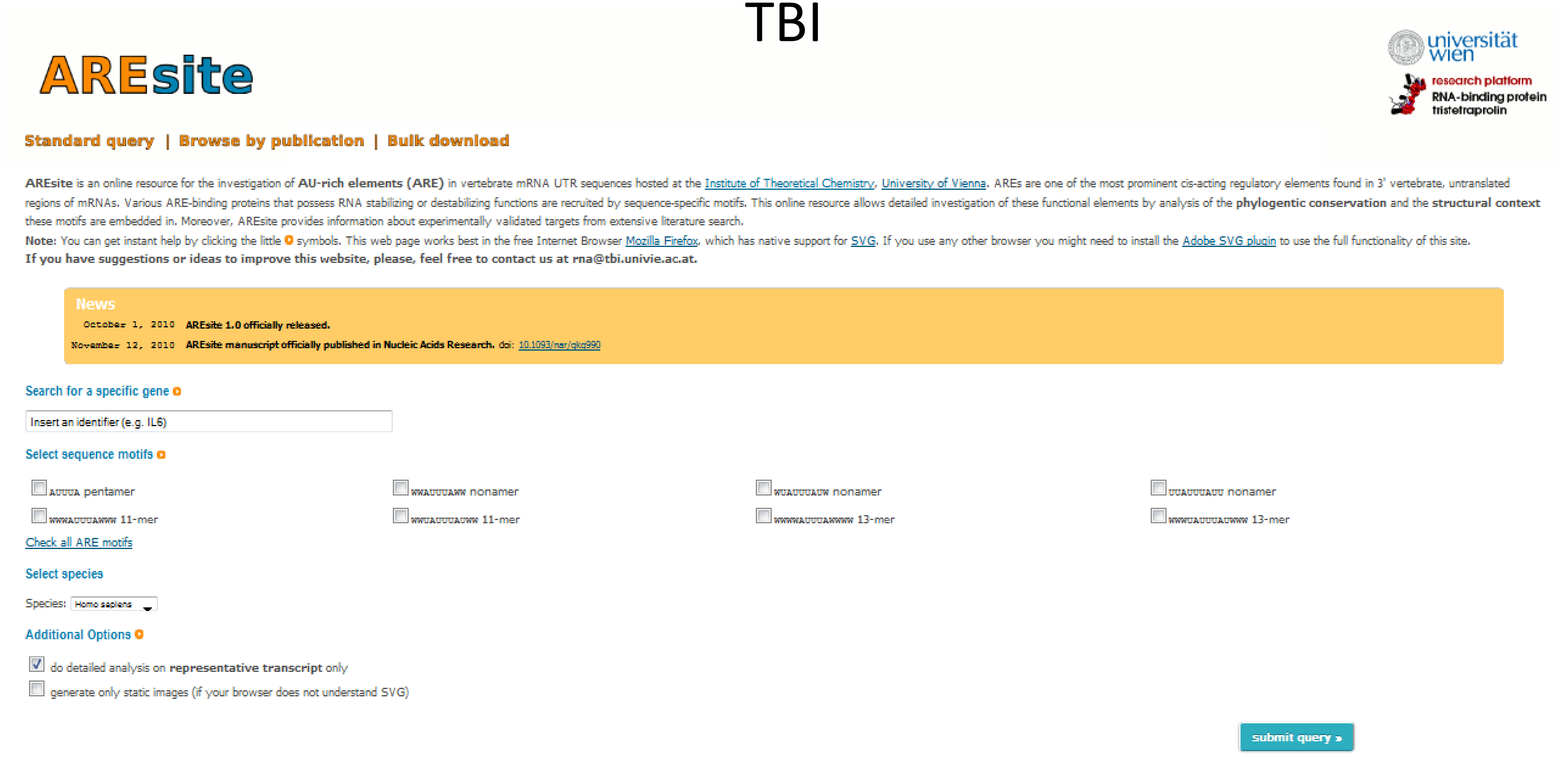
AREsite
proceedings

Jörg Fallmann
Diploma student at
TBI



AREsite

universität wien
research platform
RNA-binding protein tristetraprolin

Standard query | **Browse by publication** | **Bulk download**

AREsite is an online resource for the investigation of **AU-rich elements (ARE)** in vertebrate mRNA UTR sequences hosted at the Institute of Theoretical Chemistry, University of Vienna. AREs are one of the most prominent cis-acting regulatory elements found in 3' vertebrate, untranslated regions of mRNAs. Various ARE-binding proteins that possess RNA stabilizing or destabilizing functions are recruited by sequence-specific motifs. This online resource allows detailed investigation of these functional elements by analysis of the **phylogentic conservation** and the **structural context** these motifs are embedded in. Moreover, AREsite provides information about experimentally validated targets from extensive literature search.

Note: You can get instant help by clicking the little ● symbols. This web page works best in the free Internet Browser Mozilla Firefox, which has native support for SVG. If you use any other browser you might need to install the Adobe SVG plugin to use the full functionality of this site.

If you have suggestions or ideas to improve this website, please, feel free to contact us at rna@tbi.univie.ac.at.

