Turkish Journal of Medical Sciences

Volume 39 | Number 3 Article 3

1-1-2009

Characterization of the human GYS2 gene and its product using bioinformatic tools

FUAT DİLMEC

ALİ ATAŞ

ORIGINAL ARTICLE

Turk J Med Sci
2009; 39 (3): 353-359
© TÜBİTAK
E-mail: medsci@tubitak.gov.tr
doi:10.3906/sag-0803-33

Fuat DİLMEÇ[1]
Ali ATAŞ[2]

Characterization of the human *GYS2* gene and its product using bioinformatic tools

Aim: The *GYS2* gene, which encodes for glycogen synthase 2 (liver) (GS), is an enzyme responsible for the synthesis of 1,4-linked glucose chains in glycogen. The present study aimed to investigate the homology, conserved domain, and promoter and expression profiles of the human *GYS2* gene among various vertebrate species using bioinformatic tools.

Materials and Methods: We analyzed the homology with NCBI blast, the conserved domain with EBI ClustalW and Mega4, the promoter with Genomatix, and the expression profiles with DigiNorthern software.

Results: GS proteins and their conserved domains (Glycogen_syn) were more conserved in all the organisms investigated. There was 1 fully conserved domain (Glycogen_syn) and several truncated sub-domains. Comparative screening of the promoters showed that *GYS2* genes did not have any common conserved transcription factor binding sites.

Conclusions: This study shows that the GS molecules in various species, except *Ornithorhynchus anatinus* and *Danio rerio*, were well conserved throughout evolution.

Key Words: Glycogen synthase, genomics, evolution, promoter, expression

[1] Department of Medical Biology, Faculty of Medicine, Harran University, Şanlıurfa - TURKEY

[2] Department of Pediatrics, Faculty of Medicine, Harran University, Şanlıurfa - TURKEY

İnsan *GYS2* geni ve onun ürününün biyoinformatik araçlarla karekterizasyonu

Amaç: *GYS2* geni, glikojende 1,4-glukoz zincirlerinin sentezinden sorumlu olan glikojen sentaz 2 (karaciğer) (GS)'yi kodlar. İnsanın *GYS2* geninin çeşitli omurgalı türler arasında homolojisi, korunan bölgeleri, promotor ve ekspresyon profillerinin biyoinformatik araçlarla araştırılması amaçlandı.

Yöntem ve Gereç: Homolojiyi NCBI blast, korunan bölgeleri EBI ClustalW ve Mega4, promoter Genomatix ve ekspresyon profillerini DigiNorthern yazılımları ile analiz ettik.

Bulgular: Sonuçlarımız, GS proteinleri ve onların korunan bölgelerinin (Glycogen_syn) incelenen tüm organizmalar arasında fazlaca korunduğunu gösterdi. Tümü, bir korunan bölge (Glycogen_syn) ve bir kaç kesikli alt bölgeye sahipti. Promotorların karşılaştırılmalı taranması ile, *GYS2* genin genel olarak herhangi bir korunan transkripsiyon faktör bağlama alanın olmadığı gösterildi.

Sonuç: Bu çalışma evrimsel süreç boyunca GS proteinin, incelenen türlerde *Ornithorhynchus anatinus* ve *Danio rerio* dışında korunmuş olduğunu göstermektedir.

Anahtar Sözcükler: Glikojen sentaz, genomiks, evrim, promotor, ekspresyon

Received: March 27, 2008
Accepted: January 12, 2009

Correspondence
Fuat DİLMEÇ
Department of Medical Biology,
Faculty of Medicine,
Harran University,
Şanlıurfa - TURKEY

fdilmec@harran.edu.tr

Introduction

The human *GYS2* gene, also known as the glycogen synthase 2 (liver) gene, is a single copy gene on chromosome 12p12.2 that consists of 16 exons at least 68,431 bp in length. It encodes a glycogen synthase (GS) of approximately 80,957 Da (1). GS is the enzyme responsible for the synthesis of 1,4-linked glucose chains in glycogen. It is the rate-limiting enzyme in the synthesis of polysaccharides, and its activity is highly regulated through phosphorylation at multiple sites and by allosteric effectors, mainly glucose 6-phosphate (G6P) (http://www.ncbi.nlm.nih.gov/Structure/cdd).

