
In silico screen for structured RNAs, neuronal RNA co-expression and co-folding.

Stefan E Seemann

Division of Genetics and Bioinformatics, IBHV, University of Copenhagen

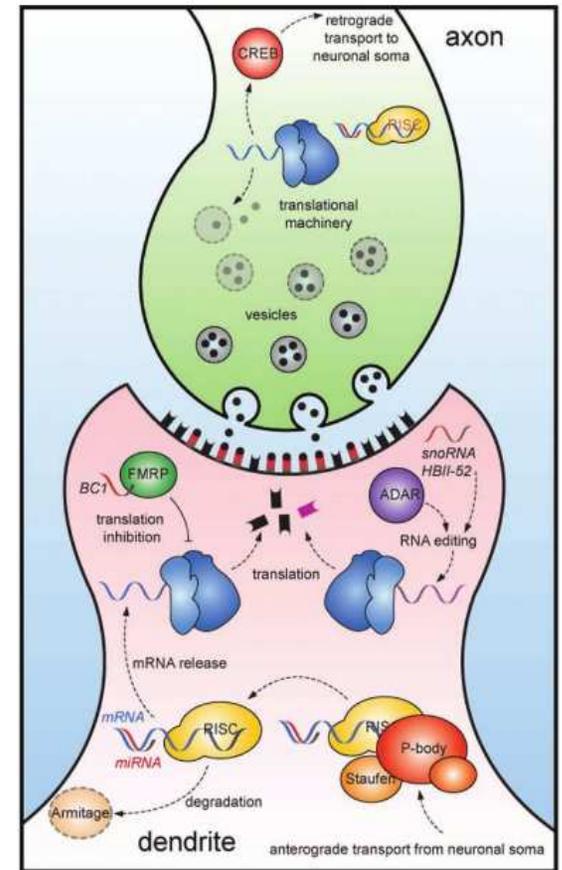
February 15, 2010

Motivation

▷ Motivation
RNA motif search
Neuronal expression
RNA co-folding
PETcofold
Summary

Regulatory ncRNAs functioning in brain:

- **lincRNAs** (large intervening ncRNAs)
HOTAIR – trans-acting gene regulation
Air and *H19* – direct the imprinting of neighboring genes
- **microRNAs:**
miR-124 – guides neuronal differentiation by targeting large number of mRNAs
miR-134 – regulates size of dendritic spines by inhibiting *Limk1*
- **snoRNAs** (small nucleolar RNAs)
HBII-52 – modifies A-I RNA editing + alternative splicing of serotonin receptor
- **siRNAs** (small interfering RNAs)
- **piRNAs** (PIWI-interacting RNAs)

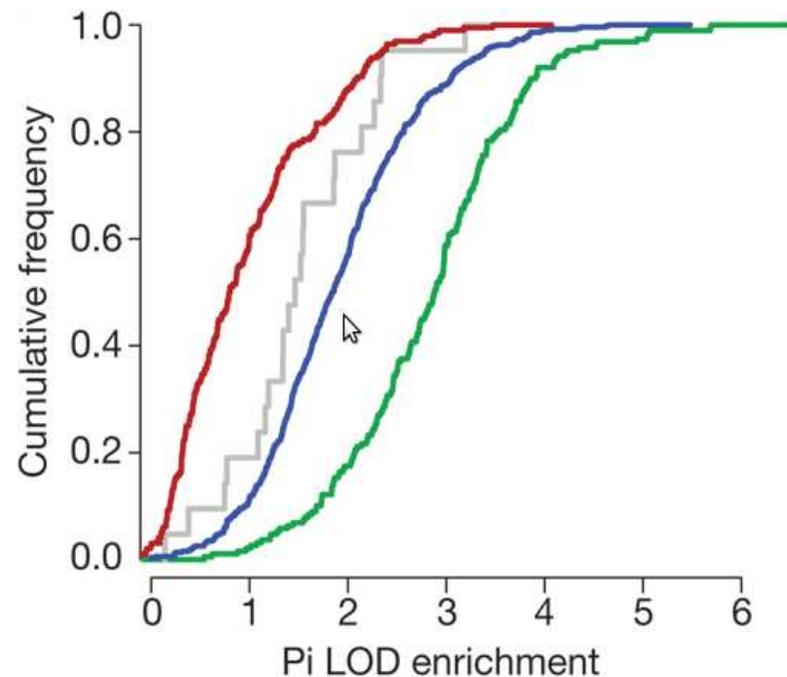


[Mercer 2008]

