Journal of Clinical and Basic Research
Online ISSN: 2538-3736

Research Article

In Silico Identification of Housekeeping Genes by Expressed Sequence Tags and Assessment of Short Tandem Repeats in Their Promoters

Mahboobeh Ramezani [1,2], Bita Javan[3], Mohammad Hasan Jafari[4], Arsham Banisadr[1,4,5], Mahdieh Pourafshar[1,5], Seyedeh Sara Raeissadati [1,5], Pardis Mokhtary [1,5], Mohammad Arefi [1,5], Azizeh Karimian[8], Morteza Oladnabi [6,7,2]*

1. Student Research Committee, Golestan University of Medical Sciences, Gorgan, Iran
2.Department of Medical Genetics, School of Advanced Technologies in Medicine, Golestan University of Medical Sciences, Gorgan, Iran
3.Medical Cellular and Molecular Research Center, Golestan University of Medical Sciences, Gorgan, Iran
4.Department of Medical Biotechnology, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran.
5.Department of Medical Biotechnology, School of Advanced Technologies in Medicine, Golestan University of Medical Sciences, Gorgan, Iran.
6.Ischemic Disorders Research Center, Golestan University of Medical Sciences, Gorgan, Iran
7.Gorgan Congenital Malformations Research Center, Golestan University of Medical Sciences, Gorgan, Iran
8.Deputy Of Research and Technology, Golestan University of Medical Sciences, Gorgan, Iran
*** Correspondence:** Dr Morteza Oladnabi , Associated Professor of Medical Genetics, Gorgan Congenital Malformations Research Center, Golestan University of Medical Sciences, Gorgan, Iran.
Email: oladnabidozin@yahoo.com

Received September 29, 2021 **Accepted** November 6, 2021

ABSTRACT

Background and objectives: In gene expression studies, to validate and obtain reliable results, normalization of qRT-PCR data by housekeeping genes (HKGs) is required. However, the expression level of these genes may vary in tissues or cells and may change under certain circumstances. Thus, selection of HKGs is critical for gene expression studies. For this purpose, we analyzed expression of protein-coding genes by expressed sequence tags (ESTs) and short tandem repeat (STR) to select suitable HKGs.

Methods: The EST profile of 17242 protein-coding genes was extracted from the UniGene database. Log2 (TPM + 2) scale was used to normalize EST counts across 16 normal and tumor tissues. Selection of HKGs was limited to genes expressed in developmental stages as well as in normal and tumor tissues. Then, the genes with no expression change between the normal and tumor tissues were selected. Finally, the STRs and the gene ontology analysis of candidate genes were performed.

Results: We found that 93 genes had no expression change between the normal and tumor tissues. The STRs analysis showed that GCGCGC repeats had the highest frequency in candidate gene promoters.

Conclusion: We introduced a new set of genes as potential HKGs, some of which can be used for a particular tissue. However, further investigations are required to confirm our findings.

Keywords: Housekeeping gene;EST; In Silico

DOI: 10.29252/Jcbr.5.4.31

This work is licensed under a Creative Commons Attribution 4.0 License.
© The authors

[DOI: 10.61186/jcbr.5.4.31]

Introduction

A quantitative study of gene expression based on the amount of RNA produced by a specific gene in biological conditions is necessary to understand the structure and function of the gene. Quantitative real-time PCR (qRT-PCR) is considered as an accurate, powerful, and reliable technique for gene expression studies. However, several factors can affect its outcomes. Therefore, qRT-PCR data need to be normalized by internal controls to remove the non-specific variability related to differences in the quantity and quality of the RNA (1).

In gene expression studies, accuracy of normalized data is highly dependent on the reliability of reference genes. Failure to select an appropriate reference gene may introduce bias in the gene expression data (2). An ideal internal control gene should be ubiquitous and constitutively expression in different cell types and tissues, regardless of the type of tissue, disease condition, developmental stages, or experimental conditions (3). These conditions concerned with reference genes are more consistent with the housekeeping genes (HKGs) as critical genes to maintain basal cellular functions and existence of cells (4). However, there is no HKG capable of stable expression in all tissue types under all experimental conditions (5). For instance, studies have indicated that commonly known HKGs, such as β-actin (ACTB) and glyceraldehyde-3- phosphate dehydrogenase (GAPDH) have variable expression levels in various developmental stages and different physiological and pathological conditions (6–8). Therefore, it is suggested to consider validation of the expression stability of HKGs prior to comparison and normalization with the target gene.

One of the efficient ways to obtain information about gene expression is by expressed sequence tags (ESTs), a high-throughput method for gene expression analysis representing the expression profile, including complexity and abundance levels of transcripts from different tissues, cell types, and developmental stages (9,10). Furthermore, short tandem repeats (STRs) are among the elements affecting gene expression. These are repetitive short sequences consisting of 1-6 bp motifs covering approximately 3% of the human genome (11,12). Studies have demonstrated that most of the STRs are located near and flanking the transcription start site (TSS) or within the gene, playing a decisive role in gene expression (13,14). Other evidence suggests that genes containing a high density of repeat sequences have a higher rate of transcriptional divergence and gene expression (15). In the present study, we introduced a new category of HKGs based on the criteria mentioned for these genes by EST data and compared the nucleotide composition, abundance, and type of STRs in these genes with those of human protein-encoding transcripts.

