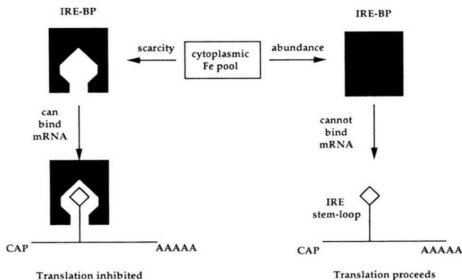# Measures to detect the effect of SNPs on RNA secondary structure

Sabarinathan Radhakrishnan

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

RNA structure and function

- Structural features are essential for the proper function of many ncRNAs and *cis*-acting regulatory elements.
- e.g. Iron responsive element (IRE)

- ▸ Structural changes in IRE - aberrant FTL gene regulation - hereditary hyperferritinemia-cataract syndrome

▸ Structural changes in IRE - aberrant FTL gene regulation - hereditary hyperferritinemia-cataract syndrome

▸ Structural changes in tRNA - wide variety of diseases including diabetes, cardiomyopathies, etc.,

Effect of SNPs on RNA structure and function

- Structural changes in IRE - aberrant FTL gene regulation - hereditary hyperferritinemia-cataract syndrome
- Structural changes in tRNA - wide variety of diseases including diabetes, cardiomyopathies, etc.,
- Mutation tolerant RNA viruses - the RNA structural change inhibit the replication and translation initiation processes.
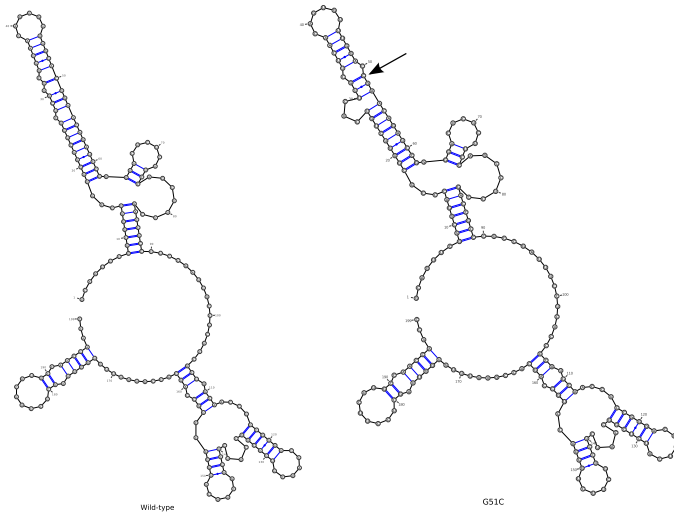
Global Structural change
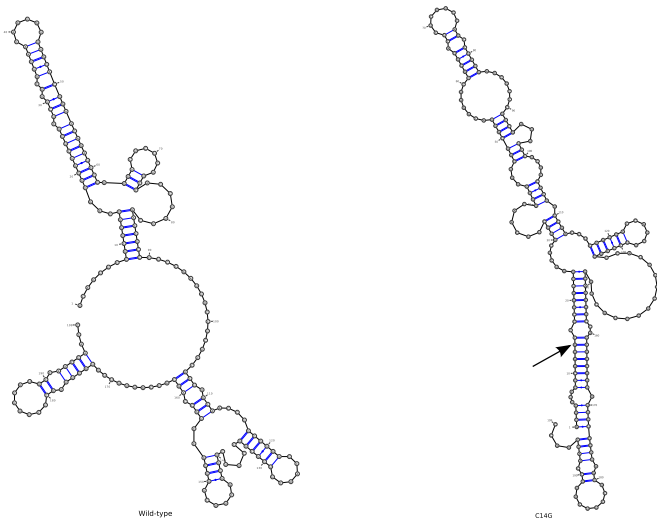
Example: Structure of 5'UTR region of FTL mRNA

Wild-type

C14G

Local Structural change

Example: Structure of 5'UTR region of FTL mRNA

Wild-type

G51C

► **Resources**
  ► `RNAmute` (Churkin and Barash, 2006; 2008)
  ► `RDMAS` (Shu et al., 2006)
  ► `RNAmutants` (Waldispuhl et al., 2008; 2009)
  ► `SNPfold` (Halvorsen et al., 2010)

