Indian Journal of Biotechnology
Vol 4, January 2005, pp 21-38

UlTRaSCAN—An algorithm for prediction of pattern(s) in untranslated regions of eukaryotic mRNAs

Rajeshwari Marikkannu[1] and P Palanivelu[2]*

[1]Bioinformatics Centre and [2]Department of Molecular Microbiology, School of Biotechnology
Madurai Kamaraj University, Madurai 625 021, India

Received 2 July 2003; revised 10 March 2004; accepted 20 March 2004

Many *cis* acting elements have been identified in the 3′ and 5′ untranslated regions (UTRs) of eukaryotic mRNAs. These *cis* acting elements are found to play vital roles in pre-mRNA processing, nucleo-cytoplasmic transport of processed mRNAs, determining the efficiency of translation of mRNAs and their stability and degradation in the cytoplasm. UTRScan utility has been used to identify these patterns in mRNAs. However, the UTRScan is not very sensitive and also not highly specific. It often generates many false positives as it scans the whole mRNA including the coding sequence. The authors have developed a new algorithm and an Internet based web application tool, named UlTRaSCAN, which overcomes these limitations and proved to be highly specific and sensitive in detecting patterns in UTRs. The new algorithm identifies these *cis* acting elements only in the 3′ and 5′ UTRs of eukaryotic mRNA/DNA sequences. The sensitivity is more than doubled and the specificity is increased, close to 100%. The UlTRaSCAN also minimized the false positives to almost 0% and the false negatives to large extent.

Keywords: eukaryotic mRNAs, 3′ untranslated region, 5′ untranslated region, UTRScan, UlTRaSCAN, UTR patterns
IPC Code: **Int. Cl.**[7] C 12 N 15/10; G 06 F 15/00

Introduction

Cells respond to various environmental changes, not only by reprogramming the expression of specific genes but also regulating the level of expression of these genes throughout the genome. The rate of transcription of a particular gene is controlled mainly by interactions of diverse group of regulatory proteins belonging to the classes of transcriptional initiators, activators and repressors. The primary control of transcription is exercised at the initiation of transcription itself. After transcription, additional levels of controls are exercised at the post-transcriptional stage. The post-transcriptional controls include, processing of the pre-mRNA transcript, export of processed mRNAs to the cytoplasm, stability and degradation of mRNAs in the cytoplasm[1] and their translational efficiency in a particular tissue[2]. Not only various binding proteins but also the inducers of a particular protein bind at the 3' untranslated regions (UTRs) and stabilize the mRNAs.

The mRNAs contain three distinct regions: the coding sequence (CDS), the 5' UTR and 3' UTR. Whereas the CDS determines the sequence of amino acids and hence, the type of protein, the 5' UTR or otherwise called "leader sequence" and its secondary structure(s) determines the translational efficiency of the mRNA and the 3' UTR otherwise known as the "tail," determines the stability and degradation of the message.

The required information for the translational efficiency of mRNAs and also for their stability and degradation mostly resides on the mRNA itself as *cis* acting elements. Many such *cis* acting elements have been identified in the 3′ and 5′ UTRs of eukaryotic mRNAs. UTRScan utility software has been used to identify such patterns in different types of DNA/mRNA sequences. The UTRScan searches user-submitted sequences for any of the previously identified UTR patterns, which are already available in UTRsite[3]. However, the UTRScan is not very sensitive and specific and often produces many false positives, as it scans any input DNA/mRNA sequence including the CDS. In this paper, the authors report the development of a new algorithm, which identifies the patterns only in the 3′ and 5′ UTR of mRNAs and, therefore, increases the sensitivity and specificity and minimizes the false positives markedly.

For these studies, 5′ UTR sequences are defined as the mRNA region spanning from the cap site to the

*Author for correspondence:
Tel: 91-452-2458208; Fax: 91-452-2459105
E-mail: PP@mrna.tn.nic.in

start codon (excluding the start codon, ATG) and 3′ UTR sequences are defined as the mRNA region spanning from the stop codon (TGA, TAA, TAG) (excluding the stop codon) to the poly-A starting site. Fig. 1 shows the transcription, splicing and translational events of a typical eukaryotic mRNA and also the CDS, translated and UTRs on the processed mRNA. In eukaryotic mRNAs, the average length of the 5′ UTR is ~200 nucleotides and the 3′ UTR is ~200 to 750 nucleotides.

Materials and Methods

Worldwide Web (WWW) Survey

Internet resources available for UTR analysis are as mentioned below:

1. UTR Database and UTRsite

These are specialized databases comprising sequences and functional elements at 5′ and 3′ UTRs of eukaryotic mRNAs.

UTR Database (UTRdb)

The UTRdb is a specialized collection of sequences derived from redundancy of 5′ and 3′ UTR sequences from eukaryotic mRNAs[4]. The UTRdb entries have been enriched with specialized information not present in the primary databases. It includes all the sequence patterns demonstrated by experimental evidences to play some functional role. Moreover, it consists of eight different sequence collections which were generated for both 5′ and 3′ UTR sequences with 120767 entries, one for each of the eukaryotic divisions of the EMBL/GenBank nucleotide database, viz., 1. Human, 2. Rodent, 3. Other mammals, 4. Other vertebrates, 5. Invertebrates, 6. Plants, 7. Fungi and 8. Patents.

UTRsite

UTRsite describes the various regulatory elements present in UTR regions whose functional role has been established on an experimental basis. Each UTRsite entry is constructed on the basis of information reported in the literature and revised by distinguished scientists experimentally analyzing on the functional characterization of the relevant UTR regulatory elements. The functional UTR patterns are defined on the basis of the information reported in the literature and/or by the scientists, experts in the field. These patterns were described by using the pattern description syntax used in the PATSCAN program[5]. This database is continuously updated with new entries describing functional patterns whose biological role has been experimentally demonstrated.

UTRScan

The UTRScan utility software allows the user to search user-submitted sequences for any of the patterns collected in UTRsite. The UTRfasta utility allows database searches against fully annotated UTRdb entries. The UTRdb are updated with new EMBL database releases and the UTRsite will be continuously updated by adding functional role to the UTRs that has been demonstrated experimentally.

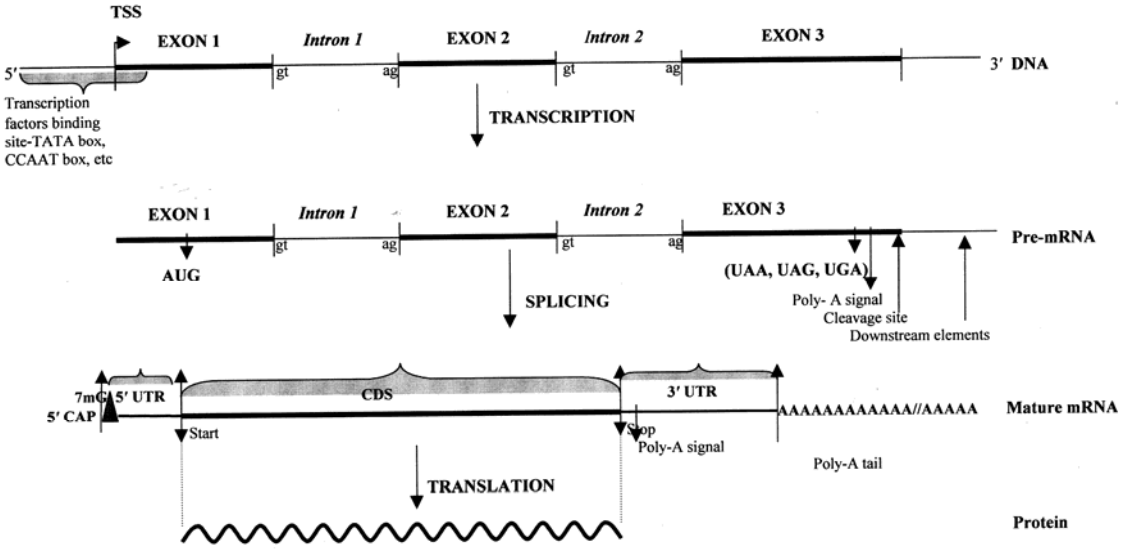

Fig.1— Pre-mRNA processing in eukaryotes and the 5′, 3′ UTRs and CDS in the processed mRNA