MATERIALS AND METHODS

In the present study, all human class II (protein-coding) genes were selected from the GeneCards database (http://www.genecards.org/List). Then, EST profiles of the genes were extracted from the UniGene database (https://www.ncbi.nlm.nih.gov/unigene). In this database, only 17,242 genes amongst the total of 130,062 genes have an EST profile. Overall, 81550 sequences corresponding to the the first 120 nucleotides of the coding genes were obtained from the ENSEMBLE database. These sequences were also evaluated for STR, type, and copy number as well as nucleotide percentage in In Silico (http://insilico.ehu.es/mini_tools/microsatellites) and ALGGEN(http://alggen.lsi.upc.es/cgibin/promo_v3/promo/promoinit.cgi?driDB=TF_8.3) databases, respectively.

In the next step, Log2 (transcript count per million + 2) scale was used to normalize EST counts. Then, the ratio of cancer-to-normal tissue expression was calculated for each tissue type. Genes that exhibited their expression variations in the range of 1± 0.01 were selected as housekeeping genes for the examined tissues. In addition, EST data on the UniGene database were divided into

[DOI: 10.61186/jcbr.5.4.31]

three categories: 45 normal tissues, 25 tumor tissues, and seven developmental stages. Finally, the genes with expression in all of these items and no variability of gene expression level between normal and tumor tissues were selected as HKGs. Next, the gene ontology (GO) analysis was performed using the Web Gestalt database (http://webgestalt.org) to determine the molecular and biological function as well as cellular component. Data were analyzed by one-way ANOVA using the SPSS 16.0 software. P-values ≥0.05 were considered statistically significant.

RESULTS

Based on the EST data on the UniGene database, only 100 genes from total of 17,242 genes were expressed in all normal and tumor tissues types. Among them, only 16 normal and tumor tissues were comparable to each other. Eventually, 93 genes with no expression variation between normal and tumor tissues as well as in different developmental conditions were selected as candidate HKGs (Figure 1). Common HKGs such as ACTB and GAPDH showed variable expression. Among 45 normal tissues in the UniGene database, GAPDH is expressed in 44 ones. However, it was not among the genes that are expressed in both normal and tumor tissues. Therefore, it was discarded at this stage of our study. Although ACTB expression was observed in all studied tissues, it expression changes were within our desirable range (1±0.01) only in three tissues. In other words, ACTB expression had significant variability in other tissues. Three genes including CALM1, MORF4L1 and HNRNPK covered more tissues. These genes are GC rich in most of their transcripts.

Figure 1. Some candidate HKGs in different tissues. Among the normal and tumor tissue data, only 16 normal and tumor tissues were comparable to each other. Some of HKGs are common in different tissues; however, some tissues have their own genes.

STR analysis

First, transcripts of 93 genes were identified, and then the 120 bp up-stream sequences to the +1TSSs of these 661 transcripts were analyzed based on STRs.

[DOI: 10.61186/jcbr.5.4.31]

Among these genes, 40.7% had no STR and 59.3% had STRs, most of which had just one (Table 1). The STRs were also studied based on the type and copy number (Table 2). The binary repetitions were the most frequent, and STRs with three copy numbers had the highest frequency. Study of the STRs nucleotide composition revealed that those with three GC repetitions had the highest percentage

Table 1. Frequency of STRs in 120 bp up-stream sequences of TSSs

Number of STRs	Frequency	Percent
0	269	40.7
1	228	34.5
2	107	16.2
3	40	6.1
4	11	1.7
5	5	0.8
6	1	0.2

Table 2. Frequency of various STRs types and the number of STR repeats

Type of STRs	Frequency	Percent	Repeats in STRs	Frequency	Percent
MONO	132	20.7	3 repeats	412	64.7
DI	393	61.7	4-5 repeats	84	13.2
TRI	69	10.8	6-7 repeats	104	16.3
TETRA	16	2.5	8-10 repeats	26	4.1
PENTA	26	4.1	>10 repeats	11	1.7
HEXA	1	0.2			

GC content

In this stage, 120 bp up-stream of each transcript were analyzed based on the percentage of purine, pyrimidine, and GC content as well as purine/pyrimidine ratio. The results indicated that almost half of the transcripts had a purine/pyrimidine ratio of higher than one (data not shown). The GC-ratio was evaluated using the following formula: G+C/A+T+C+G×100. (Figure 2) demonstrates the frequency of genes based on the GC percentage. A total of 191 transcripts belonging to 49 genes had a high GC content. Based on the results, *GNAS* with the highest GC content (> 90%) is suggested as the HKG in the pancreatic and skintissues.

