RNA Biology

ISSN: 1547-6286 (Print) 1555-8584 (Online) Journal homepage: http://www.tandfonline.com/loi/krnb20

# Two covariance models for iron-responsive elements

Stewart G Stevens, Paul P. Gardner & Chris Brown

**To cite this article:** Stewart G Stevens, Paul P. Gardner & Chris Brown (2011) Two covariance models for iron-responsive elements, RNA Biology, 8:5, 792-801, DOI: 10.4161/rna.8.5.16037

**To link to this article:** http://dx.doi.org/10.4161/rna.8.5.16037

RNA Biology 8:5, 792-801; September/October 2011; © 2011 Landes Bioscience

# Two covariance models for iron-responsive elements

Stewart G. Stevens,[1] Paul P. Gardner[2] and Chris M. Brown[1,*]

[1]Biochemistry and Genetics Otago; University of Otago; Dunedin, New Zealand; [2]Wellcome Trust Sanger Institute; Wellcome Trust Genome Campus; Hinxton, UK

**Key words:** IRE, covariance, RFAM, cis-acting, iron, regulation, DSTN, MGAT4A, VHL, ENPEP

**Abbreviations:** IRE, iron-responsive element; IRP, iron regulatory protein; UTR, untranslated region

Iron-responsive elements (IREs) function in the 5' or 3' untranslated regions (UTRs) of mRNAs as post-transcriptional structured cis-acting RNA regulatory elements. One known functional mechanism is the binding of iron regulatory proteins (IRPs) to 5' UTR IREs, reducing translation rates at low iron levels. Another known mechanism is IRPs binding to 3' UTR IREs in other mRNAs, increasing RNA stability. Experimentally proven elements are quite small, have some diversity of sequence and structure, and functional genes have similar pseudogenes in the genome. This paper presents two new IRE covariance models, comprising a new IRE clan in the RFAM database to encompass this variation without over-generalisation. Two IRE models rather than a single model is consistent with experimentally proven structures and predictions. All of the IREs with experimental support are modeled. These two new models show a marked increase in the sensitivity and specificity in detection of known iron-responsive elements and ability to predict novel IREs.





## Introduction

The cis-acting iron-responsive element (IRE) was first discovered in human ferritin mRNA (*FTH1*).[1] Since then, IREs have been found in both the 5' and 3' UTRs of diverse mRNAs over a wide phylogenetic range—mainly eukaryotic animals and also some prokaryotes, but not plants.[2] Iron regulatory proteins 1 and 2 (IRP1 and IRP2) bind to IREs in low iron conditions. In high iron conditions IRP1 binds to iron complexes, adopting conformations unsuitable for IRE binding. IRP2 is degraded in high iron conditions, making it unavailable for binding.[3] IRP binding to an IRE in the ferritin mRNA 5' UTR inhibits ferritin translation. Multiple IRPs bind to five IREs in the 3' UTR of the human transferrin receptor (*TFRC*), stabilizing these transcripts by blocking an endonucleolytic cleavage.[4] Thus, IREs represent a classic paradigm by which RNA regulatory elements can mediate the translation rate of specific mRNAs. This is required to provide a rapid response to iron, which is essential, but potentially toxic.[5]

Before experimental RNA structure data were available, predicted IRE secondary structures had invariably shown a 6 base apical hairpin loop, five based upper stem, bulged C8 and variable lower stem. It is the apical loop and mid-stem C8 bulge that are critical for IRE function and this has become a canonical model used by RFAM[6] and other databases.[7,8] However the data from an NMR study by Addess[9] and a crystal structure of IRP1 bound to a ferritin IRE by Walden[10] show that the C14 and G18 bases in the apical hairpin loop are in fact paired, producing a tri-loop (A15, G16, U17) and that two bases (a C8 and U6) are bulged from the hairpin stem. These structures[9,10] did confirm the predicted secondary structure of an A-form helical stem, interrupted by a mid-stem C8 bulge, with an apical loop that presents conserved bases interacting with IRE binding proteins.

Predicted secondary structures suggested ferritin IREs might have an additional U6 bulge in the lower stem, this was confirmed by the NMR and crystal structure data. However, IREs found in transferrin receptor mRNAs (as well as IREs in *SLC40A1*, *SDHB*, *ACO2*, *SLC11A2*) lack the bulged U6 and this region is predicted to be paired. Thus the dual mid-stem bulge distinguishes the ferritin IREs from other IREs. This division is not novel[11]—those with the additional U6 bulge would correspond to the UGC (and variants) class of Picinnelli and Samuelsson 2007.[12] Most IREs predicted secondary structures conform better to the single bulge structure (IRE Family 1). Therefore, this new RFAM IRE clan divides the known IREs into two structural families—with and without the lower bulged U6.

In addition to the positional difference between IREs located in the 5' and 3' UTR there is notable heterogeneity in regulatory mechanisms and effects. For example: (1) The IRE in the 3' UTR of *SLC11A2* was shown to have a higher affinity for IRP1 than IRP2, in contrast to the *FTH1* IRE which was shown to have similar affinities.[13] (2) Both *TFRC* and *CDC42BPA* have IREs in their 3' UTR conferring RNA stability, yet in low iron conditions the mRNA of *CDC42BPA* showed greater stability than *TFRC*.[14] (3) Regulation of splicing could have a direct impact on post-transcriptional regulation for at least two IREs

*Correspondence to: Chris M. Brown; Email: chris.brown@otago.ac.nz
Submitted: 01/31/11; Revised: 03/27/11; Accepted: 04/01/11
DOI: 10.4161/rna.8.5.16037

**Table 1.** List of known IRE-containing mRNAs used to build the covariance models. mRNAs are listed in chronologic order of discovery.