Motivation

▷ Motivation
RNA motif search
Neuronal expression
RNA co-folding
PETcofold
Summary

Cumulative distribution of sequence conservation across mammals for known non-coding RNA exons (grey), lincRNA exons (blue), protein-coding exons (green) and introns (red) [Guttman 2009]: -

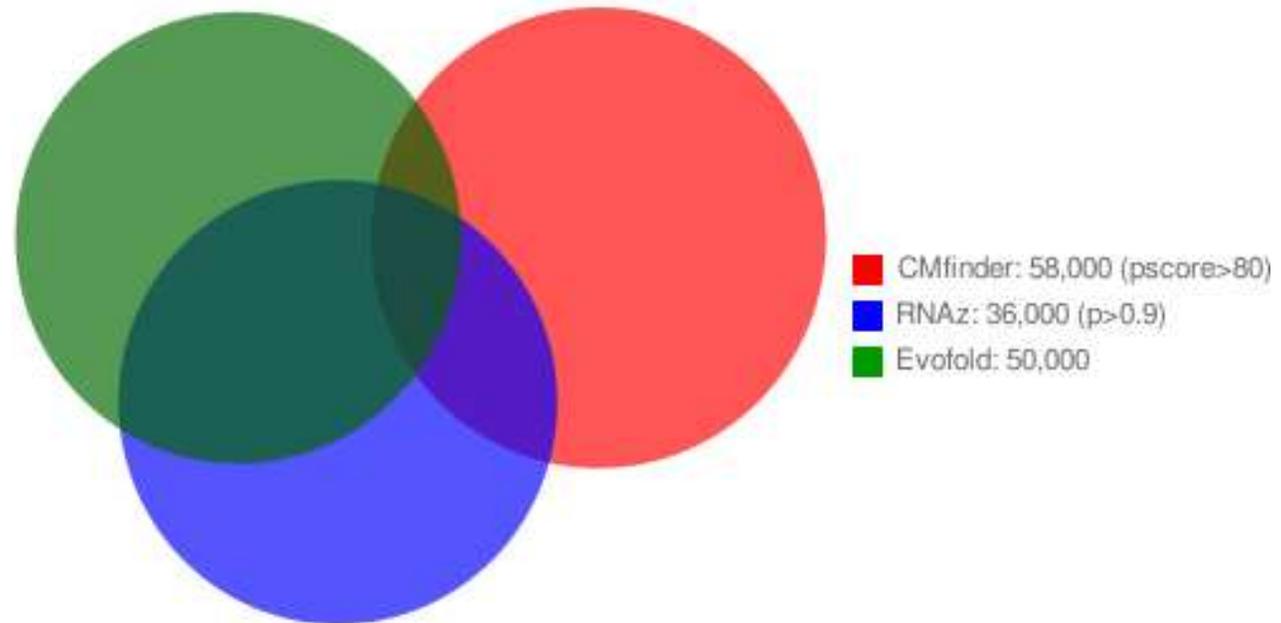


⇒ What causes purifying selection of the functions of ncRNAs if not sequence conservation?

RNA motif search

Motivation
▷ RNA motif search
Neuronal expression
RNA co-folding
PETcofold
Summary

Structural and evolutionary information are used in different flavors to predict ncRNAs.



