  541

Query  540   ACTTTTAGGAGCTCTGATCCAGCCACACGCCAAATGCTAATTGGTACGGTTCTCCATGAG  599
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  542   ACTTTTAGGAGCTCTGATCCAGCCACACGCCAAATGCTAATTGGTACGGTTCTCCATGAG  601

Query  600   GTGTTTCAAAAAGCAATAAATAATAGCTTTGCCCCAGAAAAGCTACAAGAACTTGCTTTT  659
             |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Sbjct  602   GTGTTTCAAAAAGCCATAAATAATAGCTTTGCCCCAGAAAAGCTACAAGAACTTGCTTTT  661

Query  660   CAAACAATTCAAGAA-TAAGACATTTGAAGGAAATGTACCGCTTAAATCTAAGTCAAGAT  718
             ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Sbjct  662   CAAACAATTCAAGAAATAAGACATTTGAAGGAAATGTACCGCTTAAATCTAAGTCAAGAT  721

Query  719   GAAATAAAACAAGAAGTAGAGGACTATCTTCCTTCGTTTTGTAAATGGGCAGGAGATTTC  778
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  722   GAAATAAAACAAGAAGTAGAGGACTATCTTCCTTCGTTTTGTAAATGGGCAGGAGATTTC  781

Query  779   ATGCATAAAAACACTTCGACTGACTTCCCTCAGATGCAGCTCTCTCTGCCAAGTGATAAT  838
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  782   ATGCATAAAAACACTTCGACTGACTTCCCTCAGATGCAGCTCTCTCTGCCAAGTGATAAT  841
```

PACRGL-LIMCH1

CCCCAAACGCAGGCGCTCGGTGGCGGTAGCCGCGGTTGTTGGCCGACCGAGTGCCGGTCATAAGCCCCCCCCGGTGGGGGGCAGCTGGTGTGCGGATCGCGG
GGGAGAGAGGCGCGGTAGGAACGGGTCCCCGGAGCCGTGAACCGCGGGTACAGGAGTGAAAGGGAAGCAATGCAAAAATCAAAGGGCTCTGGAGGTACACAG
TGAAAAACAGAGCAACAGGTAACTATGATCAAAGGACATCATCAAGCACACAGTTAAAACACAGGAATGCAGTTCAGGGAAGCAAATCCTCATTGTCAACCA
TTCTCCAAAGTCTGCAAGAAAACTTCATCCTAAACCAAGTGATAAACTGAACCCTAAAACAATTAATCCGGAGCACCAGCATGTTTGACATGCGGTGTGAGG
GGAGGCCGCGGTGCAGCCGCACAGCAGGGCCCGCCAGGAGCAGCTGCAGCTGATAAATAACCAGCTGAGGGAAGAGGACGACAAATGGCAAGATGACCTGGC
CGTTGGAAGAGTCGTAGAAGAAGTGTTTCTCAGGACTTAATCAAGAAAGAGGAAGAAAGGAAAAAAATGGAGAAGTTACTGGCTGGAGAAAGATGGGACAAG
GAACGAAGGAAAAGCATCA

5’ partner: PACRGL



Junction point
exon=3 444..598
/gene="PACRGL"

/gene_synonym="C4orf28; MGC29898"

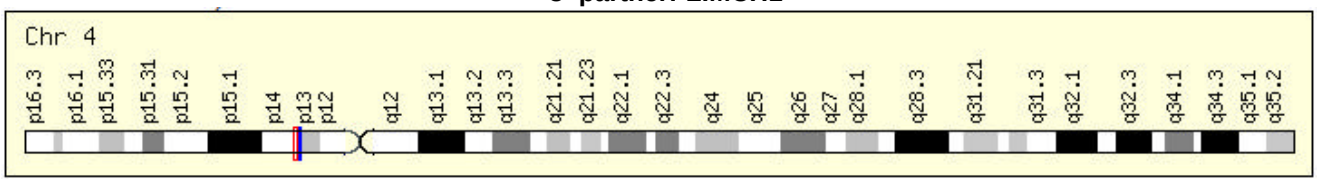
BLAST vs mRNA

>ref|NM_001130727.1| UniGene info linked to NM_001130727.1GEO profiles info linked to NM_001130727.1 Gene info linked to NM_001130727.1Genome view with mapviewer linked to NM_001130727.1 Homo sapiens PARK2 co-regulated-like (PACRGL), transcript variant 2, mRNA
Length=1796
GENE ID: 133015 PACRGL | PARK2 co-regulated-like [Homo sapiens]
Score = 394 bits (213), Expect = 1e-108
Identities = 221/225 (99%), Gaps = 0/225 (0%)

Strand=Plus/Plus

```
Query  156  GGAGTGAAAGGGAAGCAATGCAAAAATCAAAGGGCTCTGGAGGTACACAGTTGAAAAACA  215
            ||||||||||||||||||||| |||||| ||||||||||||||||||||||||||||||
Sbjct  375  GGAGTGAAAGGGAAGCAATGCAGAAATCAGAGGGCTCTGGAGGTACACAGTTGAAAAACA  434

Query  216  GAGCAACAGGTAACTATGATCAAAGGACATCATCAAGCACACAGTTAAAACACAGGAATG  275
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  435  GAGCAACAGGTAACTATGATCAAAGGACATCATCAAGCACACAGTTAAAACACAGGAATG  494

Query  276  CAGTTCAGGGAAGCAAATCCTCATTGTCAACCAGTTCTCCAAAGTCTGCAAGAAAACTTC  335
            |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Sbjct  495  CAGTTCAGGGAAGCAAATCCTCATTGTCAACCAGTTCTCCAGAGTCTGCAAGAAAACTTC  554

Query  336  ATCCTAAACCAAGTGATAAACTGAACCCTAAAACAATTAATCCGG  380
            |||||| |||||||||||||||||||||||||||||||||||||
Sbjct  555  ATCCTAGACCAAGTGATAAACTGAACCCTAAAACAATTAATCCGG  599
```

3' partner: LIMCH1



Junction point
exon=11 647..773
/gene="LIMCH1"
/gene_synonym="DKFZp434I0312; DKFZp686A01247;

BLAST vs mRNA

>ref|NM_001112720.1| UniGene info linked to NM_001112720.1GEO profiles info linked to NM_001112720.1Gene info linked to NM_001112720.1Genome view with mapviewer linked to NM_001112720.1 Homo sapiens LIM and calponin homology domains 1 (LIMCH1), transcript variant 5, mRNA
Length=5722
GENE ID: 22998 LIMCH1 | LIM and calponin homology domains 1 [Homo sapiens]
Score = 472 bits (255), Expect = 7e-132
Identities = 258/259 (99%), Gaps = 1/259 (0%)

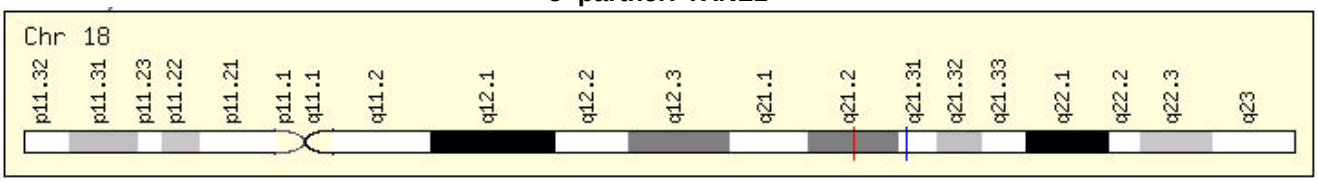
Strand=Plus/Plus

```
Query  379  GGAGCACCAGCATGTTTGACATGCGGTGTGAGGAGGAGGCCGCGGTGCAGCCGCACAGCA  438
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  646  GGAGCACCAGCATGTTTGACATGCGGTGTGAGGAGGAGGCCGCGGTGCAGCCGCACAGCA  705

Query  439  GGGCCCGCCAGGAGCAGCTGCAGCTGATAAATAACCAGCTGAGGGAAGAGGACGACAAAT  498
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  706  GGGCCCGCCAGGAGCAGCTGCAGCTGATAAATAACCAGCTGAGGGAAGAGGACGACAAAT  765

Query  499  GGCAAGATGACCTGGCTCGTTGGAAGAGTCGTAGAAGAAGTGTTTCTCAGGACTTAATCA  558
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  766  GGCAAGATGACCTGGCTCGTTGGAAGAGTCGTAGAAGAAGTGTTTCTCAGGACTTAATCA  825

Query  559  AGAAAGAGGAAGAAAGGaaaaaaaTGGAGAAGTTACTGGCTGGAGAAAGATGGGACAAGT  618
            |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Sbjct  826  AGAAAGAGGAAGAAAGGAAAAAAATGGAGAAGTTACTGGCTGGAGAA-GATGGGACAAGT  884

