Cloning and sequence analysis of the LOC339524 gene in Sprague-Dawley rats

Z.H. Long[1], **H. Li**[1], **F. Chen**[1] **and L.Y. Zou**[2]

[1]Department of Anesthesiology, Xinqiao Hospital, Third Military Medical University, Chongqing, China
[2]College of Basic Medical Sciences, Third Military Medical University, Chongqing, China

Corresponding author: H. Li
E-mail: Lh78553@163.com

Genet. Mol. Res. 14 (4): 16577-16584 (2015)
Received August 5, 2015
Accepted October 25, 2015
Published December 11, 2015
DOI http://dx.doi.org/10.4238/2015.December.11.4

ABSTRACT. We cloned the LOC339524 gene in Sprague-Dawley (SD) rats and analyzed the structure and function of the protein encoded by it. Based on the known human LOC339524 gene sequences, the full-length coding sequence of the LOC339524 gene in SD rats was cloned and amplified by the polymerase chain reaction using the complementary DNA of SD rats as a template. Bioinformatics analysis showed that the length of the cloned LOC339524 gene (GenBank accession No. KM224520) was 831 bp and it encoded a deduced protein of 276 amino acids. Sequence analysis revealed that the coded protein was identical to that produced in humans and its functional domain was located in the 138-236 amino acid fragments, a proline-rich region. Our results suggest that the encoded protein may be a significant regulator of the inflammatory response and may provide sufficient information to justify an in-depth investigation of the role of the LOC339524 gene.

Key words: Sprague-Dawley rat (*Rattus norvegicus*); LOC339524 gene; Clone; Bioinformatics

 ©FUNPEC-RP www.funpecrp.com.br

INTRODUCTION

The human (*Homo sapiens*) LOC339524 gene sequence (GenBank accession No. BC126240.1) was determined in 2004 by the Human Genome Project (Gerhard et al.,2004), and is composed of 39,440 bases. In 2012, the human LOC339524 protein was identified for the first time using the high-performance liquid chromatography (HPLC)-Chip-MS system from effluents of a perfused human heart that had experienced myocardial ischemia (Li et al., 2012). The LOC339524 gene has the chromosomal locus 1p22.3 (Ota et al., 2004). This locus has a high frequency of deletions, amplifications, and translocations (Li et al., 2013), which are associated with tumors (Balakrishnanet al., 2006; Muranenet al., 2011; Johanssonet al., 2012; de Maturana et al., 2013; Li et al., 2013; Rinket al., 2013), alcohol dependence (Gizeret al., 2011), leprosy (Liu et al., 2013), and mandibular prognathism (Ikunoet al., 2014), suggesting that the LOC339524 gene might be involved in the occurrence of these diseases and that its function is extremely important. Both the LOC339524 gene sequence of the Sprague-Dawley (SD) rat and the function of the protein it encodes are still unknown. To facilitate further study into the function of the protein, cloning and bioinformatics analysis of the LOC339524 gene in SD rats are particularly important.

Based on both the *H. sapiens* LOC339524 gene sequence and the complementary DNA (cDNA) gene of SD rats, for the first time we cloned and analyzed the LOC339524 gene sequence of SD rats, and predicted the structure and function of the protein it encodes. This study lays the foundation for further research on LOC339524 gene function and its molecular mechanisms.

MATERIAL AND METHODS

Sample preparation

The feeding of the SD rats and the collection of their hearts were conducted in accordance with the International Guiding Principles for Biomedical Research Involving Animals (http://www.cioms.ch/index.php/texts-of-guidelines) and approved by the Third Military Medical University Council on Animal Care Guidelines. The SD rats were purchased from the Animal Center of Daping Hospital (Chongqing, China) and their fresh myocardium tissues were mixed with TRIzol (Invitrogen, USA); then, they were constantly milled until the TRIzol liquid was free of fine particles.

Total RNA extraction and first-strand cDNA synthesis

Total RNA was extracted using RNAiso Plus (TaKaRa Biotechnology Co. Ltd., Dalian, China), strictly following the manufacturer instructions. The concentration of total RNA was checked using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific Inc.) at 260 nmin terms of optical density values. The RNA integrality was assessed using 1.0% agarose gel electrophoresis. First-strand cDNA synthesis was performed from 4 μg of total RNA by using an EasyScript First-Strand cDNA Synthesis SuperMix kit (TaKaRa, Japan) following the manufacturer instructions. The reverse-transcription products were then used for polymerase chain reaction (PCR) amplification and stored at -80°C.