Glycogen storage disease type 0 (GSD-0) is caused by mutations in the *GYS2* gene and is inherited in an autosomal recessive manner (1,2). The number of amino acids in human GS is identical to and the deduced amino acid sequence homology is 92% that of the rat liver enzyme. Human and rat liver GS are truncated by 34 amino acids, as compared to the human muscle enzyme, and by 32 amino acids, as compared to the rabbit muscle enzyme. The amino acid similarity between human liver and human muscle GS is only 69% (3). Human liver GS was studied and its properties were compared with those of rat liver GS. Rat and human liver GS are similar in their pH profile, kinetic constants for the substrate UDP-glucose and the activator glucose 6-phosphate, and in elution profiles from Q-Sepharose. Lastly, amino acid analysis indicates there are differences between the enzymes in the 2 species (4).

Mammals express 2 isoforms of GS, which are encoded by the *GYS1* and *GYS2* genes. *GYS1* encodes the muscle isoform of GS and is expressed in skeletal muscle, cardiac muscle, adipose tissue, kidneys, and the brain (5). *GYS2* expression has been observed only in the liver (6).

The present study aimed to analyze the *GYS2* gene in different species in silico. Specifically, its Glycogen_syn domain on *GYS2* genes, the transcription factor binding sites on its promoters, the tissue expression profile, homology level, and phylogenetic tree among vertebrates were examined using bioinformatic tools.

Materials and Methods

Homology search

The search for homologous protein sequences to human GS was carried out using a basic local alignment search tool, BLASTp software (7,8), at NCBI (http://www.ncbi.nlm.nih.gov) using the human GS amino acid sequence (GI: 11496237) as a query against the SwissProt protein databases. Full protein, and Glycogen_syn sequences of human and other species were downloaded and then aligned using ClustalW software (9,10) at the EBI site (http://www.ebi.ac.uk).

Promoter Analysis

We used Genomatix software (http://www.genomatix.de) for analysis of *GYS2* gene promoters in various species (11). These nucleotide sequences were downloaded and aligned using ClustalW software. Then common transcription factor binding sites were searched for with the DiAlign TF module in Genomatix software for all *GYS2* promoters present in the database.

Evolutionary Analysis

We used amino acid sequences of Glycogen_syn domains to construct phylogenetic trees using the neighbor-joining (NJ) method, with Jones-Taylor-Thomton (JTT) distances. NJ searches were conducted using MEGA4 molecular evolutionary genetics analysis software (12) and 500 bootstrap replicates were assessed for the reliability of internal branches; sites with gaps were ignored in this analysis.

In silico Expression Analysis

The DigiNorthern database (13) was used to analyze the expression of *GYS2* mRNA based on expressed sequence tag (EST) data. DigiNorthern collects all ESTs for a query gene and categorizes them based on the types of tissues and their histological status. Pairwise comparison of relative values was performed with the Fisher's exact test using SPSS v.11.0 for Windows.

Results

Homology Search

BLASTp results showed that the GS molecule was conserved in various species (Table 1). The homology search indicated that the GS sequences in *Pan troglodytes (P. troglodytes)* (99%) and *Macaca mulatta* (*M. mulatta)* (97%) had the highest homology to that of human GS. In contrast, *O. anatinus* had the lowest homology to the human GS protein (69%) (Table 1, Figure 1).

ClustalW alignment revealed a well-conserved domain; the Glycogen_syn domain was well conserved and had many conserved sub-domains, such as "DYEEFVRGCHLGVFPSYYEPWGYTP" (near position C, terminal side) (Figure 1). These

Table 1. BLASTp results of vertebrate GS molecules and their homology.