News

October 1, 2010	**AREsite 1.0 officially released.**
November 12, 2010	**AREsite manuscript officially published in Nucleic Acids Research.** doi: 10.1093/nar/gkq990

Search for a specific gene

Insert an identifier (e.g. IL6)

Select sequence motifs

☐ AUUUA pentamer
☐ WWAUUUAWW nonamer
☐ WUAUUUAUW nonamer
☐ UUAUUUAUU nonamer
☐ WWWAUUUAWWW 11-mer
☐ WWUAUUUAUWW 11-mer
☐ WWWWAUUUAWWWW 13-mer
☐ WWWUAUUUAUWWW 13-mer

Check all ARE motifs

Select species

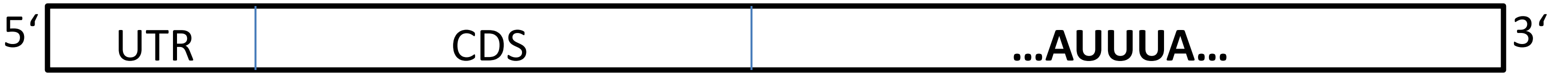
Species: Homo sapiens

Additional Options

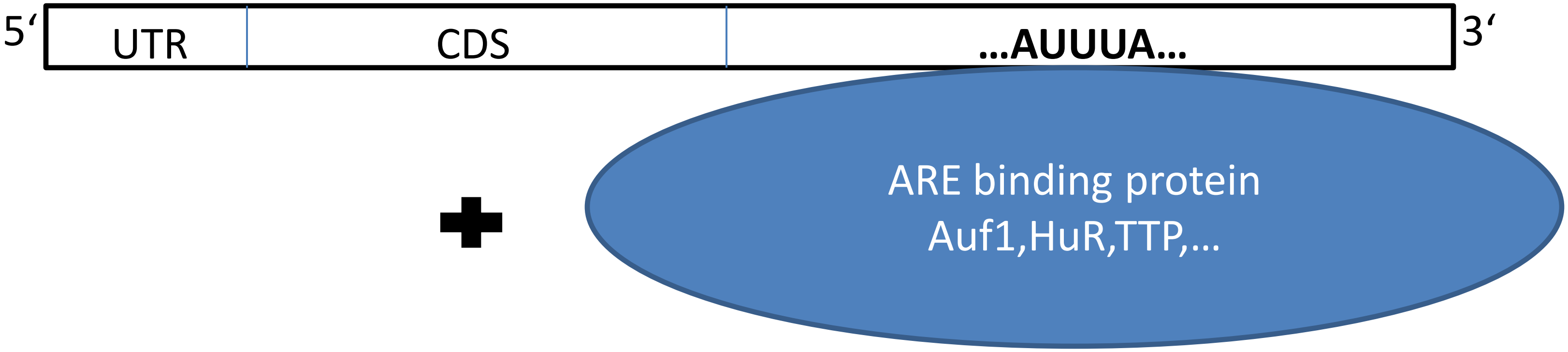
☑ do detailed analysis on **representative transcript** only
☐ generate only static images (if your browser does not understand SVG)

What is an ARE (good for?)

- AU rich element
- Cis-acting regulatory elements found at 3‘UTR
- In ~7% of human protein coding genes
- Known to play a role in mRNA stability
- Core motif AUUUA



