

RNAPLEX AND SEED SEARCH WITH SUFFIX TREE

Hakim Tafer

TBI
University of Vienna

BLED, 2007

- Why study ncRNA-RNA interaction?
- General method to search for ncRNA-RNA interaction
- RNAplex
- siRNA target site accessibility study

- Intensive development of tools for the genome-wide detection of ncRNA
- The function of the majority of the ncRNA families is unknown.
- However many ncRNA families act on RNA by RNA-RNA interaction, like miRNA, siRNA, viRNA, gRNA, snoRNA.
- We have a huge amount of molecules whose functions are unknown

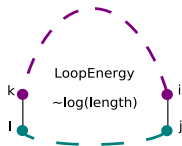
- Identify possible hybridization site
 - General search tools BLAST, FASTA.
 - fast but lack of sensitivity
 - indexing system used heavy
 - ViennaRNA RNAduplex: Slow but high sensitivity
- Find characteristics that confirm putative binding sites

- Computes the hybridization energy of two RNA sequences
- Only intermolecular interactions
- No information about target site accessibility
- Returns all possible interactions above a given threshold

- RNAPlex derives from RNAduplex
- Scan the whole genome on a PIV 3.2[GHz] for miRNA targets in 50 minutes
- 100 times faster than RNAduplex
- Returns non-overlapping interactions

RNAPLEX VS RNADUPLEX

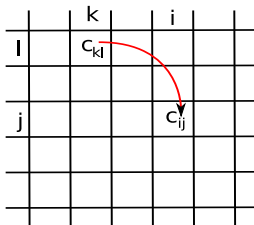
RNA Duplex



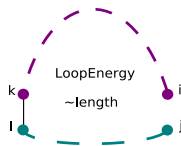
$$c[i][j] = \min(c[k][l] + \text{LoopEnergy}(i, j, k, l))$$

$$c[i][j-1] = \min(c[i][j] + p)$$

$$\text{CPU} \sim O(NMF^3)$$

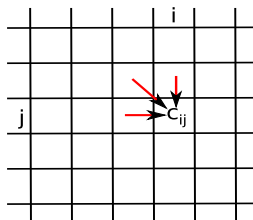


RNAplex



$$\min \begin{pmatrix} c[i-1][j] + \text{LoopEnergy}(i, j, i-1, j) \\ c[i-1][j+1] + \text{LoopEnergy}(i, j, i-1, j+1) \\ c[i][j+1] + \text{LoopEnergy}(i, j, i, j+1) \end{pmatrix}$$

$$\text{CPU} \sim O(NM)$$



- RNAPlex too slow for RNA-RNA interaction search
- 30 days to scan 70000 ncRNA on human mRNAdb on 30 cluster
- Use filter to avoid useless use of DP
- Search for seed with GUUGLE

- Report Matches of length k between the target and the query
- GU mismatches are allowed
- Algorithm based on the WOTD suffix tree construction
- FAST: 55s to scan 2-8 miR-375 against whole human 3UTR

Target

ATGAGTC

TGAGTC

GAGTC

AGTC

GTC

TC

C

Query

TACAG

ACAG

CAG

AG

G

SORT:

A<G<C<T

ATGAGTC

AGTC

GAGTC

GTC

C

TGAGTC

TC

A>G>C>T

TACAG

CAG

G

AG

ACAG

MATCH:

ATGAGTC

TACAG

AGTC

GAGTC

CAG

GTC

C

G

TGAGTC

AG

TC

ACAG

ATGAGTC

TACAG

AGTC

GAGTC

CAG

GTC

C

G

TGAGTC

AG

TC

ACAG

ATGAGTC

TACAG

AGTC

SORT:
MATCH:

AGTC

TACAG

ATGAGTC

MATCH:

ATGAGTC

TACAG

ATG
TAC

GUUGLE+RNAPLEX

ATCGTTCGTA

GCGCCCCGGGGTAGCTTCCATCCGCATGC

ATCGTTCGTA ATCGTTCGTA
GCGCCCCGGGGTAGCTTCCATCCGCATGC
 ATCGTTCGTA

GCGCCCCGGGGTAGCTTCCATCCGCATGC
 |||| |||
 ATCGTTCGTA

HYBRIDIZATION

- GUUGLE reports seed
- RNAPlex hybridizes
- Report hit above threshold

MIRNA TARGET SEARCH WITH GUUGLE+RNAPLEX

- miRNA targets retrieved from tarbase
- the smallest of the longest stretch : 7nt
- We can search for miRNA target with a 7nt seed
- RNAPlex+GUUGLE is 2-4 times faster than RNAPlex
- Are we happy?

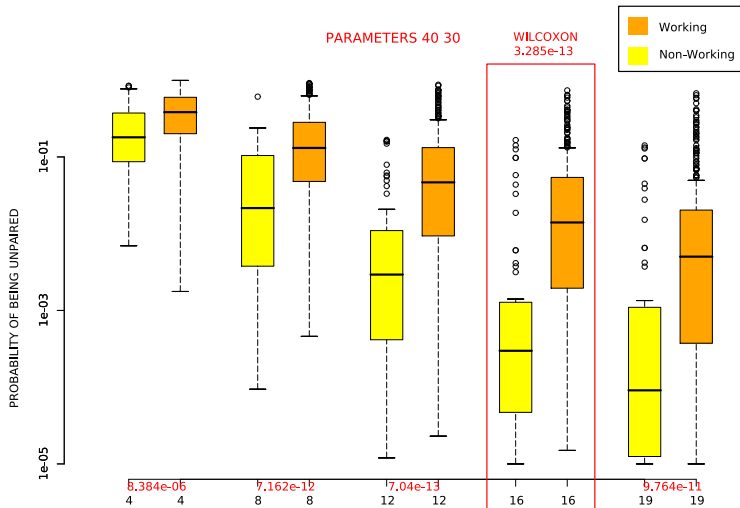
RNAPLEX+GUUGLE SENSITIVITY

- RNAPlex+GUUGLE sensitivity depends on seed size
- Seed size of 2 necessary to obtain RNAPlex sensitivity
- At this seed size RNAPlex+GUUGLE 6-10 times slower
- RNAPlex+GUUGLE/RNAPlex depends on the problem

- Schubert et al. and Overhoff et al. 2005
 - Target Accessibility study
 - The target site should be as free as possible
 - Can we detect this with our tools?
- Use RNAplfold to obtain target site accessibility

- Local version of RNAfold
- Return base pairs probabilities
- Return probability that a stretch of u nucleotides is unpaired
- Runs in $\sim O[NL^2]$ where L is the windows size
- Can be used for genome wide accessibility study

TARGET SITE ACCESSIBILITY



CONCLUSION AND OUTLOOK

- Experimentally target site accessibility plays a role in siRNA efficiency
- We can see this with RNAplfold
- RNAplfold : -80% of false positive
- Coupled with RNAplex fast, reliable siRNA prediction tool
- Look for other descriptors
- Applicable on miRNA?
 - No such correlation has been found
 - Few available data
 - Kinetics may play a bigger role than thermodynamics

THANK YOU

Ivo, Berni, Gregor, Joerg, Peter