Species	Common name	Accession no.	Protein name (in liver)	Number of amino acids	% identity with human GS
Homo sapiens	Human	NP068776	GS	703	100
Pan troglodytes	Chimpanzee	XP520790	GS	703	99
Macaca mulatta	Rhesus monkey	XP001098578	Similar to GS	703	97
Canis familiaris	Dog	XP534869	Similar to GS	703	95
Equus caballus	Horse	XP001502236	Similar to GS	703	94
Mus musculus	Mouse	NP663547	GS	704	94
Bos taurus	Cow	XP617616	Similar to GS	702	94
Rattus norvegicus	Rat	NP037221	GS	703	92
Monodelphis domestica	gray short-tailed opossum	XP001363412	Similar to GS	700	90
Danio rerio	zebrafish	NP001018199	GS	701	78
Ornithorhynchus anatinus	platypus	XP001505471	Similar to GS	733	69

conserved sub-domains were most probably functionally important and any mutations in them were deleterious, as evidenced by their evolutionary conservation.

Multiple alignment results of human GS and its homologues showed that this molecule was well conserved in the investigated species. The GS molecule in *Ornithorhynchus anatinus* (*O. anatinus*) (733 amino acids) was longer than in the other species (701-704 amino acids). The remainder of the molecules in the other species displayed a very similar pattern.

We determined that 1 conserved domain family was glycogen synthase (Glycogen_syn). This family consists of the eukaryotic GS proteins *GYS1*, *GYS2*, and *GYS3*. GS is the enzyme responsible for the synthesis of 1,4-linked glucose chains in glycogen. It is the rate-limiting enzyme in the synthesis of polysaccharides, and its activity is highly regulated through phosphorylation at multiple sites and by allosteric effectors, mainly glucose 6-phosphate (G6P). Furthermore, we detected some partial sub-domains in the Glycogen_synthase conserved domain family that were truncated near C-termini. They were ignored in the analysis due to inconsistency and truncation.

Promoter Analysis

Analyzing the promoters present in the Genomatix database we could not detect any common transcription factor binding sites (TFBs) in the *GYS2* promoters in *H. sapiens, R. norvegicus, C. lupus familiaris, P. troglodytes,* or *G. gallus*; however, we observed that the similarity (value 1.000) and the percentage of identical nucleic acids (in short sequence segments) was 98% between the *GYS2* promoters in *H. sapiens* and *P. troglodytes* for each pairwise alignment.

Evolutionary Analysis

From the phylogenetic trees constructed with MEGA4, we observed that the Glycogen_syn domains of GS proteins (gene products) were conserved in all the organisms investigated. We showed that the Glycogen_syn domains in *H. sapiens* and *P. troglodytes* were more closely grouped (scale length 0.05) (Figure 2). In contrast, the Glycogen_syn domains in *O. anatinus* were the least similar to those in humans. When we constructed the phylogenetic tree we ignored the sub-domain sequences in Glycogen_syn conserved domains in the all species due to their high diversity caused by possible truncation.

In Silico Expression Analysis

The distribution of human *GYS2* ESTs in the cDNA library database (all ESTs) was analyzed using DigiNorthern software. In columns 2 and 3 of Table 2, relative values of *GYS2* ESTs in the cDNA library from normal and tumor tissues, respectively, were both absolute numbers as well as normalized values per 10^6 cDNA. Its normal and cancerous tissue expression profiles were compared and the significance in its expression pattern was assessed with Fisher's exact test (P value 0.05) (Table 2). It