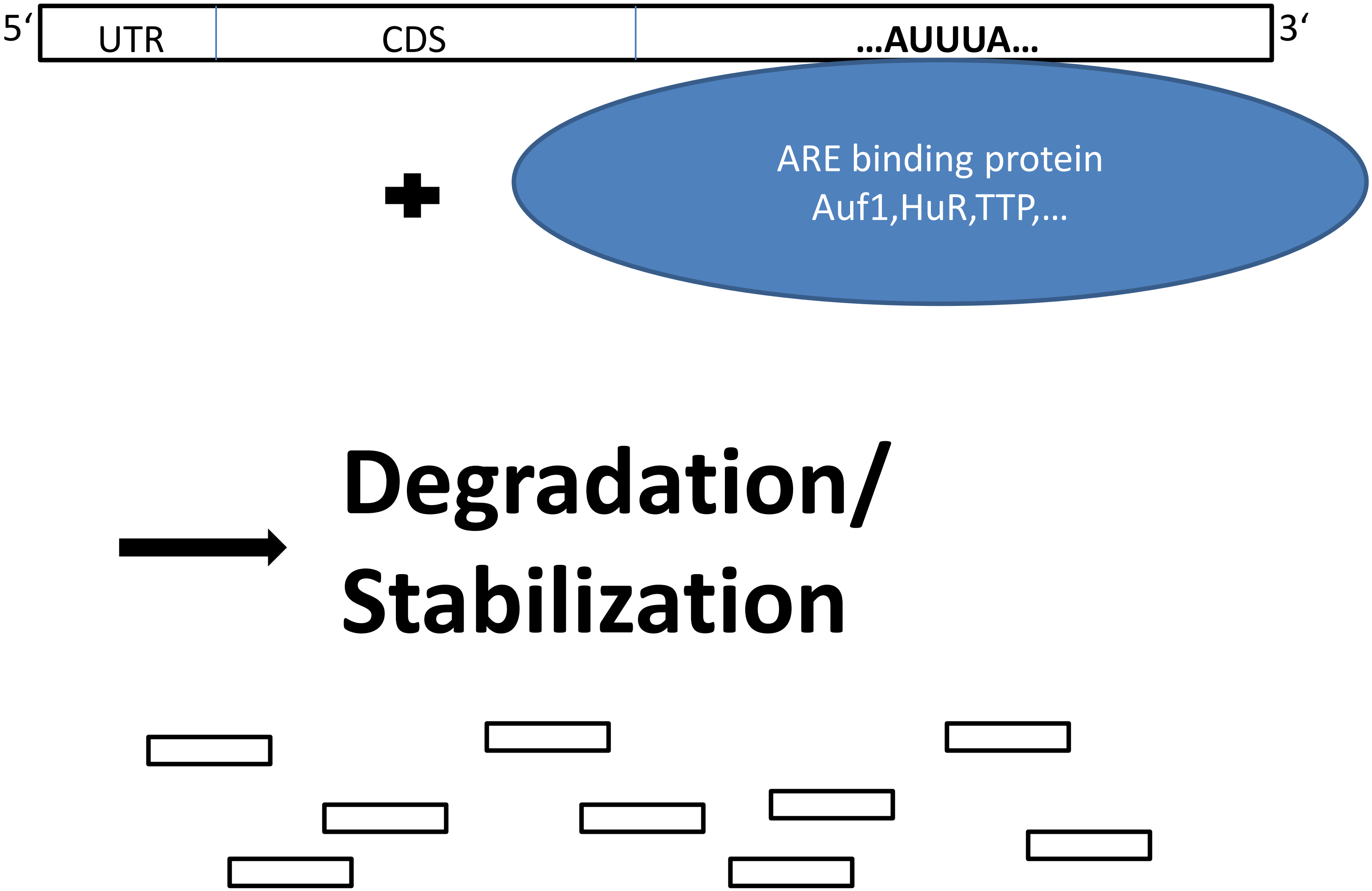
5‘ UTR CDS **…AUUUA…** 3‘



5‘ UTR CDS **…AUUUA…** 3‘

✚

ARE binding protein
Auf1,HuR,TTP,…



5‘

UTR

CDS

...AUUUA...

3‘

ARE binding protein

Auf1,HuR,TTP,...

Degradation/

Stabilization

ARE binding proteins

- TTP -> Kovarik Lab -> TTP ko mouse
- HuR, Auf1 -> best studied
- Most prominent target TNFα -> cancer

Making of

- Ensembl release 56 -> database in background

- Only genes from human and mouse with:

- At least one transcript containing a 3‘UTR
- If more -> representing transcript ->