Query  619  GAACGAAGGAAAAGCATCA  637
            ||||||||||||||||||
Sbjct  885  GAACGAAGGAAAAGCATCA  903
```

TXNL1-CDH2

TTGGCTTAGAAAATGACCTCTGAGAAGCAATGAGGACACAGATATTCCAAAAGGCTATATGGATTTAATGCCTTTTATTAACAAAGCTGGTTGTGAATGTCT
AATGAAAGTGATGAGCATGGATTTGACAACTGTTTACGAAAAGACACAACCTTCTTGGAATCTGACTGTGATGAACAGCTGCTTATTACTGTGGCATTCAATC
AACCTGTTAAGCTTTATTCCATGAAATTTCAAGGGCCAGATAATGGTCAGGGCCCTAAATATGTAAAAATTTTTATCAACCTACCCCGATCTATGGATTTTGA
AGAGGCAGAAAGAAGTGAACCAACTCAAGCTCTGGAACTGACAGAGGATGATATTAAAGAAGATGGCATTGTTCCACTTCGTTATGTTAAGTTTCAGAATGT
AACAGTGTAACTATATTTGTTCAGTCGAATCAAGGTGAAGAGGAAACAACAAGAATTTCATATTTTACTTTTATTGGTACTCCAGTCCAGGCAACAAATATGA
ATGACTTCAAACGAAAAGTGCAACAGTATACGTTAATAATTCAAGCTACAGACATGGAAGGCAATCCCACATATGGCCTTTCAAACACAGCCACGGCCGTCAT
CACAGTGACAGATGTCAATGACAATCCTCCAGAGTTTACTGCCATGACGTTTTATGGTGAAGTTCCTGAGAACAGGGTAGACATCATAGTAGCTAATCTAAC
GTGACCGATAAGATCAACCCCATACACCAGCCTGGAACGCAGTGTACAGAATCAGTGGCGGAGATCCTACTGGACGGTTCGCCATCCAGACCGACCCAAACAC
CAACGACGGGTTAGTCACCGTGGTCAAACAATCGACTTTGAAACAAATAGGATGTTTGTCCTTACTGTTTGCTGCCGAAAAATCAAGTGCCATTAGCCAAGGC
AATTCAGCACCCGCCTCAGTCAACTGCAACCGTGTCTGTTACAGTTATTGACGTAAATGAAAACCCTTATTTTGCCCCCAATCCTAAGATCATTCGCCAGAAA
AAGGCTTCATGCCGTACCATGTTGAAAAATTCACTGGCTCAGACCGAATTCGATATATGCAGGCAAAATTTAGAAACCACTTAATTTATCTGGATCCTGGCCA
TGGCCTAAATTGAGTTCCTGTGAATGACAAAATTACCTATACATGTGCCTGTTTTTGTGAGACA

5’ partner: TXNL1



Junction point
exon=8 1090..1194
/gene="TXNL1"
/gene_synonym="TRP32; Txl; TXL-1; TXNL"

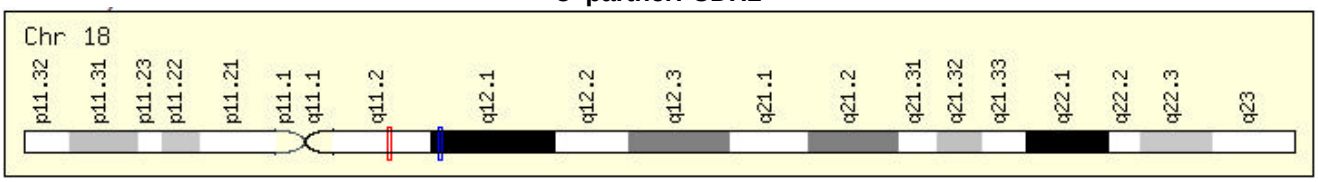
BLAST vs mRNA

>ref|NR_024546.1| Gene info linked to NR_024546.1Genome view with mapviewer linked to NR_024546.1 Homo sapiens thioredoxin-like 1 (TXNL1), transcript variant 2, non-coding RNA
Length=1434
GENE ID: 9352 TXNL1 | thioredoxin-like 1 [Homo sapiens]
Score = 957 bits (518), Expect = 0.0
Identities = 523/525 (99%), Gaps = 2/525 (0%)

Strand=Plus/Plus

```
Query  5     CTTAGAAAATGACCTCTGAGAAGCAATGAGGACACAGATATTCCAAAAGGCTATATGGAT  64
             ||||||||||||||| ||| ||||||||||||||||||||||||||||||||||||||||
Sbjct  672   CTTAGAAAATGACC-CTG-GAAGCAATGAGGACACAGATATTCCAAAAGGCTATATGGAT  729

Query  65    TTAATGCCTTTTATTAACAAAGCTGGTTGTGAATGTCTTAATGAAAGTGATGAGCATGGA  124
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  730   TTAATGCCTTTTATTAACAAAGCTGGTTGTGAATGTCTTAATGAAAGTGATGAGCATGGA  789

Query  125   TTTGACAACTGTTTACGAAAAGACACAACCTTCTTGGAATCTGACTGTGATGAACAGCTG  184
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  790   TTTGACAACTGTTTACGAAAAGACACAACCTTCTTGGAATCTGACTGTGATGAACAGCTG  849

Query  185   CTTATTACTGTGGCATTCAATCAACCTGTTAAGCTTTATTCCATGAAATTTCAAGGGCCA  244
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  850   CTTATTACTGTGGCATTCAATCAACCTGTTAAGCTTTATTCCATGAAATTTCAAGGGCCA  909

Query  245   GATAATGGTCAGGGCCCTAAATATGTAAAAATTTTTATCAACCTACCCCGATCTATGGAT  304
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  910   GATAATGGTCAGGGCCCTAAATATGTAAAAATTTTTATCAACCTACCCCGATCTATGGAT  969

Query  305   TTTGAAGAGGCAGAAAGAAGTGAACCAACTCAAGCTCTGGAACTGACAGAGGATGATATT  364
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  970   TTTGAAGAGGCAGAAAGAAGTGAACCAACTCAAGCTCTGGAACTGACAGAGGATGATATT  1029

Query  365   AAAGAAGATGGCATTGTTCCACTTCGTTATGTTAAGTTTCAGAATGTTAACAGTGTAACT  424
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1030  AAAGAAGATGGCATTGTTCCACTTCGTTATGTTAAGTTTCAGAATGTTAACAGTGTAACT  1089

Query  425   ATATTTGTTCAGTCGAATCAAGGTGAAGAGGAAACAACAAGAATTTCATATTTTACTTTT  484
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1090  ATATTTGTTCAGTCGAATCAAGGTGAAGAGGAAACAACAAGAATTTCATATTTTACTTTT  1149

Query  485   ATTGGTACTCCAGTCCAGGCAACAAATATGAATGACTTCAAACGA  529
             |||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1150  ATTGGTACTCCAGTCCAGGCAACAAATATGAATGACTTCAAACGA  1194
```

3’ partner: CDH2



Junction point
exon=8 1480..1617
/gene="CDH2"
/gene_synonym="CD325; CDHN; CDw325; NCAD"

BLAST vs mRNA

>ref|NM_001792.3| UniGene info linked to NM_001792.3GEO profiles info linked to NM_001792.3Gene info linked to NM_001792.3Genome view with mapviewer linked to NM_001792.3 Homo sapiens cadherin 2, type 1, N-cadherin (neuronal) (CDH2), mRNA
Length=4380
GENE ID: 1000 CDH2 | cadherin 2, type 1, N-cadherin (neuronal) [Homo sapiens]
Score = 1000 bits (541), Expect = 0.0
Identities = 634/673 (95%), Gaps = 29/673 (4%)

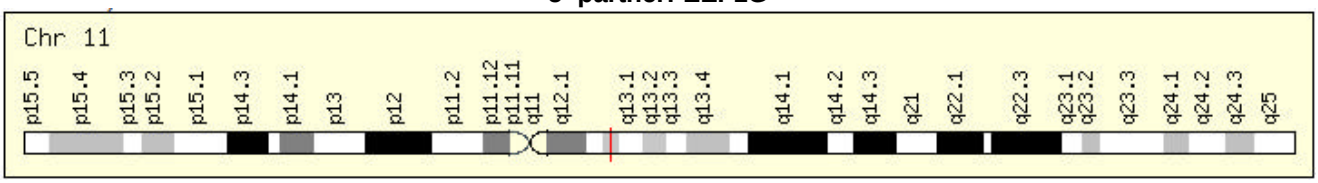
Strand=Plus/Plus

```
Query  529   AAAAGTGCAACAGTATACGTTAATAATTCAAGCTACAGACATGGAAGGCAATCCCACATA  588
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1479  AAAAGTGCAACAGTATACGTTAATAATTCAAGCTACAGACATGGAAGGCAATCCCACATA  1538

Query  589   TGGCCTTTCAAACACAGCCACGGCCGTCATCACAGTGACAGATGTCAATGACAATCCTCC  648
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539  TGGCCTTTCAAACACAGCCACGGCCGTCATCACAGTGACAGATGTCAATGACAATCCTCC  1598

Query  649   AGAGTTTACTGCCATGACGTTTTATGGTGAAGTTCCTGAGAACAGGGTAGACATCATAGT  708
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1599  AGAGTTTACTGCCATGACGTTTTATGGTGAAGTTCCTGAGAACAGGGTAGACATCATAGT  1658

Query  709   AGCTAATCTAACTGTGACCGATAA-GATCAACCCCATACACCAGCCTGGAACGCAGTGTA  767
             |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Sbjct  1659  AGCTAATCTAACTGTGACCGATAAGGATCAACCCCATACACCAGCCTGGAACGCAGTGTA  1718

Query  768   CAGAATCAGTGGCGGAGATCCTACTGGACGGTTCGCCATCCAGACCGACCCAAACAGCAA  827
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1719  CAGAATCAGTGGCGGAGATCCTACTGGACGGTTCGCCATCCAGACCGACCCAAACAGCAA  1778

