

RAIN: RNA–protein Association and Interaction Networks

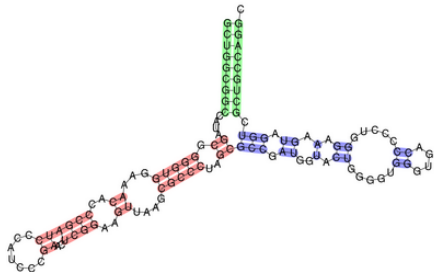
Alexander Junge (ajunge@rth.dk)

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

Computational Analysis of RNA Structure and Function
Benasque, Spain
July 28, 2015

non-coding RNAs (ncRNAs)

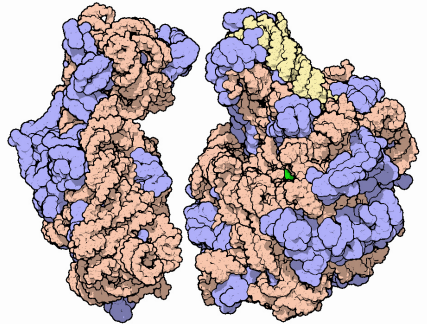
- RNAs not coding for proteins exist in all kingdoms of life
- Fulfill diverse set of biological functions
- Functions rely on **interactions** with proteins, coding RNAs and other non-coding RNAs



ncRNA-Protein Interactions

Example: **ribosomal RNAs (rRNAs)**

- Ribosome apparatus consists of two subunits
- Each subunit composed of rRNA and different proteins
- Catalytic core made of rRNAs
⇒ Ribosome is ribozyme
- Proteins form scaffold



≈80 proteins

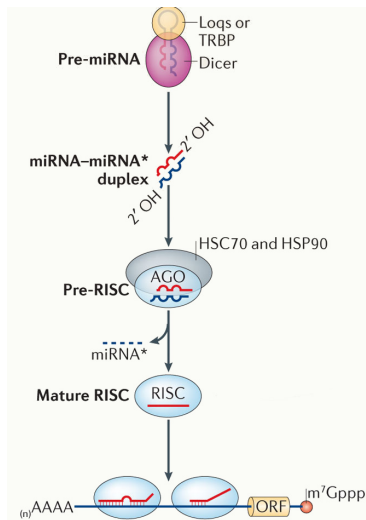
Ribosomal RNAs

PDB Molecule of the Month series by David S. Goodsell, October 2000

ncRNA-mRNA Interactions

Example: **microRNAs** (**miRNAs**)

- Short, single-stranded RNA molecules of ≈ 22 nucleotides
- Basepairing with complementary target mRNAs results in
 1. Translation inhibition or
 2. mRNA degradation



Ameres & Zamore, Nature Reviews
Molecular Cell Biology, 2013

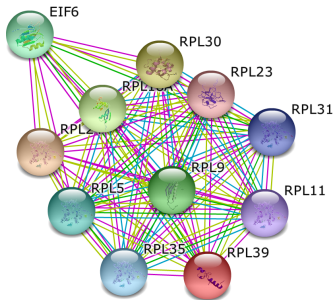
Current Understanding of ncRNAs and Proteins

Interactions of ncRNAs are often not well characterized.

What about proteins?

The STRING Database (<http://string-db.org/>)

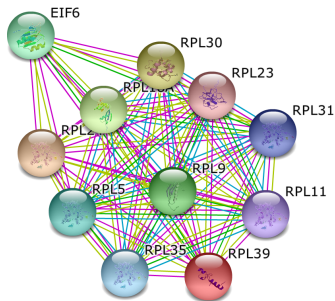
- Protein-protein interaction networks for 1133 organisms
- Integrates known and predicted protein-protein interactions
 - Experiments
 - Text mining
 - Genomic context
 - Databases
- Interactions are scored according to their reliability



Franceschini et al., Nucleic Acids Res.,
2013

The STRING Database (<http://string-db.org/>)

- Protein-protein interaction networks for 1133 organisms
- Integrates known and predicted protein-protein interactions
 - Experiments
 - Text mining
 - Genomic context
 - Databases
- Interactions are scored according to their reliability
- **No such resource for ncRNA-target interactions**



