Finding locally disrupted RNA structure from SNPs

Sabarinathan Radhakrishnan

Center for non-coding RNA in Technology and Health (RTH)
University of Copenhagen
Denmark

Bled, Slovenia February 17, 2011

Outline

- Introduction
- Motivation
- Local structure comparison
- Results
- Further work

DISEASE by RNA

- Disease associated mutations are ofenly identified in intergenic and non-coding regions - Genome-wide association studies (GWAS)
- 95% of the human genome are transcribed and of possible mutation carriers
- SNP induced structural changes in the regulatory RNAs of the human genome results in disease phenotype (Hyperferritinemia Catarct Syndrome, Retinoblastoma, etc.,)
- Also, it alters the function of replication and translation (Hepatitis C Virus) and resistance against antibiotics (Bacteria)

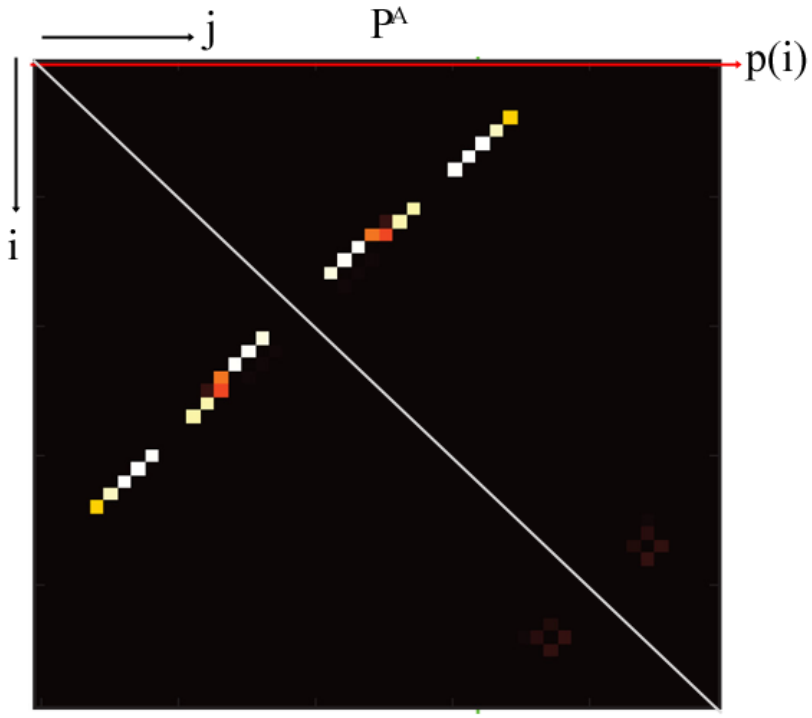
Existing algorithms and database

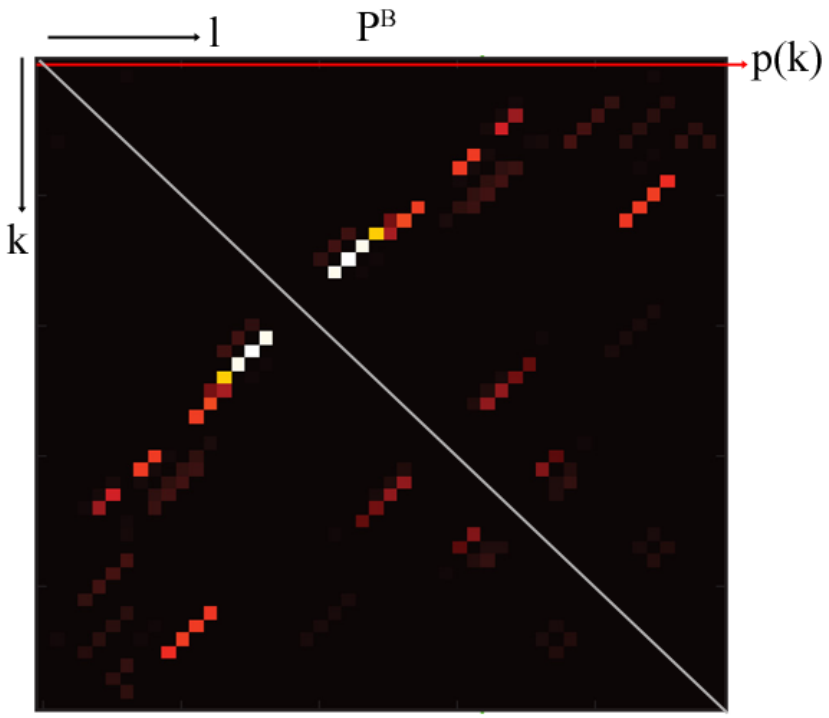
- **Resources**: RNAmute (Churkin and Barash, 2006; 2008), RDMAS (Shu et al., 2006) RNAmutants (Waldispuhl et al., 2008; 2009), SNPfold (Halvorsen et al., 2010)
- **Efficiency**: Handles Single Point Mutations / Multiple Point Mutations
- **Function**: Measures Global structural changes in RNA(Ensemble) induced by the SNP
- **Biological relevance**: Majority of mutations have small, local effects on the structure ensemble, while certain specific mutations can profoundly alter it (Halvorsen et al., 2010)
- **Requirement (??)**: Program to explore the local regions (may be RNA functional elements) distrupted by SNPs in regulatory RNA's