| Gene | Species | Alternative name | Name | Location | Ref | Date |
|---|---|---|---|---|---|---|
| ***FTH1*** | *Homo sapiens* | *FHC; FTH; PLIF; FTHL6; PIG15; MGC104426; FTH1* | ferritin, heavy polypeptide 1 | 5' UTR | 1 | 1987 |
| ***FTL*** | *Homo sapiens* | *NBIA3; MGC71996; FTL* | ferritin, light polypeptide | 5' UTR | 25 | 1987 |
| ***TFRC*** | *Homo sapiens* | *TFR; CD71; TFR1; TRFR; TFRC* | transferrin receptor (p90, CD71) | 3' UTR | 3 | 1989 |
| ***ALAS2*** | *Homo sapiens* | *ASB; ANH1; XLSA; ALASE; XLDPP; ALAS-E; FLJ93603; ALAS2* | aminolevulinate, delta-, synthase 2 | 5' UTR | 26 | 1991 |
| ***SdhB*** | *Drosophila melanogaster* | *CG3283; Dmel\CG3283; Ip; SDH; SDH-Ip; SDH-IP; sdhB; SDHb* | succinate dehydrogenase complex, subunit B, iron sulfur (Ip) | 5' UTR | 14 | 1995 |
| ***ACO2*** | *Bos taurus* | | aconitase 2, mitochondrial | 5' UTR | 27 | 1996 |
| ***Ferritin*** | *Pacifastacus leniusculus* | | | 5' UTR | 28 | 1999 |
| ***Hao1*** | *Mus musculus* | *GOX; Gox1; Hao-1; MGC141211; Hao1* | Hao1 hydroxyacid oxidase 1, liver | 3' UTR | 29 | 1999 |
| ***qoxD*** | *Bacillus subtilis* | | cytochrome aa 3-600 quinol oxidase (subunit IV) | 3' UTR | 30 | 1999 |
| ***SLC11A2*** | *Homo sapiens* | *DCT1; DMT1; NRAMP2; FLJ37416; SLC11A2* | solute carrier family 11 (proton-coupled divalent metal ion transporters), member 2 | 3' UTR | 11 | 2001 |
| ***Ferritin*** | *Manduca sexta* | | *Manduca sexta* ferritin heavy chain-like protein precursor | 5' UTR | 31 | 2001 |
| ***NDUFS1*** | *Homo sapiens* | *CI-75k; CI-75Kd; PRO1304; MGC26839; NDUFS1* | NADH dehydrogenase (ubiquinone) Fe-S protein 1, 75 kDa (NADH-coenzyme Q reductase) | 5' UTR | 32 | 2001 |
| ***Ferritin*** | *Calpodes ethlius* | | Calpodes ethlius fat body secreted ferritin S subunit precursor. | 5' UTR | 33 | 2002 |
| ***Slc40a1*** | *Mus musculus* | *MTP; Ol5; Pcm; Dusg; Fpn1; MTP1; IREG1; Slc11a3; Slc39a1; Slc40a1* | solute carrier family 40 (iron-regulated transporter), member 1 | 5' UTR | 34 | 2003 |
| ***alas2*** | *Danio rerio* | *sau; alas-e; cb1063; sauternes; alas2* | aminolevulinate, delta-, synthase 2 | 5' UTR | 35 | 2005 |
| ***CDC42BPA*** | *Homo sapiens* | *MRCK; MRCKA; PK428; FLJ23347; KIAA0451; DKFZp686L1738; DKFZp686P1738; CDC42BPA* | CDC42 binding protein kinase alpha (DMPK-like) | 3' UTR | 12 | 2006 |
| ***CDC14A*** | *Homo sapiens* | *cdc14; hCDC14; CDC14A* | CDC14 cell division cycle 14 homolog A (*S. cerevisiae*) | 3' UTR | 13 | 2006 |
| ***EPAS1*** | *Homo sapiens* | *HLF; MOP2; ECYT4; HIF2A; PASD2; bHLHe73; EPAS1* | endothelial PAS domain protein 1 | 5' UTR | 15 | 2007 |



(in *CDC14A*[15] and *SLC11A2*[13]) affected by alternative splice variants—with some transcripts omitting the element.

The IRE has a regulatory role in several mRNAs involved in iron metabolism. While the depth of knowledge regarding these genes and their products is variable, they are clearly diverse. *FTH1* and *FTL* encode subunits of the iron storage complex, ferritin.[5] *TFRC* encodes a membrane receptor for transferrin, allowing cellular uptake of iron.[5] *SLC11A2* (*DMT1*) encodes a divalent-cation transporter, a membrane protein mediating iron uptake from the intestinal lumen.[16] *SLC40A1* (*IREG1*) encodes a membrane protein transporting iron in the duodenum to the circulation.[17] *ALAS2* encodes a synthase catalysing the first step of the heme biosynthesis pathway.[18] *SDHB* encodes a subunit of a Kreb's cycle enzyme required for electron transport to quinones.[19] *ACO2* encodes an isomerase catalysing the reversible isomerisation of citrate and iso-citrate.[20] *EPAS1* encodes a transcription factor involved in complex oxygen sensing pathways by the induction of oxygen-regulated genes under low oxygen conditions.[21] *CDC42BPA* encodes a kinase with a role in cytoskeletal reorganization.[22] *CDC14A* encodes a dual-specificity phosphatase implicated in cell cycle control[23] and also interacts with interphase centrosomes.[24] **Table 1** shows a complete list of known IREs with direct experimental evidence.

Most of the IREs characterized to date have initially been identified in mammalian mRNAs. For insects the first functional IRE was found in the 5' UTR of the *SDHB* mRNA of *Drosophila melanogaster*.[25] There is no evidence of this IRE in the *SDHB* mRNA for humans or other mammals. A previously published phylogenetic analysis of iron-responsive elements has shown that the IRE of *FTH1*/*FTL* occurs in a majority of metazoa.[12] Whereas, IRE like sequences in *ALAS2* and *ACO2* are present

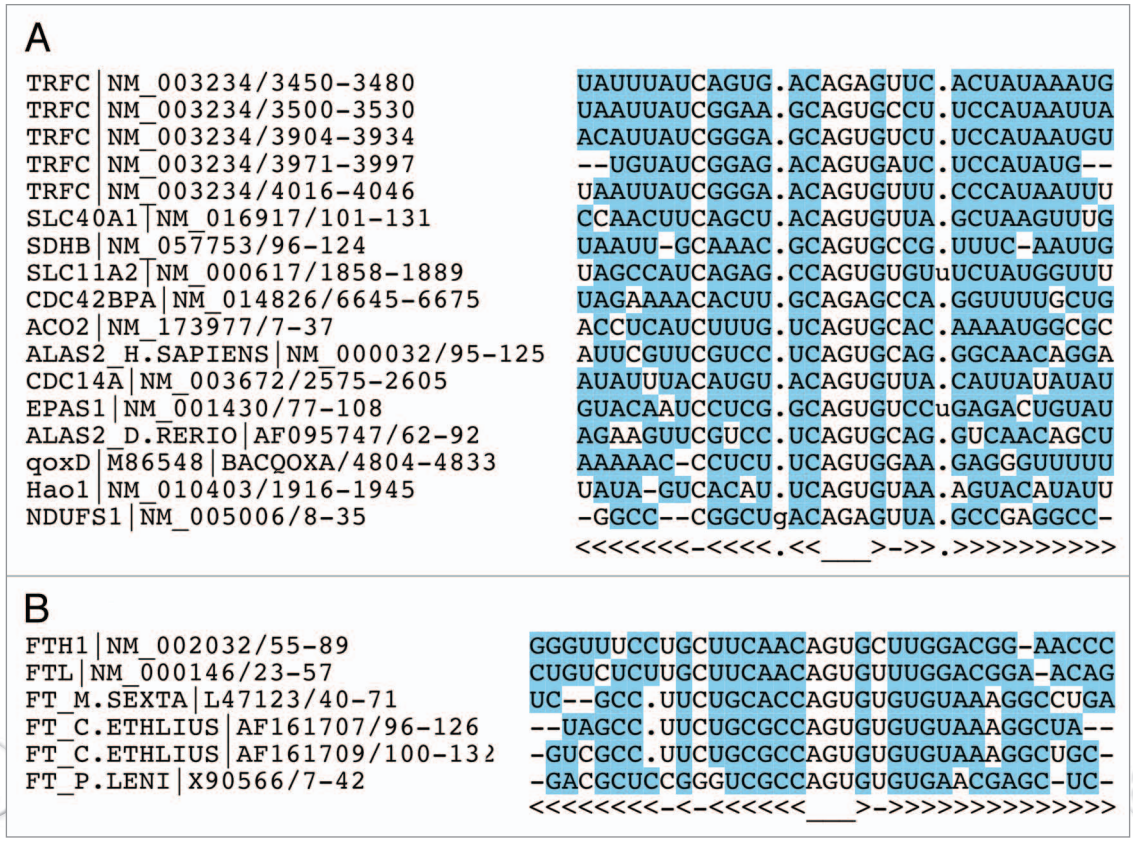


**Figure 1.** Structurally aligned IRE families. The secondary structure is shown below each family in WUSS notation and the sequences are highlighted to show base pairing conforming to the consensus structure. The identifier indicates the gene in which the sequence was found and the genbank locus with the position of the sequence. (A) IRE Family 1—with a C8 bulge between the upper and lower stems. (B) IRE Family 2—with an additional U6 bulge in the lower part of the helix.



in chordates, *SLC40A1* and *TFRC* contain IRE like sequences in vertebrates, and the IRE of *SLC11A2* (*DMT1*) is confined to mammals.[12]

Some IREs have unusual structures, for example the *EPAS1* IRE was found by immunoprecipitation and the authors of that study reported that this element could not be found by the then available in silico approaches.[26] The *EPAS1* IRE is predicted to have an additional U bulge in its upper stem and also an unpaired A opposite an unpaired C in the lower stem.