[DOI: 10.61186/jcbr.5.4.31]

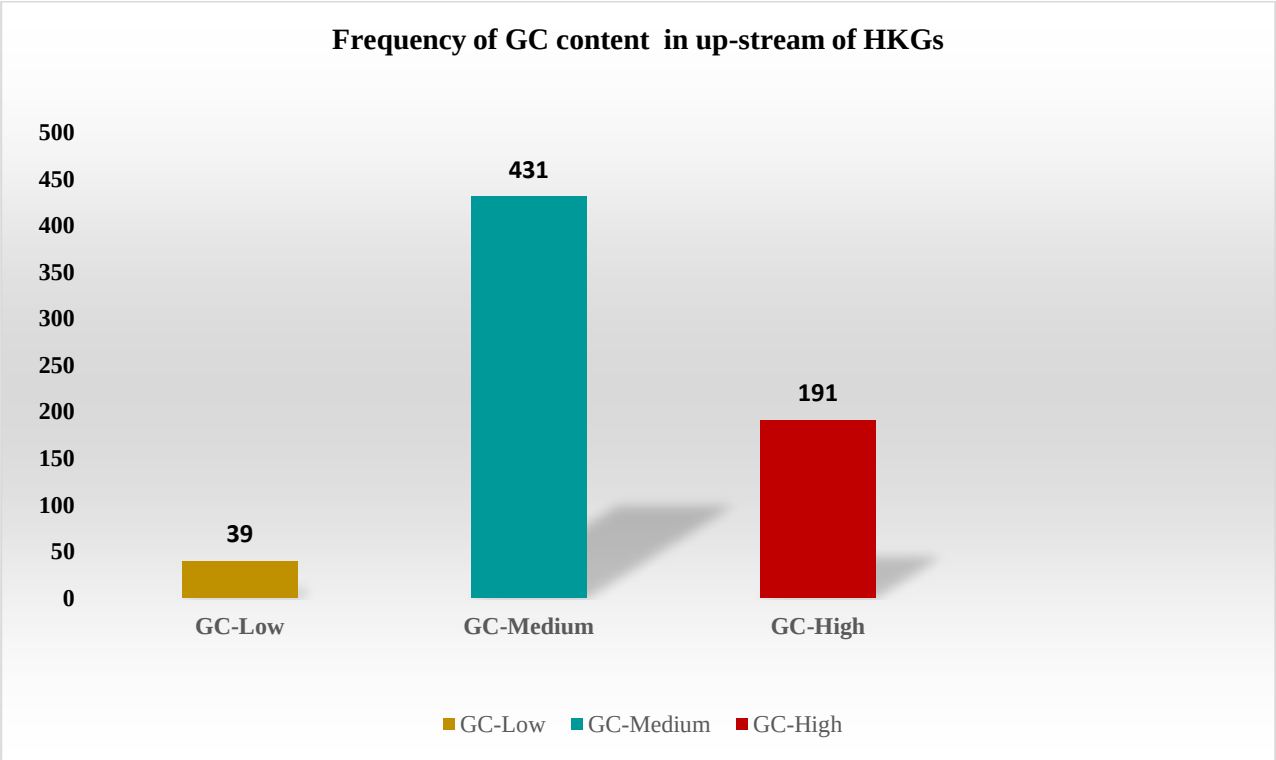

Figure 2. The frequency of genes based on GC content. The genes were classified into low GC (<40%), medium GC (40-70%), and high GC (>70%) content accordingly.

Physical interaction and pathway analysis

Evaluation of the physical interaction of candidate HKGs was performed in the STRING and GENEMANIA databases. (Figure 3) illustrates the interaction between genes according to the STRING database.

Investigation of the relationship between these genes in terms of co-expression, physical interaction, etc. in the GENEMANIA database revealed that more than half of these genes were co-expressed, which could confirm our findings

.

[DOI: 10.61186/jcbr.5.4.31]