Global structure comparison

Comparing base pair probability matrices

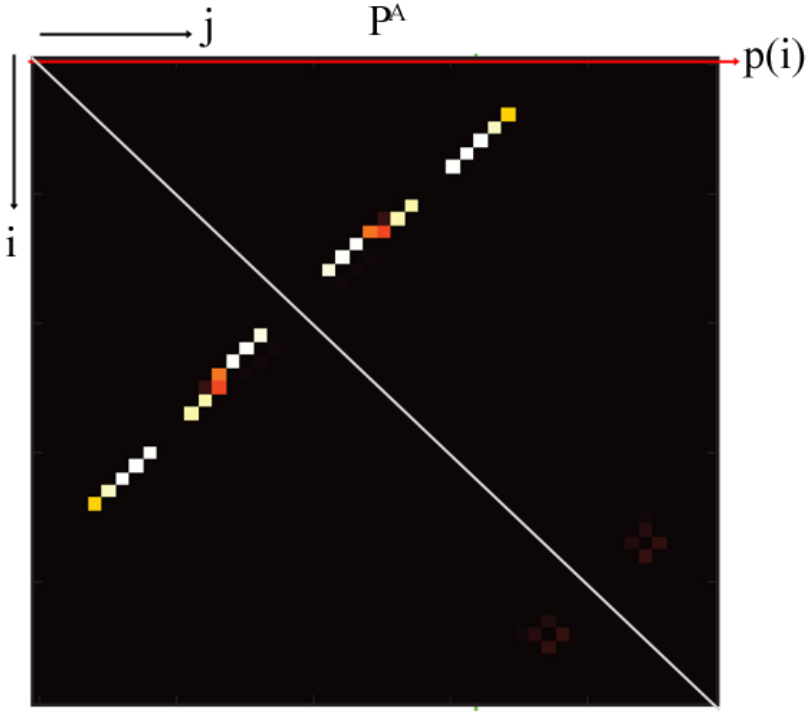
- Two sequences A (wild type) and B(mutant) with identical length
- Base pair probability P^A and P^B - partition function (RNAfold)
- $p(i)^A = \sum_{j=1}^{N^A} P_{ij}^A$ and $p(k)^B = \sum_{j=1}^{N^B} P_{kl}^B$
- r = corr(Ψ_{ij}^A and Ψ_{kl}^B) where Ψ_{ij}^A = {p(i),p(i+1),...,p(j)} and Ψ_{kl}^B = {p(k),p(k+1),...,p(l)}

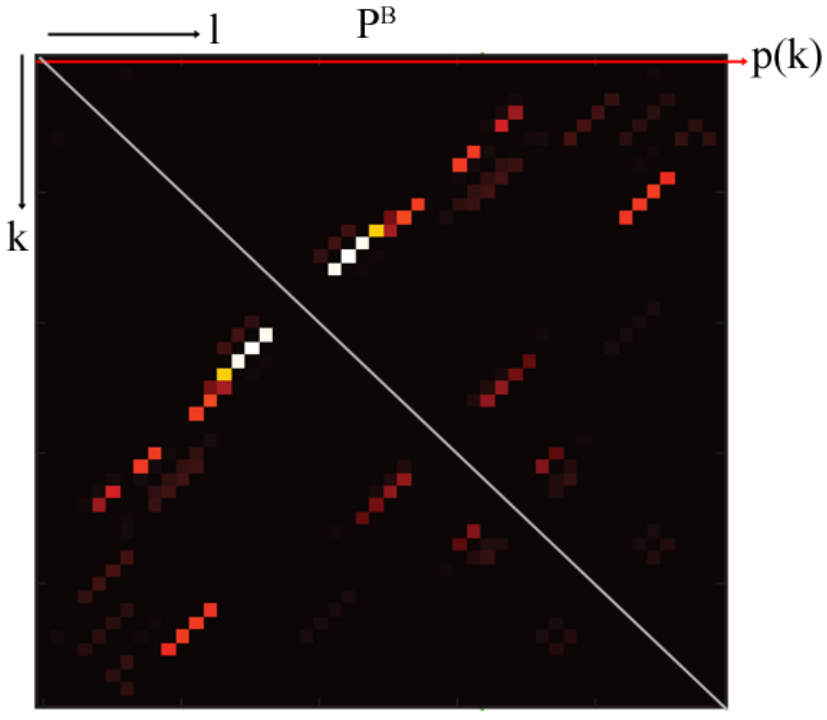




Problems

- Correlation coeffecient is inversly proportional to sequence length
- Local structure comparison - time consuming



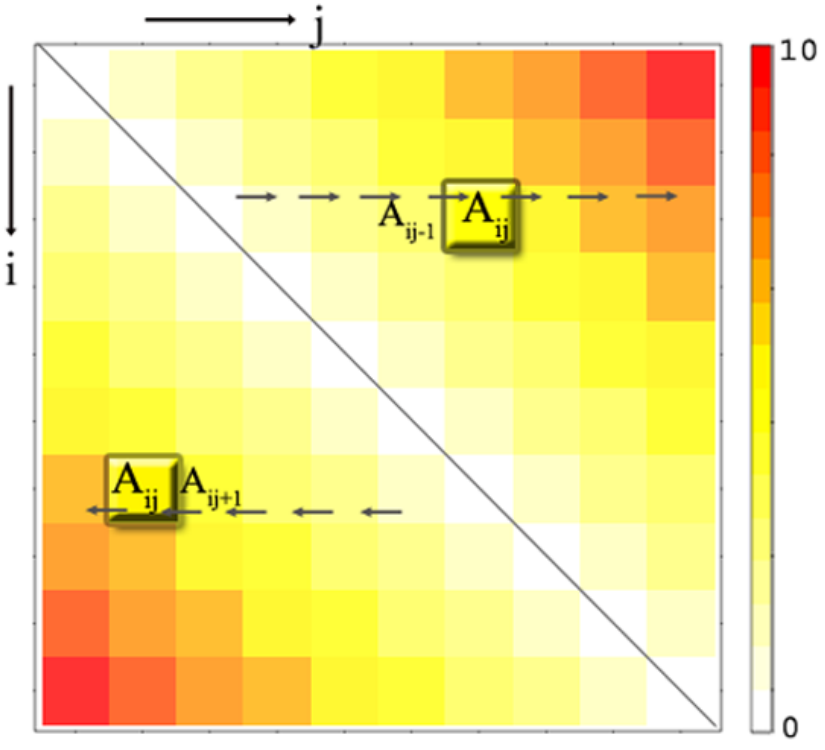


Local structure comparison

Accumulated Score matrix

pseudocode

```
upper diagonal
for (j > i):
  A_ij = A_ij−1 + P^A_ij
lower diagonal
for (j < i):
  A_ij = A_ij+1 + P^A_ij
```