SIREs (Searching for IREs)[27] is the most recently developed web-accessible bioinformatic approach to predict IREs utilizing advanced regular expressions, and thermodynamic stability. It showed improved prediction ability on mRNA sequences, but has not been tested as a genomic search tool.

An alternative approach is to use the Infernal software package[28] to build a covariance model for this structured RNA element, as is done for many elements in the RFAM database. A covariance model is a stochastic context free grammar (SCFG) designed for modeling the consensus sequence and structure of RNAs. An SCFG provides a statistical method that scores not only nucleotide residues at single stranded positions but also base pairings, insertions and deletions given an alignment to a consensus secondary structure.[28] The resulting "bit score" is the log-odds ratio of the probability of the target matching the model to the probability of target matching random sequence. The RFAM database has primarily catalogued non-coding RNAs with the use of covariance models but also contains a growing group of cis-regulatory models such as the IRE.

The covariance model provided by RFAM has not been updated for some time and the current family does not include or detect many of the known elements documented here. These new models for the IRE keep this in silico approach current and are able to identify all the IREs with experimental support to date. The models' abilities to detect putative elements in human genomic sequences were assessed. Homologous genes from non-human species were searched in an assessment of the conservation of the new putative elements.

## Results

**Models. Figure 1** shows the structurally aligned experimentally supported IRE sequences divided into two families. These were used to build two covariance models designed to represent the diversity of IREs (**Table S1**).

A collection of mRNAs containing the known IREs were searched using the new models. The new models now find known IREs in the mRNAs that were missed by the old model-*SDHB*, *SLC11A2*, *ACO2*, *EPAS1*, *CDC42BPA*, *CDC14A*, *NDUFS1* and

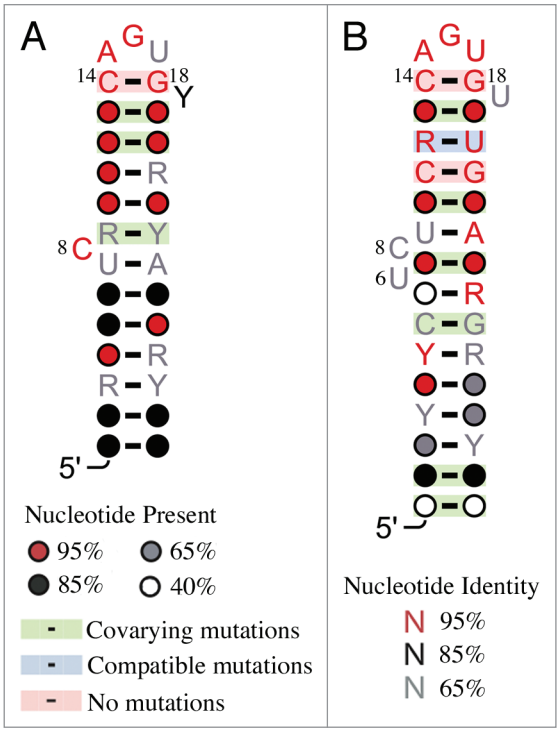


**Figure 2.** IRE Families sequence/structure. IUPAC codes corresponding to the seed sequences in each model are shown. Conserved bases and pairings are highlighted. (A) IRE Family 1. (B) IRE Family 2.



*SLC40A1* (**Table S2**). Conserved bases and pairings can be further visualized in **Figure 2**.

Reviewing the sequences as shown in **Figure 1** it is apparent that additional bulges to the consensus structures are more common in the lower part of the IRE stem. This can be further visualised in **Figure 2** where the most highly conserved pairs are in the upper part of the stem. Perfect conservation can be seen in the bases shown to interact with the iron-responsive protein in the crystal structure[10] whereas covariance consistent with maintaining structure can be seen in other bases.

The new models were used to search the entire human genome. The number of hits over a range of bit score thresholds were counted in 5' UTRs, 3' UTRs, coding regions and introns. There are several pseudogenes for some of the known IRE containing genes—e.g., ferritin.[40] The exons containing the established IREs were used in a blast search to identify regions where a likely explanation for a hit was a match against a possible pseudogene (**Table 2**).

The models were also used to search the RFAMSEQ10 database. This database includes nucleotide sequence from many species—all of the nucleotide transcript data for the EMBL species as well as data from whole genome shotguns and environmental sequence. Extraneous datasets such as ESTs and synthetic sequences are excluded.[41] In order to assess these hits, all the known protein sequences encoded by the mRNAs with experimentally established IREs were obtained and tblastn was used to search the RFAMSEQ10 database. **Table 3** shows where the new IRE Family hits were found in conjunction with these known protein matches. For IRE Family 1 at a bit score threshold of 19, 67% of hits were not closely associated with regions identified by the protein search (524/775). For IRE Family 2 at a threshold of 28 it was 75% (1,117/1,482).

The hits found by the models on the human genome may be retrieved from our companion site and visualised using the UCSC genome browser (mrna.otago.ac.nz/stevens2011a).

**Sensitivity and specificity.** The sensitivity of the new models to detect the experimentally established IREs was assessed (**Table 2**). IRE Family 1 requires a low bit score cut off of 19 to give 100% sensitivity with a bit score cut off of 25 providing a sensitivity of 93% (failing to detect only the experimentally supported *NDUFS1* IRE). For IRE Family 2 a bit score cut off of 28 gives 100% sensitivity while maintaining 100% specificity.

The specificity of the new models was assessed using similar criteria to the recently published SIREs.[27] Shuffled sequences of 150 randomly selected mRNAs were searched. SIREs reports specificity of between 91.3% and 99.3% depending on stringency using this method. We reproduced these results using the SIREs web server and our randomly selected mRNAs and found a similar 88.0% to 99.3% specificity. No hits were detected in the same random sequences using the new models published here (100% specificity at all bit scores over 10 according to this method).

Shuffled sequences for much of the human genome (2.8 Gb) were searched in order to better assess the specificity of the new covariance models. The number of hits found in random sequence in proportion to the size of genomic regions searched is shown in **Table 2**. For IRE Family 1 at the low bit score cut off of 19 (100% sensitivity to known IREs) there are 17 hits in 28 million bases of 3' UTR, 8 of which are known IREs. In shuffled sequence of the same size only 3.3 are detected on average. This indicates that the model is detecting much more than expected by chance.

The specificity of the results is better for IRE Family 2. At a bit score cut off of 28, both of the two known elements are matched whereas no chance hits are detected—even in the whole shuffled genome.