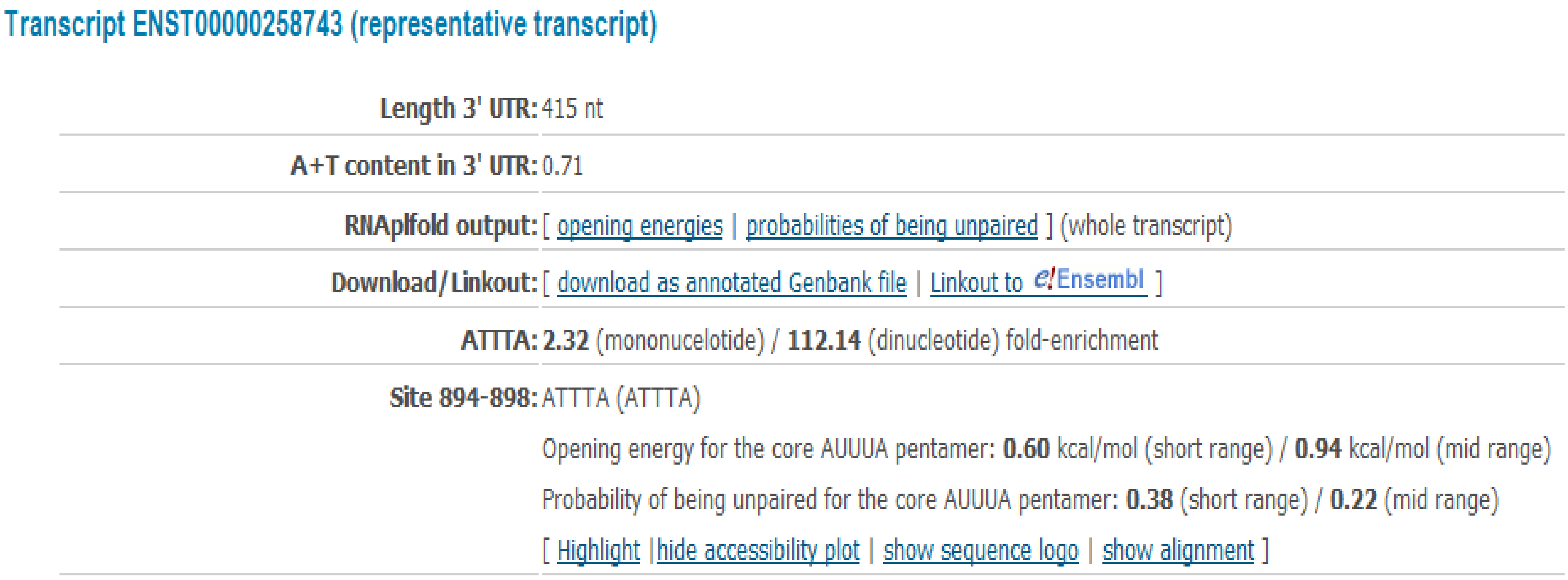
  The one with the most AUUUA counts

Accessibility

- RNAplfold: standard settings (w= 80, L= 40) for structural context in direct neighbourhood
- Settings provided by Hakim (w= 240,L= 160) for mid-range context
- SVG: accessibility for core motifs

Phylogenetic analysis

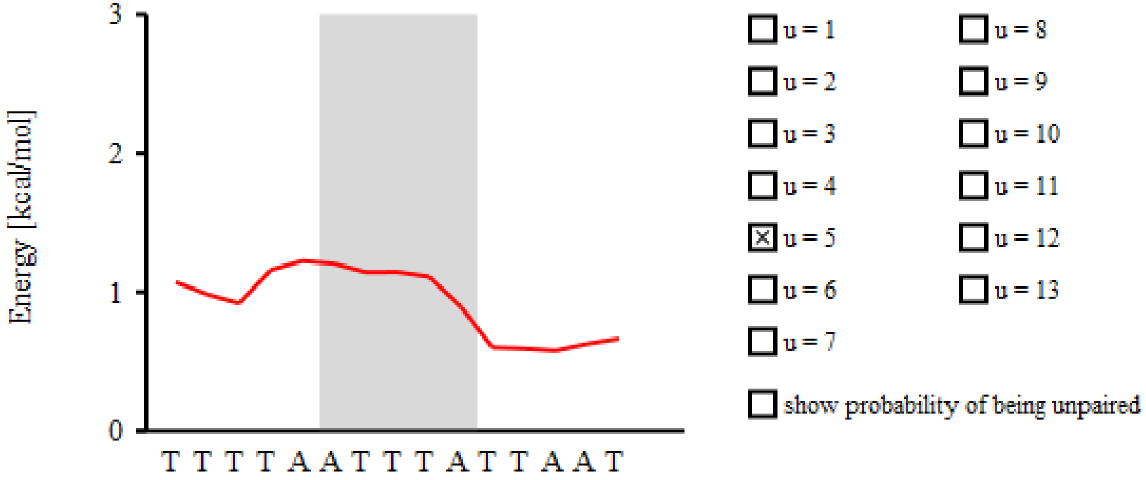
- One2one orthologs
- Genomic alignments -> maf blocks from UCSC, do not know if same transcripts in other species
- On transcript level -> Ensembl gene orthology pipeline, less data
- Phylogenetic tree



Transcript ENST00000258743 (representative transcript)

Length 3' UTR:	415 nt
A+T content in 3' UTR:	0.71
RNAplfold output:	[opening energies | probabilities of being unpaired] (whole transcript)
Download/Linkout:	[download as annotated Genbank file | Linkout to e!Ensembl]
ATTTA:	**2.32** (mononucelotide) / **112.14** (dinucleotide) fold-enrichment
Site 894-898:	ATTTA (ATTTA) Opening energy for the core AUUUA pentamer: **0.60** kcal/mol (short range) / **0.94** kcal/mol (mid range) Probability of being unpaired for the core AUUUA pentamer: **0.38** (short range) / **0.22** (mid range) [Highlight |hide accessibility plot | show sequence logo | show alignment]

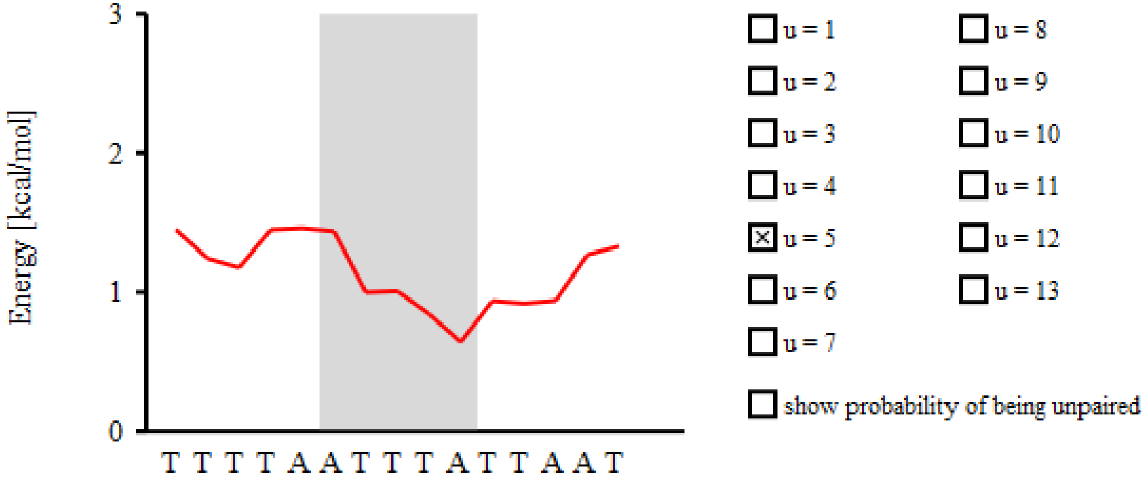
Short range interaction (W = 80, L = 40)