Evofold [Pedersen 2006]

struct. conservation
struct. probability

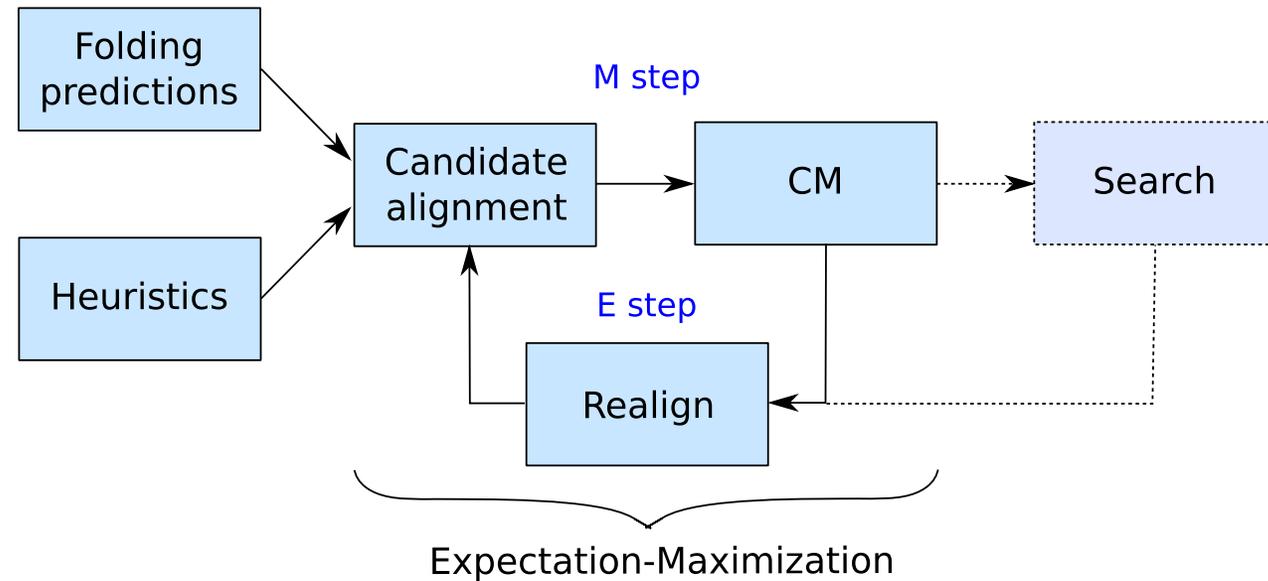
phylogenetic tree
SCF grammar

RNAz [Washietl 2005]