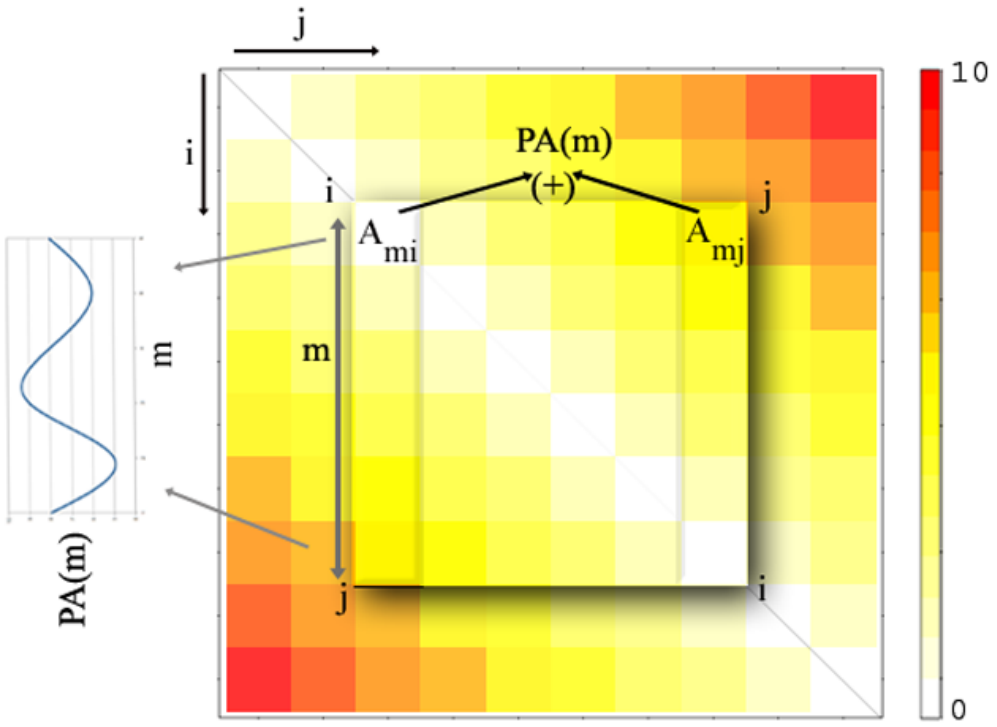


Local structure comparison

Comparing base pair probability matrices

pseudocode

```
len = length of sequence
for i = 0 to len
for j = len to i+1
 for m = i to j
 PA[m]=A[m,j] + A[m,i]
 push(PAs,PA[i])
 PB[m]=B[m,j] + B[m,i]
 push(PB',PB[i])
 endfor
 r(i,j) = corr(PAs,PBs)
 endfor
endfor
```



Material

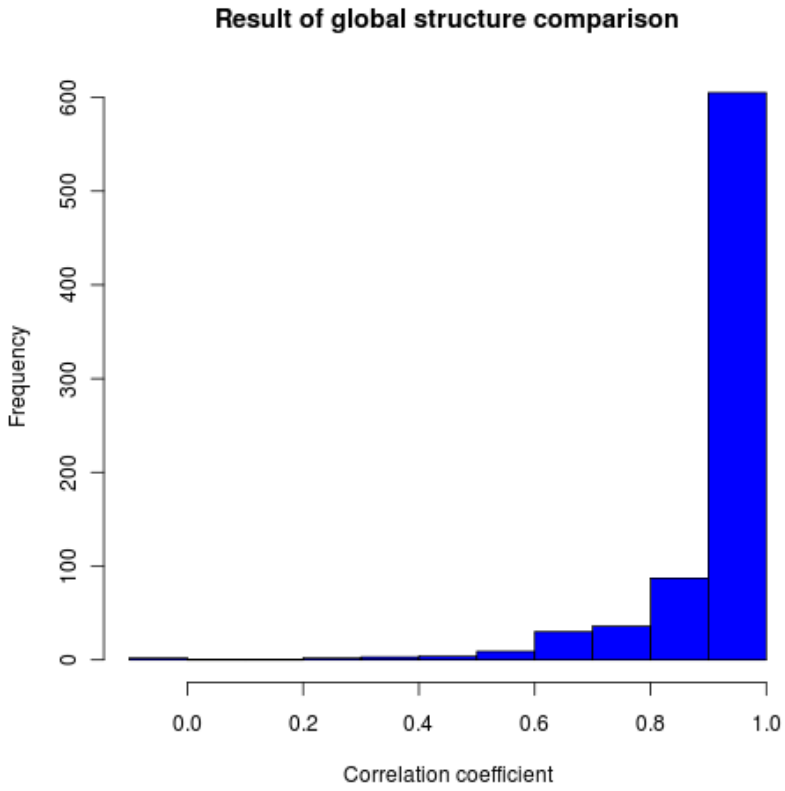
Data Description

- Genome scan of all known disease-associated SNPs in Human Gene Mutation Database (HGMD)
- 514 disease-associated SNPs in 350 regulatory RNAs (Halvorsen et al., 2010)
- Of these, 206 - 5'UTRs, 132 - 3'UTRS and 12 - ncRNAs
- SNPs were mapped only to the untranslated regions of mature mRNA and are at least 10nt away from any transcription or translation start or stop sites.