The number of IRE hits found by the models in the human genome overlapping repeat regions as predicted by RepeatMasker was determined. RepeatMasker identified repetitive sequences covering 48.8% of the whole human genome. A hit was deemed to overlap a repeat region if there was at least an 80% overlap. For IRE Family 1 at a bit score threshold of 19 there were 171 overlapping hits out of 1,020 total hits (16.8%). For the same family at a threshold of 25 there were only two overlapping hits out of 29 (8%). For IRE Family 2 at a threshold of 28 there were no overlapping hits out of 27 (0%).

**Refining hits.** It is desirable to focus on the predicted novel IREs for further investigation. To narrow down hits one approach is to combine other sources of information. The clearest criteria being that they would be in UTRs—though this annotation is not always available. In the results shown in **Table 2**, gene and UTR annotation from the Refseq genes at UCSC were used. In addition to the known IRE containing genes, the new models predicted IREs in several other genes. These are documented with a brief description in **Table 4**. Gene ontology analysis shows *VHL* in "the response to oxygen levels" category along with the known

**Table 2.** IRE hits in human genome (hg18, UCSC) using the new RFAM IRE models

| Family | Score | mRNA 3' UTRs | mRNA 5' UTRs | mRNA CDS | Pseudo | Exon ∪ Pseudo | Intergenic | Introns | Sensitivity | Specificity |
|---|---|---|---|---|---|---|---|---|---|---|
| Base coverage | | 28,130,973 | 8,316,280 | 33,733,356 | 26,763 | 68,275,928 | 5,017,407,839 | 1,129,670,779 | | |
| IRE 1 | 19 | 17 (3.3) | 7 (1.0) | 6 (4.0) | 14 (0.0) | 30 (8.1) | 794 (593.4) | 203 (133.6) | 100 | 100 |
| IRE 1 | 25 | 8 (0.1) | 5 (0.0) | 1 (0.1) | 13 (0.0) | 14 (0.1) | 13 (8.8) | 2 (2.0) | 93 | 100 |
| IRE 2 | 28 | 1 (0.0) | 2 (0.0) | 0 (0.0) | 27 (0.0) | 27 (0.0) | 0 (0.0) | 0 (0.0) | 100 | 100 |

The number of hits found at different bit score thresholds within Refseq annotated regions for 5' UTRs, 3' UTRs, coding regions and introns are shown. Pseudo refers to the regions identified by blast search with IRE containing exons as described in the Methods. Hits outside annotated genes and possible pseudogenes are recorded as intergenic. Hits found in either exons or possible pseudogenes are shown under "Exons U Pseudo." Some hits overlap annotation boundaries and so are counted in multiple columns. Sensitivity is calculated based on the ability to detect experimentally established IREs. Specificity is calculated using the SIREs method to allow comparison. Figures in brackets denote the number of hits expected in that region by searching random sequence of the same size. See **Supplemental S3** for full results over a wide range of scoring thresholds. See **Supplemental S4** for more detailed information regarding the hits within exons described in this table.





IRE containing genes, *SLC11A2*; *ALAS2*; *EPAS1* and *TFRC*. This analysis also shows that both *VHL* and *ENPEP* have biological function in blood vessel development along with the known IRE containing gene *EPAS1*.[42]

A complementary approach is to consider evolutionary conservation. Homologene (an NCBI database) and Ensembl were used initially to find the homologues of human genes with IRE hits—these were searched using the new covariance models (**Table 5**). Two of the novel candidates, *DSTN* and *MGAT4A*, had IREs predicted by the new RFAM covariance models with matches in several other species.

The SIREs web service was also used to search for IREs in these homologous sets. For 85 of the transcripts both SIREs and the covariance models presented here had a hit. In 25 transcripts only SIREs had a hit, and in 26 transcripts only the covariance models presented here had a hit.

We compared the conservation of the human IRE hits within Refseq annotated UTRs to all the human UTRs using phyloP.[54] Both the IRE Family 1 hits (bit score >19) and the IRE Family 2 hits (bit score >28) showed phyloP scores significantly higher than the UTRs (p-value $<1 \times 10^{-6}$). The IRE Family 1 hits were however significantly less conserved than all the experimentally demonstrated IREs (p-value $<1 \times 10^{-6}$). As expected IRE Family 2 hits had similar conservation to all the experimentally established IREs. (Mean phyloP scores for IRE Family 1 = 1.42; IRE Family 2 = 3.43; Experimentally supported IREs = 3.11; UTRs = 0.52).

**Table 3.** IRE hits in RFAMSEQ10 using the new RFAM IRE models

| Model | Bit score threshold | Matched to known protein | No match to known protein | Total number of hits |
|---|---|---|---|---|
| IRE Family 1 | 19 | 251 | 524 | 775 |
| IRE Family 1 | 25 | 226 | 433 | 659 |
| IRE Family 2 | 28 | 365 | 1117 | 1482 |

The proportion of hits occurring near regions that encode a protein encoded by an mRNA having an IRE. The regions identified were extended by 2 kb to include potential UTRs (see Methods). See **Supplemental S5** for a wider range of scoring thresholds.

## Discussion

**Models.** The results show these new covariance models have the ability to detect a wider range of experimentally demonstrated IREs than the previous single model. The new families include in their primary seeds a comprehensive collection of experimentally established IRE sequences. The approach of using a covariance model is quite different from the regular expression type searches of SIREs and earlier published methods. Although it is computationally more intensive this search was able to detect experimentally established elements in mRNAs such as *NDUFS1* missed by the SIREs searches.

Some IRE like elements were not included in either of the models. This was primarily due to insufficient evidence for IRP binding, but also in one case because of a highly divergent sequence. A sequence in the α-hemoglobin stabilizing protein resembles an IRE[55] but while a mutagenesis study shows regulatory involvement in high iron conditions, the study did not clearly show IRE/IRP binding. Furthermore, the sequence has a mid-stem bulged A rather than the bulged C of all other characterised human IREs. Another sequence in *Trichomonas vaginalis*[56] seems to be following an IRE-IRP like interaction and yet it shows little in common with the well-known IREs both structurally and at a sequence level. Sequences such as these cannot be included while maintaining the specificity of the models. An IRE in the ferritin mRNA of *Lampetra fluviatilis*[57] could not be included in the seed as there was no sequence record in Genbank.

**Table 4.** Genes with putative IREs identified using the new RFAM models

| | Gene | Description | Ref |
|---|---|---|---|
| 3' UTR | *HMGB1* | Transcription factor that also acts as a cytokine in immune response. | 42 |
| | *HAS3* | Transmembrane protein that synthesises hyaluronan—a polysaccharide involved in connective tissue and in cell behaviour during embryonic development and inflammation. | 43 |
| | *MGAT4A* | Glycosyltransferase with several isoforms and potential physiological roles. | 44 |
| | *VHL* | Protein targeting the hypoxia-inducible transcription factor for polyubiquitination and degradation. | 45 |
| | *LHFPL4* | Product that appears to be a tetraspan transmembrane protein. It is not well studied but does have homology with TMHS which is reported to have a role in the morphogenesis of cilia in the ear. | 46 |
| | *PLEKHA8* | Adaptor protein that transfers glucosylceramide to specific sites for the synthesis of glycosphingolipids. | 47 |
| | *FKTN* | Transmembrane protein localized to the Golgi complex with a role in glycosylation. | 48 |
| CDS | *TMEM202* | Transmembrane protein (inferred from protein domains via Uniprot) but has no known function. | |
| | *Tom1L1* | Adaptor protein thought to be involved in mitogenic signaling. | 49 |
| | *DSTN* | Protein with a functional role in actin depolymerization and filament dynamics. | 50 |
| | *ENPEP* | Protease acting in the metabolism of angiotensin. | 51 |
| | *SEC63* | Protein forming part of a transmembrane complex. This complex transports other proteins across the membrane of the endoplasmic recticulum. | 52 |
| | *FTH1P3, GUCY2GP and MYH16* | Refseq annotated pseudogenes | |
| | *LOC145783* | Refseq annotated hypothetical non-coding RNA | |

Each gene's name and the position of the IRE according to Refseq annotation are shown. A brief description of the gene's function is given.