hamming distance (SCI)
thermodynamic stability

Purpose: RNA motif finder from unaligned sequences

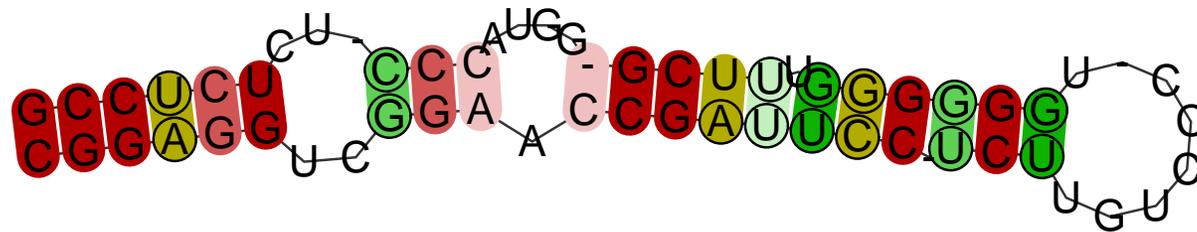
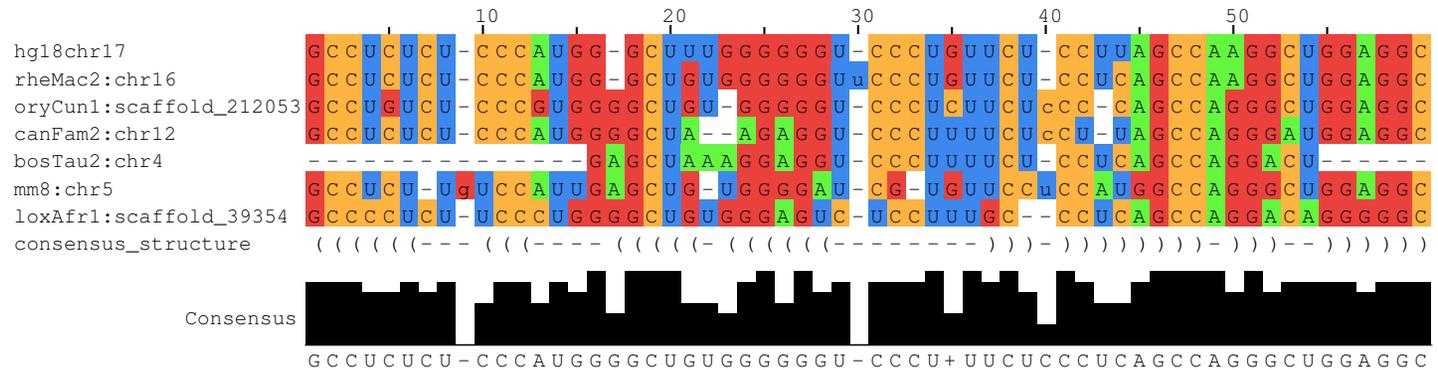
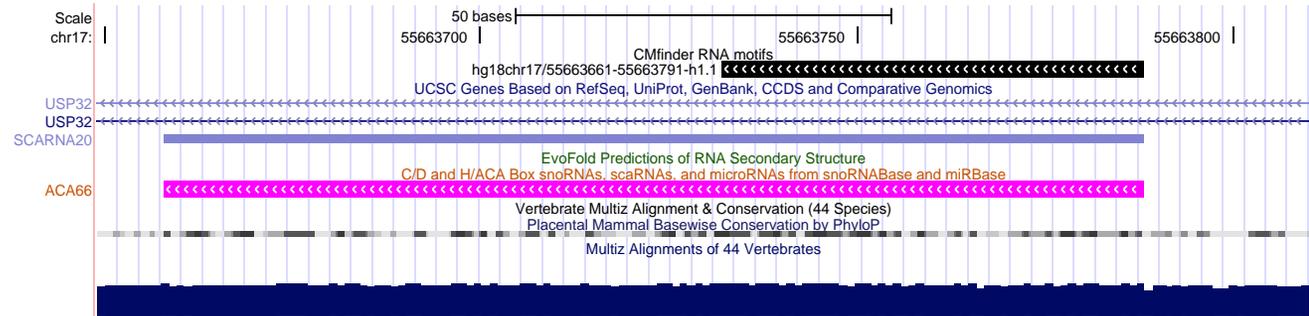
- Simultaneous *local* alignment, folding and CM-based motif description via an EM-style learning procedure