Availability

FTP: ftp://area.ba.cnr.it/pub/embnet/database/utr/

URL: http://bighost.area.ba.cnr.it/BIG/UTRHome/

2. TransTerm

This is a translational signal database, extended to include full coding sequences and UTRs. TransTerm is a database[6] of mRNA sequences and is useful for detecting translational control signals, in general. This database contains more than 1,30,000 non-redundant coding sequences with associated UTRs from over 450 species. This database includes the complete genomes of 12 prokaryotic and one eukaryotic organism. Several coding sequence parameters are available such as coding sequence length, Nc, GC3 and when it is computable, the codon adaptation index (CAI). It also includes the codon usage tables and summaries of start and stop codon contexts. TransTerm-98 has both a relational database from a www interface and a flat-file format available for internet browsers.

Availability

URL:http://biochem.otago.ac.nz:8000/Transterm/homepage. html

3. Association of Nucleotide Patterns with Gene Function Classes: (Application to Human 3′ Untranslated Sequences)

In this study Conklin *et al*[7] describes a new association rule-mining method for discovering nucleotide sequence patterns, those appear in large number of sequences. Some significant associations between nucleotide patterns and protein function classes are also presented. Among previously identified patterns, the AU-Rich Element (ARE) is found to occur within the 3′ UTRs of cytokines, providing statistical validation of an association, often reported in the literature. This method has also identified some GC-rich patterns, found to occur within the 3′ UTR of homeodomain transcription factors and nuclear proteins. This method should be applicable to identify many types of regulatory elements.

4. Identification of 3′-Terminal Exon in Human DNA

Tabaska *et al*[8] reported a new program, JTEF, for finding 3′ terminal exons in human DNA sequences. This program is based on quadratic discriminant analysis, a standard non-linear statistical pattern recognition method. The quadratic discriminant functions, used for building the algorithm were trained on a set of 3′ terminal exons of type 3′ tuexon (those containing the true stop codon). They have showed that the average predictive accuracy of JTEF is higher than the presently available best programs (GenScan and Genemark.hmm), based on a test set of 65 human DNA sequences with 121 genes. In particular, JTEF performs well on larger genomic contigs containing multiple genes and significant amounts of intergenic DNA. It is a valuable tool for genome annotation and functional genomics studies.

Availability

FTP: ftp://www.cshl.org/pub/science/mzhanglab/JTEF

URL: http://www.argon.cshl.org

5. Prediction of Eukaryotic mRNA Translational Properties

It has been well established that eukaryotic mRNAs are translated with different efficiencies, depending on their sequence characteristics. Evaluation of mRNA translatability is of primary importance especially in prediction of gene expression patterns by computer methods and to improve the recognition of mRNAs within cloned nucleotide sequences[9]. It may also be used in biotechnological experiments to optimize the expression of foreign genes in transgenic organisms. A set of 5′ UTR patterns, significantly different between mRNAs encoding abundant and scarce polypeptides is determined for mammals, dicot- and monocot plants and available in the LEADER_RNA database. Computer tools for the prediction of mRNA translatability are presented in the following databases.

Availability

URL: Monocots:http://www.mgs.bionet.nsc.ru/programs/acts2/mo_mRNA.htm

Dicots:http://www.mgs.bionet.nsc.ru/programs/acts2/mo_mRNA.htm

Mammals:http://www.mgs.bionet.nsc.ru/programs/acts2/mo_mRNA.htm

LEADER_RNAdb:http://www.mgs.bionet.nsc.ru/systems/LeaderRNA/

Measures of the Predictive Accuracy

In order to evaluate the predictive accuracy of any prediction algorithm, the predicted results are compared with the actual results (based on the experimental results) as annotated in the databases. From this comparison, the accuracy measures of the prediction algorithm are calculated. The accuracy measures of prediction algorithm are TP (true positives), TN (true negatives), FP (false positives) and FN (false negatives) and are defined as follows:

TP = the number of entries having 5′ or 3′ UTR patterns and predicted also to have 5′ or 3′ UTR patterns

TN = the number of entries not having 5′ or 3′ UTR patterns and predicted also not to have 5′ or 3′ UTR patterns

FP = the number of entries which do not have 5′ or 3′ UTR patterns, but predicted to have 5′ or 3′ UTR patterns.

FN = the number of entries which do have 5′ or 3′ UTR patterns, but predicted as not having the 5′ or 3′ UTR patterns.

Sensitivity (Sn)

Sn is defined as the measure of accuracy for including all the experimentally proved 5′ and 3′ UTRs that are correctly predicted to have UTR patterns.

$$Sn = \frac{TP}{TP + FN}$$

Specificity (Sp)

Sp is defined as the measure of accuracy for predicting the patterns only in the 3′ and 5′ UTRs and not predicting UTR patterns in the CDS.

$$Sp = \frac{TP}{TP + FP}$$

These are the widely used measurements of accuracy for prediction algorithms. Both Sn and Sp range independently over [0,1], with perfect prediction occurring only when both measures are equal to 1.

Analysis of mRNAs using UTRScan

Thirty experimentally proved nucleotide sequences from the UTRdb were retrieved and given as input to the UTRScan algorithm. The results are shown in Table 1.

Due to FP and FN, the Sp and Sn are 0.381, 0.242, respectively for the analysis of fungal entries. To eliminate the factor of biological system dependency, the analysis is extended to human and plant entries also. But again the Sp and Sn are 0.393, 0.367 and 0.647, 0.379, respectively. Thus, the above algorithm available already on the web (UTRScan) has the following limitations. It is less specific and sensitive to detect UTR patterns. Furthermore, it generates very many false positives and also predicts UTR patterns in the CDS, which are not annotated in the reference UTRdb. So, there is a need for an algorithm with higher Sp and Sn for effective analysis of UTRs and detect patterns only in the UTRs.

UlTRaSCAN — A New Algorithm

Principle and Use of UlTRaSCAN

The new algorithm, UlTRaSCAN, extracts the 5′ and 3′ UTRs from the mRNA/cDNA sequence and the start and end of the CDS and analyze it. For UlTRaSCAN analysis, the input can be from the user, in GenBank file format or FASTA file format or from user submitted sequence in FASTA text format**.** The extracted 5′ and 3′ UTRs are utilized for UTR analysis. The UlTRaSCAN results present the length of the pattern in the query sequence, the sequence of the pattern and also the start and end positions of the particular pattern. A typical sequence file and its various segments are shown in Fig. 2.

System and Program Development

UlTRaSCAN algorithm was programmed and tested using PERL (Version 5.6.0) and using CGI.pm module. The programme was developed using PENTIUM IV at 1.7 GHz system, running Red Hat Linux 8.0 with Apache web server (Version 1.3.17). The FASTA text or the FASTA file can be given as the input for the UTR analysis. The predicted patterns in the particular query sequence are displayed in the user screen. The following scheme in Fig. 3 explains the working of the UlTRaSCAN.