Cloning and sequencing of the LOC339524 gene

To extend the cDNA, the reverse transcription polymerase chain reaction (RT-PCR) primers were designed as follows: forward primer (LOC-1): 5'-ATGCTGGCCCGCCGGGACCT-3'; reverse

primer (LOC-2): 5'-TTACAAAAGGCAACTTCTGGGGA-3', according to the mRNA sequence of the LOC339524 gene of *H. sapiens* and using Primer5.0 software. The reaction system (25 μL) comprised 2 μL first-strand cDNA as a template, 4 μL dNTP (2 mM), 1 μL knockout drops (KOD-FX), 2 x 12.5 μL KOD buffer, 0.5 μL LOC-1 (10 mM), 0.5 μL LOC-2 (10 mM), and 4.5 μL double-distilled H_2O. Samples were run at 98°C for 8 min, followed by 35 cycles of 98°C for 30 s, and 68°C for 90 s (Bao et al., 2014). The RT-PCR products were purified on 1.0% agarose gel. After adding an A-tail, the product was connected to *Escherichia coli* pTA2 and transformed into pTA2-competent cells. Finally, positive clones were sent to Invitrogen Biotech Co. Ltd. (Shanghai, China) for sequencing.

Bioinformatic analysis

The open reading frame (ORF) of the LOC339524 gene was analyzed using the Biotechnology Information (NCBI) ORF Finder ((http://www.ncbi.nlm.nih.gov/gorf/gorf.html). The encoding protein's structure and physicochemical properties were predicted using ExPASy (http://www.expasy.org/vg/index/Protein). The N-terminal signal peptide structure was predicted using the SignalP 4.0 Server program. The transmembrane regions were analyzed using ExPASy Proteomics Server software (http://www.expasy.org/resources), i.e., TMpred and TopPred. Alignment of amino acid sequences was performed using the Basic Local Alignment Search Tool (BLAST)server combined with the DNAMAN8 program. The phylogenetic tree was ultimately constructed by adopting version 6 of the Molecular Evolutionary Genetics Analysis (MEGA) program.

RESULTS

Cloning and sequencing of the LOC339524 gene

The RT-PCR product of the LOC339524 gene was a pure specific DNA fragment of nearly 800 bp (Figure 1). The sequencing results showed that the cloned LOC339524 gene was 831 bp in length (Figure 2). Both results are consistent.