Query  828   CGACGGGTTAGTCACCGTGGTCAAAC-AATCGACTTTGAAACAAATAGGATGTTTGTCCT  886
             |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Sbjct  1779  CGACGGGTTAGTCACCGTGGTCAAACCAATCGACTTTGAAACAAATAGGATGTTTGTCCT  1838

Query  887   TACTGTTTGCTGCCGAAAAATCAAGTGCCATTAGCCAAGGGAATTCAGCACCCGCCTCAG  946
             ||||||| ||||| ||||| ||||||||||||||||||||||||||||||||||||||||
Sbjct  1839  TACTGTT-GCTGCAGAAAA-TCAAGTGCCATTAGCCAAGGGAATTCAGCACCCGCCTCAG  1896

Query  947   TCAACTGCAACCGTGTCTGTTACAGTTATTGACGTAAATGAAAACCCTTATTTTGCCCCC  1006
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1897  TCAACTGCAACCGTGTCTGTTACAGTTATTGACGTAAATGAAAACCCTTATTTTGCCCCC  1956
```

EEF1G-AHNAK

TTGGGGCCCAGTTTGATGCTAAAAAGTTGCAGAGACCCAACCTAAAAAGGACACACCACGGAAAGAGAAGGGTTCACGGGAAGAGAAGCAGAAGCCCCAGGC
GAGCGGAAGGAGGAGAAAAAGGCGGCTGCCCCTGCTCCTGAGGAGGAGATGGATGAATGTGAGCAGGCGCTGGCTGCTGAGCCCAAGGCCAAGGACCCCTTC
CTCACCTGCCCAAGAGGACTGTAGAAGCGGCCAGGAAGAAAACCACCCCCTTTTAAGGTTGTTTTTGTGACCGTTTTTTGGAGCATTGTTCTAAAAATGGGAA
ATTACATATTGCTGTGCCAAGGGCAACAAACACCTGCAGTTAAAGGAATACCTTCCGCGAGGCGGCTTTTCGGAGCATGCATGTTTATAGCTCCAGCCAGGC
AGACCGAGGGCTGCTGCATAAGCCCTGCTTGGTGCATTTCTTTACTTGCAAGGGGACAGAGTGTGGGCTTAGGTTTGGGACTAGAGGGGGCTTTGGCAACTA
GGTGCTCAGGTGATTATCCTTCGCTCGTTTATCCAATAAACATTTATCAAGC

5’ partner: EEF1G



Junction point
exon=7 799..1003

/gene="EEF1G"
/gene_synonym="EF1G; GIG35"

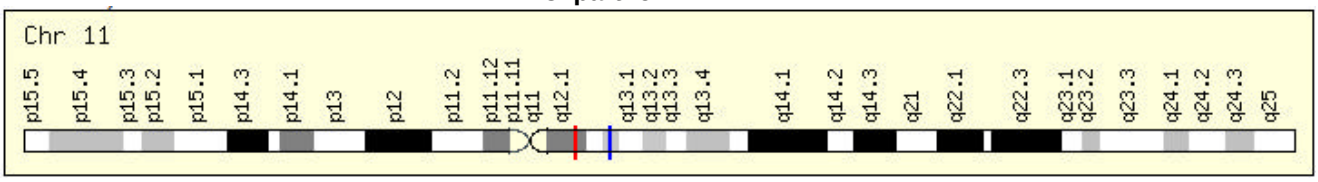
BLAST vs mRNA

>ref|NM_001404.4| UniGene info linked to NM_001404.4GEO profiles info linked to NM_001404.4Gene info linked to NM_001404.4Genome view with mapviewer linked to NM_001404.4 Homo sapiens eukaryotic translation elongation factor 1 gamma (EEF1G), mRNA
Length=1552
GENE ID: 1937 EEF1G | eukaryotic translation elongation factor 1 gamma [Homo sapiens]
Score = 398 bits (215), Expect = 1e-109
Identities = 218/219 (99%), Gaps = 1/219 (0%)

Strand=Plus/Plus

```
Query  5     GGCCCAGTTTGATGCTAAAAAG-TTGCAGAGACCCAACCTAAAAAGGACACACCACGGAA  63
             ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Sbjct  785   GGCCCAGTTTGATGCTAAAAAGTTTGCAGAGACCCAACCTAAAAAGGACACACCACGGAA  844

Query  64    AGAGAAGGGTTCACGGGAAGAGAAGCAGAAGCCCCAGGCTGAGCGGAAGGAGGAGAAAAA  123
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  845   AGAGAAGGGTTCACGGGAAGAGAAGCAGAAGCCCCAGGCTGAGCGGAAGGAGGAGAAAAA  904

Query  124   GGCGGCTGCCCCTGCTCCTGAGGAGGAGATGGATGAATGTGAGCAGGCGCTGGCTGCTGA  183
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  905   GGCGGCTGCCCCTGCTCCTGAGGAGGAGATGGATGAATGTGAGCAGGCGCTGGCTGCTGA  964

Query  184   GCCCAAGGCCAAGGACCCCTTCGCTCACCTGCCCAAGAG  222
             |||||||||||||||||||||||||||||||||||||||
Sbjct  965   GCCCAAGGCCAAGGACCCCTTCGCTCACCTGCCCAAGAG  1003
```

3' partner: AHNAK



Junction point
exon=7 743..1090
/gene="AHNAK"
/gene_synonym="AHNAKRS; MGC5395"

BLAST vs mRNA

>ref|NM_024060.2| UniGene info linked to NM_024060.2GEO profiles info linked to NM_024060.2Gene info linked to NM_024060.2Genome view with mapviewer linked to NM_024060.2 Homo sapiens AHNAK nucleoprotein (AHNAK), transcript variant 2, mRNA
Length=1108
GENE ID: 79026 AHNAK | AHNAK nucleoprotein [Homo sapiens]
Score = 630 bits (341), Expect = 9e-180
Identities = 345/347 (99%), Gaps = 0/347 (0%)

Strand=Plus/Plus

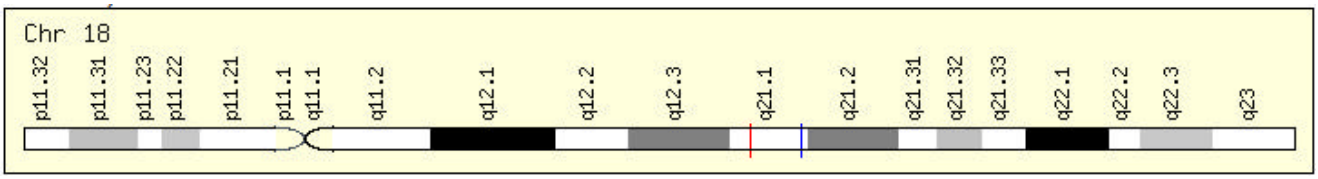
```
Query  221   AGGACTGTAGAAGCGGCCAGGAAGAAAACCACCCCCTTTTAAGGTTGTTTTTGTGACCGT  280
             |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Sbjct  741   AGGACTGTAGAAGCGGCCAGGAAGAAAACCACCCCCTCTTAAGGTTGTTTTTGTGACCGT  800

Query  281   TTTTTGGAGCATTGTTCTAAAAATGGGAAATTACATATTGCTGTGCCAAGGGCAACAAAC  340
             | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  801   TCTTTGGAGCATTGTTCTAAAAATGGGAAATTACATATTGCTGTGCCAAGGGCAACAAAC  860

Query  341   ACCTGCAGTTAAAGGAATACCTTCCGCGAGGCGGCTTTTCGGAGCATGCATGTTTATAGC  400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  861   ACCTGCAGTTAAAGGAATACCTTCCGCGAGGCGGCTTTTCGGAGCATGCATGTTTATAGC  920

Query  401   TCCAGCCAGGCCAGACCGAGGGCTGCTGCATAAGCCCTGCTTGGTGCATTTCTTTACTTG  460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  921   TCCAGCCAGGCCAGACCGAGGGCTGCTGCATAAGCCCTGCTTGGTGCATTTCTTTACTTG  980

Query  461   CAAGGGGACAGAGTGTGGGCTTAGGTTTGGGACTAGAGGGGGCTTTGGCAACTATGGTGC  520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  981   CAAGGGGACAGAGTGTGGGCTTAGGTTTGGGACTAGAGGGGGCTTTGGCAACTATGGTGC  1040
```