Prediction of UTR Patterns using UlTRaSCAN

The new software UlTRaSCAN is tested for its accuracy and specificity by analyzing a randomly selected thirty sequences from fungal, plant and

Table 1— Comparison of calibration results of UTRScan and UlTRaSCAN

Measure of accuracy	UTRScan (30)*			UlTRaSCAN (30)*		
	Fungi	Human	Plant	Fungi	Human	Plant
Sensitivity (Sn)	0.242	0.367	0.379	0.810	0.644	0.661
Specificity (Sp)	0.381	0.393	0.647	0.971	1.000	1.000
False positives	14	16	12	2	1	0
False negatives	36	23	26	16	8	22

*The number of sequences analyzed from the UTRdb for calibration

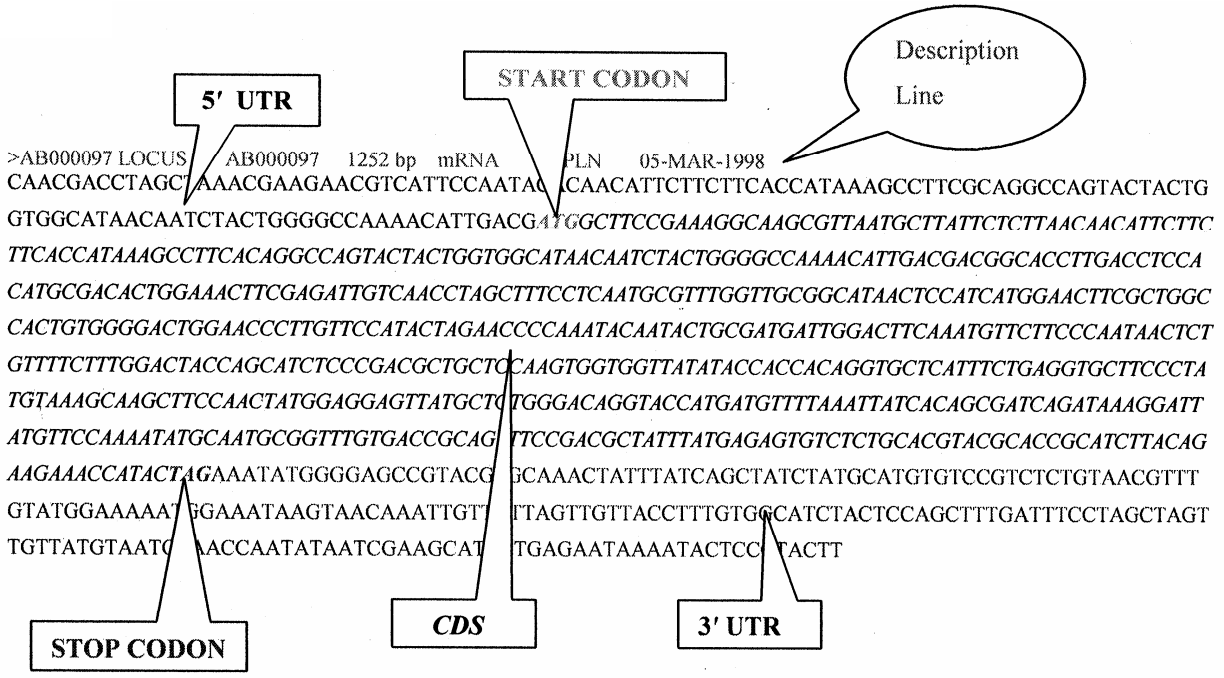

Fig. 2 —Extraction of 5′ and 3′ UTRs from input DNA/mRNA sequence and start and stop codon information

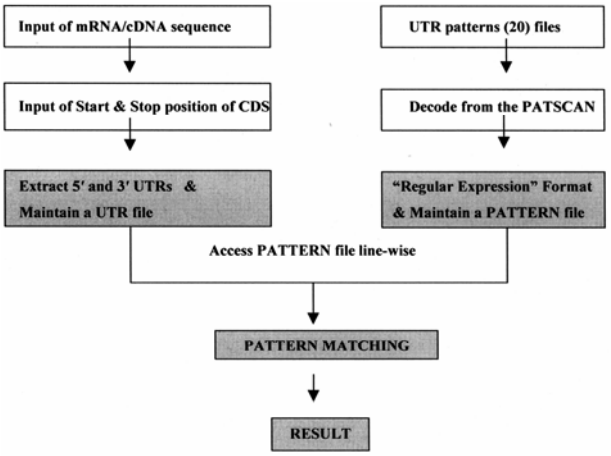

Fig. 3— Scheme for the analysis of 3′ and 5′ UTR patterns using UlTRaSCAN

human entries available in the UTRdb. The results are compared with UTRscan and shown in Table 1.

Results and Discussion

Analysis of Patterns in UTRs by UTRScan and UlTRaSCAN

The existing software UTRScan and the new software UlTRaSCAN were compared for their accuracy and specificity by analyzing thirty randomly selected sequences from fungi, plants and human, for their efficiency in correctly and completely predicting patterns in the 3′ and 5′ UTRs. The results are shown in Table 1. It is clear from the table that the prediction specificity is increased to almost 100% and the false positives are minimized to 0% by using the UlTRaSCAN. Since the 5′ and 3′ UTRs are extracted and then fed for the pattern-searching algorithm, all the false positives are almost completely eliminated in the UlTRaSCAN analysis. The patterns IRE, SECIS, TOP, IRES, and Upstream ORF are predicted to be more prevalent in the mRNAs of all eukaryotic systems.

Tables 2 to 4 present the data on the complete analysis of the 30 randomly selected sequences from fungi, human and plants both by UTRScan and UlTRaSCAN. There seems to be no pattern(s) specific either to 3′ or 5′ UTR; All the above patterns are found both in 3′ and 5′ UTRs in different classes of mRNA. However, a closer look at the patterns, reveals that only few are dominant among them and widely distributed in 5′ and 3′ UTRs. Tables 5 and 6 show that they are detected better by UlTRaSCAN rather by UTRScan. Furthermore, it is clear from Table 6 that the analysis and pattern finding in UTRs of eukaryotic mRNAs is more complete with the UlTRaSCAN algorithm rather than with UTRScan. In some instances, both the algorithms predict a particular pattern but at different regions on the same mRNA.

A consolidated account of the results of analysis from all three organisms is presented in Table 5. The dominant patterns are indicated in bold. Some of these patterns appear to be mRNA-specific, whereas others are found in all mRNAs. It is clear from the above table that the most abundant patterns in eukaryotic