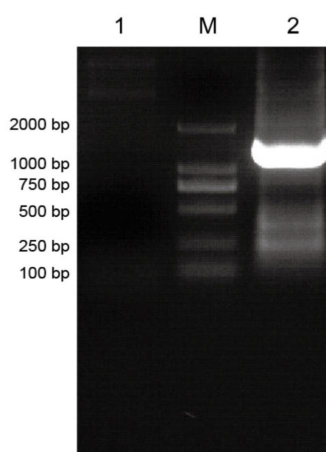

Figure 1. Electrophoresis identification of the reverse transcription polymerase chain reaction (RT-PCR) amplification product of the LOC339524 gene from Sprague-Dawley (SD) rats. *Lanes 1*, *2*, and *M* show the negative control, the target gene, and the DL2000 DNA marker, respectively.

```
1   atgctggcccgccgggacctgggacttgtgccacatggagtgtcg
    M  L  A  R  R  D  L  G  L  V  P  H  G  V  S
46  ggagtctccattgccgcgagttctacaccacagggccaggctgtt
    G  V  S  I  A  A  S  S  T  P  Q  G  Q  A  V
91  tgctccccatcggtcgctgcccccagcaccctgttgttattaagg
    C  S  P  S  V  A  A  P  S  T  L  L  L  L  R
136 actcatttgcttggagcggcatcattacaagggtgtggggtacta
    T  H  L  L  G  A  A  S  L  Q  G  C  G  V  L
181 catatactccctatttttctatttttcgaaaggctgcaggcgcgat
    H  I  L  P  I  F  L  F  S  K  G  C  R  R  D
226 gcacagtgcgcttgcacggtggggcctagtgctagcccgaggagc
    A  Q  C  A  C  T  V  G  P  S  A  S  P  R  S
271 ggacgggggccgggcagggggcggtgggcgccggcctcgtctcggt
    G  R  G  P  G  R  G  G  G  R  R  P  R  L  G
316 gctgctcggtcaggctgtcccggcgcggcggccgcgggaggccct
    A  A  R  S  G  C  P  G  A  A  A  A  G  G  P
361 gccgtccttcacccctggaggcgggctgggggcgggtgcggggc
    A  V  L  H  P  W  R  R  A  G  G  R  V  R  G
406 gcttcccctcctcagggccctcaaaccgcaagggggtttccgctt
    A  S  P  P  Q  G  P  Q  T  A  R  G  F  P  L
451 cccagtcgatggtcgtcctctcccatccccggctgcatctccatt
    P  S  R  W  S  S  S  P  I  P  G  C  I  S  I
496 tacccgtcgcccatttcctttgcccatccaggctccttggctcca
    Y  P  S  P  I  S  F  A  H  P  G  S  L  A  P
541 ctggggtctccgttcccttctcccggtccccccctccaggtcgcgg
    L  G  S  P  F  P  S  P  G  P  P  S  R  S  R
586 ctcctttgtccaggactacgcaggggcttgaccccagggcgctgg
    L  L  C  P  G  L  R  R  G  L  T  P  G  R  W
631 tttaggccggatctggggtcccttgtcactcccaggcttcttcca
    F  R  P  D  L  G  S  L  V  T  P  R  L  L  P
676 cttccgaattctggagaaccgggaatcaagccctgcgcgttcctc
    L  P  N  S  G  E  P  G  I  K  P  C  A  F  L
721 ttcttcctccttcgtgccgaaagcacgcttcatgtctgccagggc
    F  F  L  L  R  A  E  S  T  L  H  V  C  Q  G
766 atcagttctgaaagtgagcggagaacaaggagtttctttttcttc
    I  S  S  E  S  E  R  R  T  R  S  F  F  F  F
811 cccagaagttgccttttgtaa 831
    P  R  S  C  L  L  *
```

Figure 2. Nucleotide and deduced amino acid sequence of the LOC339524 protein. The upper row showsthe nucleotide sequence, and the lower row is the matching amino acid sequence. The first deduced amino acid was methionine (M). The only transmembrane region is demarcated by the black horizontal line. An asterisk marks the TAA stop codon.

Analysis of the LOC339524 gene

The full length ORF was obtained by PCR using gene-specific primers that were designed using sequences of predicted full-length LOC339524 cDNA. The most likely ORF, 831 bp, coded for 267 amino acids. This also further confirmed the LOC339524 sequence. The clone in our study was then submitted to the GenBank database (accession No. KM224520). The homology analysis was performed by NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The results illustrated that the full-length coding sequence of the LOC339524 gene from the SD rats was 100% identical to *H. sapiens* (GenBank:BC126240.1).

Bioinformatic analysis of the amino acid sequence

ExPASy online software was used to analyze the LOC339524 protein. The results indicated that its predicted molecular mass was 28.7 kDa and the instability index (II) was computed to be 77.59; the aliphatic index was 79.96; the grand average of hydropathicity (GRAVY) was -0.01; and the N-terminal of the sequence considered was methionine (Met).

The SignalP 4.0 Server program was utilized to analyze the N-terminal signal peptide structure of *Rattus norvegicus* LOC339524 protein, and there was no significant N-terminal signal peptide cleavage site. TMpred and TopPred transmembrane prediction software were used to predict the transmembrane region of the coding protein, and found that there might be one helical transmembrane domain signal (Figure 3A, B) at residues 47-68. Protcomp Version 9.0 subcellular localization prediction indicated that the LOC339524 protein was located in the cell membrane.

ExPASy analyzed the secondary structure of the encoded protein (Li et al., 2014), and showed that there were three alpha helices and five extension structures (Figure 4). A conserved domain was not found by the NCBI online Conserved Domain Search. The SWISS-MODEL software was adopted for tertiary structure homology modeling of the LOC339524 protein (Schwede et al., 2003; Arnold et al., 2006), and no suitable structure templates were found. Further analysis using PROSITE online software (http://prosite.expasy.org/) found that the functional domain of the protein was a proline-rich region located at residues 138-236.