☐ u = 1
☐ u = 2
☐ u = 3
☐ u = 4
☒ u = 5
☐ u = 6
☐ u = 7
☐ u = 8
☐ u = 9
☐ u = 10
☐ u = 11
☐ u = 12
☐ u = 13

☐ show probability of being unpaired

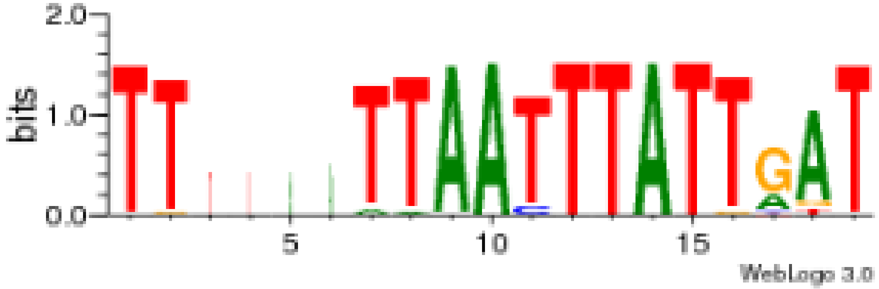
Mid range interaction (W = 240, L = 160)



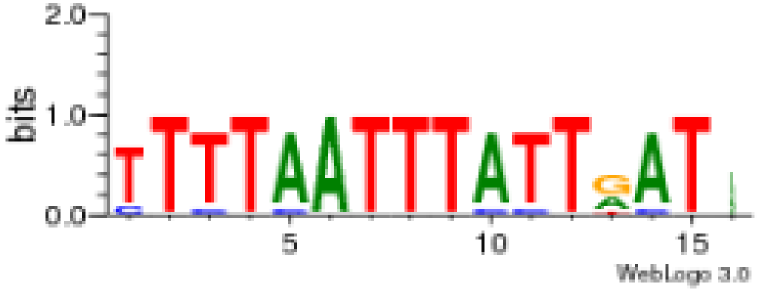
☐ u = 1
☐ u = 2
☐ u = 3
☐ u = 4
☒ u = 5
☐ u = 6
☐ u = 7
☐ u = 8
☐ u = 9
☐ u = 10
☐ u = 11
☐ u = 12
☐ u = 13