Table 2— Calibration of UTRScan and UlTRaSCAN using fungal entries

No.	UTRAcc.No.	EMBL Acc. No.	Length (bp)	CDS	UTR Type & Length	UTRScan Results	UlTRaSCAN Results
1	3RMI000006	A02536	1205	10..1101	3'(1102..1205)104	P7(212226)=FP P7(834-853) P12(11171205)=TP	P12(1102..1206)=TP P17(1122..1165)
					5' (1..9)9	FN	FN
2	3HIN000003	A21793	1060	10..927	3'(928..1060)133	FN	P12(928..1029)=TP P17(941..1038)
					5'(1..9)9	FN	FN
3	3FOX000013	A21795	1473	97..1227	3'(1228..1473)246	P7(261-280)=FP P7(777-790)	P17(1229..1456)=TP P18(1361..1367)
					5'(1..96)96	FN	P12(1..96)=TP
4	3ANI000032	A35002	1319	26..1114	3'(1115..1319)205	FN	P12(1115..1218)=TP P17(1125..1296) P18(1274..1280)
					5'(1..25)25	FN	FN
5	3ANI000033	A35006	1174	30..1025	3'(1026..1174)149	P12(10821174)=TP	P12(1026..1127)=TP P17(1029..1126) P18(1064..1070)
					5'(1..29)29	P7(239-257)=FP P7(623-636) P7(668..687)	FN
6	3SCE000220 5SCE000322	A38763	1747	259.1296	3'(1297..1747)451 5'(1..258)258	P7(1487-1506)=TP	P3(1553..1652)=TP P9=(76..91)=FP P12(0..104) P17(51..238)
7	3SCE000221 5SCE000323	A38765	3086	593..2368	3'(2369..3086)718	P7(2132..2146)=TP P12(3001..3086)	P2(2600..2633)=TP P18(2543..2622)
					5'(1..592)592	FN	P3(446..550)=FP P12(0..102) P17(1..579)
8	3SCE000222 5SCE000324	A38767	1529	523..1149	3'(1150..1529)380	FN	P3(1394..1485)=TP P9(1246..1261)
					5'(1..522)522	FN	P12(0..102)=TP P17(1..484) P18(434..440)
9	3SCE000223 5SCE000325	A38769	1300	470..979	3'(980..1300)321 5'(1..469)469	FN FN	FN P3(132..229)=TP P12(1..101) P17(4..463)
10	3SCE000224 5SCE000326	A38771	1879	180..896	3'(897..1879)983	FN	P3(1734..1839)=TP P18(1275..1354) P20(1041..1051)
					5'(1..179)179	FN	P9(33..48) P12(0..102) P17(60..175)=TP
11	3SCE000225 5SCE000327	A38773	2365	1..1344	3'(2080..2365)286	P3(764-827)=FP P7(424-437)	P3(2046..2148)=TP P12(1345..1447) P17(1346..2363)
					5'(1..1344)1344	FN	FN

Contd.

Table 2— Calibration of UTRScan and UlTRaSCAN using fungal entries—*Contd*

No	UTRAcc.No.	EMBL Acc. No.	Length (bp)	CDS	UTR Type & Length	UTR Scan Results	UITRa Scan Results
12	3EGO000001 5EGO000001	A46558	1329	243..1148	3'(1149..1329)181	FN	P2(1243..1276)=TP
					5'(1..242)242	FN	P12(0..103)=TP P17(4..236) P18(161..167)
13	3EGO000002 5EGO000002	A46560	2627	451..2280	3'(2281..2627)347	P12(2524-2627)=FP	P18(2337..2343)=TP
					5'(1..450)450	FN	P3(41..139) P12(0..104)=TP P17(1..446)
14	3EGO000003 5EGO000003	A46562	1082	315..953	3'(954..1082)129	FN	FN
					5'(1..314)314	P9(1-9)=TP	P3(19..221)=TP P12(0..101) P17(7..283)
15	3EGO000004 5EGO000004	A46564	996	271..789	3'(790..996)207	FN	FN
					5'(1..270)270	FN	P3(0..87)=TP P12(0..101) P17(0..262)
16	3EGO000005	A46566	1511	525..1232	3'(1233..1511)279	P7(1054-1067)=FP	FN
					5'(1..524)524	FN	P3(87..120)=TP P12(0..102) P17(0..523)
17	3EGO000006	A46568	1596	353..1093	3'(1094..1596)503	P12(1519-1596)=TP P9(1398..1413) P3(1012-1084)=FP	P7(1342-1361)=TP
					5'(1..352)352	FN	P12(0..101) P17(27..269)=TP
18	3EGO000008 5EGO000007	A94699	1380	210..1013	3'(1014..1380)367	FN	FN
					5'(1..199)199	P16(168..175)=TP	P12(0..101)=TP P16(167..175) P17(185..199)
19	3ANI000060	AXO30039	3108	327..2743	3'(2744..3108)365	P7(561..576)=FP P7(2246..2264)	P3(2966..3068)=TP P9(2902..2914)
					5'(1..326)326	P7(108..126)=TP	P12(0..103)=TP P17(2..309)
20	3MGR000019	AXO58235	3979	1637..3214	3'(3215..3711)497	P14(3428..3491)=TP	P9(3446..3461)=TP P14(3428..3492) P18(3295..3374)
					5'(1..1636)1636	P12(880,3979)=FP	P3(1220..1323)=TP P12(0..101) P17(4..1612)
21	3TVE000021	AX078057	3448	1133..1877	3'(1878..3448)1571	FN	P18(3090..3169)=TP
					5'(1..1132)1132	P7(1293..1308)FP	P2(22..54)=TP P3(843..947) P9(1067..1079) P12(0..102) P17(14..1062)
22	3AAC000001	U49378	1195	32..1027	3'(1028..1195)168	FN	P2(1067..1100)=TP P9(1108..1120) P12(1028..1129) P17(1030..1152)
					5'(1..31)31	FN	FN

Contd.

Table 2— Calibration of UTRScan and UlTRaSCAN using fungal entries—*Contd*

No	UTRAcc.No.	EMBL Acc. No.	Length (bp)	CDS	UTR Type & Length	UTRScan Results	UITRa SCAN Results
23	3AAL000001	X84217	1532	11..1312	3'(1313..1532)220	P12(444-1532)=FP FN	P9(1463..1475)=TP P12(1313..1416) P17(1314..1496)
					5'(1..10)10	FN	FN
24	3AAL000002	X78227	1647	59..1546	3'(1547..1647)101	P7(175-188)=FP	P17(1553..1628)=TP
					5'(1..58)58	FN	P12(0..58)=TP
25	3AAL000004	X78225	949	23..637	3'(638..949)312	P3(536-596)=FP P7(147-161)	P12(638..742)=TP P17(643..936)
					5'(1..22)22	FN	FN
26	3AAC000035	D64088	2812	111..2693	3'(2694..2812)119	P7(1393-1412)=FP P12(2782-2872)=TP	FN
					5'(1..110)110	FN	P12(0..105)=TP P17(3..71)
27	3AAC000003	X52525	932	50..763	3'(764..932)169	FN	P3(830..933)=TP P12(764..865)
28	3AAU000014	D00573	1342	80..1021	3'(1022..1342)321	P12(1251,1342)=TP	P9(1107..1122)=TP P18(1250..1329)
					5'(1..79)79	P9(1-5)=TP	P12(0..79)=TP P17(3..57)
29	3AAC000004	L34599	1266	28..1080	3'(1081..1266)186	FN	P12(1081..1185)=TP P17(1103..1221)
					5'(1..27)27	FN	FN
30	3AMU000002	AAJ10141	988	232..786	3'(787..988)202 5'(1..231)231	FN FN	FN P12(0..103)=TP P17(4..167)

Present & Predicted = TP; Present & Not predicted = FN;
Absent & Predicted = FP; Absent & Not predicted = TN;
UTR Patterns:
P1 = HISTONE 3. P2 = IRE. P3 = SECIS. P4 = APP. P5 = CPE. P6 = TGE. P7 = 15-LOX-DICE.
P8 = ARE 2. P9 = TOP. P10 = GLUT1 RNA. P11 = TNF. P12 = IRES. P13 = MSL2-5UTR.
P14 = MSL2-3UTR. P15 = RPMS12_TCE. P16 = ADH_DRE. P17 = UpstreamORF.
P18 = NANOS_TCE. P19 = VIMENTIN. P20 = BRE.