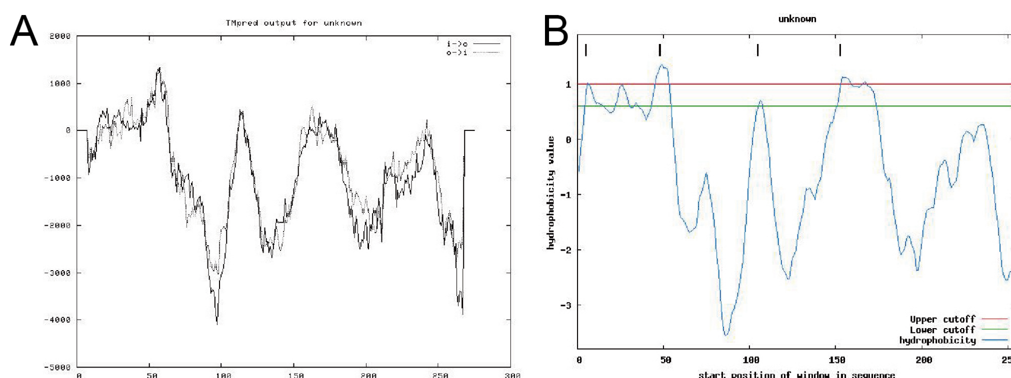

Figure 3. A. TMpred software-analyzed transmembrane region of the LOC339524 protein. **B.** TopPred software-analyzed transmembrane region of the deduced amino acid sequence of the LOC339524 protein.

1 M L A R R D L G L V P H G V S G V S I A A S S T P Q G Q A V C S P S V A A P S T L L L L R T H L L G 50
51 A A S L Q G C G V L H I L P I F L F S K G C R R D A Q C A C T V G P S A S P R S G R G P G R G G G R 100
101 R P R L G A A R S G C P G A A A A G G P A V L H P W R R A G G R V R G A S P P Q G P Q T A R G F P L 150
151 P S R W S S S P I P G C I S I Y P S P I S F A H P G S L A P L G S P F P S P G P P S R S R L L C P G 200
201 L R R G L T P G R W F R P D L G S L V T P R L L P L P N S G E P G I K P C A F L F F L L R A E S T L 250
251 H V C Q G I S S E S E R R T R S F F F F P R S C L L

KEY	Helix	Sheet	Disordered	Disordered protein binding	Dompred Boundary	DomSSEA Boundary
Annotations	M	L	E	E	A	D

Figure 4. Secondary structure of the LOC339524 protein from Sprague-Dawley (SD) rats.

Six primates with the LOC339524 protein were found using BLASTP. They were *H. sapiens*, *Pan troglodytes*, *Pongo abelii*, *Nomascus leucogenys*, *Callithrix jacchus*, and *Macaca fascicularis*. The sequence alignment was analyzed utilizing DNAMAN8 software (Figure 5) and showed that the homology between the LOC339524 protein from the SD rats and humans was 100%. This was highly consistent with the phylogenetic tree (Figure 6) that was constructed using the MEGA6 program.