```
Query  521   TCAGGTGATTATCCTTCGCTCGTTTATCCAATAAACATTTATCAAGC  567
              |||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1041  TCAGGTGATTATCCTTCGCTCGTTTATCCAATAAACATTTATCAAGC  1087
```

MBD1-CCDC11

TGGTGCCTGGCTGCCCTAGCAAGGCAGTAGACCCAGGCCTGCCTTCTGTGAAGCAAGAGCACTGACCCAGAGGAGGACAAGGAGGAGAACAAGGATGATTCTC
CCTCCAAATTGGCCCCAGAGGAAGAGGCAGGAGGGGCTGGCACACCCGTGATCACGGAGATTTTCAGCCTGGGTGGAACCCGCTTCCGAGATACAGCAGTCTC
GTTGCCAAGAGATCCAAGCCTCCTAAAGGCCAAGGAGCTGAGCACCATCTAGAAAGAATCCGACGCAGCCATCAGAAGCATAATGCTATTTTGGCTTCCATTA
AGTCAAGTGAGCGGGATCGCTTGAAAGCTGAGTGGGACCAGCACAATGACTGCAAGATTTTGGACAGCCTTGTGCGAGCAAGAATCAAGGATGCTGTGCAAGC
GTTTATCATTAACATTGAAGAAAGACGAAATAAGTAAGGCTCACAAACCTGGGACATAAACCATGAAATATTCCTAAAACCACAATTTGAGATTGACCAGTCA
AAAATCAAATTTAGCAAGGATCATGGCTGAGATATTTATATTAAATCCATAGCTATTGCTGGTTAATATTTTTAATATTTTTATTTGTGATAGCCTGTATTTT
GATGAAAAACAATATATTGTTTCATATTTCATT

5’ partner: MBD1

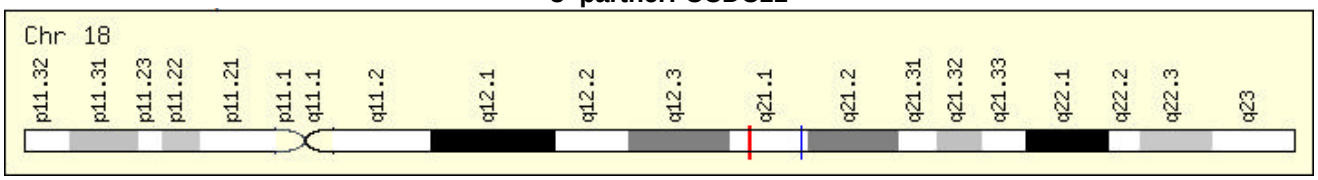


Junction point
exon=16 2004..2065
/gene="MBD1"
/gene_synonym="CXXC3; PCM1; RFT"

BLAST vs mRNA

```
>ref|NM_001204141.1| Gene info linked to NM_001204141.1 Homo sapiens methyl-CpG binding domain protein
1 (MBD1), transcript variant 10, mRNA
Length=2860
GENE ID: 4152 MBD1 | methyl-CpG binding domain protein 1 [Homo sapiens]
Score =  388 bits (210),  Expect = 7e-107
Identities = 215/217 (99%), Gaps = 2/217 (0%)

Strand=Plus/Plus

Query  1     TGGTGCCTGGCTGCCCTAGCAAGGCAGTAGACCCAGGCCTGCCTTCTGTGAAGCAAGAG-  59
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1849  TGGTGCCTGGCTGCCCTAGCAAGGCAGTAGACCCAGGCCTGCCTTCTGTGAAGCAAGAGC  1908

Query  60    CA-CTGACCCAGAGGAGGACAAGGAGGAGAACAAGGATGATTCTGCCTCCAAATTGGCCC  118
             || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1909  CACCTGACCCAGAGGAGGACAAGGAGGAGAACAAGGATGATTCTGCCTCCAAATTGGCCC  1968

Query  119   CAGAGGAAGAGGCAGGAGGGGCTGGCACACCCGTGATCACGGAGATTTTCAGCCTGGGTG  178
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1969  CAGAGGAAGAGGCAGGAGGGGCTGGCACACCCGTGATCACGGAGATTTTCAGCCTGGGTG  2028

Query  179   GAACCCGCTTCCGAGATACAGCAGTCTGGTTGCCAAG  215
             |||||||||||||||||||||||||||||||||||||
Sbjct  2029  GAACCCGCTTCCGAGATACAGCAGTCTGGTTGCCAAG  2065
```

3’ partner: CCDC11



Junction point
exon=2 161..390
/gene="CCDC11"
/gene_synonym="FLJ32743"

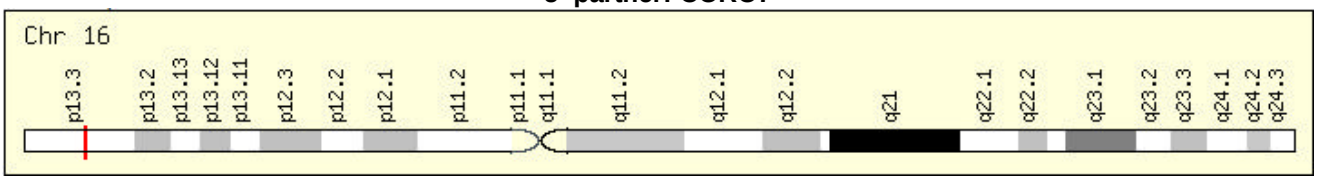
BLAST vs mRNA

```
>ref|NM_145020.3| UniGene info linked to NM_145020.3GEO profiles info linked to NM_145020.3Gene info
linked to NM_145020.3Genome view with mapviewer linked to NM_145020.3 Homo sapiens coiled-coil domain
containing 11 (CCDC11), mRNA
Length=1837
GENE ID: 220136 CCDC11 | coiled-coil domain containing 11 [Homo sapiens]
Score = 429 bits (232),  Expect = 4e-119
Identities = 232/232 (100%), Gaps = 0/232 (0%)

Strand=Plus/Plus

Query  215  GAGATCCAAGCCTCCTAAAGGCCAAGGAGCTGAGCACCATCTAGAAAGAATCCGACGCAG  274
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  160  GAGATCCAAGCCTCCTAAAGGCCAAGGAGCTGAGCACCATCTAGAAAGAATCCGACGCAG  219

Query  275  CCATCAGAAGCATAATGCTATTTTGGCTTCCATTAAGTCAAGTGAGCGGGATCGCTTGAA  334
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  220  CCATCAGAAGCATAATGCTATTTTGGCTTCCATTAAGTCAAGTGAGCGGGATCGCTTGAA  279

Query  335  AGCTGAGTGGGACCAGCACAATGACTGCAAGATTTTGGACAGCCTTGTGCGAGCAAGAAT  394
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  280  AGCTGAGTGGGACCAGCACAATGACTGCAAGATTTTGGACAGCCTTGTGCGAGCAAGAAT  339

Query  395  CAAGGATGCTGTGCAAGGGTTTATCATTAACATTGAAGAAAGACGAAATAAG  446
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  340  CAAGGATGCTGTGCAAGGGTTTATCATTAACATTGAAGAAAGACGAAATAAG  391
```

CORO7-MAGMAS

TTGGGGCGCTAATGGGCAGCCCTGGCTTCTCAGCCTGCAGCCTCCTGACATGAGCCCAGTGAGCCAAGCCCCCCGAGAGGCCCCTGCTCGTCGGGCCCCATC
TCAGCGCAGTACCTGGAAGAAAAGTCTGACCAGCAAAAGAAGGAGGAGCTGCTGAATGCCATGGTGGCAAAACTGGGGAACCGGGAGGACCCACTCCCCCAG
ACTCCTTTGAAGGCGTGGACGAGGACGAGTGGGCCAAGTACCTGGCCCAGATCATTGTGATGGGCGTGCAGGTGGTGGGCAGGGCCTTTGCACGGGCCTTGC
GCAGGAGTTTGCAGCCAGCCGGGCCGCAGCTGATGCCCGAGGACGCGCTGGACACCGGTCTGCAGCCGCTTCCAACCTCTCCGGCCTCAGCCTCCAGGAGGC
CAGCAGATTCTCAACGTGTCCAAGCTGAGCTCTGAGGAGGTCCAGAAGACTATGAACACTTATTTAACGTGAATGATAAATCCGTGGGTGGCTCCTTCTACC
GCAGTCAAAGGTGGTCCGCGCAAAGGAGCGCCTGGATGAGGAACTCAAAATCCAGGCCCAGGAGACAGAGAAAAAGGCAGATGCCTCATACGTGACTGCTCG
TCCCCCGCCACCCGCGCCTCTAATTAATAGCTGTAATAATTCTTTTCTGCACGTAAAAAAGCCACTTTCTGTACAAGTTGGCATAAAAGAAGCATGCTTTCA
TGTGCACGAACGTCCTAACAGTCAATAATCGTATGCCATCCGCCTGATCCTAAGACGATAATGCCACGTTTCGGACCTGCCGTTCCATCTAAGTTCTTGCAA
AAAATCTCTGAACATACGTCGCAAAACAGTGATATACGGTTAGCATACCGAACTGAGGCGATATCCAGACGATATGGAAGTCA

5’ partner: CORO7

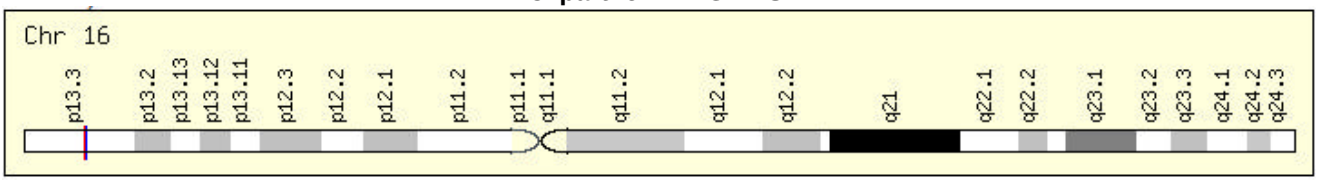


Junction point
exon=27 2874..2960
/gene="CORO7"
/gene_synonym="0610011B16Rik; CRN7; FLJ22021; FLJ44188; POD1"

BLAST vs mRNA

```
>ref|NM_001201473.1| Gene info linked to NM_001201473.1 Homo sapiens coronin 7 (CORO7), transcript
variant 3, mRNA
Length=3599
GENE ID: 79585 CORO7 | coronin 7 [Homo sapiens]
Score =  435 bits (235),  Expect = 1e-120
Identities = 235/235 (100%), Gaps = 0/235 (0%)