mRNAs are the P12, internal ribosome entry site, (IRES) and P17, upstream open reading frame (uORF) (even though the pattern is called uORF, it includes the downstream ORFs in the 3′ UTR also). Though these two patterns are found both in the 5′ and 3′ UTRs, yet they are widespread at the 5′ UTRs. It is interesting to note that both the patterns are present in almost all the eukaryotic mRNAs analyzed. For example, P12 is present in all (30/30) fungal entries, 27/30 of the human entries and 23/30 of the plant entries and the P17 is present in 28/30 of the fungal entries, 28/30 of the human entries and 28/30 of the plant entries. Approximately 95% of the analyzed mRNAs contain both the patterns and the rest 5% contain at least one of the patterns. Thus, the occurrence of both or at least one of the patterns in all eukaryotic mRNAs suggests an important role for these two patterns. Furthermore, analysis of some well-known proteins for dominant patterns again proved that these two patterns are the most abundant-

Table 3— Calibration of UTRScan and UlTRaSCAN using plant entries

No.	UTR Acc. No.	EMBL Acc. No.	Length (bp)	CDS	UTR Type & Length	UTRScan Results	UlTRaSCAN Results
1	5AAN000013 3AAN000013	U14625	1980	84. .1766	5'(1..83)83 3'(1767..1980)214	FN P12(1884..1980)=TP P9(1913..1925)	P12(0..83)=TP P17(1774..1969)=TP
2	5AAN000014 3AN000014	U36376	1337	74. .1105	5'(1..73)73 3'(1106..1337)232	FN P3(925..944) P12(1239..1337)	P12(0..73)=TP P17(56..69) FN
3	5AAC000034 3AAC000037	AB003937	736	91. .585	5'(1..90) 3'(586..736)151	P12(0..90)=TP P5(600..736)=FP P3(649..737)=TP P5(600..737)	P17(13..88)=TP P18(633..639)=TP
4	5AAC000035 3AAC000038	AB003938	825	68. .598	5'(1..67)67 3'(599..825)227	P10(107..116)=FP P12(729..825)=TP	P12(0..67)=TP P17(33..57) FN
5	5AAC000012 3AAC000012	D50528	2169	75. .1916	5'(1..74)74 3'(1917..2169)253	P9(18..30)=TP P12(2061..2169)=TP	P12(0..74)=TP P17(28..68) FN
6	5AAC000028 3AAC000028	D50529	2084	46. .1887	5'(1..45)45 3'(1888..2084)197	P9(1..5)=TP P12(1981..2084)=TP	P17(3..29)=TP P12(1888..1991)=TP
7	5AAN000033 3AAN000045	AJ249561	1893	34. .1677	5'(1..33)33 3'(1676..2106)431	FN P12(1804..1893)=TP	FN P11(1678..1780)=TP P17(1692..1881)
8	5AAN000037	AJ251751	2106	34. .1675	5'(1..33)33 3'(1676..2106)197	FN P12(2007..2106)=TP	FN P12(1676..1779)=TP P17(1689..2073) P18(1952..1958)
9	5AAC000036 3AAC000039	D50531	1726	87. .1565	5'(1..86)86 3'(1566..1726)161	FN P7(1134..1151)=FP P12(1561..1666)=TP	P12(0..86)=TP P17(52..84) FN
10	5AAC000037 3AAC000040	D50530	1666	23. .1501	5'(1..22) 3'(1502..1666)165	FN P12(1620..1726)=TP	P17(10..20)=TP P11(1502..1609)=TP
11	5GMA000274 3GMA000345	AB000097	1252	40. .1041	5'(1..39) 3'(1042..1252)211	P5(710..1252)=FP P12(1174..1252)=TP	FN P12(1042..1143)=TP P17(1047..1239)
12	5ATH000001 3ATH000002	AB000797	2073	18. .1148	5'(1..17) 3'(1149..2073)925 P12(1970..2073)=TP P2(1723..1756)	FN P3(1818..1921) P12(1149..1250)	FN P17(1153..2068)=TP P18(1939..1945)
13	5ATH000002 3ATH000004	AB000799	2157	42. .1997	5'(1..41) 3'(1998..2157)160	FN P12(2062..2157)=TP	FN P12(1998..2100)=TP P17(1998..2153)

contd

Table 3— Calibration of UTRScan and UlTRaSCAN using plant entries—Contd

No	UTR Acc. No.	EMBL Acc. No.	EMBL Acc. No.	CDS	UTR Type & Length	UTRScan Results	UlTRaSCAN Results
14	5GEC000015	AB001380	1920	72. .1643	5'(1..71)	FN	P12(0..71)=TP
	3GEC000025				3'(1644..1920)277		P17(31..66)
						FN	P9(1714..1729)=TP
15	5LES000314	AB001389	1929	205. .1722	5'(1..204)	FN	P12(0..108)=TP
	3LES000374						P17(3..202)
					3'(1723..1929)207	FN	FN
16	5ATH001543	AB001568	803	113. .622	5'(1..112)112	P16(392..399)=FP	P12(0..102)=TP
							P17(27..99)
17	5INI000001	AB001818	1438	81. .1247	5'(1..80)80	FN	P12(0..80)=TP
	3INI000001						P17(34..47)
					3'(1248..1438)191	P7(528..545)=FP	P9(1396..1411)=TP
						P7(761..779)	
18	5INI000002	AB001819	1401	37. .1206	5'(1..36)36	P9(1..5)=TP	FN
	3INI000002						
					3'(1207..1401)195	P12(1305,1401)=TP	P12(1206..1307)
							P17(1223..1393)
							P18(1215..1221)
19	5OSA000578	AB001882	1497	124. .1335	5'(1..123)123	FN	P17(0..32)=TP
	3OSA000646						P12(0..103)
					3'(1336..1497)162	FN	FN
20	5OSA000579	AB001883	987	23. .796	5'(1..22)22	FN	FN
	3OSA000647						
					3'(797..987)191	P12(900..987)=TP	P12(797..898)=TP
							P17(799..965)
21	5SGI000005	AB003138	2555	126..2319	5'(1..125)125	FN	P9(44..59)=TP
	3SGI000006						P12(0..102)
							P17(37..123)
					3'(2320..2555)236	P7(496..509)=FP	FN
22	5OSA000581	AB001885	1197	120. .935	5'(1..119)119	P7(371..386)=FP	P12(0..70)=TP
	3OSA000649					P9(1..10)=TP	
					3'(936..1197)262	FN	P17(1156..1415)=TP
23	5OSA000582	AB001886	1418	71. .1153	5'(1..70)70	FN	FN
	3OSA000650				3'(1154..1418)265	FN	P9(1348..1360)=TP
							P12(1037..1138)
							P17(1046..1335)
							P18(1128..1134)
24	5OSA000583	AB001887	1361	29. .1036	5'(1..28)28	P7(712..730)=FP	FN
	3OSA000651				3'(1037..1361)325	FN	FN
25	5OSA000584	AB001888	2147	261. .1484	5'(1..260)	FN	FN
	3OSA000652						
					3'(1485..2047)563	P12(1937..2047)=TP	FN
26	5PPY000013	AB002139	952	16. .702	5'(1..15)15	FN	FN
	3PPY000018						
					3'(703..952)250	P12(860..952)=TP	P12(703..804)=TP
							P17(793..933)
							P18(781..787)

Contd.