A clear result from this study is that due to sequence and structural similarity it is impossible to distinguish good matches in pseudogenes (e.g., *FTH1*, *FTL*) from true positives. In well-annotated genomes hits in pseudogenes are more readily identified. Even for these cases pseudogenes that consist largely of UTR fragments cannot be identified as such using protein based similarity searches. In databases broad enough to include large amounts of unannotated sequence (e.g., RFAMSEQ10) it is particularly difficult to distinguish hits in genes or pseudogenes from hits in intergenic regions. Matches in the IRE families' full alignments that are semi-automatically generated from RFAMSEQ10 undoubtedly contain hits that will correspond to pseudogenes and intergenic regions. For this reason when following an iterative approach to extend the primary seed it is probably best to limit this to sequences closely associated with genes homologous to those with experimental evidence. This observation will apply to cis-acting elements other than the IRE.

Another point to note is that when dealing with large and developing datasets such as the annotation of the human genome there is likely to be the odd unusual feature. The IRE Family Model 2 results showed a high scoring hit in the 3' UTR—this turned out to be *FTH1P3*—a ferritin pseudogene with a mis-annotated 3' UTR. Also in the search of RFAMSEQ10 there was a hit in *Vaccinia virus GLV* that was due to the engineering of a human *TFRC* sequence into this viral vector.

**Sensitivity and specificity.** Inevitably lowering bit score cut offs yields more potential IRE candidates with the sacrifice of specificity. Simply limiting search results to gene exons, particularly UTRs, gives greater certainty that a hit is not due to random occurrence. We have calculated specificity according to the method used by SIREs in order to allow comparison.[27] According to this measure the models presented here offer improved specificity. These models also detect an experimentally proven IRE not matched by SIREs (in *NDUFS1*). The SIREs webserver does however detect more IREs in homologues of *CDC42BPA*. These methods may be used together in a complementary approach. Another useful direction may be to extend the IRE models presented here beyond the experimental data. This could be done by including currently weak scoring IREs from some of the homologues of genes known to contain IREs.

The hits produced by the models have been shown to match the human genome much more frequently than in random sequence. However, the models could also be matching similar, but as yet unidentified, enriched functional elements other than the IRE in the genome. The analysis with RepeatMasker showed that only a small fraction of the hits found by IRE Family 1 are in repeat regions and IRE Family 2 results do not coincide with repeat regions at all.

**Novel predicted IRE motifs conserved across species.** Our results from searching the RFAMSEQ10 database shows IRE hits in many species. These results are consistent with the study by Piccinelli and Samuelsson in not identifying IREs in taxa such as Acoelomata. This earlier work additionally identifies IREs for ferritin orthologues in diverse taxa such as Crustacea and Porifera. At the thresholds used here these new models do not identify these IREs. The models presented here could be employed for further evolutionary study. In an evolutionary analysis the IRE sequences searched may be targeted to known IRE containing genes using protein homology. A lower bit score threshold would be justified due to the smaller size of the search space and the homology information. The cautious addition of plausible seed sequences lacking experimental support could also be considered. We have started with a set of experimentally supported seed sequences and a higher more conservative bit score

**Table 5.** Screening for IREs in homologous genes (from Homologene and Ensembl) of the predicted human IRE-containing mRNAs by the new RFAM IRE models

| IRE Model | Location | Hit | Bit Score | Number of Species | Number of Species with Hits | | *Homo sapiens* | *Pan troglodytes* | *Gorilla gorilla* | *Macaca mulatta* | *Callithrix jacchus* | *Mus musculus* | *Rattus norvegicus* | *Bos taurus* | *Equus caballus* | *Canis lupus familiaris* | *Sus scrofa* | *Gallus gallus* | *Ciona intestinalis* | *Xenopus tropicalis* | *Danio rerio* | *Gasterosteus aculeatus* | *Drosophila melanogaster* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | RFAM | SIREs | | | | | | | | | | | | | | | | | |
| IRE 1 | 3'UTR | ***CDC14A*** | **28** | **11** | 2 | 3 | | †‡ | | | | | †‡ | | | | | | ‡ | | | | |
| | | ***CDC42BPA*** | **28** | **12** | 1 | 5 | †‡ | | | ‡ | | | | | | | ‡ | | | | ‡ | | ‡ |
| | | *FKTN* | **21** | **12** | 1 | 3 | †‡ | ‡ | ‡ | | | | | | | | | | | | | | |
| | | *HAS3* | **21** | **11** | 2 | 5 | †‡ | †‡ | ‡ | | ‡ | | | | | | | | | | ‡ | | |
| | | *HMGB1* | **19** | **12** | 3 | 0 | † | † | | † | | | | | | | | | | | | | |
| | | *LHFPL4* | **20** | **12** | 2 | 1 | † | † | | | | ‡ | | | | | | | | | | | |
| | | *MGAT4A* | **20** | **13** | 5 | 2 | †‡ | †‡ | | | † | | | | † | † | | | | | | | |
| | | *PLEKHA8* | **21** | **9** | 0 | 2 | | | | | ‡ | ‡ | | | | | | | | | | | |
| | | ***SLC11A2*** | **29** | **12** | 8 | 8 | †‡ | †‡ | | †‡ | †‡ | †‡ | †‡ | †‡ | | †‡ | | | | | | | |
| | | ***TFRC*** | **38** | **11** | 7 | 7 | †‡ | †‡ | †‡ | †‡ | †‡ | †‡ | | | | | | | | | †‡ | | |
| | | *VHL* | **19** | **10** | 2 | 3 | †‡ | | †‡ | | | ‡ | | | | | | | | | | | |
| | 5'UTR | ***ACO2*** | **28** | **12** | 8 | 8 | †‡ | | | †‡ | †‡ | †‡ | | †‡ | | †‡ | | †‡ | †‡ | | | | |
| | | ***ALAS2*** | **26** | **14** | 9 | 10 | †‡ | †‡ | †‡ | ‡ | †‡ | †‡ | | †‡ | | | †‡ | | | | †‡ | †‡ | |
| | | ***EPAS1*** | **28** | **12** | 4 | 5 | †‡ | †‡ | | | †‡ | | | | | | | | ‡ | | †‡ | | |
| | | ***NDUFS1*** | **19** | **13** | 2 | 1 | † | † | | | | | | | | | | ‡ | | | | | |
| | | ***SLC40A1*** | **35** | **10** | 8 | 8 | †‡ | †‡ | | †‡ | †‡ | †‡ | †‡ | | | | | †‡ | | | †‡ | | |
| | CDS | *DSTN* | **23** | **16** | 11 | 2 | † | † | † | † | † | †‡ | † | †‡ | † | † | † | | | | | | |
| | | *ENPEP* | **25** | **10** | 2 | 4 | †‡ | †‡ | | ‡ | | | | | | | | | | | | | ‡ |
| | | *SEC63* | **20** | **12** | 0 | 3 | ‡ | ‡ | | | | | ‡ | | | | | | | | | | |
| | | *TMEM202* | **20** | **7** | 3 | 4 | †‡ | †‡ | | †‡ | | | ‡ | | | | | | | | | | |
| | | *TOM1L2* | **19** | **11** | 5 | 1 | † | † | † | † | † | | | | | | | ‡ | | | | | |
| IRE 2 | 5'UTR | ***FTH1*** | **44** | **13** | 12 | 12 | †‡ | †‡ | †‡ | | †‡ | †‡ | | †‡ | | †‡ | | †‡ | | †‡ | †‡ | †‡ | †‡ |
| | | ***FTL*** | **43** | **15** | 12 | 12 | †‡ | †‡ | †‡ | †‡ | †‡ | †‡ | †‡ | †‡ | | | | | | †‡ | †‡ | †‡ | †‡ |
| | CDS | *EP400* | **23** | **6** | 2 | 1 | † | | † | | | | | | | ‡ | | | | | | | |
| | | Total: | | | 111 | 110 | | | | | | | | | | | | | | | | | |