Strand=Plus/Plus

Query  5     GGCGCTAATGGGCAGCCCTGGCTTCTCAGCCTGCAGCCTCCTGACATGAGCCCAGTGAGC  64
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2727  GGCGCTAATGGGCAGCCCTGGCTTCTCAGCCTGCAGCCTCCTGACATGAGCCCAGTGAGC  2786

Query  65    CAAGCCCCCCGAGAGGCCCCTGCTCGTCGGGCCCCATCCTCAGCGCAGTACCTGGAAGAA  124
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2787  CAAGCCCCCCGAGAGGCCCCTGCTCGTCGGGCCCCATCCTCAGCGCAGTACCTGGAAGAA  2846

Query  125   AAGTCTGACCAGCAAAAGAAGGAGGAGCTGCTGAATGCCATGGTGGCAAAACTGGGGAAC  184
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2847  AAGTCTGACCAGCAAAAGAAGGAGGAGCTGCTGAATGCCATGGTGGCAAAACTGGGGAAC  2906

Query  185   CGGGAGGACCCACTCCCCCAGGACTCCTTTGAAGGCGTGGACGAGGACGAGTGGG  239
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2907  CGGGAGGACCCACTCCCCCAGGACTCCTTTGAAGGCGTGGACGAGGACGAGTGGG  2961
```

3’ partner: MAGMAS



Junction point
exon=2 142..226
/gene="PAM16"
/gene_synonym="CGI-136; MAGMAS; TIM16; TIMM16"

BLAST vs mRNA

>ref|NM_016069.9| Gene info linked to NM_016069.9 Homo sapiens presequence translocase-associated motor 16 homolog (S. cerevisiae) (PAM16), nuclear gene encoding mitochondrial protein, mRNA
Length=600
GENE ID: 51025 PAM16 | presequence translocase-associated motor 16 homolog (S. cerevisiae) [Homo sapiens]
Score = 721 bits (390), Expect = 0.0
Identities = 428/444 (97%), Gaps = 12/444 (2%)

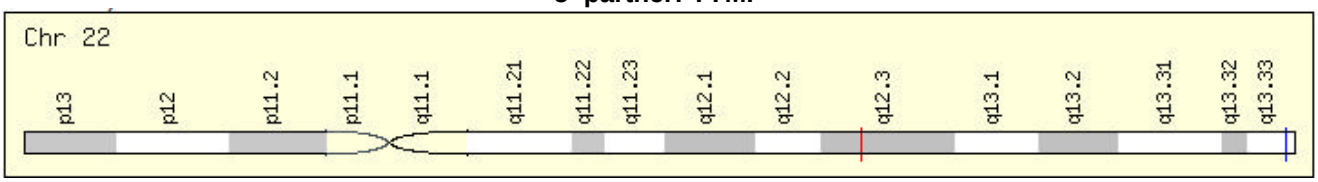
Strand=Plus/Plus

```
Query  238  GGCCAAGTACCTGGCCCAGATCATTGTGATGGGCGTGCAGGTGGTGGGCAGGGCCTTTGC  297
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  141  GGCCAAGTACCTGGCCCAGATCATTGTGATGGGCGTGCAGGTGGTGGGCAGGGCCTTTGC  200

Query  298  ACGGGCCTTGCGGCAGGAGTTTGCAGCCAGCCGGGCCGCAGCTGATGCCCGAGGACGCGC  357
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  201  ACGGGCCTTGCGGCAGGAGTTTGCAGCCAGCCGGGCCGCAGCTGATGCCCGAGGACGCGC  260

Query  358  TGGACACCGGTCTGCAGCCGCTTCCAACCTCTCCGGCCTCAGCCTCCAGGAGGCACAGCA  417
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  261  TGGACACCGGTCTGCAGCCGCTTCCAACCTCTCCGGCCTCAGCCTCCAGGAGGCACAGCA  320

Query  418  GATTCTCAACGTGTCCAAGCTGAGCTCTGAGGAGGTCCAGAAGA-CTATGAACACTTATT  476
            |||||||||||||||||||||||| |||||||||||||||||| ||||||||||||||||
Sbjct  321  GATTCTCAACGTGTCCAAGCTGAGCCCTGAGGAGGTCCAGAAGAACTATGAACACTTATT  380

Query  477  TAACGTGAATGATAAATCCGTGGGTGGCTCCTTCTACCTGCAGTCAAAGGTGGTCCGCGC  536
            ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  381  TAAGGTGAATGATAAATCCGTGGGTGGCTCCTTCTACCTGCAGTCAAAGGTGGTCCGCGC  440

Query  537  AAAGGAGCGCCTGGATGAGGAACTCAAAATCCAGGCCCAGGAG-ACAGAGAAAAAGG-CA  594
            |||||||||||||||||||||||||||||||||||||||||||| |||||||||||| ||
Sbjct  441  AAAGGAGCGCCTGGATGAGGAACTCAAAATCCAGGCCCAGGAGGACAGAGAAAAAGGGCA  500

Query  595  GATGCCTCATACGTGACTGCTCG-CTCCCCC-GCC-ACCC-GC-GCCTCTAATTAATAGC  649
            |||||| |||||||||||||||| ||||||| ||| |||| || |||||||||| |||||
Sbjct  501  GATGCCCCATACGTGACTGCTCGGCTCCCCCCGCCCACCCCGCCGCCTCTAATTTATAGC  560

Query  650  T-G-TAATAA-TT-CTTTTCTGCA  669
            | | |||||| || ||||||||||
Sbjct  561  TTGGTAATAAATTTCTTTTCTGCA  584
```

TYMP-SCO2

TTGGCGCTCTGGTGGTGGACGTTAAGTTCGGAGGGGCCGCCGTCTTCCCCAACCAGGAGCAGGCCCGGGAGCTGGCAAAGACGCTGGTTGGCGTGGGAGCCAC
CCTAGGGCTTCGGGTCGCGGCAGCGCTGACCGCCATGGACAAGCCCCTGGGTCGCTGCGTGGGCCACGCCCTGGAGGTGGAGGAGGCGCTGCTCTGCATGGAC
GGCGCAGGCCCGCCAGACTTAAGGGACCTGGTCACCACGCTCGGGGGCGCCCTGCTCTGGCTCAGCGGACACGCGGGGACTCAGGCCCAGGGCGCTGCCCGGC
TGGCCGCGGCGCTGGACGACGGCTCGGCCCTTGGCCGCTTCGAGCGGATGCTGGCGGCGCAGGGCGTGGATCCCGGTCTGGCCCGAGCCCTGTGCTCGGGAAC
TCCCGCAGAACGCCGGCAGCTGCTGCCTCGCGCCCGGGAGCAGGAGGAGCTGCTGGCGCCCGCAGATGGAGCATCAGATCCATGCTGCTGCTGACTCGGAGCC
CCACAGCTTGGCACAGGCTCTCTCAGCTCAAGCCTCCGGTCCTCCCTGGGACCCTGGGAGGCCAGGCCCTGCATCTGAGGTCCTGGCTTTTGTCAAGGCAGGC
CCCTGCAGAGACAGGTGGGCAGGGCCAGCCCCAGGGCCCTGGGCTTCGAACCCGGCTGCTGATCACAGGCCTGTTCGGGGCTGGACTCGGTGGGGCCTGGCTC
GCCCTGAGGGCTGAGAAGGAGAGGCTGCAGCAGCAAAAAGCGAACAGAAGCCCTGCGCCGGGCAGCTGTGGGCCAGGGGCGACTTCCACCTGCTGGATCACAC
ATGCCCGGGCTCGCTGCCAGGCTGACTTCCCGGGGCCAGTGGGTGCTGATGTACTTTGGCTTCCCTCCACTGCCCCTACATCTGCCCAAAACAAACTGGAAAA
AGCTTGGTGCAAGGTGGTGCGGGCAGCTGGGAATACAAAGCCTGGTTTTGCCTCCAATGCAAGCCTGTCTTTCATCCATGTGAAACCAGAGCCGGGAACAAAC
TTGAAATCATGGCCCCGCATACCTCAGGAAACTTCCACCCCAAAACTGGTTGGGGTTCTGATACC

5’ partner: TYMP



Junction point
exon=8 1049..1279
/gene="TYMP"
/gene_synonym="ECGF; ECGF1; hPD-ECGF; MEDPS1; MNGIE; MTDPS1; PDECGF; TP"

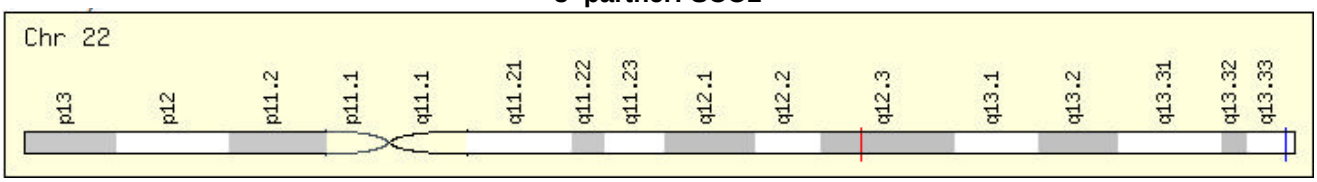
BLAST vs mRNA

>ref|NM_001113756.1| Homo sapiens thymidine phosphorylase (TYMP), transcript variant 3, mRNA
Length=1587
GENE ID: 1890 TYMP| thymidine phosphorylase [Homo sapiens]
Score = 881 bits (477), Expect = 0.0
Identities = 477/477 (100%), Gaps = 0/477 (0%)

Strand=Plus/Plus