Table 3— Calibration of UTRScan and UlTRaSCAN using plant entries—Contd

No	UTR Acc. No.	EMBL Acc. No.	EMBL Acc. No.	CDS	UTR Type & Length	UTRScan Results	UlTRaSCAN Results
27	5PPY0000163 PPY000021	AB002142	831	58. .747	5'(1..57)57	FN	P12(0..57)=TP P17(29..44)
					3'(748..831)84	P12(860..952)=TP	FN
28	5PPY000017 3PPY000022	AB002143	890	62. .742	5'(1..61)61	FN	P12(0..61)=TP P17(33..47)
					3'(743..890)	P7(689..705)=FP	P18(786..792)=TP
29	5OSA000446 3OSA000492	AB002266	5190	113. .5521	5'(1..112)112	P3(2143..2204)=FP P7(59..77)=TP	P12(0..108)=TP
					3'(5522..5910)389	P12(5821..5910)=TP	P17(5530..5870)=TP P2(5663..5696)
30	5PTI000008 3PTI000009	AB003089	1794	42. .1577	5'(1..41)41	FN	FN
					3'(1578..1794)217	P7(768..784)=FP	P12(1578..1679)=TP P17(1599..1776) P18(1616..1622) P20(1632..1642)

For legends refer. Table 2

Table 4.— Calibration of UTRScan and UlTRaSCAN using human entries

No.	UTR Acc. No.	EMBL Acc. No.	EMBL Acc. No	CDS	UTR Type & Length	UTRScan Results	UlTRaSCAN Results
1	5HSA014151	A03911	1260	25..1218	5'(1..24)24	P9(95..109)=FP	P17(10..22)TP
					3'(1219..1260)42	FN	FN
2	5HSA014155	A06977	2258	76..1905	5'(1..75)75	FN	P12(0..75)=TP P17(4..27)
					3'(1906..2258)353	P12(1825..1832)=FP P12(2170..2258)=TP	P9(1982..1994)=TP
3	5HSA014159	A07801	4035	132..245	5'(1..131)131	FN	P9(100..112)=TP P12(0..102)
					3'(246..4035)3790	FN	P2(2932..2965)=TP P3(3271..3376) P17(267..2029) P18(2858..2937)
4	5HSA014162	A07809	Nil	Nil	5'(1..119)119	FN	FN
					3'(1107..1437)331	FN	FN
5	5HSA014167	A12295	1475	103..1248	5'(1..102)102	FN	P12(0..101)=TP
					3'(1249..1475)227	FN	P9(1264..1279))=TP P17(1253..1449 P18(1370..1376) P20(1368..1378)
6	5HSA014174	A16796	1264	328..1137	5'(1..327)327	FN	P2(204..237)=TP P12(0..104) P17(28..305)
					3'(1138..1264)	P14(1111,1118)=FP	P18(1152..1528)=TP

Contd

Table 4. Calibration of UTRScan and UlTRaSCAN using human entries—Contd

No.	UTR Acc. No.	EMBL Acc. No	EMBL Acc. No	CDS	UTR Type & Length	UTRScan Results	UlTRaSCAN Results
7	5HSA014175	A17362	2188	438..1658	5'(1..437)437	P9(310..323)=TP	P12(0..101)=TP P17(5..425)
					3'(1659..2188)530	P9(1394..1412)=FP P9(2157..2175)=TP P12(2103..2188)	P2(2152..2185)=TP
8	5HSA014178	A18397	1964	100..1395	5'(1..99)99	P9(40..58)=TP P9(84..102)	P12(0..99)=TP P17(22..81)
					3'(1396..1964)569	P9(390,408)=FP P9(567..586) P9(995..1008) P12(1885..1964)=TP	FN
9	5HSA014179	A18411	1645	322..771	5'(1.31)31	P9(660..673)=FP P9(1402..1417)=TP	P9(237..249)=FP P12(0..101)
					3'(772..1645)874	P9(771,786)=TP P9(1253..1269)	P17(781..1635)=TP P2(1413..1446)
10	5HSA014180	A18585	1659	34. .1167	5'(1..33)33	FN	FN
					3'(1168..1659)492	P12(1569..1659)=TP	P2(1251..1284)=TP P12(1168..1269) P17(1178..1653)
11	5HSA014188	A21240	1482	22. .1170	5'(1..21)21	P9(103..119)=FP FN	P17(4..21)=TP
					3'(1171..1482)312	FN	P12(1171..1275)=TP
12	5HSA014193	A23337	1630	208. .1623	5'(1..207)207	FN	P2(94..127)=TP P12(0..101) P17(13..205)
					3'(1624..1630)7	FN	FN
13	5HSA014199	A28102	1637	87. .1565	5'(1..86)86	P9(618..635)=FP P9(697..710) P9(1328..1341)	P9(26..41)=TP P12(0..86) P17(9..82)
					3'(1566..1637)72	FN	FN
14	5HSA014201	A28106	1408	27. .1388	Nil	-	-
15	5HSA014202	A28108	1866	225. .1649	5'(1..224)24	FN	P12(0..101)=TP P17(35..220)
					3'(1650..1866)217	FN	P18(1751..1830)=TP
16	5HSA014204	A32135	1452	58. .1425	5'(1..57)57	P9(447..466)=FP P9(1262..1275) P9(1284..1303)	P12(0..57)=TP P17(8..20)
					3'(1426..1452)27	FN	FN
17	5HSA014209	A35395	2296	70. .1365	5'(1..69)69	P9(54..72)=TP	P12(0..69)=TP P!7(14..51)
					3'(1366..2296)931	P12(2218..2296)=TP P9(360..378)=FP P9(537.556) P9(965..978)	P18(1526..1605)=TP
18	5HSA017632	A75732	2150	1810. .1983	5'(1..1809)1809	P9(443..457)=TP P9(988..1003)=TP	P9(606..618)=TP P12(0..104) P17(6..354) P18(582..661)

Contd

Table 4. Calibration of UTRScan and UlTRaSCAN using human entries—Contd

No.	UTR Acc. No.	EMBL Acc. No	EMBL Acc. No	CDS	UTR Type & Length	UTRScan Results	UlTRaSCAN Results
19	5HSA017638	A92931	2616	57. .2426	5'(1..56)56	P9(2531..2546)	P12(0..56)=TP P9(2515..2530) P17(2433..2607)
20	5HSA017639 3HSA020007	A94121	2783	361. .2652	5'(1..360)360	P9(347..360)=TP	P12(0..104)=TP P17(8..379)
					3'(2650..2783)134	FN	P18(2693..2772)=TP
21	5HSA017641 3HSA020009	A94721	2180	74. .1015	5'(1..73)73 3'(1016..2180)1165	FN FN	P12(0..73)=TP P2(1618..1651)=TP P3(1963..2065) P9(1305..1317) P17(1017..2149) P20(2016..2026)
22	5HSA024617	AX033848	3457	403. .2925	5'(1..402)402	FN	P2(296..329)=TP P9(2987..2999) P12(0..101) P17(16..401)
23	5HSA024619	AX039604	10682	187. .5637	5'(1..186)186	P9(3143..3162)=FP P9(3674..3691) P9(6198..6217) P9(9457..9472) P9(10309..10325) P9(4666..4726) P7(9091..10682)=TP	P12(4..105)=TP P17(8..185) P2(7700..7733) P3(8114..8213) P5(9091..10683) P9(5753..5765) P20(8653..8663)
24	5HSA024622	AX045251	4077	564. .3395	5'(1..844)844	P14(766..773)=FP	P3(87..193)=TP P12(0..101) P17(11..527)
25	5HSA024624	AX054697	4077	564. .3395	5'(1..563)563	FN	P2(517..550)=TP P9(268..280) P12(0..102) P17(5..544) P18(3844..3850)
26	5HSA024625	AX063465	1879	30. .1121	5'(1..29)29	P9(964..979)=FP	P2(0..28)=TP P17(5..21) P3(1343..1437) P9(1813..1825) P12(1154..1260) P18(1484..1563)
27	5HSA024628	AX067150	6782	113. .6547	5'(1..112)112	P9(143..159)=FP P9(842..856) P9(2007..2025) P9(5853..5866)	P12(0..105)=TP P17(3..110) P18(6682..6761)
28	5HSA024633	AX067158	515	1. .384	5'(1..384)384	P12(419..515)=TP	P12(385..487) P17(386..504)
29	5HSA024634	AX067159	304	174..>304	5'(1..173)173	P12(208..304)=FP	P12(0..101) P17(11..126)=TP
30	5HSA024635	AX067160	473	343..473	5'(1..342)342	P12(377..473)=FP	P3(19..115)=TP P12(0..106) P17(11..295)

For legends refer Table 2

(Table 7), i.e., P12 is present in all the 20 mRNAs except β-tubulin from *O. volvulus* (but it contains P17) and P17 is present in all the 20 mRNAs.