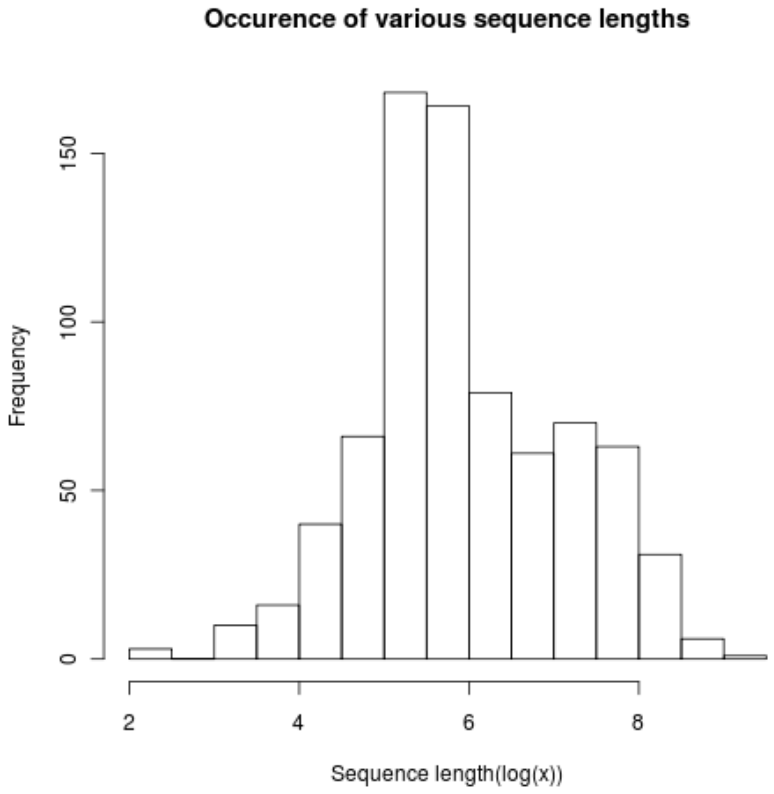
Results of global comparison

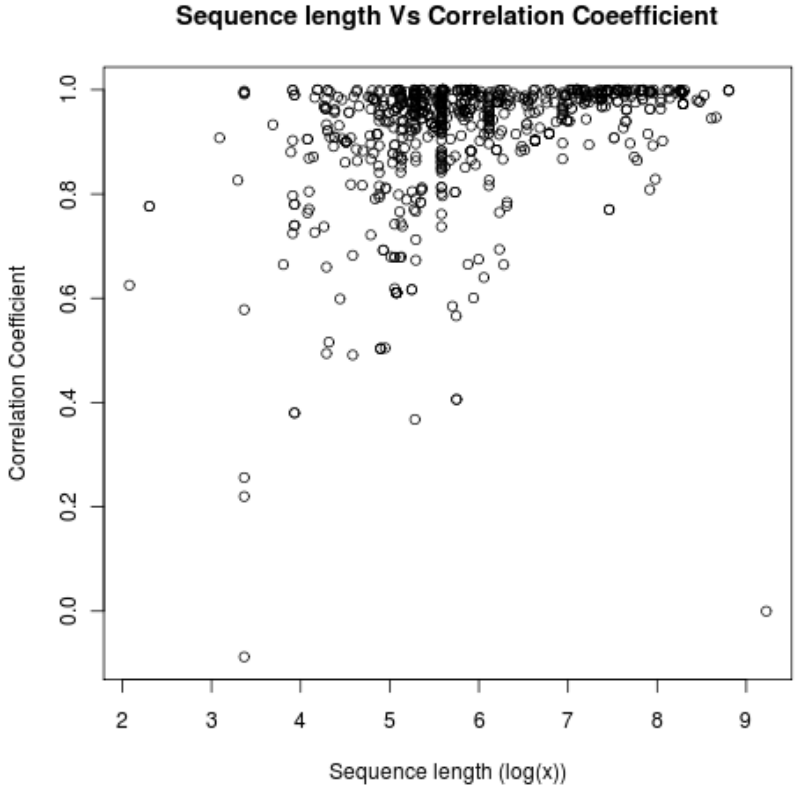
Impact of SNPs in RNA structure

Among the 514 disease-associated SNPs, majority of SNPs impart less global conformational changes (r = 0.9 - 1), that represents the impact of local conformational changes in regulatory RNAs