```
Query  5     CGCTCTGGTGGTGGACGTTAAGTTCGGAGGGGCCGCCGTCTTCCCCAACCAGGAGCAGGC  64
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  804   CGCTCTGGTGGTGGACGTTAAGTTCGGAGGGGCCGCCGTCTTCCCCAACCAGGAGCAGGC  863

Query  65    CCGGGAGCTGGCAAAGACGCTGGTTGGCGTGGGAGCCAGCCTAGGGCTTCGGGTCGCGGC  124
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  864   CCGGGAGCTGGCAAAGACGCTGGTTGGCGTGGGAGCCAGCCTAGGGCTTCGGGTCGCGGC  923

Query  125   AGCGCTGACCGCCATGGACAAGCCCCTGGGTCGCTGCGTGGGCCACGCCCTGGAGGTGGA  184
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  924   AGCGCTGACCGCCATGGACAAGCCCCTGGGTCGCTGCGTGGGCCACGCCCTGGAGGTGGA  983

Query  185   GGAGGCGCTGCTCTGCATGGACGGCGCAGGCCCGCCAGACTTAAGGGACCTGGTCACCAC  244
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  984   GGAGGCGCTGCTCTGCATGGACGGCGCAGGCCCGCCAGACTTAAGGGACCTGGTCACCAC  1043

Query  245   GCTCGGGGGCGCCCTGCTCTGGCTCAGCGGACACGCGGGGACTCAGGCCCAGGGCGCTGC  304
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1044  GCTCGGGGGCGCCCTGCTCTGGCTCAGCGGACACGCGGGGACTCAGGCCCAGGGCGCTGC  1103

Query  305   CCGGGTGGCCGCGGCGCTGGACGACGGCTCGGCCCTTGGCCGCTTCGAGCGGATGCTGGC  364
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1104  CCGGGTGGCCGCGGCGCTGGACGACGGCTCGGCCCTTGGCCGCTTCGAGCGGATGCTGGC  1163

Query  365   GGCGCAGGGCGTGGATCCCGGTCTGGCCCGAGCCCTGTGCTCGGGAAGTCCCGCAGAACG  424
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1164  GGCGCAGGGCGTGGATCCCGGTCTGGCCCGAGCCCTGTGCTCGGGAAGTCCCGCAGAACG  1223

Query  425   CCGGCAGCTGCTGCCTCGCGCCCGGGAGCAGGAGGAGCTGCTGGCGCCCGCAGATGG  481
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1224  CCGGCAGCTGCTGCCTCGCGCCCGGGAGCAGGAGGAGCTGCTGGCGCCCGCAGATGG  1280
```

3’ partner: SCO2



Junction point
exon=2 146..1002
/gene="SCO2"
/gene_synonym="MGC125823; MGC125825; SCO1L"

BLAST vs mRNA

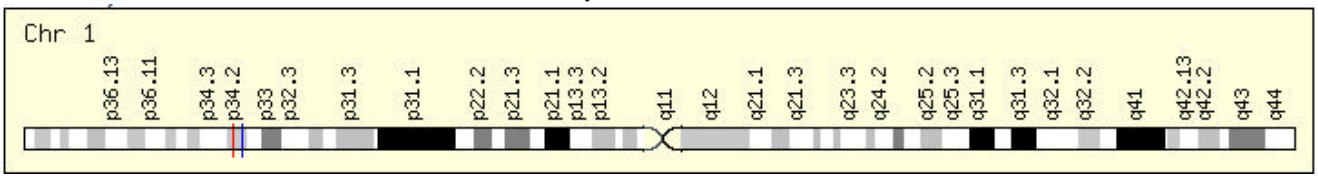
>ref|NM_001169110.1| Homo sapiens SCO cytochrome oxidase deficient homolog 2 (yeast) (SCO2), nuclear gene encoding mitochondrial protein, transcript variant 3, mRNA
Length=1002
GENE ID: 9997 SCO2| SCO cytochrome oxidase deficient homolog 2 (yeast) [Homo sapiens]
Score = 828 bits (448), Expect = 0.0
Identities = 560/609 (92%), Gaps = 28/609 (4%)

```
Strand=Plus/Plus

                 ↓
Query  479   TGGAGCATCAGATCCATGCTGCTGCTGACTCGGAGCCCCACAGCTTGGCACAGGCTCTCT  538
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  144   TGGAGCATCAGATCCATGCTGCTGCTGACTCGGAGCCCCACAGCTTGGCACAGGCTCTCT  203

Query  539   CAGCTCAAGCCTCCGGTCCTCCCTGGGACCCTGGGAGGCCAGGCCCTGCATCTGAGGTCC  598
             |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Sbjct  204   CAGCTCAAGCCTCGGGTCCTCCCTGGGACCCTGGGAGGCCAGGCCCTGCATCTGAGGTCC  263

Query  599   TGGCTTTTGTCAAGGCAGGGCCCTGCAGAGACAGGTGGGCAGGGCCAGCCCCAGGGCCCT  658
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  264   TGGCTTTTGTCAAGGCAGGGCCCTGCAGAGACAGGTGGGCAGGGCCAGCCCCAGGGCCCT  323

Query  659   GGGCTTCGAACCCGGCTGCTGATCACAGGCCTGTTCGGGGCTGGACTCGGTGGGGCCTGG  718
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  324   GGGCTTCGAACCCGGCTGCTGATCACAGGCCTGTTCGGGGCTGGACTCGGTGGGGCCTGG  383

Query  719   CTGGCCCTGAGGGCTGAGAAGGAGAGGCTGCAGCAGCAAAAAGCGAACAGAAGCCCTGCG  778
             |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Sbjct  384   CTGGCCCTGAGGGCTGAGAAGGAGAGGCTGCAGCAGCAAAA-GCGAACAGAAGCCCTGCG  442

Query  779   CCGGGCAGCTGTGGGCCAGGGGCGACTTCCACCTGCTGGATCACAGATGCCCGGGCTCGC  838
             || |||||||||||||||| ||||||||||||||||||||||| ||| ||||||||||||
Sbjct  443   CCAGGCAGCTGTGGGCCAGGG-CGACTTCCACCTGCTGGATCACAGAGGCC-GGGCTCGC  500

Query  839   TGCCAGGCTGACTTCCCGGGGCCAGTGGGTGCTGATGTACTTTGGCTTCCCTCCACTGCC  898
             ||| |||||||||||| ||||||||||||||||||||||||||||| ||| |||||||||
Sbjct  501   TGCAAGGCTGACTTCC-GGGGCCAGTGGGTGCTGATGTACTTTGGCTTCACTC-ACTGCC  558

Query  899   CCT-ACATCTGCCCAAAACAAACTGGAAAAAGCTTGGTGCAAGGTGGTGCGGGCAGCTGG  957
             | | |||||||||||| | | | ||||| || ||| |||||| |||||||||| |||||||
Sbjct  559   C-TGACATCTGCCCAGA-CGAGCTGGAGAA-GCT-GGTGCA-GGTGGTGCGG-CAGCTGG  612
```

PPCS-LOC100507214

GCACGAGGCCGCTTCCCACCCCAGACTTGGCTGTCCGCTCTGCGGCCTTCGGGCCCAGCCCTTTCGGGCTTGCTGAGCCTGGAGGCCGAGGAGAATGCACTTC
CGGGTTTTGCTGAGGCTCTGAGGAGCTACCAGGAGGCTGCGGCTGCAGGCACCTTCCTGGCAGTAGAGTTCACCACTTTGGCGGACTATTTGCATCTGTTGCA
GGCTGCGGCCCAGGCACTCAATCCGCTAGGCCCTTCTGCGATGTTTTACCTGGCTGCGGCTGTGTCAGATTTCTATGTTCCTGTCTCTGAAATGCCTGAACAC
AAGATCCAGTCATCTGGGGGCCCACTGCAGGGAAAAGTTCAGTTAGAAGACATACTTCACCATCTTGAAAAAGAAGAAATCAATCCCCTTGCTACTACAGAAC
AACAACTCTGTTTGGTGCTTATTCCAGCCAGCACAGTGAAGACAGGCTGAGGACTGCTACCACAGATGTAGAAGAGCTTATAGTGAAGCACATGGGTGAAACA
AAAGAAGTGAGAACTAATAGCATAGAATTTTAAAGACACCTGTGATTTTGTTCATTGCCCTTCATTAAATTAACATATTAAAAACTAATGTTTGGCTATCACT
GTATAGTTTGAAAAGC

5’ partner: PPCS



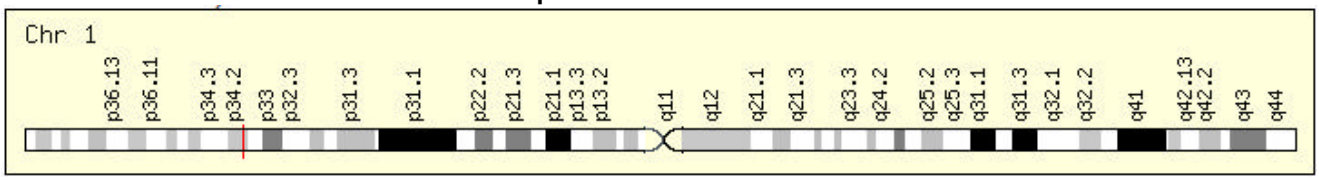
Junction point
exon=2 573..676
/gene="PPCS"
/gene_synonym="FLJ11838; MGC117357; MGC138220;
RP11-163G10.1"

BLAST vs mRNA