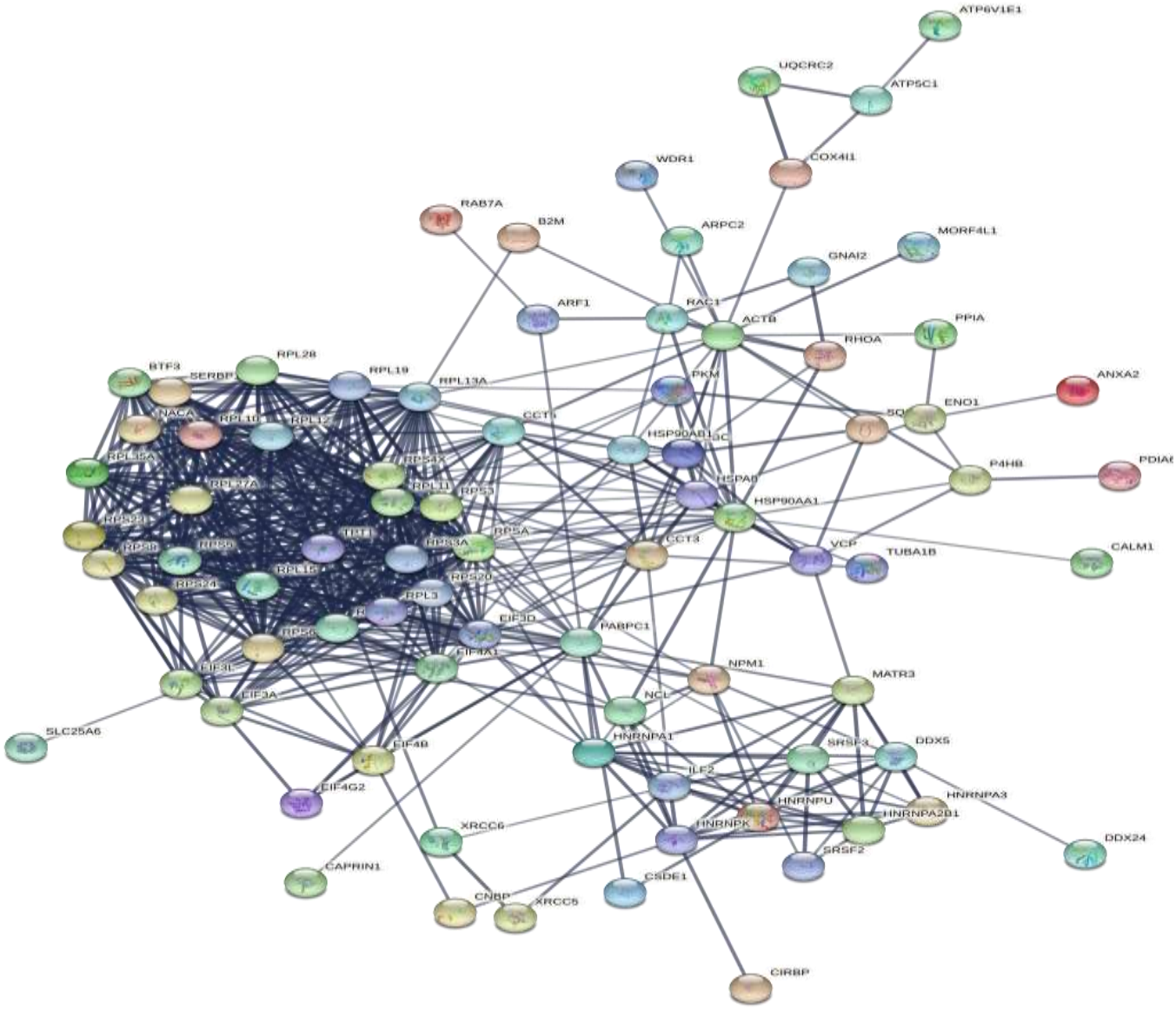

Figure 3. The output of interactions between candidate HKGs according to the STRING database. The line thickness indicates the strength of data support at high confidence (0.7). PPI enrichment p-value < 1.0e-16.

Biological process

Evaluation of molecular and biological function as well as cellular component of candidate HKGs were performed in the Web Gestalt database (Figure 4). These genes encode proteins that are often involved in the structure of ribosomal subunits, RNA processing, transcription, translation, protein localization, ubiquitination, cellular structure, and skeletal proteins. Most of these proteins are located in the cytoplasm, nucleus, and membranes that are involved in transporting materials. Among all candidate HKGs in this study, CALM1, HNRNPK, and MORF4L1 covered more tissues. These genes are involved in formation of calmodulin-calcium complex, regulation of transcription, and chromatin and protein N-terminus binding, respectively

.

[DOI: 10.61186/jcbr.5.4.31]