Franceschini et al., *Nucleic Acids Res.*,
2013

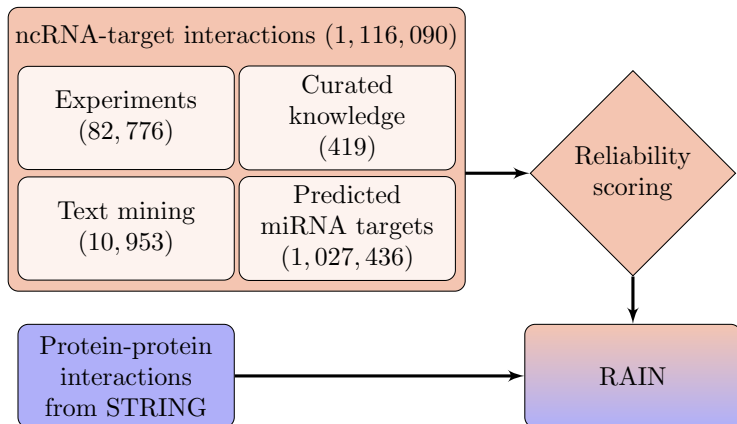
Main Project Goals

Aim: Integrate ncRNAs into STRING by

1. **Collecting** ncRNA-RNA and ncRNA-protein interactions from various sources
2. **Unifying** ncRNA-target interactions and protein-protein interactions

→ Facilitate understanding of the cell's complex interaction network

Sources for RNA-target Interactions



Experimentally validated interactions from

- StarBase [Li et al., Nucleic Acids Res., 2014]
- CLASH [Helwak et al., Cell, 2013]
- miRTarBase [Hsu et al., Nucleic Acids Res., 2014]
- NPInter [Yuan et al., Nucleic Acids Res, 2014]

Predicted miRNA-target interactions from

- miRanda [John et al., PLoS Biol., 2005]
- PicTar [Krek et al., Nature Genetics, 2005]
- TargetScan [Garcia et al., Nat Struct Mol Biol., 2011]
- STarMirDB [Rennie et al., Nucleic Acids Res, 2014]

From Raw Interaction Scores to Confidences

| miRNA | Target mRNA | Raw Score |
|------------------|-----------------|-----------|
| hsa-miR-4685-5p | ENSP00000407818 | 16.8 |
| hsa-miR-548at-3p | ENSP00000406043 | 13.2 |
| hsa-miR-205-3p | ENSP00000407858 | 2.0 |
| ⋮ | ⋮ | ⋮ |

→ translate raw scores into reliability score $\in [0, 1]$

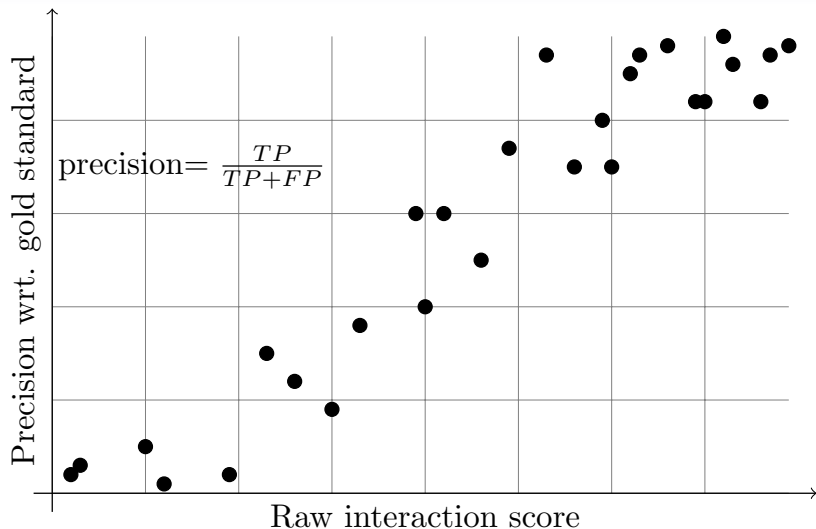
From Raw Interaction Scores to Confidences

| miRNA | Target mRNA | Raw Score |
|------------------|-----------------|-----------|
| hsa-miR-4685-5p | ENSP00000407818 | 16.8 |
| hsa-miR-548at-3p | ENSP00000406043 | 13.2 |
| hsa-miR-205-3p | ENSP00000407858 | 2.0 |
| ⋮ | ⋮ | ⋮ |