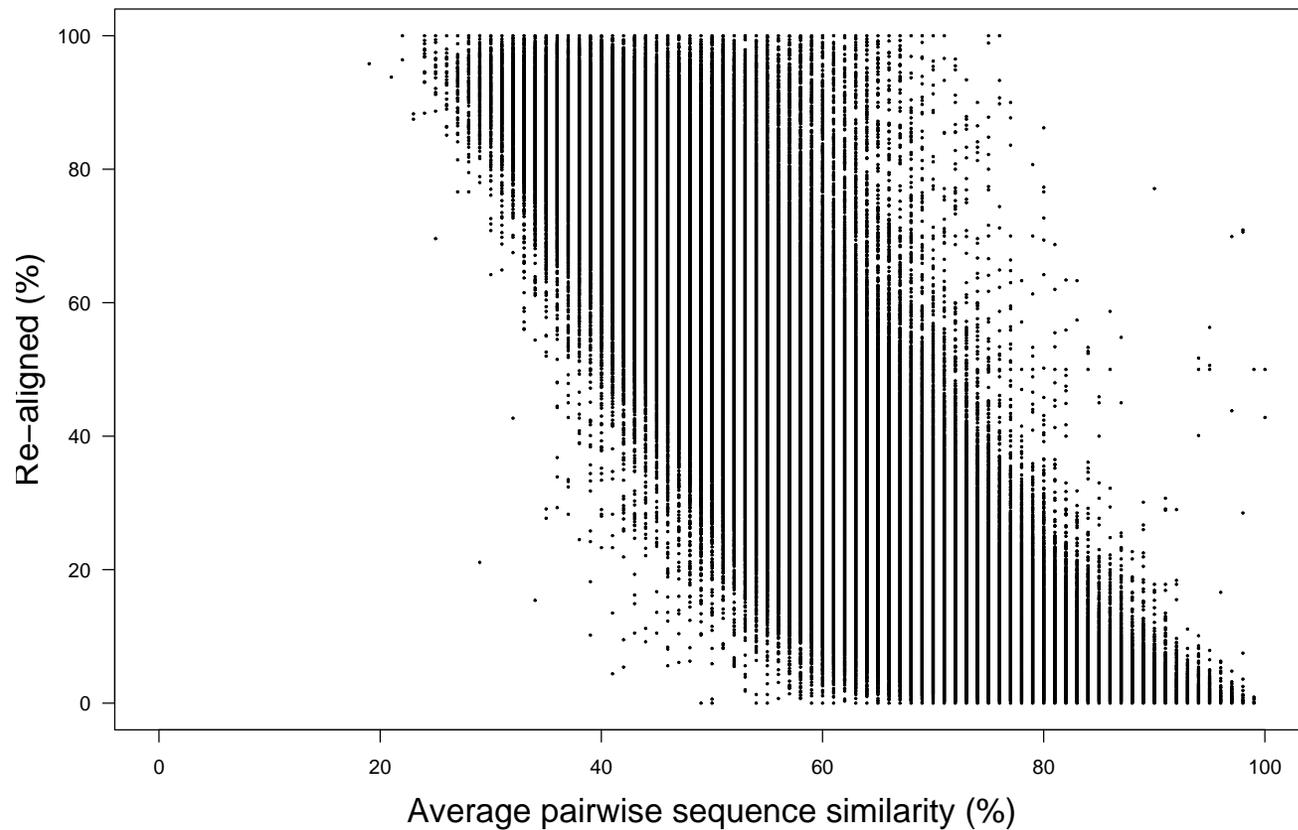
- 17-way UCSC MULTIZ-blocks cover 50% of Hg18 human genome
- 600,000 non-overlapping RNA motifs with pscore > 40 (430,000 RNA motifs in mouse)
- 2.88% of human input sequences consist of candidate regions
- 51% of candidate regions are intergenic, 33% are intronic, 9% are located in 5' UTRs, 4% in 3' UTRs and 3% are exonic.
- 27.0% (1,357) of known ncRNAs (mirBase 14, snoRNABase 3, Rfam 9.1, JonesEddy's list) are covered:
 - 479 microRNAs,
 - 470 tRNAs (including 450 mitochondrial tRNAs),
 - 260 snoRNAs,
 - 51 spliceosomal RNAs,
 - 14 rRNAs

- Motivation
- ▷ RNA motif search
- Neuronal expression
- RNA co-folding
- PETcofold
- Summary

CMfinder finds the 5' stem of the small Cajal body specific RNA 20 (scaRNA20 or ACA66):



Degree of realignment correlates with sequence similarity
(Pearson correlation coefficient ρ of -0.71):



Correlation to Rfam structures:

- 2,074 Rfam structures in input set
- 350 items are overlapped by 780 CMfinder motifs (by 75% in average)
- 71% mean structural correlation (MCC) for 100% overlap
- one reason for good performance is re-alignment of input
⇒ correlation to sequence similarity $\rho = -0.84$