- **Resources**
  - `RNAmute` (Churkin and Barash, 2006; 2008)
  - `RDMAS` (Shu et al., 2006)
  - `RNAmutants` (Waldispuhl et al., 2008; 2009)
 - `SNPfold` (Halvorsen et al., 2010)

- **Limitations**
  - Uses the global structure prediction method
  - Long-range base pairs have limited accuracy
  - Less importance for local structural changes

- ▸ Local structural change in binding regions (Westerhout et al., 2005; Abbink et al., 2008; Hemert et al., 2008; Grover et al., 2011).
- ▸ e.g., structural change affects the binding of siRNA in Nef gene in human immunodeficiency virus (HIV-1) (Westerhout et al., 2005).
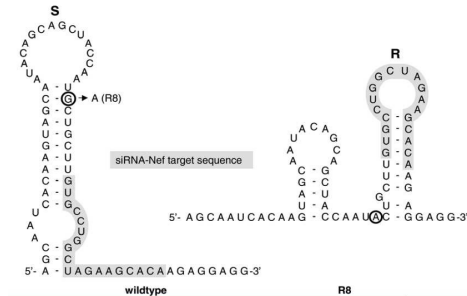
- Develop a method to detect SNP induced structural change in local regions of an RNA structure.
- Employ different measures to compute the difference between the base-pair probabilities of wild-type and mutant structures.
- Pipeline for genome-wide analysis

RNA Folding

- Base-pair probabilities (P) of ensemble secondary structures - partition function (McCaskill, 1993) – `RNAfold` and `RNAplfold`
- P_{ij} are the probabilities of nucleotides i and j form a base pair

- Difference between wild-type (P) and mutant (P^*) can be measured by,
  - Euclidean base pairing distance,

$$d^2(P, P^*) = \sum_{i<j}(P_{ij} - P^*_{ij})$$

  - Pearson correlation coefficient,

$$r(P, P^*) = cov(P, P^*)/\sqrt{var(P)var(P^*)}$$

▸ Difference between wild-type (P) and mutant (P^*) can be measured by,

    ▸ Euclidean base pairing distance,

$$d^2(P, P^*) = \sum_{i<j} (P_{ij} - P^*_{ij})$$

position-wise pairing probabilities (π),

$$\pi_i = \sum_j P_{ij}$$

    ▸ Pearson correlation coefficient,

$$r(P, P^*) = cov(P, P^*) / \sqrt{var(P)var(P^*)}$$

- Difference between wild-type (P) and mutant (P^*) can be measured by,
  - Euclidean base pairing distance,

$$d^2(P, P^*) = \sum_{i<j} (P_{ij} - P_{ij}^*)$$

  position-wise pairing probabilities (π),

$$\pi_i = \sum_j P_{ij}$$

$$d^2(\pi, \pi^*) = \sum_{i=1}^{n} (\pi_i - \pi_i^*)$$

  - Pearson correlation coefficient,

$$r(P, P^*) = cov(P, P^*) / \sqrt{var(P)var(P^*)}$$

$$r(\pi, \pi^*) = cov(\pi, \pi^*) / \sqrt{var(\pi)var(\pi^*)}$$

▸ compute the difference for all sequence intervals [k,l] in the given sequence $1 \leq k \leq l \leq n$,

$$\pi_i[k, l] = \sum_{j=k}^{l} P_{ij}$$

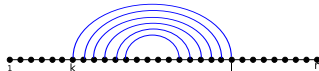- compute the difference for all sequence intervals [k,l] in the given sequence $1 \leq k \leq l \leq n$,
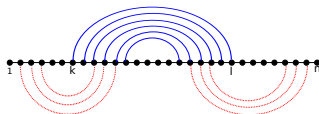
$$\pi_i[k, l] = \sum_{j=k}^{l} P_{ij}$$

compute $d^2_{[k,l]}(\pi, \pi^*)$ and $r_{[k,l]}(\pi, \pi^*)$

Local Structural (Dis)similarities

▸ compute the difference for all sequence intervals [k,l] in the given sequence $1 \leq k \leq l \leq n$,

$$\pi_i[k, l] = \sum_{j=k}^{l} P_{ij}$$

compute $d^2_{[k,l]}(\pi, \pi^*)$ and $r_{[k,l]}(\pi, \pi^*)$
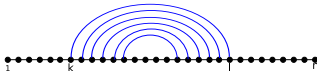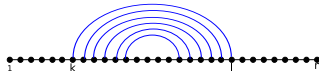
▸ max $d^2_{[k,l]}(\pi, \pi^*)$ and min $r_{[k,l]}(\pi, \pi^*)$ are selected as best hit local region.