Thus, an almost 100% occurrence of these two patterns in all eukaryotic mRNAs either in combination or in single suggests the possibility that these patterns may be playing some critical role in eukaryotic mRNAs. As these additional regulatory patterns viz., additional ribosome entry site(s) and additional uORF(s) on the same mRNA may likely to serve as 'identification markers', which could possibly involve in the transport of eukaryotic mRNAs from the nucleus to cytoplasm using ribosome or ribosomal subunits, as the site of synthesis of both ribosomes and mRNAs is the nucleus. Under nuclear environment the mRNAs may bind and wrap around the ribosome or ribosomal subunits using these IRES and uORF sequences and get co-transported into the cytoplasm. Under cytoplasmic environment the regular ribosome binding site and ORF is used for

Table 5— Comparative analysis of common patterns in fungi, plant and human mRNA sequences by UlTRASCAN

Pattern	Fungi			Human			Plant		
	Total	**5'**	**3'**	**Total**	**5'**	**3'**	**Total**	**5'**	**3'**
P1 HISTONE3	0	0	0	0	0	0	0	0	0
P2 IRE	**4**	**1**	**3**	**8**	**5**	**3**	**1**	**0**	**1**
P3 SECIS	**13**	**6**	**7**	**7**	**4**	**3**	**0**	**0**	**0**
P4 APP	0	0	0	0	0	0	0	0	0
P5 CPE	0	0	0	1	1	0	0	0	0
P6 TGE	0	0	0	0	0	0	0	0	0
P7 15-LOX-DICE	1	0	1	0	0	0	0	0	0
P8 ARE2	0	0	0	0	0	0	0	0	0
P9 TOP	**9**	**3**	**6**	**11**	**8**	**3**	**3**	**1**	**2**
P10 GLUT1 RNA	0	0	0	0	0	0	0	0	0
P11 TNF	0	0	0	0	0	0	2	0	2
P12 IRES	**30**	**20**	**10**	**27**	**25**	**2**	**23**	**14**	**9**
P13 MSL2-5UTR	0	0	0	0	0	0	0	0	0
P14 MSL2-3UTR	0	0	0	0	0	0	0	0	0
P15 RPMS12_TCE	0	0	0	0	0	0	0	0	0
P16 ADH_DRE	1	1	0	0	0	0	0	0	0
P17 Upstream ORF	**28**	**17**	**11**	**28**	**23**	**5**	**28**	**15**	**13**
P18 NANOS_TCE	**11**	**2**	**9**	**9**	**1**	**8**	**8**	**0**	**8**
P19 VIMENTIN	0	0	0	0	0	0	0	0	0
P20 BRE	1	0	1	2	1	1	1	0	1

Table 6— Comparative analysis on the UTR patterns from fungi, plant and human mRNA sequences by UlTRaSACN and UTRScan

Pattern	UlTRaSCAN			UTR Scan			UlTRaSCAN			UTR Scan			UlTRaSCAN			UTR Scan		
	T	5'	3'	T	5'	3'	T	5'	3'	T	5'	3'	T	5'	3'	T	5'	3'
P1 HISTONE3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P2 IRE	4	1	3	0	0	0	8	5	3	0	0	0	1	0	1	0	0	0
P3 SECIS	13	6	7	1	1	0	7	4	3	0	0	0	0	0	0	2	0	2
P4 APP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P5 CPE	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
P6 TGE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P7 15-LOX-DICE	1	0	1	3	1	2	0	0	0	0	0	0	0	0	0	1	1	0
P8 ARE2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P9 TOP	9	3	6	3	2	1	11	8	3	9	6	3	3	1	2	4	4	0
P10 GLUT1 RNA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P11 TNF	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	0	0	0
P12 IRES 30	20	10	6	0	6	27	25	2	6	1	5	23	14	9	19	1 18		
P13 MSL2-5UTR	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	0	0	0
P14 MSL2-3UTR	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	0	0	0
P15 RPMS12_TCE	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	0	0	0
P16 ADH_DRE	1	1	0	0	0	0	0	0	0	0	0	0	1	0	2	0	0	0
P17 Upstream ORF	28	17	11	0	0	0	28	2 3	5	0	0	0	28	15	13	0	0	0
P18 NANOS_TCE	11	2	9	0	0	0	9	1	8	0	0	0	8	0	8	0	0	0
P19 VIMENTIN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P20 BRE	1	0	1	0	0	0	2	1	1	0	0	0	1	0	1	0	0	0

translation of the mRNAs. Secondly, these motifs may also serve as markers for mRNA degradation pathways (possibly by an exosome-mediated pathway). That is, once the function of a particular mRNA species is over and no longer required in the cell, the exosome/ribonucleases may possibly recognize and bind IRES and/or uORF region(s) for cleaving the mRNAs leading to further complete degradation in the cytoplasm. It should be noted that the degradation of rRNAs in the cytoplasm is also exosome mediated.

Table 7— Analysis of UTRs of some well-known proteins using UlTRaSCAN

No.	GenEMBL Acc. No.	Length (bp)	CDS	UlTRaSCAN Results
1	AB003287	1952	115..1458	P12(1..101)
	Bombyx mori mRNA			P17(50..104)
	(β-tubulin)			P3(1536..1569)
				P3(1585..1686)
				P16(1699..1707)
				P18(1929..1935)
2	AB011069	1943	93..1436	P12(0..92)
	Bombyx mori mRNA			P17(24..86)
	(β-tubulin)			P5(1763..1944)
				P9(1881..1893)
				P16(1771..1779)
				P18(1531..1537)
3	AB035080	1196	64. .816	P12(0..63)
	Canis familiaris mRNA			P17(29..61)
	(β-casein)			
4	AF019885	1420	37..1383	
	Onchocerca volvulus mRNA			P17(1392..1408)
	(β-tubulin)			
5	AF082877	1827	1. .1365	
	Cryptosporidiumparvum mRNA			P12(1366..1467)
	(α-tubulin)			P17(1373..1818)
6	AF128397	1097	35. .634	P17(13..29)
	Trichosurus vulpecula mRNA			
	(α-casein)			P3(743..849)
				P9(653..665)
				P12(635..737)
7	AF128398	1278	34. .891	
	Trichosurus vulpecula			P9(989.1001)
	(β-casein)			P12(892..995)
				P17(906..1238)
8	J00895	9206	2996..3163,	P2(1318..1351)
	Gallus gallus,		3415..3465,	P3(2748..2839)
	Ovalbumin		4047..4175,	P5(1616..2644)
			4576..4693,	P12(1..101)
			5652..5794,	P17(1..2978)
			6126..6281,	P18(253..259)
			7864. .8259	P20(514..524)
9	J00922	8372	3435..3602,	P2(302..335)
	Gallus gallus,		4148..4198,	P3(754..854)
	Ovalbumin		4554..4682,	P5(842..3329)
			5530..5647,	P9(969..981)
			5869..6011,	P12(1..101)
			6093..6248,	P17(23..3420)
			7125..7526	P20(21..31)
				P14(7790..7865)

contd

Table 7— Analysis of UTRs of some well-known proteins using UlTRaSCAN—Contd.