→ translate raw scores into reliability score $\in [0, 1]$

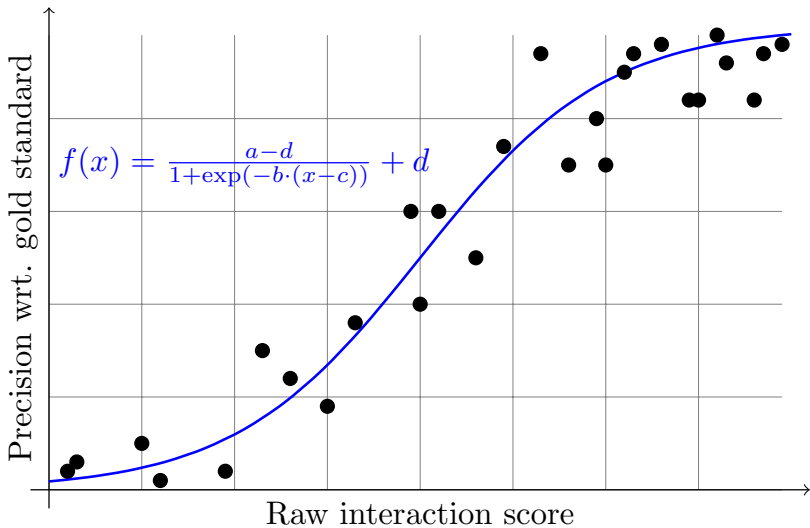


From Raw Interaction Scores to Confidences



Gold standard: 152 expert curated miRNA-mRNA interactions from [Croft et al., BMC Syst Biol, 2012]

From Raw Interaction Scores to Confidences



RAIN webinterface

<http://rth.dk/resources/rain>

Interactions with reliability score > 0.15 per organism

| Organisms | Curated | Experiments | Predictions | Text mining | Total |
|---------------|---------|-------------|-------------|-------------|---------|
| Homo sapiens | 419 | 70.984 | 612.342 | 9.775 | 688.659 |
| Mus musculus | 0 | 6.042 | 332.776 | 1.025 | 339.248 |
| Danio rerio | 0 | 98 | 82.318 | 54 | 82.442 |
| S. cerevisiae | 0 | 5.098 | 0 | 1 | 5.099 |

RAIN webinterface

<http://rth.dk/resources/rain>

RTH
CENTER FOR NON-CODING RNA
IN TECHNOLOGY AND HEALTH

ATCTCTGTCCCTGCAACCGTTGTTAAGGAT
GCTATTCTGGTAAGGGTTTACAAGACAACGTA
AATATAATGTATAAAGAGTATCTTCAGGAGACG
TAATAATGTAAAAATGATGCTCAATAGAAATTA
AGCTGAGGCTCAAAATAGTCCCAATCCTTGC
AAGAAGGCTCGAGAAGTCATTAGATTTGGGGG
GTATGGGTATTCCATTGAACTTGGTAGGA
AGAAACCTTGGGGGAATCTTACATCATTTC
CTCTACCTACTCTAAAGGAACTTTAAAAAT
CTTTATCTTGTCTTTATCTTAATACTGATA
CATTATTAGAGAGAATATCTTGGTTAATTTT

HOME RESEARCH RESOURCES PUBLICATIONS ABOUT PEOPLE EVENTS NEWS JOBS CONTACT

RAIN: RNA-protein Association and Interaction Networks

Search

- Single name
- Multiple names

Help

Download

Search for interactions by non-coding RNA (ncRNA) or protein name in an organism of interest.

ncRNA or protein name: (examples: [#1](#) [#2](#))

[Help \[?\]](#)

organism:

RAIN developers

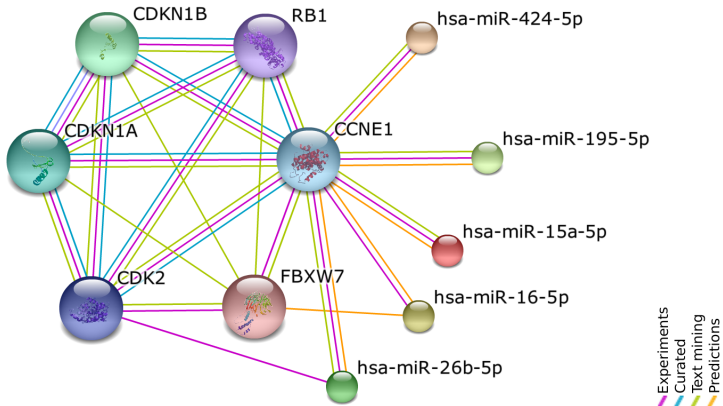
RAIN is being developed at [CBS](#), [CPR](#), [RTH](#), [SIB](#), [DTU](#), [KU](#) and [UZH](#).