Results of global comparison

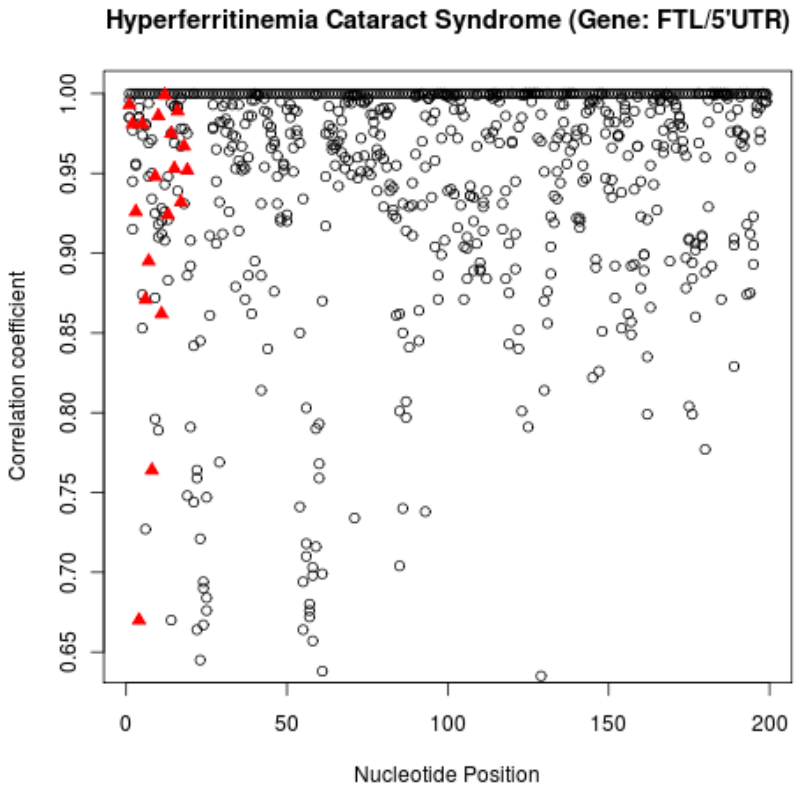


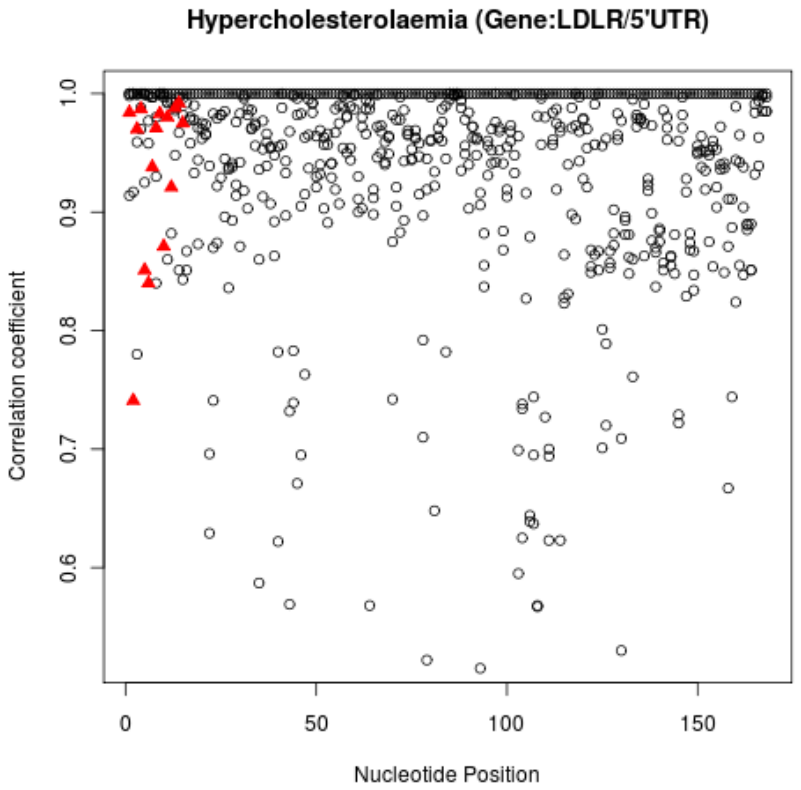


Results of global comparison

Analysis of SNP's in selective RNAs

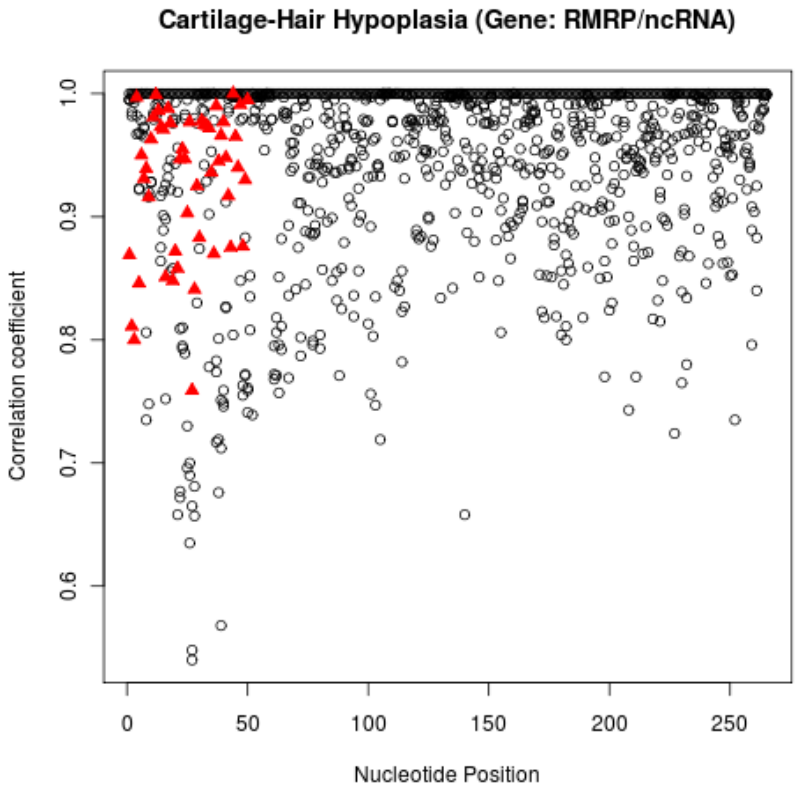
Result of global structure comparison from all (N X 3) possible SNPs in a RNA. The selective RNAs given below are the ones having higher number (> 10) of known SNPs from HGMD, that are profoundly associated with local conformational changes (highlighted in red triangle)