No.	GenEMBL Acc. No.	GenEMBL Acc. No.	CDS	UlTRaSCAN Results
10	M16339 *Entamoeba histolytica* mRNA (actin)	1534	207. .1337	P3(11..104) P12(1..102) P17(20..204)
11	M16340 *Entamoeba histolytica* mRNA (actin)	1294	11. .1141	P12(1142..1243) P17(1164..1291)
12	M19146 *Plasmodium falciparum* mRNA (actin)	1334	131..1261	P12(1..104) P17(3..128) P18(35..41)
13	M22718 *Plasmodium falciparum* mRNA (actin)	2513	392.839, 1209. .1891	P5(69..327) P12(1..101) P17(5..391) P18(12..18) P20(284..294) P9(2040..2055)
14	M32364 *H. attenuata* mRNA (actin)	2536	262..516, 999. .1874	P12(1..101) P17(67..256) P18(1941..1947)
15	M33123 Bovine mRNA (α-s1-casein)	1123	64. .708	P12(1..63) P17(718..1121)
16	M55158 Bovine, β-casein	10338	3714..3764, 4489..4515, 4628..4654, 6550. .6573, 6666..6707, 8028..8525, 9127. .9132	P2(377..410) P3(540..627) P9(741..756) P12(1..101) P17(23..3671) P18(521..527)
17	X02009 Chicken mRNA (ovotransferrin - conalbumin)	2376	77. .2194	P12(1..76) P17(2214..2364)
18	X16482 Sheep mRNA (β-casein)	1088	61. .729	P3(956..1056) P12(730..831) P17(141..1085)
19	X53964 Japanese quail mRNA (ovalbumin)	1879	54. .1205	P12(1..53) P17(6..20) P2(1375..1408) P3(1408..1485) P20(1745..1755)
20	Z38129 *L. polyphemus* mRNA (actin)	1716	56. .1186	P12(1..55) P17(6..51) P5(1409..1717) P18(15687..1573)

For legends refer Table 2

The pattern P2, an iron responsive element (IRE) is found both towards 5′ and 3′ UTRs in human and fungal entries. In human mRNAs it is localized more on the 5′ UTR than on 3′ UTR, whereas it is *vice versa* in fungi. However, only 1/30 was found in plants that too at 3′ UTR. These results suggest that the animal and fungal mRNAs possess more iron responsive elements than plant mRNAs.

The next dominant pattern P3, a selenocysteine insertion sequence (SECIS), is most prevalent in fungi and almost equally distributed on 5′ and 3′ UTRs. In human there are only 7 out of 30 sequences analyzed and again almost equally distributed at 5′ and 3′ UTRs. Interestingly, no SECIS pattern is found in plant mRNAs corroborating the experimental evidence that only very few proteins are selenoproteins in plants as compared to microbes and animals.

The next dominant pattern was P9, an oligopyrimidines tract (TOP). It is required for the translational control of mRNAs of ribosomal proteins in eukaryotes. The TOP pattern was more on 5′ UTRs on human (8/11) but towards 3′ UTRs in fungi (6/9). Only very few (3/30) were detected in plants. (However, the UTRScan detected 4 in the 5' UTR).

The next important pattern was P12, an internal ribosome entry site (IRES). This is an alternative mechanism of translation initiation to the conventional 5′ cap dependent ribosome scanning mechanism. IRES is widely distributed in fungi, plants and human and as expected usually widespread at 5′ UTRs and its significance is discussed elsewhere.

The next dominant pattern found was P17, an uORF sequence, which is also widely distributed like IRES and its significance is discussed elsewhere.

The last important pattern, P18, was NANOS_TCE, a translational control element, which consists of a 90-nucleotide region located in 3′ UTR of Nanos mRNAs, which is able to fold into a bipartite secondary structure. As predicted, this was found mostly in 3′ UTRs of fungal, human and plant mRNAs.

The patterns P19 and P20 were rare and appear to be mRNA-specific.

Merits of UlTRaSCAN

From the above discussion, it is clear that UlTRaSCAN is a better algorithm for detection of UTR patterns in eukaryotic mRNAs. Since the 5' and 3' UTRs are extracted and then fed into the pattern-searching algorithm, all the false positives are almost completely eliminated**.** The sensitivity is more than doubled and the specificity is increased close to 100%. The UlTRaSCAN also minimized the false positives to almost to 0% and the false negatives to large extent. Furthermore, UlTRaSCAN shows the type, length and the sequence of the pattern and the range of the pattern (start and end positions) that is detected in the query sequence. In a modified recent version of the algorithm (which is also available on the website) the number of repeats of a particular pattern in the query mRNA is also predicted.

Limitations and Exceptions

UlTRaSCAN is useful only for finding the 5' and 3' UTR patterns in the gene sequences that are annotated with the CDS information and therefore, automatic annotation of anonymous sequences generated by sequencing projects is not dependable.

Conclusions

The present study was aimed at developing a new and effective algorithm to analyze the UTRs in eukaryotic mRNAs. UlTRaSCAN nearly completely eliminated all the false positives and proved to have higher specificity than the UTRScan. So UlTRaSCAN is recommended to be a highly efficient algorithm with ~100% specificity and independent of the source of the eukaryotic system.

Availability: *The algorithm is available on* ***http://gene.tn.nic.in/~raje/***
*Supplementary Materials****: Source code of UlTRaSCAN: http://gene.tn.nic.in/~raje/ultra/toolkit***

Acknowledgement

The authors wish to thank Prof. S Krishnaswamy and Mr N Jeyakumar of Bioinformatics Center, Madurai Kamaraj University, for critical evaluation and suggestions on the manuscript and for the help in the PERLSCRIPT, respectively.

References

1 Decker C J & Parker R, Mechanisms of mRNA degradation in eukaryotes, *Trends Biochem Sci*, 19 (1994) 336-340.
2 Sonenberg N, mRNA translation: Influence of the 5′ and 3′ untranslated regions, *Curr Opin Gen Dev,* 4 (1994) 310-315.
3 Pesole G, Liuni S, Grillo G & Saccone C, Structural and compositional features of untranslated regions of eukaryotic mRNAs, *Gene,* 205 (1997) 95-102.
4 Pesole G, Liuni S, Grillo G, Licciulli F, Larizza A *et al*, UTRdb and UTRsite: Specialized databases of sequences and functional elements of 5′ and 3′ untranslated regions of eukaryotic mRNAs, *Nucleic Acids Res*, 18 (2000) 193-196.

5 Overbeek R, Larsen N & D'souza M, Searching for patterns in genomic data, *Trends Genet*, 13 (1997) 497-498.
6 Dalphin M E, Stockwell P A, Tate W P & Brown C M, TransTerm, the translational signal database extended to include full coding sequences and untranslated regions, *Nucleic Acids Res,* 27 (1999) 293-294.
7 Conklin D, Jonassen I, Aasland R & Taylor W R, Association of nucleotide patterns with gene function classes: Application to human 3′ untranslated sequences, *Bioinformatics*, 18 (2002) 182-189.
8 Tabaska J E, Davulri R V & Zhang M Q, Identifying the 3′-terminal exon in human DNA, *Bioinformatics*, 17 (2001) 602-607.
9 Kochetov A V, Ponmarenko M P, Frolov A S, Kisselev L L & Kolchanov N A, Prediction of eukaryotic mRNA translational properties, *Bioinformatics,* 15 (1999) 704-712.