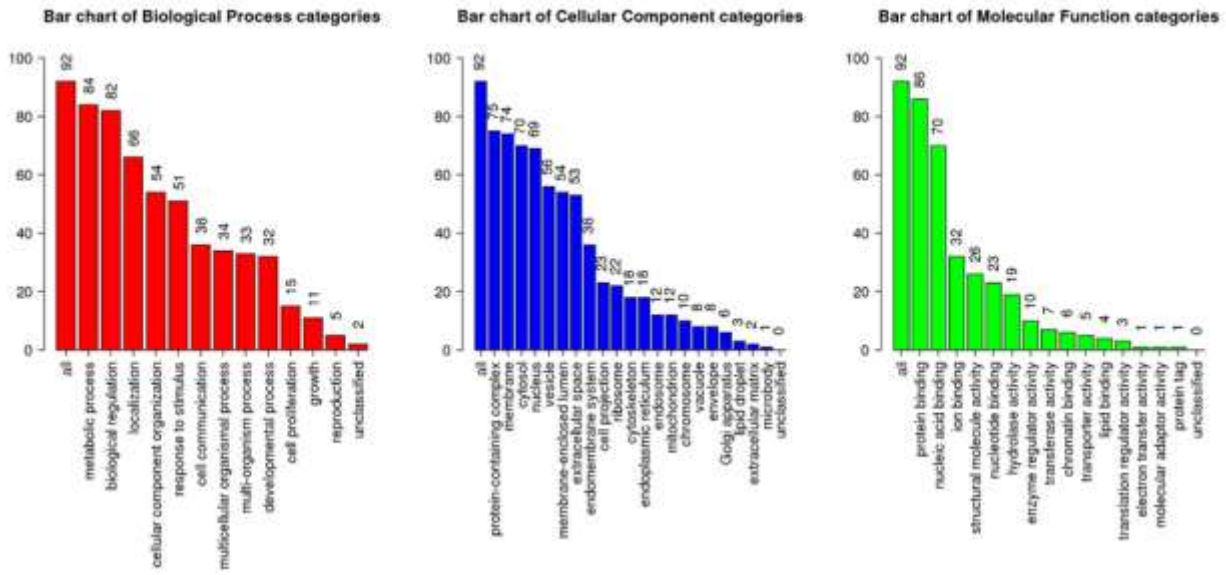

Figure 4. The gene ontology analysis of candidate genes. The Web Gestalt database was used to perform the GO enrichment analysis and to categorize genes in biological process (Red bar charts), cellular component (Blue bar charts), and molecular function ontology (Green bar charts).

Table 3. List of the candidate HKGs found in this study

Gene	Ensemble Transcript ID	GO Term (biological process)	pu/py	STR
ADAR	ENSG00000160710	RNA processing	0.85	3(CCT), 6(A),10(T)
ANAPC5	ENST00000261819	Protein phosphatase binding	0.67	
ANXA2	ENST00000396024	Membrane raft assembly	1.45	4(GGCGG)
ARF1	ENST00000541182	Post-Golgi vesicle-mediated transport	0.43	
ARPC2	ENST00000295685	Positive regulation of actin filament polymerization	1.93	3(CG)
ATP5C1	ENST00000356708	ATP biosynthetic process	2.08	3(AG)
ATP6V1E1	ENST00000253413	ATP hydrolysis coupled proton transport	1.26	6(A) 3(CT)
B2M	ENST00000558401	ATP hydrolysis coupled proton transport	1.11	4(CT)
BTF3	ENST00000380591	Transcription by RNA polymerase II	0.94	3(CG)
CALM1	ENST00000356978	Detection of calcium ion	1.45	3(GC)
CAPRIN1	ENST00000341394	Negative regulation of translation	0.94	
CCT3	ENST00000295688	Protein folding	0.90	3(AC)
CCT5	ENST00000280326	Protein folding	0.79	3(CCT), 22(A)
CIRBP	ENST00000589710	Negative regulation of translation	1.22	3(GC)

[DOI: 10.61186/jcbr.5.4.31]

CNBP	ENST00000422453	Negative regulation of transcription by RNA polymerase II	1.07	3(GC), 3(CG)
COX4I1	ENST00000562336	Electron transport chain	0.67	
CSDE1	ENST00000610726	Regulation of transcription, DNA-templated	1.07	
DDX24	ENST00000621632	RNA secondary structure unwinding/ ATP-dependent RNA helicase activity	1.07	
DDX5	ENST00000225792	Alternative mRNA splicing, via spliceosome	1.40	
EIF3A	ENST00000369144	Translation initiation factor activity	0.82	3(GC), 3(GC)
EIF3D	ENST00000216190	Translation initiation factor activity	1.55	3(GC)
EIF3L	ENST00000624234	Translation initiation factor activity	1.07	
EIF4A1	ENST00000293831	Nucleic acid binding and hydrolase activity	1.35	
EIF4B	ENST00000262056	Nucleic acid binding and RNA binding	0.79	6(A), 3(AGC)
EIF4G2	ENST00000526148	Regulation of translation	1.07	
ENO1	ENST00000234590	Transcription corepressor activity	1.79	
GHITM	ENST00000372134	Apoptotic process	0.67	
GLO1	ENST00000373365	Carbohydrate metabolic process	0.94	3(AG)
GNAI2	ENST00000313601	GTP binding	0.67	
GNAS	ENST00000371100	GTP binding and obsolete signal transducer activity	1.07	
HDLBP	ENST00000391975	Lipid transport	1.22	6(A), 3(GGGC)
HMGN1	ENST00000380749	Chromatin binding and nucleosomal DNA binding.	1.73	16(G)
HNRNPA1	ENST00000546500	mRNA processing	0.62	3(CCG), 6(T)
HNRNPA2B1	ENST00000354667	mRNA processing	1.50	6(G)
HNRNPA3	ENST00000411529	mRNA processing	0.54	3(GC)
HNRNPK	ENST00000376263	Regulation of transcription	0.87	
HNRNPU	ENST00000640218	Chromatin organization	1.03	3(GC)
HSP90AA1	ENST00000334701	G2/M transition of mitotic cell cycle	0.97	
HSP90AB1	ENST00000371554	Protein folding	1.00	
HSPA8	ENST00000534624	Ubiquitin protein ligase binding	0.97	4(CG)
ILF2	ENST00000615950	Double-stranded RNA binding	1.18	6(T)
MAPRE1	ENST00000375571	Cell cycle	0.97	3(GC), 3(CT)
MATR3	ENST00000394805	Posttranscriptional regulation of gene expression	2.00	4(GC), 4(GC)
MORF4L1	ENST00000331268	Chromatin binding and protein N-terminus binding	1.31	3(GGGGT), 16(G)
NACA	ENST000005509	Protein transport/transcription coactivator	0.74	