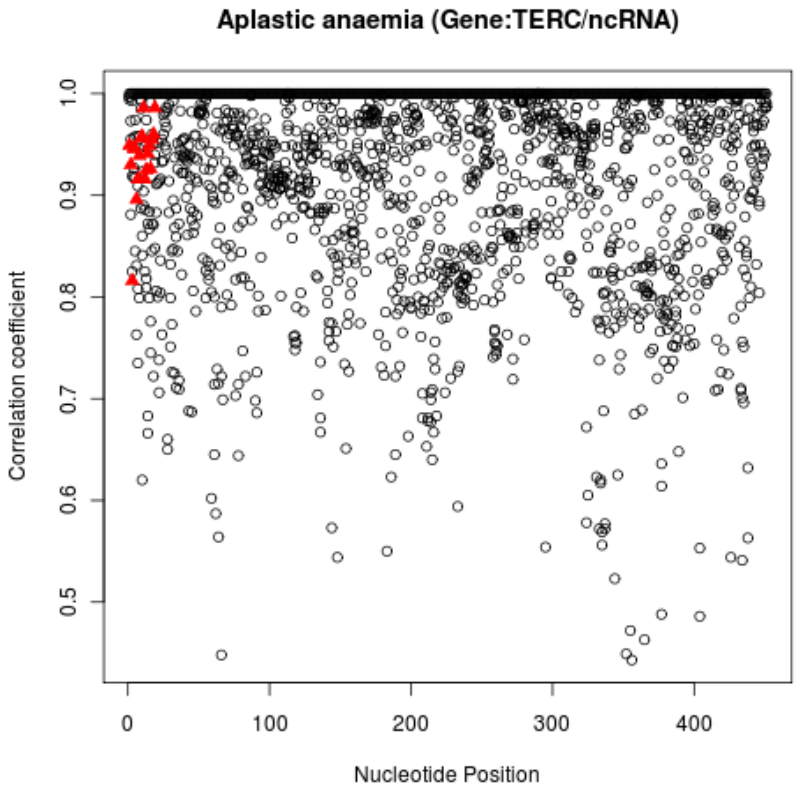




Results of global comparison

Analysis of SNP's in selective RNAs





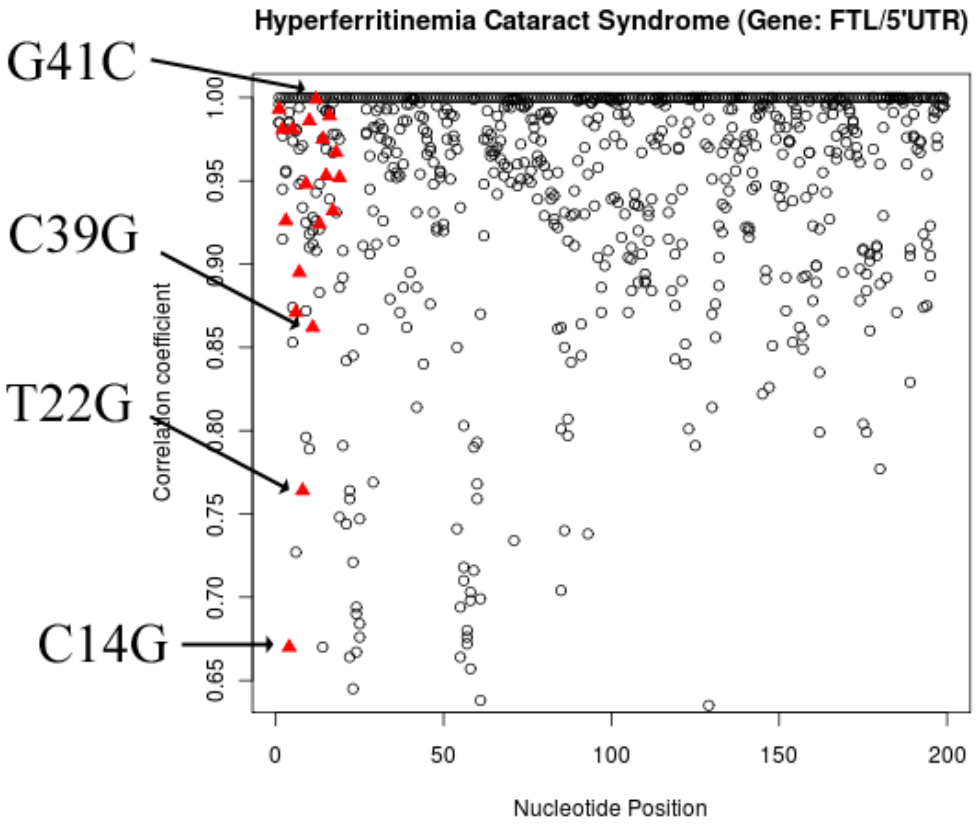
Local structure comparison

Case study: Hyperferritinemia Cataract Syndrome

- Hereditary Hyperferritinimeia Cataract Syndrome is an autosomic dominant disorder caused by heterogeneous mutations on the iron-responsive element (IRE) of ferritin L-chain mRNA.
- The mutations in 5'UTR regions distrubs the structure of IRE which alters the binding affinity of IRP (Iron Response Protein) leading to aberrant FTL regulation.
- In wildtype sequence, the position of IRE element is predicted between 30 to 52 bases using UTRscan.

Local structure comparison

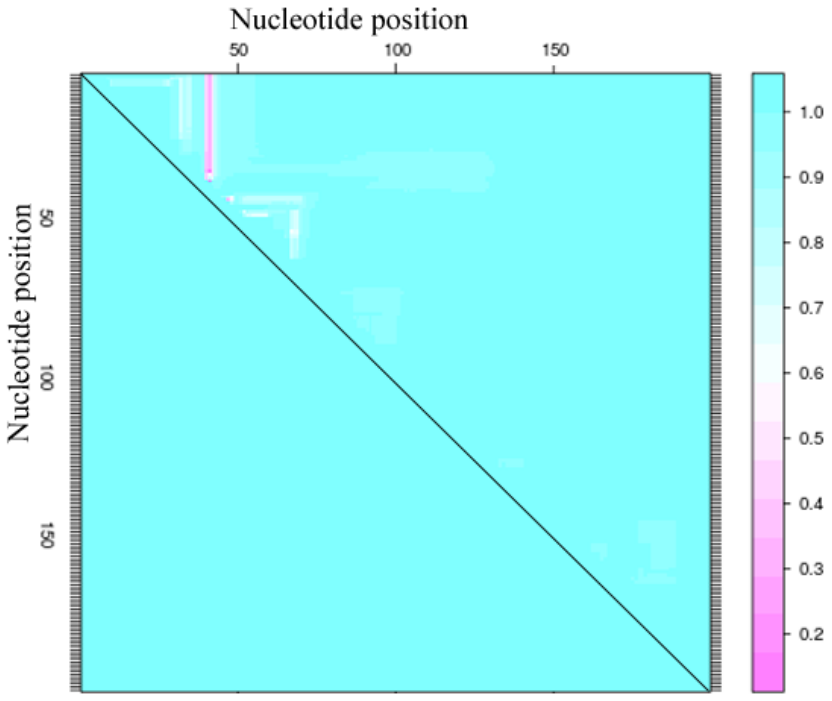
Selecting some known SNPs to analyse locally distrupted regions using the proposed method