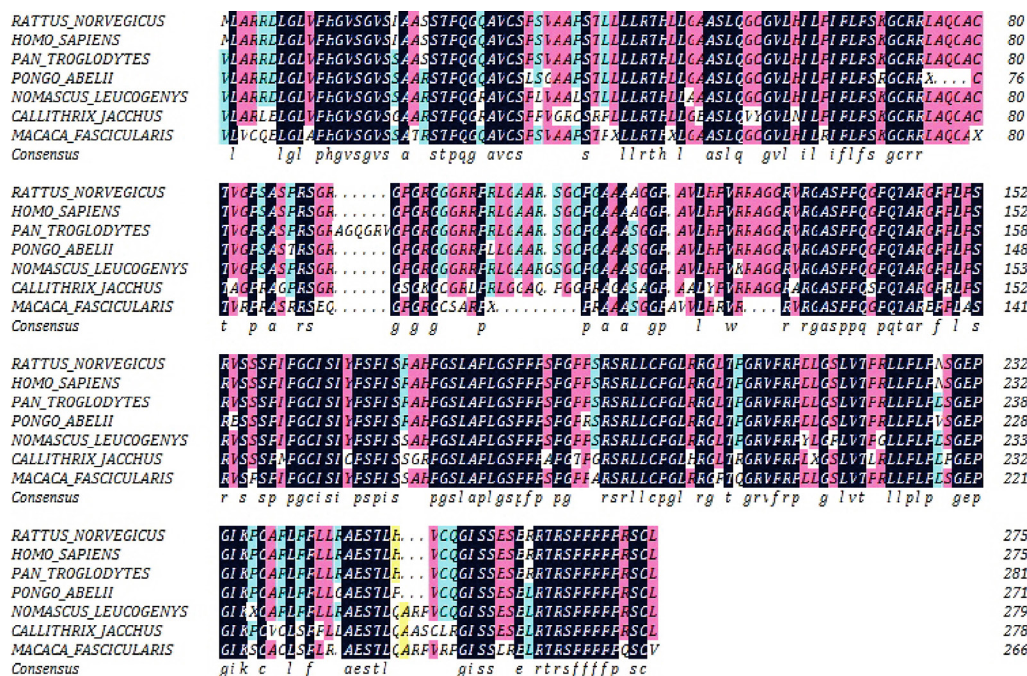

Figure 5. Multiple alignment of predicted amino acid sequences of the LOC339524 protein from Sprague-Dawley (SD) rats with other species.

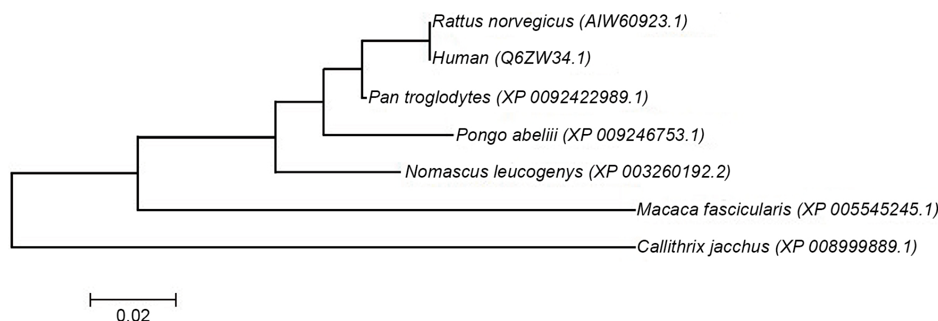

Figure 6. Phylogenetic analysis of the LOC339524 protein relative to the homologs of other organisms. Based on the neighbor-joining method, sequence alignment was processed using the MEGA 6.0 program.

DISCUSSION

This is the first time that the LOC339524 gene from SD rats has been cloned successfully, based on rat cDNA and known *H. sapiens* LOC339524 gene sequences. ExPASy ProtParam online analysis showed that the LOC339524 protein instability index was so large that it could be classified as an unstable protein, suggesting that the protein is active and important. Transmembrane prediction software indicated that at residues 47-68, the protein may have a transmembrane helix, which is consistent with the results reported by Ota et al. (2004), suggesting that the protein is a single transmembrane protein without a putative signal peptide, just like hypoxia and Akt-induced stem cell factor (HASF) (Beigi et al., 2013). Further analysis showed that the functional domain of the LOC339524 protein is located at residues 138-236, a region that is likely to induce and regulate cytokine production (Boissanet al., 2014; Turnquistet al., 2014), thereby regulating the immune system. The LOC339524 protein may be a regulator of inflammatory response (Liu et al., 2014), and the proline-rich region may be a receiver of regulatory factors.

Although the LOC339524 protein's homology with those of the six primates is extremely high, suggesting that it is highly conserved during evolution and that its function is crucial, predictions were made based on the uncharacterized protein and this is the direct reason for the failed tertiary structure homology modeling. This shows that the protein is brand-new, is different from all known proteins, and has great research value.

We successfully cloned the LOC339524 gene in SD rats and analyzed its structure and function. These results should help understand the role of the gene at chromosomal locus 1p22.3, as well as related diseases.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

The authors would like to thank H. Yu from the Central Laboratory of Xinqiao Hospital, Dr. D.G. Chen from the oncology department of Xinqiao Hospital, and X.Y. Fan from Xi'nan University for kindly providing support during the bioinformatics analysis. Research partially supported by both the National Natural Science Foundation of China (Grant #81070094) and the Natural Science Foundation Project of CQ CSTC (#2013C292).