[DOI: 10.61186/jcbr.5.4.31]

	52	activity and TBP-class protein binding.		
NCL	ENST00000322723	Nucleic acid binding and identical protein binding	0.58	
NPM1	ENST00000296930	Ribosomal large subunit export from nucleus	2.16	3(GC)
P4HB	ENST00000331483	Protein folding	1.50	
PABPC1	ENST00000318607	Nuclear-transcribed mRNA poly(A) tail shortening	1.07	3(CG)
PDIA6	ENST00000404371	Isomerase activity and protein disulfide	0.97	
PKM	ENST00000319622	Glucose metabolic process/ATP biosynthetic process	0.82	
PPIA	ENST00000468812	RNA-dependent DNA biosynthetic process	1.61	
PTTG1IP	ENST00000330938	Positive regulation of protein ubiquitination	0.45	
RAB7A	ENST00000265062	Protein targeting to lysosome	1.18	
RAC1	ENST00000348035	GTP binding and enzyme binding	2.08	3(GCG)
RHOA	ENST000004181115	Cell morphogenesis	1.26	
RPL10	ENST00000424325	rRNA processing	0.76	3(CT)
RPL11	ENST00000374550	Ribosomal large subunit assembly	0.82	3(CT)
RPL12	ENST00000361436	Translational initiation	0.90	3(CG)
RPL13A	ENST00000391857	Regulation of translation	1.18	
RPL15	ENST00000456530	Cytoplasmic translation	1.31	3(AG), 3(TC)
RPL19	ENST00000579260	Structural constituent of ribosome	1.26	
RPL27A	ENST00000314138	Translational initiation	1.45	
RPL28	ENST00000344063	Signal recognition particle-dependent co-translational protein targeting to membrane	0.90	
RPL3	ENST00000216146	Structural constituent of ribosome.	0.60	3(GT)
RPL35A	ENST00000464167	Translational initiation	0.76	
RPL4	ENST00000307961	Structural constituent of ribosome	1.07	3(AG), 6(T)
RPS20	ENST00000519807	Structural constituent of ribosome	0.76	3(TA)
RPS23	ENST00000296674	Structural constituent of ribosome	0.79	
RPS24	ENST00000440692	Structural constituent of ribosome	1.07	
RPS3	ENST000005311888	Nuclear-transcribed mRNA catabolic process	0.52	
RPS3A	ENST00000274065	Structural constituent of ribosome	1.22	3(CG), 3(CA), 3(CT)
RPS4X	ENST00000316084	Structural constituent of ribosome	1.22	3(GC)
RPS5	ENST00000596046	Structural constituent of ribosome	0.82	3(GC)
RPS6	ENST000003803	Structural constituent of ribosome	0.60	

[DOI: 10.61186/jcbr.5.4.31]

	94			
RPS8	ENST00000396651	Structural constituent of ribosome	1.03	
RPSA	ENST00000301821	Ribosome binding	0.67	
SARS	ENST00000369923	RNA binding and aminoacyl-tRNA ligase activity.	1.03	3(GC)
SERBP1	ENST00000370994	Regulation of mRNA stability	1.03	3(TG)
SERP1	ENST00000479209	Cellular protein modification process	1.03	
SLC25A6	ENST00000381401	Transporter activity	1.31	3(AG) 8(C)
SQSTM1	ENST00000389805	Homodimerization activity	1.93	3(AG)
SRSF2	ENST00000392485	Termination of RNA polymerase II transcription	1.40	
SRSF3	ENST00000373715	mRNA export from nucleus	0.97	
TPT1	ENST00000616577	Transcription factor binding	0.52	3(CG)
TUBA1B	ENST00000336023	Cytoskeleton organization	1.00	3(GC), 3(GCCC)
UBC	ENST00000536769	MAPK cascade	0.64	3(TG) 6(T)
UQCRC2	ENST00000268379	Mitochondrial electron transport	0.76	3(AC)
VCP	ENST00000358901	Signaling receptor binding	0.67	3(TG)
WDR1	ENST00000499869	Cytoskeleton organization	1.18	3(TA), 4(CG)
XRCC5	ENST00000392133	Transcription regulatory region DNA binding	1.07	5(CA)
XRCC6	ENST00000359308	Protein C-terminus binding.	1.11	
YWHAB	ENST00000372839	Protein targeting	1.14	