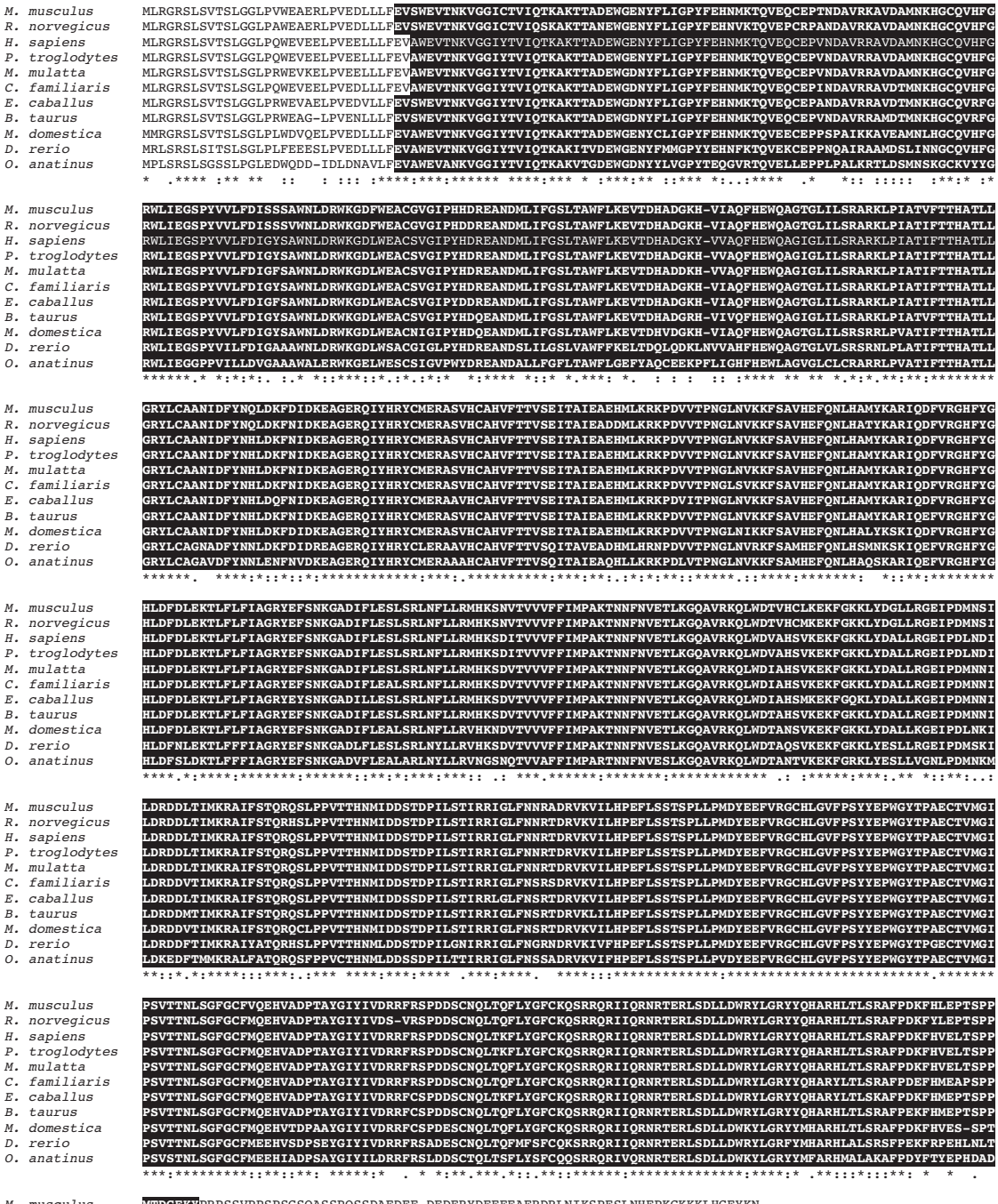

Figure 1. Multiple alignment of vertebrate GS proteins. The Glycogen_syn domains are highlighted with a black background. The conserved amino acid residues are shown by an asterisk (*) and amino acids with similar properties are shown by a semi-colon under the alignment.

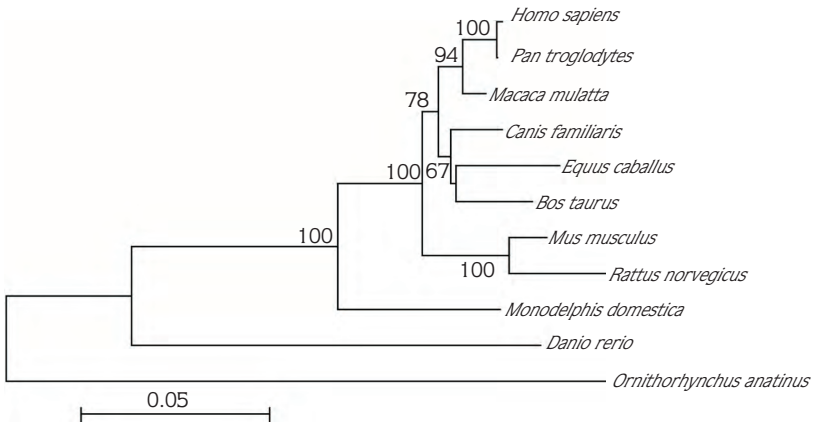

Figure 2. Phylogenetic tree of Glycogen_syn conserved domains in the vertebrate species studied. Phylogenetic trees were constructed using MEGA4 software. Species names are indicated in the figure. Branch lengths indicate evolutionary relationships.