The bit score shown is from the initial search of the human genome. Bit score thresholds used are 19 for IRE family 1 and 23 for IRE family 2. Shaded areas indicate the identification of a homologous RNA transcript. The genomic locations (UTR/CDS) are for the hit in the human gene. Hits in other species are considered anywhere in the transcript. Results are only shown for selected species. Complete results are available from mrna.otago.ac.nz/stevens2011a. Gene names shown in bold have mRNAs containing experimentally demonstrated IREs. Note that for CDC14A Homologene links to a transcript variant for *Homo sapiens* that lacks the experimentally demonstrated IRE sequence and is not detected. †, IRE match found using the new RFAM IRE families. ‡, IRE match found using the SIREs web server.

threshold for the purposes of helping to find the most likely novel IREs.

There are several novel IREs predicted in mRNAs with high bit scores and conservation across species (**Table 5**). Several have scores higher than the lowest true positive (*NDUFS1*, bit score 19). The 3' UTR of *MGAT4A* contained an IRE like element with a bit score of 20 (1.8 of 14 hits predicted to be due to chance—**Table S2**). mRNAs that contain one or multiple IREs in their 3' UTRs are expected to show decreased expression under high iron conditions. High throughput mRNA expression studies from the GEO database that tested iron effects were examined. In a study of rat duodenal mucosa in response to dietary iron, covering post-natal development of rats[58] there was decreased *MGAT4A* mRNA for high iron conditions but only at the 6–12 week stage of development. This change was similar but smaller to the change seen for *TFRC* expression where IRE/IRP binding in the 3' UTR has been shown to prevent endonucleolytic cleavage of the *TFRC* mRNA.[4] Other high throughput studies showed no significant difference in iron dose related *MGAT4A* mRNA expression.[59-61] The limited developmental window could explain why the IRE

motif hasn't so far been identified in studies involving cultured cells or mature organisms.

The predictions were extended to the coding parts of mRNAs, although a role for IREs in these regions has not been tested or demonstrated. Of these one of the highest scoring was an IRE like element in the coding region of *DSTN* (Score 23, 0.3 of 2 hits due to chance)—it is found in eleven homologous mRNAs. This hit is interesting given recently published evidence suggesting that *CDC42BPA* regulates iron uptake via transferrin mediated endocytosis dependent on cytoskeletal structure.[62] DSTN has a functional role in actin depolymerisation and filament dynamics.[51] A hypothesis is that this element is part of a molecular localization mechanism for the *DSTN* mRNA.

The refinement of the IRE families in this study provides testable predictions as to the functional parts of the IRE. The predictions of novel IREs using these models suggest avenues for further experimentation.

## Materials and Methods

**Models.** A literature search was conducted to include all IREs for which there was direct experimental evidence. The data was audited to establish the experimental support for each element to be included in the families. Demonstration of an IRP-IRE binding was considered sufficient for inclusion of the element sequence though most studies also included assays showing a regulatory effect. The sequences corresponding to the established elements were obtained from Genbank.[63] The list of known IREs with experimental evidence shown in **Table 1** was used to construct the new families.

The full Stockholm alignment comprising the new RFAM IRE families can be found as **Sup S1** and has been submitted to the RFAM database.

The entire assembled human genome (hg18-obtained from UCSC) was searched using cmsearch (part of the Infernal package). Both strands were considered for matches to the covariance models of both IRE Family 1 and IRE Family 2. Refseq gene annotations were obtained from UCSC via the Table browser interface. The annotations for complete genes, 3' UTRs, 5' UTRs, coding regions and introns were all separately obtained as BED intervals. Results obtained from searching using the covariance models were intersected with these intervals—taking into account the strand of the identified regions.

In order to identify possible pseudogenes for IRE containing genes (e.g., *FTH1*, *FTL*), exons with experimentally demonstrated IREs in *Homo sapiens* were first extracted from NCBI. These sequences were used as queries for blastn in searching the human genome (hg18-obtained from UCSC) using a word size of 8. A BED interval file was created corresponding to the blastn hits (**Sup. S6**). This method aims to provide some explanation for IRE hits in intergenic and intronic regions rather than to formally identify pseudogenes.

To identify sequence regions in RFAMSEQ10 corresponding to genes with known IREs, protein sequences were first obtained from NCBI corresponding to the nucleotide records containing experimentally established IREs. These protein sequences were then used to search the RFAMSEQ10 nucleotide database with tblastn. A 2 kb flanking region was added to each end of the tblastn search results so as to capture the corresponding UTRs. These regions were then intersected with hits for the IRE Families generated using the RFAM pipeline.

**Sensitivity and specificity.** Experimentally established IREs were identified with reference to the original publications (see **Table 1**). Where applicable, their genomic locations on version hg18 of the human genome (from UCSC) were recorded in BED interval files for each of the IRE families (**Sup. S7**). These data were used in assessing sensitivity of results from cmsearch—calculated by dividing the number of results intersecting the true positives by the number of true positives.

Chromosomes 1–22, X and Y of the human genome (hg18) were shuffled into random order and searched on one strand. This provided 2,858,013,655 random bases with base frequencies identical to the human genome. The new IRE models were used to search this random sequence and thereby arrive at an expected number of hits by chance at each bit score cut off. The size of the genomic locations identified by the annotations used in the intial search of the human genome was determined along with the size of the possible pseudogene sequences and intergenic sequence outside all these regions. The number of hits expected in random sequence of similar size was calculated for each of the genomic regions of interest—allowing for the fact that the search was conducted on both strands.

Repeat regions obtained from the UCSC browser track, "RepMask 3.2.7" via the Table browser interface were intersected with IRE hits found by the new RFAM IRE Family models. The count of unique results determined the number of IRE hits corresponding to repeat regions.

**Hits in orthologous mRNAs.** Bit score cut offs of 19 for IRE Family 1 and 23 for IRE Family 2 were applied to the results from the covariance model search of the human genome. A similar methodology was followed for both IRE families. Using genes identified with hits in exonic regions (5' UTR, 3' UTR or CDS), corresponding Homologene records at NCBI were obtained. From these associated nucleotide sequences were retrieved for all taxonomies—limiting the results to "biomol mrna" [prop]. In addition homologous transcript records with annotated UTRs were obtained from Ensembl. The retrieved sequences were searched using cmsearch using the same bit score cut off as in the initial search of the human genome for the corresponding family. Matches were recorded for all available species for each gene of interest.