RAIN reference

to be announced

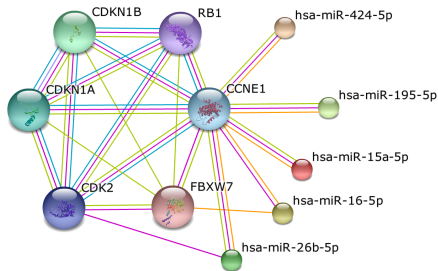
RAIN webinterface

<http://rth.dk/resources/rain>



RAIN webinterface

<http://rth.dk/resources/rain>



Experiments
Curated
Text mining
Predictions

hsa-miR-195-5p

description: Homo sapiens miR-195 mature from the 5' arm
genotype: miRNA

CCNE1 [ENSP0000026243]

cyclin E1; Essential for the control of the cell cycle at the G1/S (start) transition

Evidence suggesting a functional link:

| | |
|-----------------------------------|---|
| Neighborhood in the Genome: | - disabled - |
| Gene Fusions: | - disabled - |
| Cooccurrence Across Genomes: | - disabled - |
| Co-Expression: | - disabled - |
| Experimental/Biochemical Data: | yes (score 0.311). Show |
| Association in Curated Databases: | none / insignificant. |
| Co-Mentioned in PubMed Abstracts: | yes (score 0.583). Show |

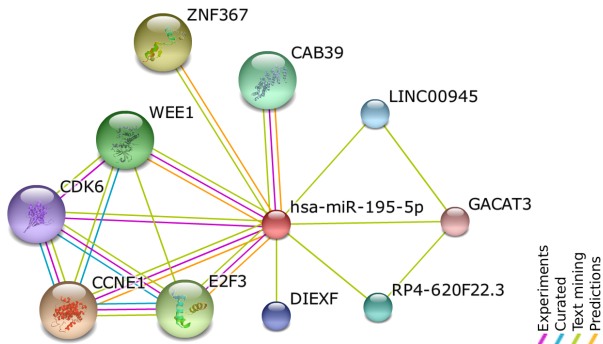
Additional data from

| | | |
|--------------------|-------|----------------------|
| External data: | 0.228 | Show |
| Experimental data: | 0.311 | Show |
| Text-mining: | 0.583 | Show |

Combined Score: 0.758

RAIN webinterface

<http://rth.dk/resources/rain>



Summary

- **Integrate** different sources of ncRNA-target interactions
- **Convert** raw interaction scores into confidences
- **Provide** research community with more complete picture
 - Protein-protein interactions
 - ncRNA-protein interactions
 - ncRNA-RNA interactions

Summary

- **Integrate** different sources of ncRNA-target interactions
- **Convert** raw interaction scores into confidences
- **Provide** research community with more complete picture
 - Protein-protein interactions
 - ncRNA-protein interactions
 - ncRNA-RNA interactions

Future Work:

- *More* gold standard interactions
- *More* sources of evidence
- Cover *more* organisms

Acknowledgements

- Jan C. Refsgaard²
- Christian Garde^{1,3}
- Xiaoyong Pan^{1,2}
- Alberto Santos²
- Ferhat Alkan¹
- Christian Anthon¹
- Christian von Mering⁴
- Christopher T. Workman^{1,3}
- Lars Juhl Jensen^{1,2}
- Jan Gorodkin¹

¹Center for non-coding RNA in Technology and Health, University of Copenhagen

²Novo Nordisk Foundation Center for Protein Research, University of Copenhagen

³Center for Biological Sequence Analysis, Technical University of Denmark

⁴Institute of Molecular Life Sciences and Swiss Institute of Bioinformatics, University of Zurich

Funding: Danish Council for Independent Research, Danish Center for Scientific Computing, Innovation Fund Denmark, Novo Nordisk Foundation [NNF14CC0001], Danish Cancer Society

Acknowledgements

- Jan C. Refsgaard²
- Christian Garde^{1,3}
- Xiaoyong Pan^{1,2}
- Alberto Santos²
- Ferhat Alkan¹
- Christian Anthon¹
- Christian von Mering⁴
- Christopher T. Workman^{1,3}
- Lars Juhl Jensen^{1,2}
- Jan Gorodkin¹

¹Center for non-coding RNA in Technology and Health, University of Copenhagen

²Novo Nordisk Foundation Center for Protein Research, University of Copenhagen

³Center for Biological Sequence Analysis, Technical University of Denmark

⁴Institute of Molecular Life Sciences and Swiss Institute of Bioinformatics, University of Zurich

Funding: Danish Council for Independent Research, Danish Center for Scientific Computing, Innovation Fund Denmark, Novo Nordisk Foundation [NNF14CC0001], Danish Cancer Society

Thank you for your attention!

<http://rth.dk/resources/rain> Feedback? ajunge@rth.dk