▸ compute the difference for all sequence intervals [k,l] in the given sequence $1 \le k \le l \le n$,

$$\pi_i[k, l] = \sum_{j=k}^{l} P_{ij}$$

compute $d^2_{[k,l]}(\pi, \pi^*)$ and $r_{[k,l]}(\pi, \pi^*)$

▸ max $d^2_{[k,l]}(\pi, \pi^*)$ and min $r_{[k,l]}(\pi, \pi^*)$ are selected as best hit local region.

▸ **Identify the potential SNPs on genome**

  ▸ Compute the local base pairing probabilities using `RNAplfold` with the parameters W=200 and L=120.

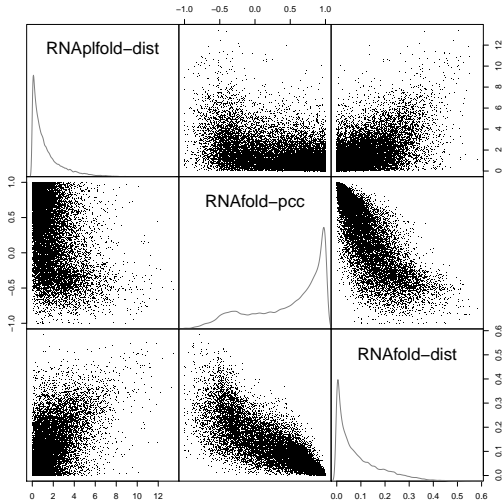 ▸ Use the distance measure with fixed window length (h'=200 and h''=20)

$$d_{(k)}^2(P, P^*) = \sum_{i=k}^{k+h'} \sum_{j=i}^{i+h''} (P_{ij} - P_{ij}^*)^2$$

  ▸ RNAplfold-dist score - $\max_k$ $d_{(k)}$

▸ **Identify the potential SNPs on genome**

    ▸ Compute the local base pairing probabilities using `RNAplfold` with the parameters W=200 and L=120.

    ▸ Use the distance measure with fixed window length (h'=200 and h''=20)

$$d_{(k)}^2(P, P^*) = \sum_{i=k}^{k+h'} \sum_{j=i}^{i+h''} (P_{ij} - P_{ij}^*)^2$$

    ▸ RNAplfold-dist score - $\max_k d_{(k)}$

▸ **Identify the local structural change**

    ▸ Consider the sequences flanking (+/- 200) to SNP position and compute base pairing probabilities using `RNAfold`

    ▸ Compute the difference using $d_{[k,l]}^2(\pi, \pi^*)$ (RNAfold-dist) and $r_{[k,l]}(\pi, \pi^*)$ (RNAfold-pcc) for all sequence intervals[k,l].

- Collected a set of disease-associated SNPs (dSNPs) and predicted microRNA target sites on 3'UTRs of human genes - miRdSNP database (Bruno et al., 2012).
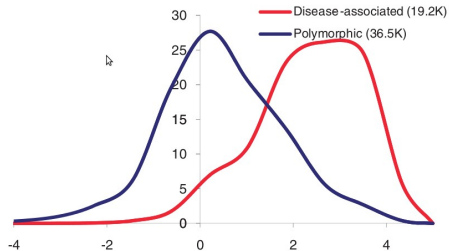
	dSNPs	polymorphic
No. of variants in 3'UTR	611[a]	1433[b]
No. of miRNA target sites mapped[c]	546	689
Local structural change in miRNA target sites	?	?