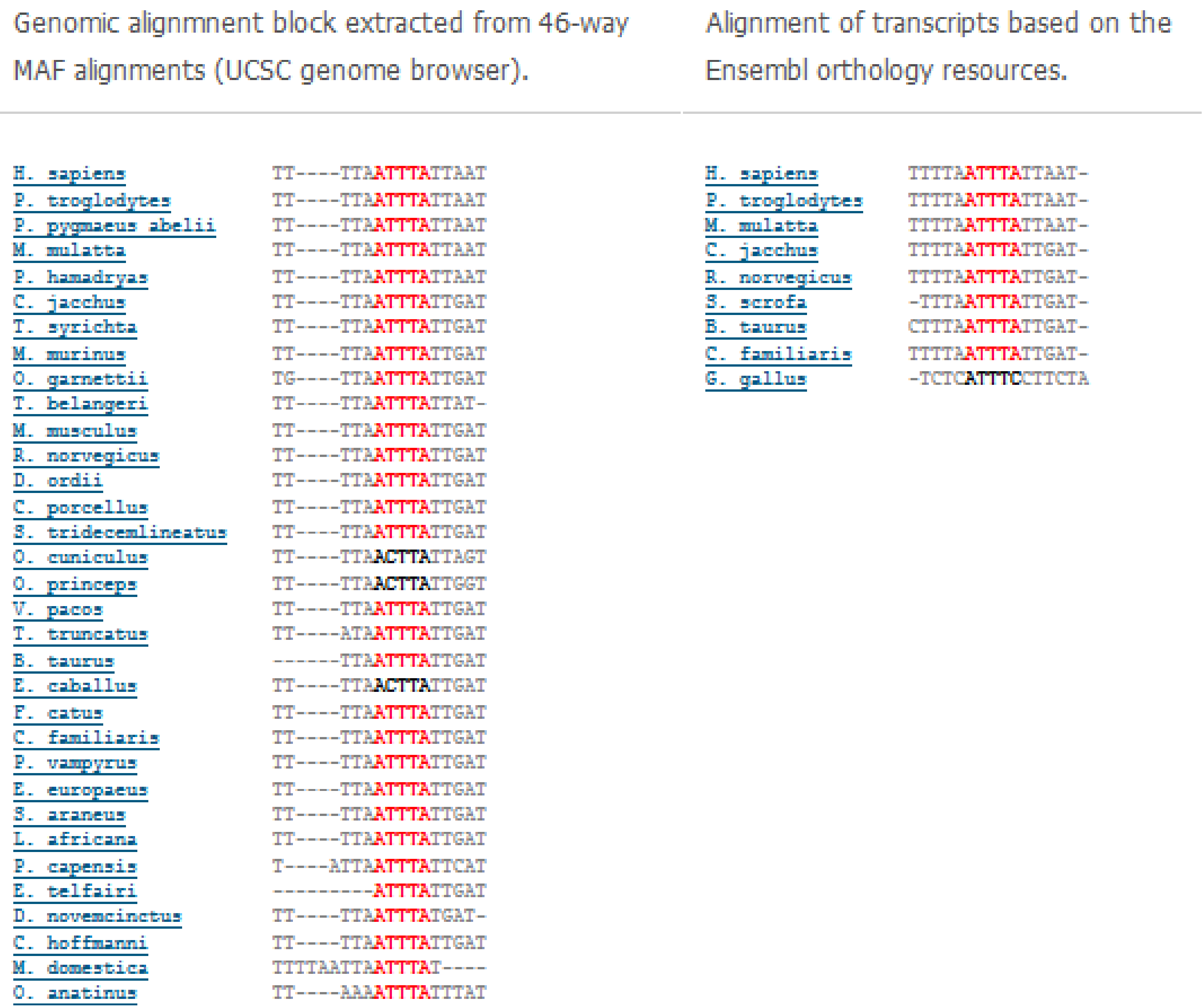
☐ show probability of being unpaired

Genomic alignmnent block extracted from 46-way MAF alignments (UCSC genome browser).



Alignment of transcripts based on the Ensembl orthology resources.





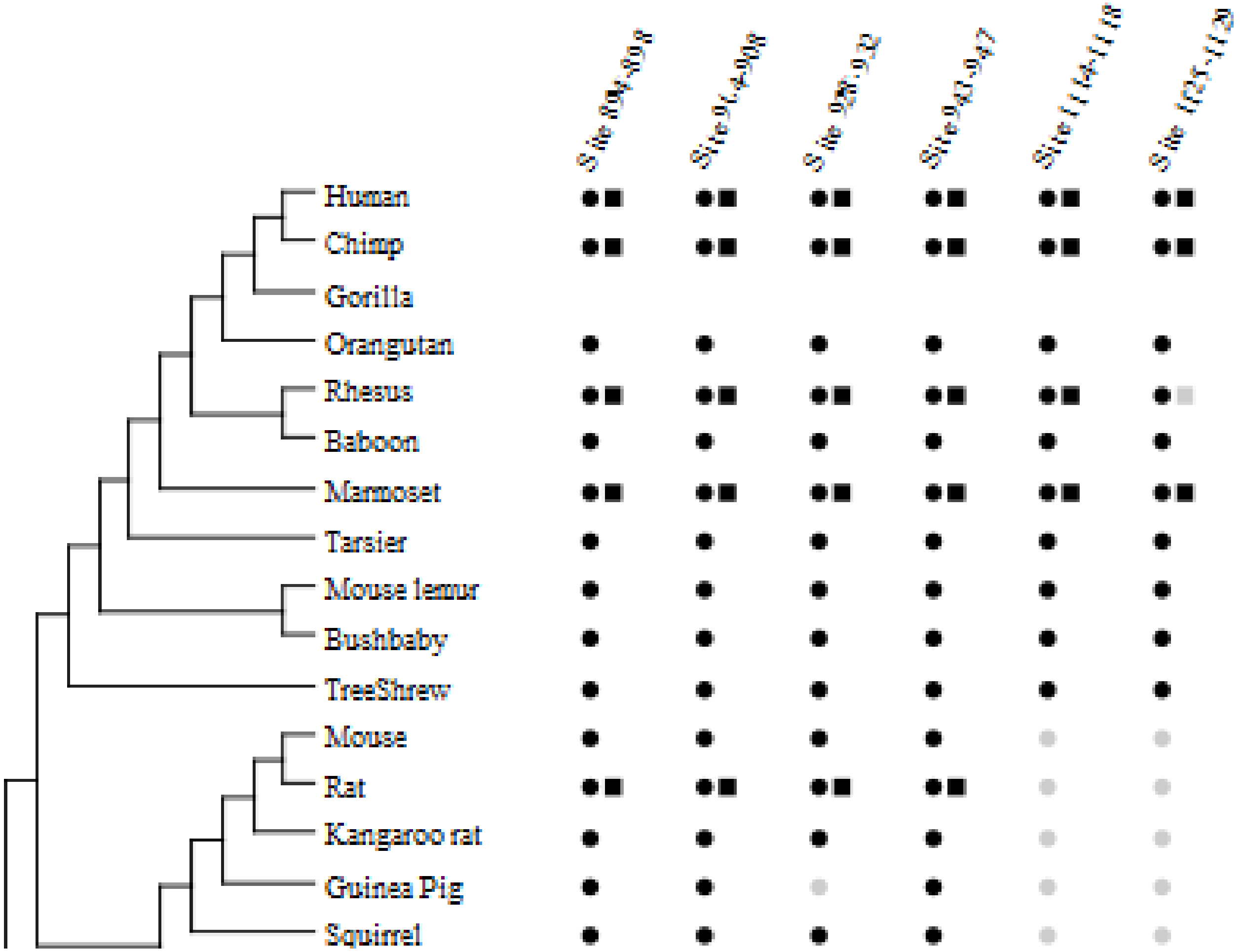
Genomic alignment block extracted from 46-way MAF alignments (UCSC genome browser).