DISCUSSION

Ideally, a HKG should be expressed in all normal and tumor tissues as well as in different developmental conditions, with a fairly constant expression level (3). Given that the EST frequency can be an approximate index of gene expression level in a specific cell or tissue, EST data can be an appropriate resource for gene expression analysis (16). In the present study, based on the EST data of 16 normal and tumor tissues, we identified 93 HKGs that are expressed in all tissues. In previous studies, HKGs have been selected only based on identical expression in both tumor and normal tissue. However, in this research, the third criterion i.e., identicial expression in different stages of development was also considered. Our findings showed that some of HKGs are common between tissues and some of them (18 genes) are exclusive for one tissue. Among the studied tissues, esophagus with 30 genes and brain with four genes had the highest and lowest number of HKGs candidates, respectively. Moreover, CALM1, MORF4L1, and HNRNPK covered more tissues, so they can be considered as the most suitable candidate HKGs. In contrast, our results indicated that common HKGs such as ACTB and GAPDH had variable expression in different tissue types and are not a suitable candidate to normalize expression data in gene expression studies. Previous studies have also reported the variable expression of these genes in tissues (7,8,18).

In addition, the type and nucleotide

[DOI: 10.61186/jcbr.5.4.31]

composition of STRs in the promoter region of these genes were investigated for the first time in this study. The core promoter sequences contain fundamental motifs for the expression of the downstream genes. Several studies have demonstrated that STRs affect chromatin organization, regulation of gene activity, modulation of gene expression, etc. (11,15,19). In our previous study, we analyzed the nucleotide composition of the 120-bp flanking sequences to the +1 TSS of human protein-coding genes and showed that approximately 25% of these genes have at least STR of 3-repeats in their core promoters and GA-repeats play a decisive role in the regulation of transcription (20,21).

In the present study, we analyzed the 120-bp immediate upstream sequences to the +1TSSs of candidate HKGs and compared it with other protein-coding genes in order to find a particular pattern for predicting HKGs according to their core promoter STRs. However, we found no difference between HKGs and other protein-coding genes in their core promoter STRs composition. In both groups, most genes had only one STR of which the binary repetitions were the most frequent. In addition, most genes had STRs with three copy numbers and a high-GC content.

CONCLUSION

Our results suggest a novel set of genes as potential HKGs and confirm that some of the common HKGs such as GAPDH and ACTB are not suitable for gene expression normalization. In addition, some genes can be suggested as potential HKGs for a specific tissue; however, this requires further investigations and other high-throughput data since low abundance is a limitation of EST. In addition, core promoter STRs composition and frequency of these genes are the same as other protein-coding genes and GC repeats have the highest frequency in core promoter of these genes.

ACKNOWLEDGMENTS

We would like to acknowledge the Student Research Committee of Golestan University of Medical Sciences for funding this study.

DECLARATIONS

Funding

This research has been supported and funded by the Student Research Committee, Golestan University of Medical Sciences, Gorgan, Iran (project code: 110634).

Ethics approvals and consent to participate

The study was approved by the ethics committee of the Golestan University of Medical Sciences, Iran (ethical code: IR.GOUMS.REC.1397.297).

Conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this article.

REFERENCES

1. Gachon C, Mingam A, Charrier B. Real-time PCR: what relevance to plant studies?. Journal of experimental botany. 2004 Jul 1;55(402):1445-54. [View at Publisher] [DOI] [PMID] [Google Scholar]

2. Saviozzi S, Cordero F, Iacono ML, Novello S, Giorgio SV, Calogero RA. Selection of suitable reference genes for accurate normalization of gene expression profile studies in non-small cell lung cancer. BMC cancer. 2006 Dec;6(1):1-0. [View at Publisher] [DOI] [PMID] [PMCID] [Google Scholar]

3. Casadei R, Pelleri MC, Vitale L, Facchin F, Lenzi L, Canaider S, Strippoli P, Frabetti F. Identification of housekeeping genes suitable for gene expression analysis in the zebrafish. Gene Expression Patterns. 2011 Mar 1;11(3-4):271-6. [View at Publisher] [DOI] [PMID] [Google Scholar]

4. Eisenberg E, Levanon EY. Human housekeeping genes, revisited. TRENDS in Genetics. 2013 Oct 1;29(10):569-74. [View at Publisher] [DOI] [PMID] [Google Scholar]

[DOI: 10.61186/jcbr.5.4.31]

5. Kozera B, Rapacz M. Reference genes in real-time PCR. Journal of applied genetics. 2013 Nov;54(4):391-406. [View at Publisher] [DOI] [PMID] [PMCID] [Google Scholar]

6. Mathur D. Selection of suitable housekeeping genes for expression analysis in glioblastoma using quantitative RT-PCR. Annals of neurosciences. 2014 Apr;21(2):62. [DOI] [PMID] [PMCID] [Google Scholar]