Table 2. Comparison of the relative values of human *GYS2* cDNA from specific normal and tumor tissues in the cDNA library database.

Tissue/organ type	The number of ESTs		
	Normal	Cancer	P
Kidney	3/74,917 (40)	0/96,375 (0)	0.084
Liver	12/73,021 (164)	16/81,780 (196)	0.708
Pooled tissue	10/373,366 (27)	0/55,060 (0)	0.627
Uncharacterized tissue	0/88,784 (0)	0/105,216 (0)	1.000
Whole body	1/73,648 (14)	0/0 (0)	1.000
Total no. of ESTs found	26/683,736 (38)	16/338,431 (47)	0.514

Absolute numbers and the relative values as normalized per 10^6 cDNA (in parentheses) in normal and cancerous tissues are shown. P values are for comparison of relative values of *GYS2* ESTs in normal versus tumorous tissues, based on Fisher's exact test (last column).

was determined that the human *GYS2* gene was expressed at low or high levels in some tissues, but not expressed in others. The tissue distribution and differential expression patterns in normal and cancerous human tissues were different. Expression of the *GYS2* gene in some of the normal and cancerous human tissues, such as uncharacterized tissue, was not detected. In contrast, its expression seemed to have been lost in some of the cancerous tissues, such as kidney, pooled, and whole body. Compared to normal tissues, its expression was not significantly different ($P > 0.05$) in cancerous tissues (Table 2).

Discussion

GS is the enzyme responsible for the synthesis of 1,4-linked glucose chains in glycogen (14). Our BLASTp results indicate that the GS molecule occurred in various species of vertebrates and that these molecules had a 69%-99% conservation degree in total amino acid sequences (Table 1). The human GS molecule had the highest homology to that of *P. troglodytes* (chimpanzee) (99%) and *R. norvegicus* (rat) (92%), and the lowest homology to that of *O. anatinus* (69%). Therefore, these results indicate that the *GYS2* gene has been evolutionarily well conserved, and played the same role in different organisms. Recently, it

has been shown that human GS is homologous to that of the rat (92%) (3), which is in agreement with the results of the present study. A recent study indicates that the amino acid similarity between human liver and muscle GS is only 69% (3). In addition, another study reported that GS is an important enzyme in liver glycogen synthesis and that characterization of this enzyme in humans will help provide insight regarding human liver glycogen synthesis. In this way the vertebrate *GYS2* gene appears to be relatively conserved throughout evolution (4).

We also examined the phylogenetic tree of Glycogen_syn conserved domains of GS in different species using MEGA4 software. We observed that human Glycogen_syn conserved domains were closest in homology to those of *P. troglodytes*, but were not homologous to those of *O. anatinus* (Figure 2). These domains are very important for their functions. When we compared the Glycogen_syn domains in different species, a great deal of conserved motifs on this conserved domain were observed. The longest of them was "DYEEFVRGCHLGVFPSYYEPWGYTP", and it was determined to be well conserved in all the investigated species (Figure 1). The importance of these and other motif sequences needs to be experimentally defined.

The expression of the human *GYS2* gene in different tissues was analyzed using DigiNorthern software (Table 2). Its expression patterns in normal and cancerous human tissues were different. In some tissues, such as cancerous kidney, pooled, and uncharacterized tissues, it was not expressed; its expression in normal and cancerous liver tissue was not significantly different (P = 0.708) (Table 2). Thus, this gene was not considered a candidate for cancer development. Immunoblot analysis indicates that L-type GS is found in the liver, but not in skeletal muscle, the heart, fat, kidneys, or the brain (6). Furthermore, we showed that GS was expressed in kidney and pooled tissues, as well as in the liver. It has been reported that the availability of comprehensive data generated by high-throughput functional genomics techniques, primarily EST and serial analysis of gene expression (SAGE), facilitates gene expression analysis (15).