Clustering of known RNAs:

- ~ 150 families with paralogs
- 115 known RNA families are merged in clusters with 0.86 sensitivity and an average PPV of 0.80

Neuronal RNA expression

Motivation
RNA motif search
 Neuronal
▷ expression
RNA co-folding
PETcofold
Summary

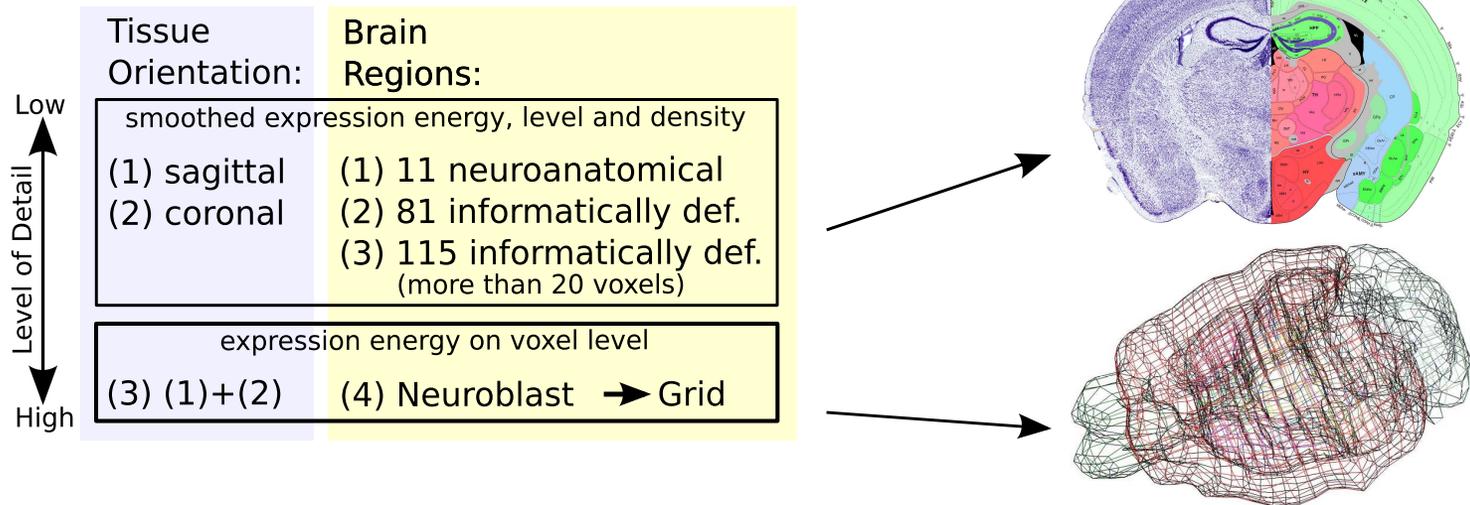
Expression of CMfinder motifs by using available data:

- ~ 200,000 candidate regions overlap transcribed genomic regions in human [Kapranov 2007, Nakaya 2007, Bertone 2004]
- Solexa transcriptome data of rat retinal neurons from Niels Tommerups group
- Allen Brain Atlas (ABA)
 - *In situ* hybridization (ISH) data.
 - Displays the expression of 21,000 RNA transcripts in the adult mouse brain at cellular resolution.
 - 16,900 ABA probes exhibit cellular expression above background.