```
>ref|NM_024664.2|  Homo sapiens phosphopantothenoylcysteine synthetase (PPCS), transcript variant 1,
mRNA
Length=1489
GENE ID: 79717 PPCS| phosphopantothenoylcysteine synthetase [Homo sapiens]
Score =  612 bits (331),  Expect = 4e-174
Identities = 331/331 (100%), Gaps = 0/331 (0%)

Strand=Plus/Plus

Query  9     CCGCTTCCCACCCCAGACTTGGCTGTCCGCTCTGCGGCCTTCGGGCCCAGCCCTTTCGGG  68
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  346   CCGCTTCCCACCCCAGACTTGGCTGTCCGCTCTGCGGCCTTCGGGCCCAGCCCTTTCGGG  405

Query  69    CTTGCTGAGCCTGGAGGCCGAGGAGAATGCACTTCCGGGTTTTGCTGAGGCTCTGAGGAG  128
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  406   CTTGCTGAGCCTGGAGGCCGAGGAGAATGCACTTCCGGGTTTTGCTGAGGCTCTGAGGAG  465

Query  129   CTACCAGGAGGCTGCGGCTGCAGGCACCTTCCTGGCAGTAGAGTTCACCACTTTGGCGGA  188
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  466   CTACCAGGAGGCTGCGGCTGCAGGCACCTTCCTGGCAGTAGAGTTCACCACTTTGGCGGA  525
```

```
Query  189  CTATTTGCATCTGTTGCAGGCTGCGGCCCAGGCACTCAATCCGCTAGGCCCTTCTGCGAT  248
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  526  CTATTTGCATCTGTTGCAGGCTGCGGCCCAGGCACTCAATCCGCTAGGCCCTTCTGCGAT  585

Query  249  GTTTTACCTGGCTGCGGCTGTGTCAGATTTCTATGTTCCTGTCTCTGAAATGCCTGAACA  308
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  586  GTTTTACCTGGCTGCGGCTGTGTCAGATTTCTATGTTCCTGTCTCTGAAATGCCTGAACA  645

Query  309  CAAGATCCAGTCATCTGGGGGCCCACTGCAG  339
            ||||||||||||||||||||||||||||||
Sbjct  646  CAAGATCCAGTCATCTGGGGGCCCACTGCAG  676
```

3' partner: LOC100507214



BLAST vs mRNA

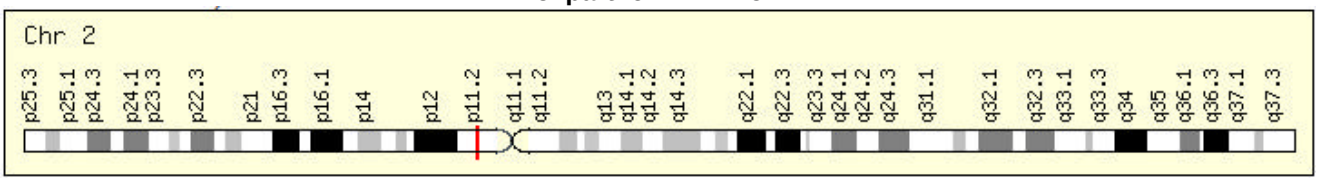
```
>ref|XR_113349.1| PREDICTED: Homo sapiens hypothetical LOC100507214 (LOC100507214), partial miscRNA
Length=668
GENE ID: 100507214 LOC100507214| hypothetical LOC100507214 [Homo sapiens]
Score =  544 bits (294),  Expect = 1e-153
Identities = 296/297 (99%), Gaps = 0/297 (0%)
Strand=Plus/Plus
```

```
Query  338  AGGGAAAAGTTCAGTTAGAAGACATACTTCACCATCTTGAAAAAGAAGAAATCAATCCCC  397
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  371  AGGGAAAAGTTCAGTTAGAAGACATACTTCACCATCTTGAAAAAGAAGAAATCAATCCCC  430

Query  398  TTGCTACTACAGAAGAACAACTCTGTTTGGTGCTTATTCCAGCCAGCACAGTGAAGACAG  457
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  431  TTGCTACTACAGAAGAACAACTCTGTTTGGTGCTTATTCCAGCCAGCACAGTGAAGACAG  490

Query  458  GCTGAGGACTGCTACCACAGATGTAGAAGAGCTTATAGTGAAGCACATGGGTGAAACAAA  517
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  491  GCTGAGGACTGCTACCACAGATGTAGAAGAGCTTATAGTGAAGCACATGGGTGAAACAAA  550

Query  518  AGAAGTGAGAACTAATAGCATAGAATTTTAAAGACACCTGTGATTTTGTTCATTGCCCTT  577
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  551  AGAAGTGAGAACTAATAGCATAGAATTTTAAAGACACCTGTGATTTTGTTCATTGCCCTT  610

Query  578  CATTAAATTAACATATTAAAAACTAATGTTTGGCTATCACTGTATAGTTTGAAAAGC  634
            ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  611  CATTAAATTGACATATTAAAAACTAATGTTTGGCTATCACTGTATAGTTTGAAAAGC  667
```

RMND5A-ANAPC1

```
TTGGCGTGAAAACAAAAACAATCGGCCGCGCCGTCGCAGGCACCCGAACGTCGCGAGCGGGGCCTGGGGACGCGGGGCCGAGTGCAGCGAGCGAACGGGAGC
GCGGCGACTCGCCAGGGGGCTAGGGCGCCATGGGGCAGGCGGGCTCCGGCTGCGCGGGGCTCCCCCGGCGCCGCGGCTAGTGCGCCCGCCGCCTCGGCCGCC
CAGCCTCCCGCGCCGCCCGCTTGGGGAACGAGGAGCAGGACGCGGCCTCGGCGGGGCCCGGGCCGAACGGCTGCGGACACCTGGGCGCCGAGGAGCCGAGCG
CGCCGCCTCCGGCATGGATCAGTGCGTGACGGTGGAGCGCGAGCTGGAGAAGGTGCTGCACAAGTTCTCGGGCTACGGGCAGCTGTGCGAGCGCGGCCTGGA
GAGCTCATCGACTACACCGGCGGTCTCAAGCACCAGATCCTGCAGAGCCACGGCCAAGATGCTGAATTATCAGGGACACTTTCACTTGTTTTGACACAGTGC
GTAAAAGAATAAAGGATACTGTTCAAAAATTGGCCTCCGACCACAAAGACATCCACAGCAGTGTTTCTCGGGTTGGAAAAGCCATTGATAAGGATTCACTTTA
AGAGATTTGGAAACTCTTCCCTTTGGAATTGCTCTTCCCATCAGAGATGCAATTTATCACTGTCGTGAACAGCCTGCCTCAGACTGGCCAGAAGCTGTCTGTC
TCTTGATTGGACGTCAGGATCTTTCCAAGCAGGCCTGCGAAGGAAACTTACTCAAAGGAGTCTATGTTCCTTCAGAACAGAAACTGAGAGGAAGATGACGGCA
TGAATGACATGAATCACGAGTCATGTCATTAATATGGAGTGAGATTTATGTGCAGGTGTGCGAAGGCTTCTTCAGAGTGCGCATCTGTCCGTGTCATGTAGTC
CAGTACCCGAGCTCATGACCACGAGTCACTCGAAGGAAAGGAATCGATGCCTCAATGGGTCACGACTATGTCTTCGTAGACAGGAGTTACTGATCGTACATCT
GTTCACGAGCATGCATCTAATTGAACTGACTGCAGTCCTCCGGAACCAGTGACTAATGAACTCGAAGTCTCCAATGGAACCCGTGCACCTTCAAG
```

5' partner: RMND5A



Junction point
exon=2 520..662
/gene="RMND5A"

/gene_synonym="CTLH; FLJ12753; FLJ13910; FLJ21795; MGC78451; p44CTLH; RMD5"

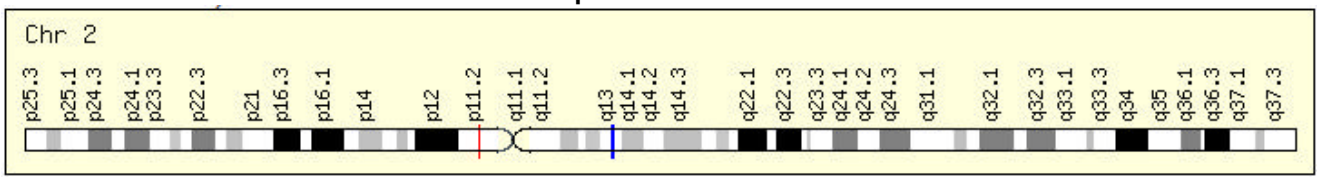
BLAST vs mRNA

>ref|NM_022780.3| UniGene info linked to NM_022780.3GEO profiles info linked to NM_022780.3Gene info linked to NM_022780.3Genome view with mapviewer linked to NM_022780.3 Homo sapiens required for meiotic nuclear division 5 homolog A (S. cerevisiae) (RMND5A), mRNA
Length=6201
GENE ID: 64795 RMND5A | required for meiotic nuclear division 5 homolog A (S. cerevisiae) [Homo sapiens]
Score = 1083 bits (586), Expect = 0.0
Identities = 600/607 (99%), Gaps = 0/607 (0%)

Strand=Plus/Plus