Results of local structure comparison

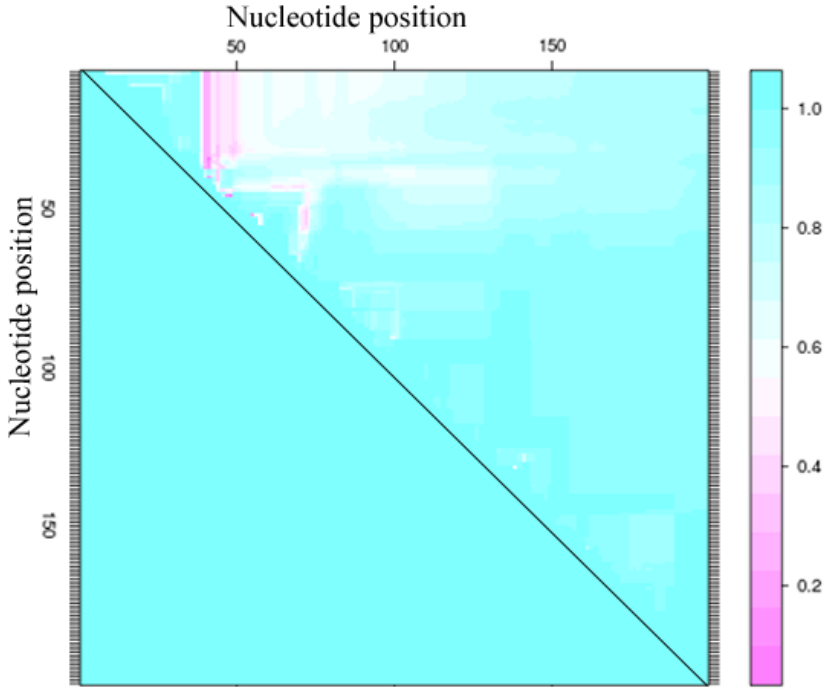
Hyperferritinemia Cataract Syndrome (Gene: FTL/5'UTR)

G41C



Global=0.999
Local=0.342 (substructure:1-42)

C39G

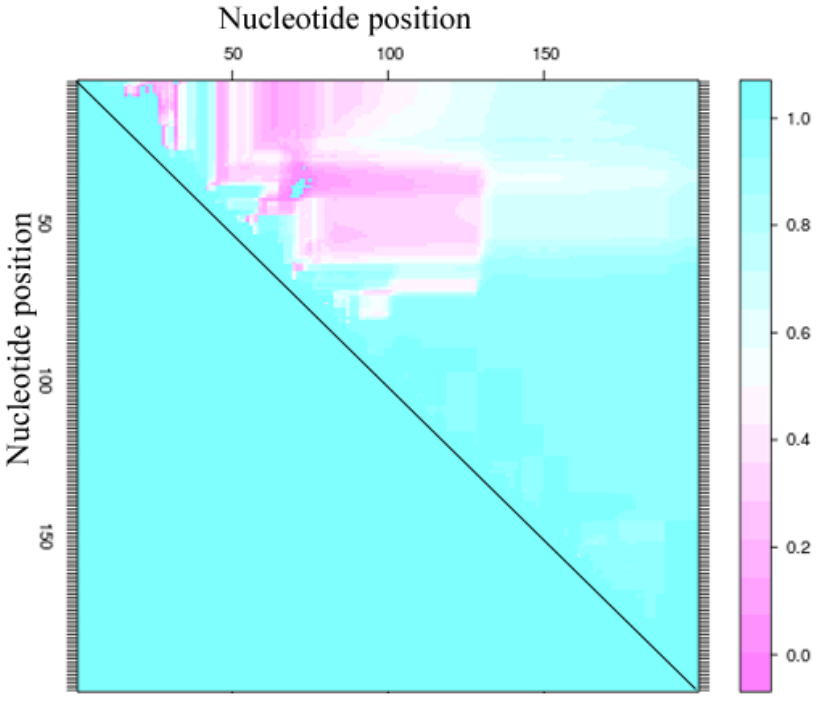


Global=0.862
Local=0.247 (substructure: 1-40)

Results of local structure comparison

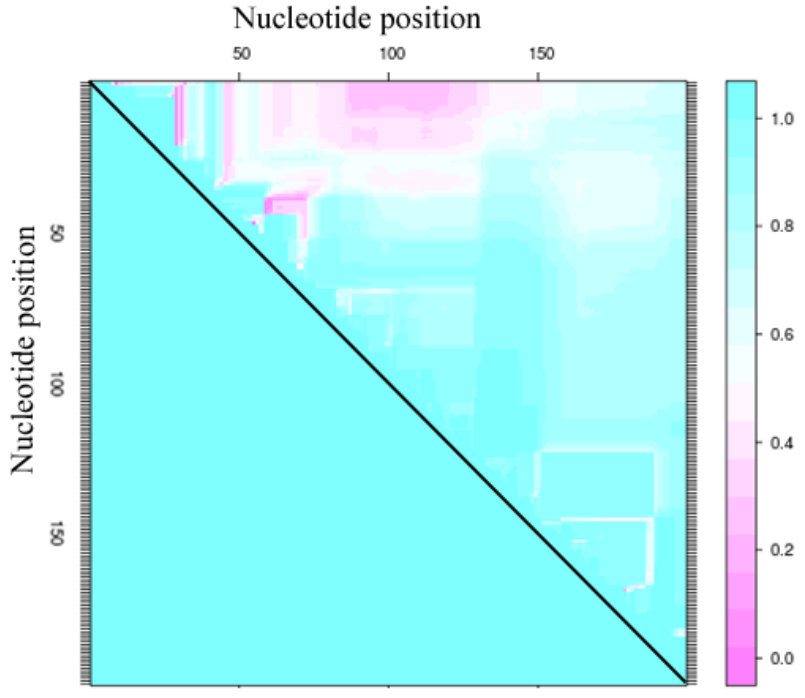
Hyperferritinemia Cataract Syndrome (Gene: FTL/5'UTR)

