snoRNA Target Interaction and Prediction
part of “interactions of RNAs and proteins”

Sonja Prohaska

Computational EvoDevo
University Leipzig

May 19, 2017
Structure of box H/ACA snoRNAs

Function: guide pseudouridination of uridine by dyskerin

- has two sequence motifs:
  - H box: 5’-ANANNA-3’, ACA box: 5’-ACA-3’
- target RNA binds to an internal loop of the snoRNA
- only few mismatches and symmetric internal loops of length 2 or 4
- target position for pseudouridylation between the two duplexes
- enzyme: DKC1 = Dyskerin, pseudouridine synthase
- makes complex with: GAR1, NHP2, NOP10, DKC1
Classes of snoRNAs

names of proteins are taken from yeast
Prediction of H/ACA box snoRNA targets

**RNAsnoop**

- two stem-loop structures
- dynamic programming, thermodynamic folding
- calculate (unbranched) hairpin loop (H)
- calculate left duplex between snoRNA and target RNA (L)
- calculate right duplex between snoRNA and target RNA (R)
- machine learning trained on set of functional snoRNAs

**Constraints:**

- take T as given
- duplexes allow only single and tandem mismatches
- target position (i-2) of U to be modified follows left duplex
Prediction of H/ACA box snoRNA targets

- use pattern search and RNA folding to find the 'ACA' and 'ANNANA' motifs and the two stem-loop components
- compute interaction structure separately for the two structures
- calculate left side $L$, calculate unbranched fold $M$ then calculate right side $R$
consider the *intra*-molecular interaction of subsequence $y[p, q]$ (same as $(y_p...y_q)$) of the snoRNA sequence $y$

build the best stem-loop structure $M_{p,q}$ from

- a hairpin loop $H$
- and interior loops (incl. stacked basepairs) $I$
- if $k = 1$ and $l = 1$ the “interior loop” is a stacked basepair
- if $k > 1$ or $l > 1$ it is an interior loop
RNAsnoop: unbranched fold $M$

$$M_{p,q} = \min \left\{ \mathcal{H}(y[p, q]) \right.$$ \hspace{1cm} \left. \min_{k,l} M_{p-k, q+l} + \mathcal{I}(y[p - k, p], y[q, q + l]) \right\}$$

- either start with the hairpin loop $\mathcal{H}$
- or extend the stem with an interior loop $\mathcal{I}$
look at basepair $y_i, x_j$

- $x$ is the target RNA, $y$ is the snoRNA
- index $i$ runs along the target RNA $x$ (from 5’ to 3’)
- index $j$ runs along the target RNA $y$ (from 5’ to 3’)
- only symmetric interior loops of length 2 or 4 are allowed in $L$
- to $L_{i-k, j+k}$ we add the an interior loop $\mathcal{I}$
**RNAsnoop: Left side**

- **stacked base pairs**
  - \(k=1\)
  - \(j\) \(i\) \(j+k\) \(i-k\)
  - \(y\) \(x\)

- **interior loop**
  - **length 2**
    - \(k=2\)
    - \(j\) \(i\)
    - \(j+k\) \(i-k\)
  - **length 4**
    - \(k=3\)
    - \(j\) \(i\)
    - \(j+k\) \(i-k\)

- **k \(!=l\)**
  - \(j\) \(i\)
  - \(j+k\) \(i-l\)

\[ L_{i,j} = \min_{k=1,2,3} L_{i-k,j+k} + I(x[i-k,i], y[j,j+k]) \]  \(\text{(2)}\)
**RNAsnooP: Right side**

- Look at closing basepair $y_i, x_j$
- We combine the left side $L$, the stem-loop $M$ and the pseudouridine-loop
- The pseudouridinylation site at $i - 2$ and a nucleotide at $i - 1$ need to be unpaired ($\Psi$-$N$)
- $x_{i-3}$ has to contribute a basepair to $L$
- $x_i$ has to contribute a basepair to $R$
\[ R_{i,j} = \min \begin{cases} \min_{k,l \leq 2} R_{i-k,j+l} + \mathcal{I}(x[i-k,i], y[j,j+l]) \\ \min_{l \in [3, |y| - j]} L_{i-3,j+l+1} + M_{j+1,j+l} \\ \text{if } x_{i-2} = 'U' \end{cases} \quad (3) \]

- the second term starts the right duplex by adding the left side \( L \) and the stem-loop \( M \)
- the first term continues the right side with an interior loop (or stacked pair) that has the same constraints on loop sizes as the left side
- notice the constraints on the positioning of \( \Psi \) at \( x_{i-2} \)
How to validate the target prediction

▶ rank predicted targets according to their MFE
▶ see if two or more H/ACA structures are adjacent
▶ check overlap with existing annotation
▶ varify predictions in the lab with a functional test

Outlook
Use a reverse strategy to design a snoRNA that converts a specific U in a target RNA into a pseudouridine.

Williams GT and Farzaneh F. (2012) *Are snoRNAs and snoRNA host genes new players in cancer?* Nature reviews 12,84-88