The SIREs web server (ccbg.imppc.org/sires/index.html) was used to search these same sequences in order to obtain a comparison of results.

The phyloP scores for genomic regions corresponding to Refseq UTRs and the IRE hits located within these were obtained for hg18 from the UCSC Conservation track (44-way vertebrate). A Student t-test was used to compare the scores from the different regions.

## Acknowledgments

This work was supported by a Human Frontier Science Program grant to Ian Macara, Anne Spang & C.M.B. (RGP0031_2009) and by the Wellcome Trust in a grant to P.P.G. (WT077044/Z/05/Z).

## Note

Supplemental materials can be found at:
www.landesbioscience.com/journals/rnabiology/article/16037

## References

1. Hentze MW, Caughman SW, Rouault TA, Barriocanal JG, Dancis A, Harford JB, et al. Identification of the iron-responsive element for the translational regulation of human ferritin mRNA. Science 1987; 238:1570-3.
2. Leipuviene R, Theil EC. The family of iron responsive RNA structures regulated by changes in cellular iron and oxygen. Cell Mol Life Sci (CMLS) 2007; 64:2945-55.
3. Muckenthaler MU, Galy B, Hentze MW. Systemic iron homeostasis and the iron-responsive element/iron-regulatory protein (IRE/IRP) regulatory network. Annu Rev Nutr 2008; 28:197-213.
4. Koeller DM, Casey JL, Hentze MW, Gerhardt EM, Chan LN, Klausner RD, et al. A cytosolic protein binds to structural elements within the iron regulatory region of the transferrin receptor mRNA. Proc Natl Acad Sci USA 1989; 86:3574-8.
5. Hentze MW, Muckenthaler MU, Galy B, Camaschella C. Two to tango: regulation of Mammalian iron metabolism. Cell 2010; 142:24-38.
6. Gardner PP, Daub J, Tate J, Moore BL, Osuch IH, Griffiths-Jones S, et al. Rfam: Wikipedia, clans and the "decimal" release. Nucleic Acids Res 2011; 39:141-5.
7. Jacobs GH, Chen A, Stevens SG, Stockwell PA, Black MA, Tate WP, et al. Transterm: a database to aid the analysis of regulatory sequences in mRNAs. Nucleic Acids Res 2009; 37:72-6.
8. Grillo G, Turi A, Licciulli F, Mignone F, Liuni S, Banfi S, et al. UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. Nucleic Acids Res 2009; 38:75-80.
9. Addess KJ, Basilion JP, Klausner RD, Rouault TA, Pardi A. Structure and dynamics of the iron responsive element RNA: implications for binding of the RNA by iron regulatory binding proteins. J Mol Biol 1997; 274:72-83.
10. Walden WE, Selezneva AI, Dupuy J, Volbeda A, Fontecilla-Camps JC, Theil EC, et al. Structure of dual function iron regulatory protein 1 complexed with ferritin IRE-RNA. Science 2006; 314:1903-8.
11. Hentze MW, Kuhn LC. Molecular control of vertebrate iron metabolism: mRNA-based regulatory circuits operated by iron, nitric oxide and oxidative stress. Proc Natl Acad Sci USA 1996; 93:8175-82.
12. Piccinelli P, Samuelsson T. Evolution of the iron-responsive element. RNA 2007; 13:952.
13. Gunshin H, Allerson CR, Polycarpou-Schwarz M, Rofts A, Rogers JT, Kishi F, et al. Iron-dependent regulation of the divalent metal ion transporter. FEBS Lett 2001; 509:309-16.
14. Cmejla R, Petrak J, Cmejlova J. A novel iron responsive element in the 3'UTR of human MRCKalpha. Biochem Biophys Res Commun 2006; 341:158-66.
15. Sanchez M, Galy B, Dandekar T, Bengert P, Vainshtein Y, Stolte J, et al. Iron regulation and the cell cycle: identification of an iron-responsive element in the 3'-untranslated region of human cell division cycle 14A mRNA by a refined microarray-based screening strategy. J Biol Chem 2006; 281:22865.
16. Nunez MT, Tapia V, Rojas A, Aguirre P, Gomez F, Nualart F. Iron supply determines apical/basolateral membrane distribution of intestinal iron transporters DMT1 and ferroportin 1. Am J Physiol Cell Physiol 2009; 298:477-85.
17. McKie AT, Marciani P, Rolfs A, Brennan K, Wehr K, Barrow D, et al. A novel duodenal iron-regulated transporter, IREG1, implicated in the basolateral transfer of iron to the circulation. Mol Cell 2000; 5:299-309.
18. Harigae H, Suwabe N, Weinstock PH, Nagai M, Fujita H, Yamamoto M, et al. Deficient heme and globin synthesis in embryonic stem cells lacking the erythroid-specific delta-aminolevulinate synthase gene. Blood 1998; 91:798-805.
19. Melefors O. Translational regulation in vivo of the *Drosophila melanogaster* mRNA encoding succinate dehydrogenase iron protein via iron responsive elements. Biochem Biophys Res Commun 1996; 221:437-41.
20. Gruer MJ, Artymiuk PJ, Guest JR. The aconitase family: three structural variations on a common theme. Trends Biochem Sci 1997; 22:3-6.
21. Majmundar AJ, Wong WJ, Simon MC. Hypoxia-inducible factors and the response to hypoxic stress. Mol Cell 2010; 40:294-309.
22. Leung T, Chen XQ, Tan I, Manser E, Lim L. Myotonic dystrophy kinase-related Cdc42-binding kinase acts as a Cdc42 effector in promoting cytoskeletal reorganization. Mol Cell Biol 1998; 18:130-40.
23. Bembenek J, Yu H. Regulation of the anaphase-promoting complex by the dual specificity phosphatase human Cdc14a. J Biol Chem 2001; 276:48237-42.
24. Mailand N, Lukas C, Kaiser BK, Jackson PK, Bartek J, Lukas J. Deregulated human Cdc14A phosphatase disrupts centrosome separation and chromosome segregation. Nat Cell Biol 2002; 4:317-22.
25. Kohler SA, Henderson BR, Kühn LC. Succinate dehydrogenase b mRNA of *Drosophila melanogaster* has a functional iron-responsive element in its 5-untranslated region. J Biol Chem 1995; 270:30781.
26. Sanchez M, Galy B, Muckenthaler MU, Hentze MW. Iron-regulatory proteins limit hypoxia-inducible factor-2alpha expression in iron deficiency. Nat Struct Mol Biol 2007; 14:420-6.
27. Campillos M, Cases I, Hentze MW, Sanchez M. SIREs: searching for iron-responsive elements. Nucleic Acids Res 2010; 38:360-7.
28. Nawrocki EP, Kolbe DL, Eddy SR. Infernal 1.0: inference of RNA alignments. Bioinformatics 2009; 25:1335-7.
29. Aziz N, Munro HN. Iron regulates ferritin mRNA translation through a segment of its 5'untranslated region. Proc Natl Acad Sci USA 1987; 84:8478-82.
30. Dandekar T, Stripecke R, Gray N, Goossen B, Constable A, Johansson H, et al. Identification of a novel iron-responsive element in murine and human erythroid delta-aminolevulinic acid synthase mRNA. EMBO J 1991; 10:1903.
31. Gray NK, Pantopoulos K, Dandekar T, Ackrell BAC, Hentze MW. Translational regulation of mammalian and Drosophila citric acid cycle enzymes via iron-responsive elements. Proc Natl Acad Sci USA 1996; 93:4925-30.
32. Huang TS, Melefors O, Lind MI, Söderhäll K. An atypical iron-responsive element (IRE) within crayfish ferritin mRNA and an iron regulatory protein 1 (IRP1)-like protein from crayfish hepatopancreas. Insect Biochem Mol Biol 1999; 29:1-9.
33. Kohler SA, Menotti E, Kuhn LC. Molecular cloning of mouse glycolate oxidase. High evolutionary conservation and presence of an iron-responsive element-like sequence in the mRNA. J Biol Chem 1999; 274:2401-7.
34. Alén C, Sonenshein AL. *Bacillus subtilis* aconitase is an RNA-binding protein. Proc Natl Acad Sci USA 1999; 96:10412-7.
35. Zhang D, Albert DW, Kohlhepp P, D-Pham DQ, Winzerling JJ. Repression of *Manduca sexta* ferritin synthesis by IRP1/IRE interaction. Insect Mol Biol 2001; 10:531-9.
36. Lin E, Graziano JH, Freyer GA. Regulation of the 75-kDa subunit of mitochondrial complex I by iron. J Biol Chem 2001; 276:27685-92.
37. Nichol H, Winzerling J. Structured RNA upstream of insect cap distal iron responsive elements enhances iron regulatory protein-mediated control of translation. Insect Biochem Mol Biol 2002; 32:1699-710.
38. Lymboussaki A, Pignatti E, Montosi G, Garuti C, Haile D, Pietrangelo A. The role of the iron responsive element in the control of ferroportin1/IREG1/MTP1 gene expression. J Hepatol 2003; 39:710.
39. Wingert RA, Galloway JL, Barut B, Foott H, Fraenkel P, Axe JL, et al. Deficiency of glutaredoxin 5 reveals Fe-S clusters are required for vertebrate haem synthesis. Nature 2005; 436:1035-9.
40. Costanzo F, Colombo M, Staempfli S, Santoro C, Marone M, Frank R, et al. Structure of gene and pseudogenes of human apoferritin H. Nucleic Acids Res 1986; 14:721-36.
41. Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, et al. Rfam: updates to the RNA families database. Nucleic Acids Res 2009; 37:136-40.
42. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 2009; 37:1-13.
43. Lotze MT, Tracey KJ. High-mobility group box 1 protein (HMGB1): nuclear weapon in the immune arsenal. Nat Rev Immunol 2005; 5:331-42.
44. Toole BP. Hyaluronan: from extracellular glue to pericellular cue. Nat Rev Cancer 2004; 4:528-39.
45. Lopez-Orduna E, Cruz M, Garcia-Mena J. The transcription of MGAT4A glycosyl transferase is increased in white cells of peripheral blood of type 2 diabetes patients. BMC Genet 2007; 8:73.
46. Kaelin WG Jr. The von hippel-lindau tumor suppressor protein: an update. Methods Enzymol 2007; 435:371-83.
47. Longo-Guess CM, Gagnon LH, Cook SA, Wu J, Zheng QY, Johnson KR. A missense mutation in the previously undescribed gene Tmhs underlies deafness in hurry-scurry (hscy) mice. Proc Natl Acad Sci USA 2005; 102:7894-9.
48. Yamaji T, Kumagai K, Tomishige N, Hanada K. Two sphingolipid transfer proteins, CERT and FAPP2: their roles in sphingolipid metabolism. IUBMB Life 2008; 60:511-8.
49. Percival JM, Froehner SC. Golgi complex organization in skeletal muscle: a role for Golgi-mediated glycosylation in muscular dystrophies? Traffic 2007; 8:184-94.
50. Franco M, Furstoss O, Simon V, Benistant C, Hong WJ, Roche S. The adaptor protein Tom1L1 is a negative regulator of Src mitogenic signaling induced by growth factors. Mol Cell Biol 2006; 26:1932-47.
51. Carlier MF, Laurent V, Santolini J, Melki R, Didry D, Xia GX, et al. Actin depolymerizing factor (ADF/cofilin) enhances the rate of filament turnover: implication in actin-based motility. J Cell Biol 1997; 136:1307-22.
52. Zini S, Fournie-Zaluski MC, Chauvel E, Roques BP, Corvol P, Llorens-Cortes C. Identification of metabolic pathways of brain angiotensin II and III using specific aminopeptidase inhibitors: predominant role of angiotensin III in the control of vasopressin release. Proc Natl Acad Sci USA 1996; 93:11968-73.
53. Meyer HA, Grau H, Kraft R, Kostka S, Prehn S, Kalies KU, et al. Mammalian Sec61 is associated with Sec62 and Sec63. J Biol Chem 2000; 275:14550-7.
54. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res 2010; 20:110-21.