7. Ersahin T, Carkacioglu L, Can T, Konu O, Atalay V, Cetin-Atalay R. Identification of novel reference genes based on MeSH categories. PLoS One. 2014;9(3). [View at Publisher] [DOI] [PMID] [PMCID] [Google Scholar]

8. Lin J, Redies C. Histological evidence: housekeeping genes beta-actin and GAPDH are of limited value for normalization of gene expression. Development genes and evolution. 2012 Nov;222(6):369-76. [View at Publisher] [DOI] [PMID] [Google Scholar]

9. Ewing RM, Kahla AB, Poirot O, Lopez F, Audic S, Claverie JM. Large-scale statistical analyses of rice ESTs reveal correlated patterns of gene expression. Genome research. 1999 Oct 1;9(10):950-9. [View at Publisher] [DOI] [PMID] [PMCID] [Google Scholar]

10. Funari VA, Voevodski K, Leyfer D, Yerkes L, Cramer D, Tolan DR. Quantitative gene expression profiles in real time from expressed sequence tag databases. Gene Expr. 2010;14(6):321-36. [View at Publisher] [DOI] [PMID] [PMCID] [Google Scholar]

11. Bagshaw AT. Functional mechanisms of microsatellite DNA in eukaryotic genomes. Genome biology and evolution. 2017 Sep;9(9):2428-43. [View at Publisher] [DOI] [PMID] [PMCID] [Google Scholar]

12. Gemayel R, Vinces MD, Legendre M, Verstrepen KJ. Variable tandem repeats accelerate evolution of coding and regulatory sequences. Annual review of genetics. 2010 Dec 1;44:445-77. [View at Publisher] [DOI] [PMID] [Google Scholar]

13. Bolton KA, Ross JP, Grice DM, Bowden NA, Holliday EG, Avery-Kiejda KA, Scott RJ. STaRRRT: a table of short tandem repeats in regulatory regions of the human genome. BMC genomics. 2013 Dec;14(1):1-6. [View at Publisher] [DOI] [PMID] [PMCID] [Google Scholar]

14. Gymrek M, Willems T, Guilmatre A, Zeng H, Markus B, Georgiev S, Daly MJ, Price AL, Pritchard JK, Sharp AJ, Erlich Y. Abundant contribution of short tandem repeats to gene expression variation in humans. Nature genetics. 2016 Jan;48(1):22-9. [View at Publisher] [DOI] [PMID] [PMCID] [Google Scholar]

15. Fotsing SF, Margoliash J, Wang C, Saini S, Yanicky R, Shleizer-Burko S, Goren A, Gymrek M. The impact of short tandem repeat variation on gene expression. Nature genetics. 2019 Nov;51(11):1652-9. [View at Publisher] [DOI] [PMID] [PMCID] [Google Scholar]

16. Milnthorpe AT, Soloviev M. The use of EST expression matrixes for the quality control of gene expression data. PLoS One. 2012 Mar 8;7(3):e32966. [View at Publisher] [DOI] [PMID] [PMCID] [Google Scholar]

17. Waxman S, Wurmbach E. De-regulation of common housekeeping genes in hepatocellular carcinoma. BMC genomics. 2007 Dec;8(1):1-9. [View at Publisher] [DOI] [PMID] [PMCID] [Google Scholar]

18. Aithal MG, Rajeswari N. Validation of housekeeping genes for gene expression analysis in glioblastoma using quantitative real-time polymerase chain reaction. Brain tumor research and treatment. 2015 Apr 1;3(1):24-9. [View at Publisher] [DOI] [PMID] [PMCID] [Google Scholar]

19. Gymrek M. A genomic view of short tandem repeats. Current opinion in genetics

[DOI: 10.61186/jcbr.5.4.31]

& development. 2017 Jun 1;44:9-16. [View at Publisher] [DOI] [PMID] [Google Scholar]

20. Valipour E, Kowsari A, Bayat H, Banan M, Kazeminasab S, Mohammadparast S, Ohadi M. Polymorphic core promoter GA-repeats alter gene expression of the early embryonic developmental genes. Gene. 2013 Dec 1;531(2):175-9. [View at Publisher] [DOI] [PMID] [Google Scholar]

21. Darvish H, Nabi MO, Firouzabadi SG, Karimlou M, Heidari A, Najmabadi H, Ohadi M. Exceptional human core promoter nucleotide compositions. Gene. 2011 Apr 15;475(2):79-86. [View at Publisher] [DOI] [PMID] [Google Scholar]

How to Cite: Ramezani M, Javan B, Jafari M H, Banisadr A, pourafshar M, Raeissadati S S, et al . In Silico Identification of Housekeeping Genes by Expressed Sequence Tags and Assessment of Short Tandem Repeats in Their Promoters. Journal of Clinical and Basic Research. 2021; 5 (4) :31-43

[DOI: 10.61186/jcbr.5.4.31]