[a]461 - Genes, 231 diseases
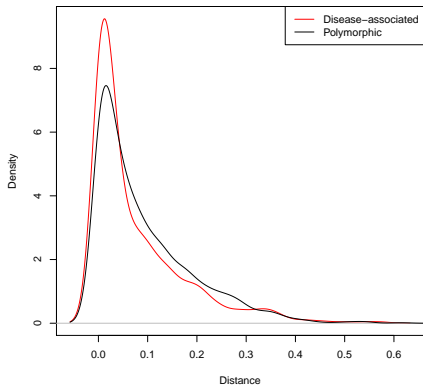[b]present in flanking regions(+/-200) of dSNPs
[c]target sites are predicted using `TargetScan5.0` and `PicTar`

dSNPs Vs polymorphic variants

dSNPs Vs polymorphic variants
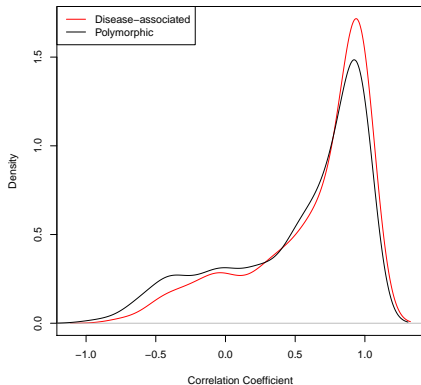
- Collected a set of disease-associated SNPs (dSNPs) and predicted microRNA target sites on 3'UTRs of human genes - miRdSNP database (Bruno et al., 2012).

	dSNPs	polymorphic
No. of variants in 3'UTR	611[a]	1433[b]
No. of miRNA target sites mapped[c]	546	689
Local structural change in miRNA target sites	102	159

[a]461 - Genes, 231 diseases
[b]present in flanking regions(+/-200) of dSNPs
[c]target sites are predicted using `TargetScan5.0` and `PicTar`
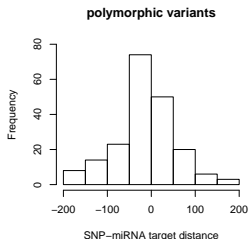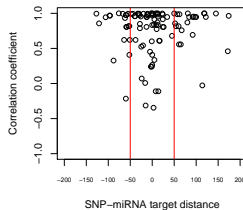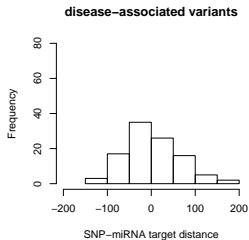
Structural changes in miRNA target sites

Distance between the SNP position and miRNA target site

disease−associated variants
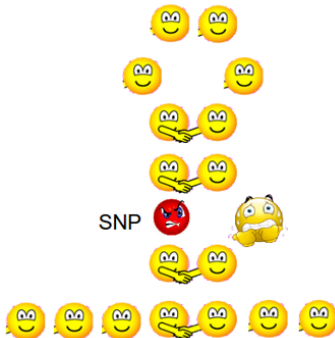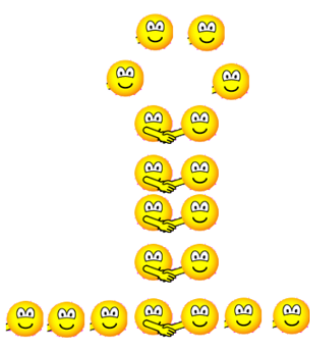
polymorphic variants
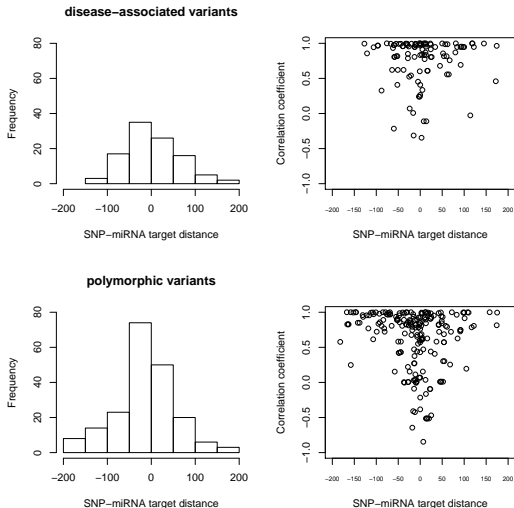
Structural changes in miRNA target sites

Distance between the SNP position and miRNA target site
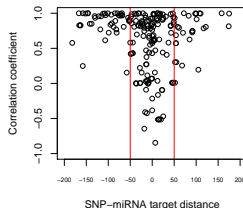
- ► Stefan Seemann
- ► Hakim Tafer
- ► Peter F. Stadler
- ► Jan Gorodkin
- ► All colleagues of RTH

SNP

Thanks for your attention