T22G



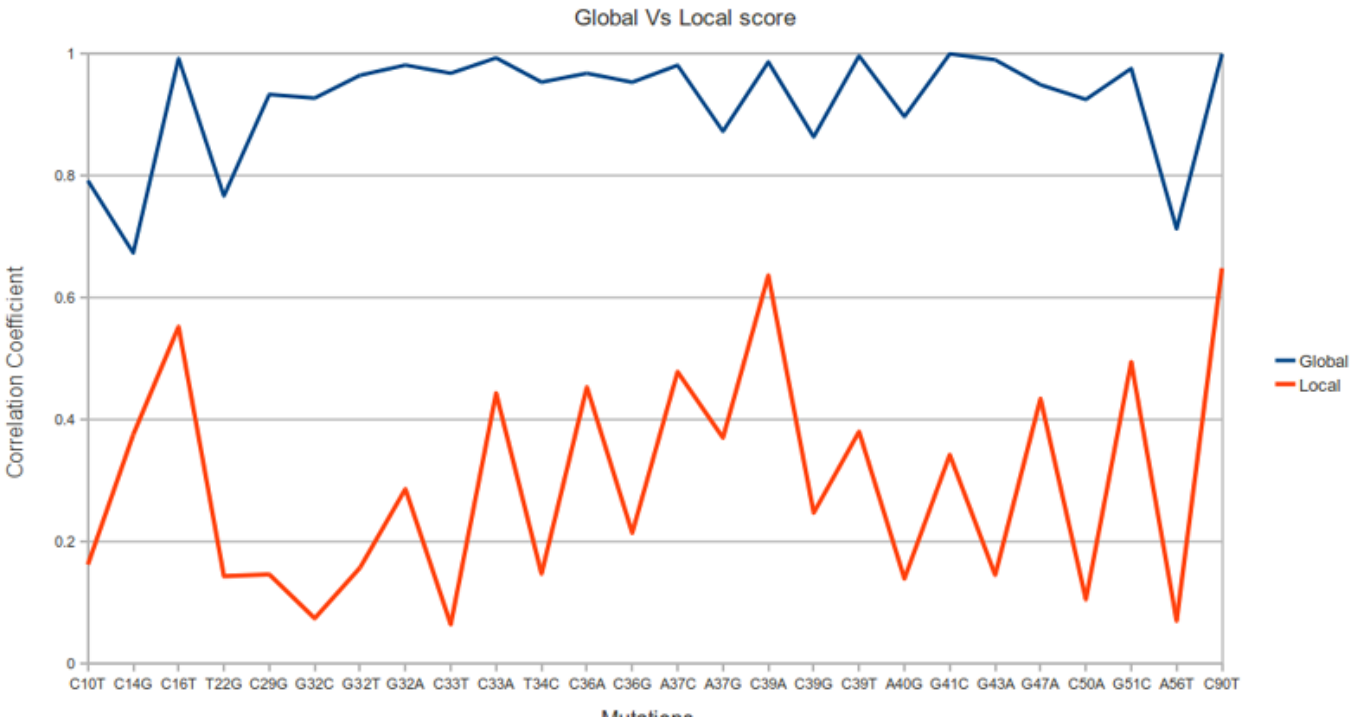
Global=0.764
Local=0.221 (substructure: 1-72)

C14G



Global=0.670
Local=0.375 (substructure: 1-139)

Hyperferritinemia Cataract Syndrome (Gene: FTL/5'UTR)



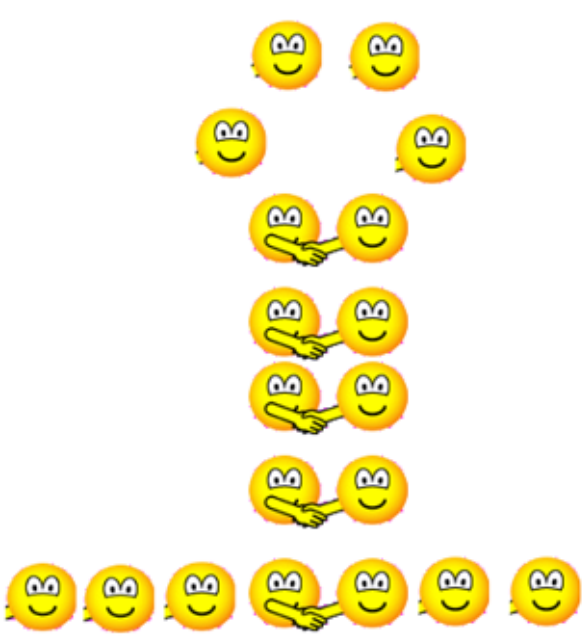
Further work

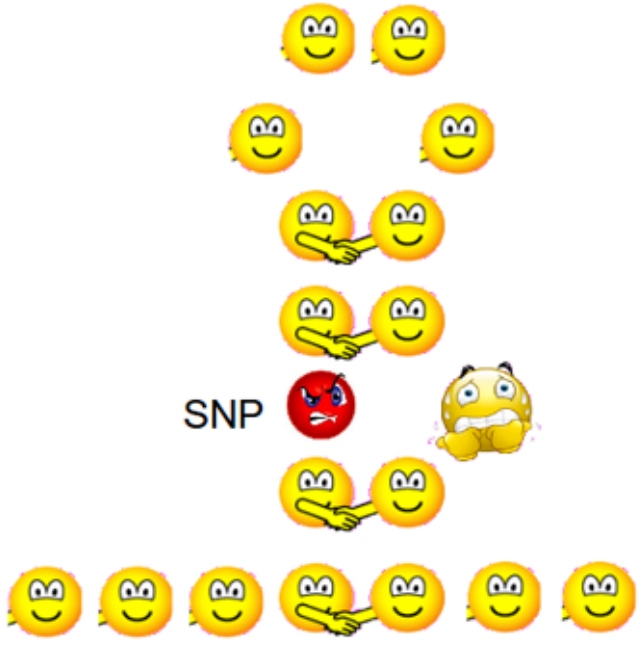
More !! to do..

- Test with more data's for Optimization
- Extend this method to identify the impact of SNPs in RNA structure prediction from multiple sequence alignment
- Genome wide study of SNP associated phenotypes in PIG Genome

Acknowledgement

- Jan Gorodkin
- Stefan and other colleagues of RTH group
- Faculty of Life Sciences, KU
- Alain Laederach





Thank you for your attention