We used the DiAlign TF module in Genomatix software to predict transcription factor binding sites (transcriptional elements) of all orthologous *GYS2* promoters present in the database. DiAlign TF results show that *GYS2* orthologous promoters had no common conserved transcriptional elements. The conservation of transcriptional elements in promoter sequences may provide further evidence in support of functional conservation (11,16,17); however, the element in the promoters or their vicinity might be more mobile than the genes themselves. Our results indicate that the binding sites of different transcription factors might have been located on different parts of the promoter or promoter vicinity in various species.

These findings provide the foundation for future investigations of the molecular mechanisms underlying the heterogeneity of GS activity in humans. Basic bioinformatic techniques are powerful tools, in terms of leading to the discovery and analysis of novel genes (18). Even though the results from bioinformatic studies were very helpful in directing and designing the present experiments, they need to be supported and confirmed by additional experimental data.

Acknowledgement

We thank Mustafa Zerin (associate professor) for critically reading the manuscript.

References

1. Orho M, Bosshard NU, Buist NR, Gitzelmann R, Aynsley-Green A, Blumel P et al. Mutations in the liver glycogen synthase gene in children with hypoglycemia due to glycogen storage disease type 0. J Clin Invest 1998; 102: 507-15.
2. Weinstein DA, Correia CE, Saunders AC, Wolfsdorf JI. Hepatic glycogen synthase deficiency: an infrequently recognized cause of ketotic hypoglycemia. Mol Genet Metab 2006; 87: 284-288.
3. Nuttall FQ, Gannon MC, Bai G, Lee EY. Primary structure of human liver glycogen synthase deduced by cDNA cloning. Arch Biochem Biophys 1994; 311: 443-449.
4. Westphal SA, Nuttall FQ. Comparative characterization of human and rat liver glycogen synthase.Arch Biochem Biophys 1992; 292: 479-486.

5. Kaslow HR, Lesikar DD. Isozymes of glycogen synthase. FEBS Lett 1984; 172: 294-298.

6. Kaslow HR, Lesikar DD, Antwi D, Tan AW. L-type glycogen synthase. Tissue distribution and electrophoretic mobility. J Biol Chem 1985; 260: 9953-956.

7. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997; 25: 3389-402.

8. Altschul SF, Wootton JC, Gertz EM, Agarwala R, Morgulis A, Schäffer AA et al. Protein database searches using compositionally adjusted substitution matrices. FEBS J 2005; 272: 5101-9.

9. Thompson JD, Higgins DG, Gibson TJ. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994; 22: 4673-80.

10. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H et al. Clustal W and Clustal X version 2.0. Bioinformatics 2007; 23: 2947-948.

11. Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A et al. MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics 2005; 21: 2933-42.

12. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 2007; 24: 1596-9.

13. Wang J, Liang P. DigiNorthern, digital expression analysis of query genes based on ESTs. Bioinformatics 2003; 19: 653-4.

14. Garcia-Rocha M, Roca A, De La Iglesia N, Baba O, Fernandez-Novell JM, Ferrer JC et al. Intracellular distribution of glycogen synthase and glycogen in primary cultured rat hepatocytes. Biochem J 2001; 357: 17-24.

15. Lash AE, Tolstoshev CM, Wagner L, Schuler GD, Strausberg RL, Riggins GJ et al. SAGEmap: a public gene expression resource. Genome Res 2000; 10: 1051-60.

16. Vanpoucke G, Goossens S, De Craene B, Gilbert B, van Roy F, Berx G. GATA-4 and MEF2C transcription factors control the tissue specific expression of the alphaT-catenin gene CTNNA3. Nucleic Acids Res 2004; 32: 4155-65.

17. Doan LL, Porter SD, Duan Z, Flubacher MM, Montoya D, Tsichlis PN et al. Targeted transcriptional repression of Gfi1 by GFI1 and GFI1B in lymphoid cells. Nucleic Acids Res 2004; 32: 2508-19.

18. Hu Z, Chen K, Wang L, Yao Q. Identification and characterization of Bombyx mori eIF5A gene through bioinformatics approaches. In Silico Biol 2005; 5: 573-80.