REFERENCES

Arnold K, Bordoli L, Kopp J and Schwede T (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22: 195-201.

Balakrishnan A, von Neuhoff N, Rudolph C, Kamphues K, et al. (2006). Quantitative microsatellite analysis to delineate the commonly deleted region 1p22.3 in mantle cell lymphomas. *Genes Chromosomes Cancer* 45: 883-892.

Bao C, Bao W, Guo Z, He Q, et al. (2014). Cloning and tissue-specific Expression of TCTP Gene in Inner Mongolia cashmere goat (*Capra hircus*). *Biotechnol. Bull.* 7: 125-130.

Beigi F, Schmeckpeper J, Pow-Anpongkul P, Payne JA, et al. (2013). C3orf58, a novel paracrine protein, stimulates cardiomyocyte cell-cycle progression through the PI3K-AKT-CDK7 pathway. *Circ. Res.* 113: 372-380.

Boissan M, Montagnac G, Shen Q, Griparic L, et al. (2014). Membrane trafficking. Nucleoside diphosphate kinases fuel dynamin superfamily proteins with GTP for membrane remodeling. *Science* 344: 1510-1515.

de Maturana EL, Ye Y, Calle ML, Rothman N, et al. (2013). Application of multi-SNP approaches Bayesian LASSO and AUC-RF to detect main effects of inflammatory-gene variants associated with bladder cancer risk. *PLoS One* 8: e83745.

Gerhard DS, Wagner L, Feingold EA, Shenmen CM, et al. (2004). The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). *Genome Res.* 14: 2121-2127.

Gizer IR, Ehlers CL, Vieten C, Seaton-Smith KL, et al. (2011). Linkage scan of alcohol dependence in the UCSF Family Alcoholism Study. *Drug Alcohol Depend.* 113: 125-132.

Ikuno K, Kajii TS, Oka A, Inoko H, et al. (2014). Microsatellite genome-wide association study for mandibular prognathism. *Am. J. Orthod. Dentofacial Orthop.* 145: 757-762.

Johansson M, Roberts A, Chen D, Li Y, et al. (2012). Using prior information from the medical literature in GWAS of oral cancer identifies novel susceptibility variant on chromosome 4 - the AdAPT method. *PLoS One* 7: e36888.

Li H, Li J, Wang Y and Yang T (2012). Proteomic analysis of effluents from perfused human heart for transplantation: identification of potential biomarkers for ischemic heart damage. *Proteome Sci.*10: 21.

Li P, Lin Y, Zhang Y, Zhu Z, et al. (2013). SSX2IP promotes metastasis and chemotherapeutic resistance of hepatocellular carcinoma. *J. Transl. Med.* 11:52.

Li W, Liu Y, Sheng X, Yin P, et al. (2014). Structure and mechanism of a type III secretion protease, NleC. *Acta. Crystallogr. D. Biol. Crystallogr.* 70: 40-47.

Liu H, Bao F, Irwanto A, Fu X, et al. (2013). An association study of TOLL and CARD with leprosy susceptibility in Chinese population. *Hum. Mol. Genet.* 22: 4430-4437.

Liu Y, Carlsson R, Comabella M, Wang J, et al. (2014). FoxA1 directs the lineage and immunosuppressive properties of a novel regulatory T cell population in EAE and MS. *Nat. Med.* 20: 272-282.

Muranen TA, Greco D, Fagerholm R, Kilpivaara O, et al. (2011). Breast tumors from CHEK2 1100delC-mutation carriers: genomic landscape and clinical implications. *Breast Cancer Res.* 13: R90.

Ota T, Suzuki Y, Nishikawa T, Otsuki T, et al. (2004). Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nat. Genet.* 36: 40-45.

Rink L, Ochs MF, Zhou Y, von Mehren M, et al. (2013). ZNF-mediated resistance to imatinib mesylate in gastrointestinal stromal tumor. *PLoS One* 8: e54477.

Schwede T, Kopp J, Guex N and Peitsch MC (2003). SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res.* 31: 3381-3385.

Turnquist C, Wang Y1, Severson DT1, Zhong S, et al. (2014). STAT1-induced ASPP2 transcription identifies a link between neuroinflammation, cell polarity, and tumor suppression. *Proc. Natl. Acad. Sci. U. S. A.* 111: 9834-9839.

 ©FUNPEC-RP www.funpecrp.com.br