H. sapiens	TT----TTAATTTATTAAT
P. troglodytes	TT----TTAATTTATTAAT
P. pygmaeus abelii	TT----TTAATTTATTAAT
M. mulatta	TT----TTAATTTATTAAT
P. hamadryas	TT----TTAATTTATTAAT
C. jacchus	TT----TTAATTTATTGAT
T. syrichta	TT----TTAATTTATTGAT
M. murinus	TT----TTAATTTATTGAT
O. garnettii	TG----TTAATTTATTGAT
T. belangeri	TT----TTAATTTATTAT-
M. musculus	TT----TTAATTTATTGAT
R. norvegicus	TT----TTAATTTATTGAT
D. ordii	TT----TTAATTTATTGAT
C. porcellus	TT----TTAATTTATTGAT
S. tridecemlineatus	TT----TTAATTTATTGAT
O. cuniculus	TT----TTAACTTATTAGT
O. princeps	TT----TTAACTTATTGGT
V. pacos	TT----TTAATTTATTGAT
T. truncatus	TT----ATAATTTATTGAT
B. taurus	------TTAATTTATTGAT
E. caballus	TT----TTAACTTATTGAT
F. catus	TT----TTAATTTATTGAT
C. familiaris	TT----TTAATTTATTGAT
P. vampyrus	TT----TTAATTTATTGAT
E. europaeus	TT----TTAATTTATTGAT
S. araneus	TT----TTAATTTATTGAT
L. africana	TT----TTAATTTATTGAT
P. capensis	T----ATTAATTTATTCAT
E. telfairi	---------ATTTATTGAT
D. novemcinctus	TT----TTAATTTATGAT-
C. hoffmanni	TT----TTAATTTATTGAT
M. domestica	TTTTAATTAATTTAT----
O. anatinus	TT----AAAATTTATTTAT

Alignment of transcripts based on the Ensembl orthology resources.

H. sapiens	TTTTAATTTATTAAT-
P. troglodytes	TTTTAATTTATTAAT-
M. mulatta	TTTTAATTTATTAAT-
C. jacchus	TTTTAATTTATTGAT-
R. norvegicus	TTTTAATTTATTGAT-
S. scrofa	-TTTAATTTATTGAT-
B. taurus	CTTTAATTTATTGAT-
C. familiaris	TTTTAATTTATTGAT-
G. gallus	-TCTCATTTCCTTCTA

The tree below summarizes the conservation pattern of the detected ARE motifs. Circles indicate genomic MAF alignments, while boxes are used transcript alignments. Signs in grey indicate that the sequence is present in the alignment, but the correpsonding ARE pattern was not detected.



So far

- Release 1 done
- Published in NAR

Gruber AR, Fallmann J, Kratochvill F, Kovarik P, Hofacker IL (2010). "AREsite: a database for the comprehensive investigation of AU-rich elements.". *Nucleic Acids Res* **39** (Database issue): D66-9. doi:10.1093/nar/gkq990. PMID 21071424.

ToDo

- Generate some (relevant) information
  - Statistics on my data
  - Influence of already bound proteins or miRNA on accessibility -> constraint folding with RNAplfold
  - Literature search for proteins with ARE binding domains, which, where and how much
  - Have a look at crystal structures to see how they bind, may be similar motifs -> HuR seems to simply bind U-oligos

Efficient use of accessibility in microRNA target prediction

- Published online 30 August 2010 Nucleic Acids Research, 2011, Vol. 39, No. 1 19–29
- doi:10.1093/nar/gkq768
- Ray M. Marín and Jirí Vaníček

Hypothesis

- ARE binding proteins bind to single strand motifs
- Higher accessibility means better target
- Implement algorithm of Marín and Vaníček to get one comparable number for accessibility and count of motif sites

Summary

- Partial complementarity sufficient

  => target prediction non-trivial

Strategies so far:

- Hybridization energy
- Conservation among 3‘ UTRs
- Accessibility

Hybridization energy

- Selection of strongest interaction
- Strongest physical interaction is most likely to be functional?

Conservation

- If important -> conserved
- Sometimes not enough, miRNAs can act on non conserved regions

Accessibility

- Without (partial) accessibility no interaction
- TotalFreeEnergy TFE = opening + hybridization
- Does not incorporate protein – RNA energy contributions to interactions -> ToDo?!