55. dos Santos CO, Dore LC, Valentine E, Shelat SG, Hardison RC, Ghosh M, et al. An iron responsive element-like stem-loop regulates alpha-hemoglobin-stabilizing protein mRNA. J Biol Chem 2008; 283:26956-64.
56. Solano-González E, Burrola-Barraza E, León-Sicairos C, Avila-González L, Gutiérrez-Escolano L, Ortega-López J, et al. The trichomonad cysteine proteinase TVCP4 transcript contains an iron-responsive element. FEBS Lett 2007; 581:2919-28.
57. Andersen O, Pantopoulos K, Kao HT, Muckenthaler M, Youson JH, Pieribone V. Regulation of iron metabolism in the sanguivore lamprey *Lampetra fluviatilis*-molecular cloning of two ferritin subunits and two iron-regulatory proteins (IRP) reveals evolutionary conservation of the iron-regulatory element (IRE)/IRP regulatory system. Eur J Biochem 1998; 254:223-9.
58. Collins JF, Franck CA, Kowdley KV, Ghishan FK. Identification of differentially expressed genes in response to dietary iron deprivation in rat duodenum. Am J Physiol Gastrointest Liver Physiol 2005; 288:964-71.
59. Chicault C, Toutain B, Monnier A, Aubry M, Fergelot P, Le Treut A, et al. Iron-related transcriptomic variations in CaCo-2 cells, an in vitro model of intestinal absorptive cells. Physiol Genomics 2006; 26:55-67.
60. Rodriguez A, Hilvo M, Kytomaki L, Fleming RE, Britton RS, Bacon BR, et al. Effects of iron loading on muscle: genome-wide mRNA expression profiling in the mouse. BMC Genomics 2007; 8:379.
61. Ohyashiki JH, Kobayashi C, Hamamura R, Okabe S, Tauchi T, Ohyashiki K. The oral iron chelator deferasirox represses signaling through the mTOR in myeloid leukemia cells by enhancing expression of REDD1. Cancer Sci 2009; 100:970-7.
62. Cmejla R, Ptackova P, Petrak J, Savvulidi F, Cerny J, Sebesta O, et al. Human MRCKalpha is regulated by cellular iron levels and interferes with transferrin iron uptake. Biochem Biophys Res Commun 2010; 395:163-7.
63. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. GenBank. Nucleic Acids Res 2008; 36:25-30.