```
Query  5    CGTGAAAACAAAAACAATCGGCCGCGCCGTCGCAGGCACCCGAACGTCGCGAGCGGGGCC  64
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  60   CGTGAAAACAAAAACAATCGGCCGCGCCGTCGCAGGCACCCGAACGTCGCGAGCGGGGCC  119

Query  65   TGGGGACGCGGGGCCGAGTGCAGCGAGCGAACGGGAGCAGCGGCGACTCGCCAgggggct  124
            |||||||||||| ||||||||||||||||||||||||||||||||||| |||||||
Sbjct  120  TGGGGACGCGGAGCCGAGTGCAGCGAGCGAACGGGAGCAGCGGCGACTCGCCGGGGGGCT  179

Query  125  agggcgccatggggcaggcgggctccggctgcgcggggctcccccggcgccgcggctagt  184
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  180  AGGGCGCCATGGGGCAGGCGGGCTCCGGCTGCGCGGGGCTCCCCCGGCGCCGCGGCTAGT  239

Query  185  gcgcccgccgcctcggccgcctcagcctcccgcgccgcccgcTTGGGGAACGAGGAGCAG  244
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  240  GCGCCCGCCGCCTCGGCCGCCTCAGCCTCCCGCGCCGCCCGCTTGGGGAACGAGGAGCAG  299

Query  245  GACGCGGCCTCGGCGGGGCCCGGGCCGAACGGCTGCGGACACCTGGGCGCCGAGGAGCCG  304
            |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Sbjct  300  GACGCGGCCTCGGTGGGGCCCGGGCCGAACGGCTGCGGACACCTGGGCGCCGAGGAGCCG  359

Query  305  AGCGCCGCCGCCTCCGGCATGGATCAGTGCGTGACGGTGGAGCGCGAGCTGGAGAAGGTG  364
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  360  AGCGCCGCCGCCTCCGGCATGGATCAGTGCGTGACGGTGGAGCGCGAGCTGGAGAAGGTG  419

Query  365  CTGCACAAGTTCTCGGGCTACGGGCAGCTGTGCGAGCGCGGCCTGGAGGAGCTCATCGAC  424
            |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Sbjct  420  CTGCACAAGTTCTCAGGCTACGGGCAGCTGTGCGAGCGCGGCCTGGAGGAGCTCATCGAC  479

Query  425  TACACCGGCGGTCTCAAGCACCAGATCCTGCAGAGCCACGGCCAAGATGCTGAATTATCA  484
            |||||||||| ||||||||| ||||||||||||||||||||||||||||||||||||||
Sbjct  480  TACACCGGCGGCCTCAAGCACGAGATCCTGCAGAGCCACGGCCAAGATGCTGAATTATCA  539

Query  485  GGGACACTTTCACTTGTTTTGACACAGTGCTGTAAAAGAATAAAGGATACTGTTCAAAAA  544
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  540  GGGACACTTTCACTTGTTTTGACACAGTGCTGTAAAAGAATAAAGGATACTGTTCAAAAA  599

Query  545  TTGGCCTCCGACCACAAAGACATCCACAGCAGTGTTTCTCGGGTTGGAAAAGCCATTGAT  604
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  600  TTGGCCTCCGACCACAAAGACATCCACAGCAGTGTTTCTCGGGTTGGAAAAGCCATTGAT  659

Query  605  AAGGATT  611
            ||| |||
Sbjct  660  AAGAATT  666
```

3' partner: ANAPC1



Junction point
exon=25 3060..3238
/gene="ANAPC1"
/gene_synonym="APC1; MCPR; TSG24"

BLAST vs mRNA

>ref|NM_022662.2| UniGene info linked to NM_022662.2GEO profiles info linked to NM_022662.2Gene info linked to NM_022662.2Genome view with mapviewer linked to NM_022662.2 Homo sapiens anaphase promoting complex subunit 1 (ANAPC1), mRNA
Length=6329
GENE ID: 64682 ANAPC1 | anaphase promoting complex subunit 1 [Homo sapiens]
Score = 510 bits (276), Expect = 2e-143
Identities = 359/393 (92%), Gaps = 30/393 (7%)

Strand=Plus/Plus

```
Query  606   AGGATTCACTTTAAGAGATTTGGAAACTCTTCCCTTTGGAATTGCTCTTCCCATCAGAGA  665
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3058  AGGATTCACTTTAAGAGATTTGGAAACTCTTCCCTTTGGAATTGCTCTTCCCATCAGAGA  3117

Query  666   TGCAATTTATCACTGTCGTGAACAGCCTGCCTCAGACTGGCCAGAAGCTGTCTGTCTCTT  725
             |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Sbjct  3118  TGCAATTTATCACTGTCGTGAGCAGCCTGCCTCAGACTGGCCAGAAGCTGTCTGTCTCTT  3177

Query  726   GATTGGACGTCAGGATCTTTCCAAGCAGGCCTGCGAAGGAAACTTACTCAAAGG-A-GTC  783
             |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||| | |||
Sbjct  3178  GATTGGACGTCAGGATCTTTCCAAGCAGGCCTGCGAAGGAAACTTACCCAAAGGGAAGTC  3237

Query  784   T--------AT--G-T-T-CCTTCAG-AACAGAAACTGAG-AGGAAGATGACGGCATGAA  828
             |        ||  | | | ||||||| |||||||||||| ||||||||||||||||||||
Sbjct  3238  TGTGCTCTCATCAGATGTTCCTTCAGGAACAGAAACTGAGGAGGAAGATGACGGCATGAA  3297

Query  829   TGACATGAATCACGAG-TCATGTCATTAATATGGAGTGA-GATTTATG--TGCAGG-TGT  883
             ||||||||||||||| |||||||||||||||||||||| |||||| |  |||||| |||
Sbjct  3298  TGACATGAATCACGAGGTCATGTCATTAATATGGAGTGAAGATTTAAGGGTGCAGGATGT  3357

Query  884   GCGAAGGCTTCTTCAGAGTGCGCATC-TGTCCGTGTCAT-GTAGTGCAGTACCC-GAGCT  940
             |||||||||||||||||||||||||| ||||||||||  |||||||||||||| |||||
Sbjct  3358  GCGAAGGCTTCTTCAGAGTGCGCATCCTGTCCGTGTCAACGTAGTGCAGTACCCAGAGCT  3417

Query  941   CA-TGACCACGAGT-CACTCGAAGGAAA-GGAA  970
             || |||||||||| || |||| ||||| ||||
Sbjct  3418  CAGTGACCACGAGTTCA-TCGA-GGAAAAGGAA  3448
```

Supplemental References

[1] Wilming LG, Gilbert JG, Howe K, Trevanion S, Hubbard T, and Harrow JL. The vertebrate genome annotation (Vega) database. *Nucleic Acids Res.* 2008; 36:D753-60.

[2] Pruitt KD, Tatusova T, Klimke W, and Maglott DR. NCBI Reference Sequences: current status, policy and new initiatives. *Nucleic Acids Res.* 2009; 37:D32-6.

[3] Karolchik D, Kuhn RM, Baertsch R, Barber GP, Clawson H, Diekhans M, Giardine B, Harte RA, Hinrichs AS, Hsu F et al. The UCSC Genome Browser Database: 2008 update. *Nucleic Acids Res.* 2008; 36:D773-9.

[4] Romani A, Guerra M, Trerotola M, and Alberti S. Detection and analysis of spliced chimeric mRNAs in sequence databanks. *Nucleic Acids Res.* 2003; 31:1-8.

[5] Calin GA, Sevignani C, Dumitru CD, Hyslop T, Noch E, Yendamuri S, Shimizu M, Rattan S, Bullrich F, Negrini M et al. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci U S A.* 2004; 101:2999-3004.

[6] Guerra E, Trerotola M, Dell' Arciprete R, Bonasera V, Palombo B, El-Sewedy T, Ciccimarra T, Crescenzi C, Lorenzini F, Rossi C et al. A bi-cistronic CYCLIN D1-TROP2 mRNA chimera demonstrates a novel oncogenic mechanism in human cancer. *Cancer Res.* 2008; 68:8113-8121.

[7] Li H, Wang J, Mor G, and Sklar J. A neoplastic gene fusion mimics trans-splicing of RNAs in normal human cells. *Science.* 2008; 321:1357-61.

[8] Terrinoni A, Dell'Arciprete R, Fornaro M, Stella M, and Alberti S. The Cyclin D1 gene contains a cryptic promoter that is functional in human cancer cells. *Genes Chromosomes Cancer.* 2001; 31:209-20.

[9] Maher CA, Kumar-Sinha C, Cao X, Kalyana-Sundaram S, Han B, Jing X, Sam L, Barrette T, Palanisamy N, and Chinnaiyan AM. Transcriptome sequencing to detect gene fusions in cancer. *Nature.* 2009; 458:97-101.

[10] Rickman DS, Pflueger D, Moss B, VanDoren VE, Chen CX, de la Taille A, Kuefer R, Tewari AK, Setlur SR, Demichelis F et al. SLC45A3-ELK4 is a novel and frequent erythroblast transformation-specific fusion transcript in prostate cancer. *Cancer Res.* 2009; 69:2734-8.

[11] Communi D, Suarez-Huerta N, Dussossoy D, Savi P, and Boeynaems J-M. Cotranscription and Intergenic Splicing of Human P2Y11 and SSF1 Genes. *J. Biol. Chem.* 2001; 276:16561-16566.