New approach

- Rank of motifs according to their over – representation and accessibility

$$P_{SH} = \sum_{i=c}^{l-n+1} \binom{l-n+1}{i} P^{i} (1-P)^{l-n+1-i}$$

Where:
P = prob to find n-mer by chance at least c times -> MM based on composition
l = length of 3‘ UTR
n = number of nucleotides in the seed (= motif length)
c = occurence of the motif

the **p-value** is the probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true

$$P_{SH} = \sum_{i=c_{access}}^{t_{access}} \binom{t_{access}}{i} P^{i} (1-P)^{t_{access}-i}$$

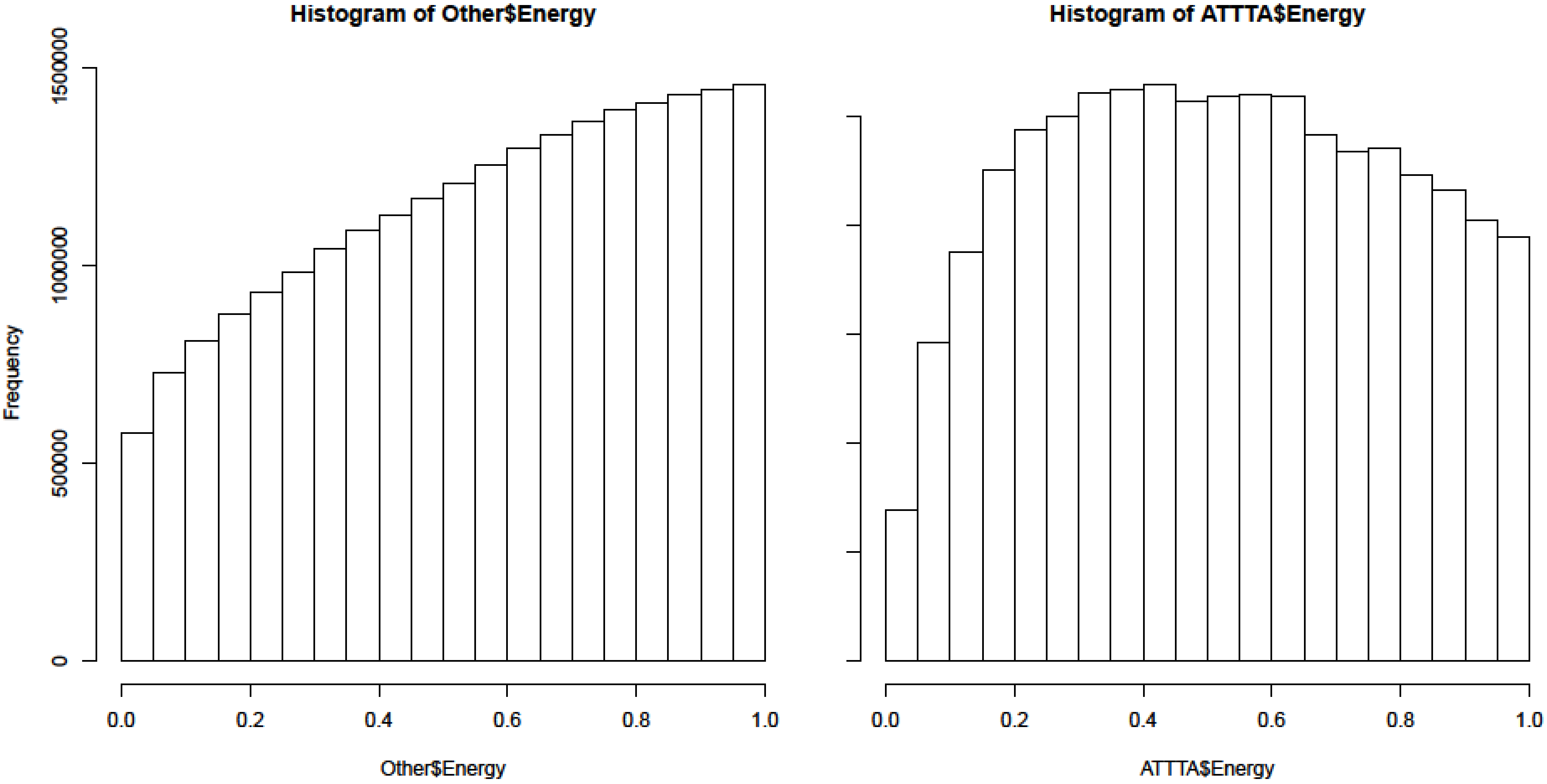
c_{access} = occurence of accessible motifs (u5 < 1kJ now)

t_{access} = total number of accessible nucleotides with u5 < 1kJ

Results

- We get a ranking of transcripts
- Low P_{SH} should mean highly regulated by AREs

- A lot of statistics to do


Histogram of Other$Energy
Frequency
0
500000
1000000
1500000
0.0
0.2
0.4
0.6
0.8
1.0
Other$Energy
Histogram of ATTTA$Energy
0.0
0.2
0.4
0.6
0.8
1.0
ATTTA$Energy

Thanks to

- you
- Andreas Gruber for ongoing support
- Ivo, Xtof & The TBI crew
- Greetings to all from Andreas