Allen Brain Atlas (ABA) [Lein 2007]

Motivation
 RNA motif search
 Neuronal
 ▷ expression
 RNA co-folding
 PETcofold
 Summary

ISH images are mapped to neuroanatomical regions:



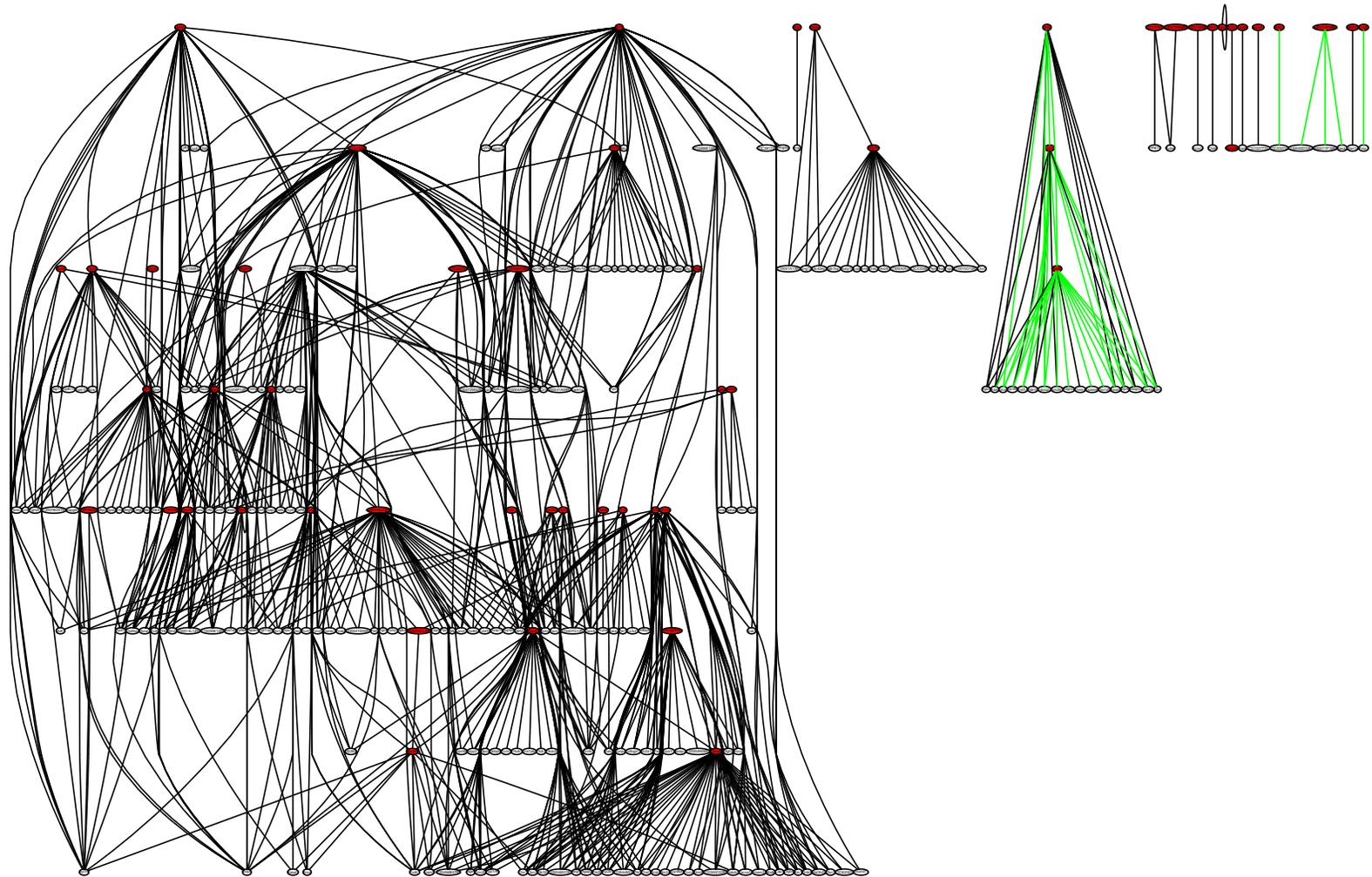
ABA probes with RNA motif, CDS-free and expression over background:

	Intergenic	Intronic	5' UTR	3' UTR	Σ
CMfinder	303	139	17	1,024	1,483
RNAz	200	124	21	889	1,234
Σ	425	213	31	1,442	2,111

Neuronal RNA co-expression

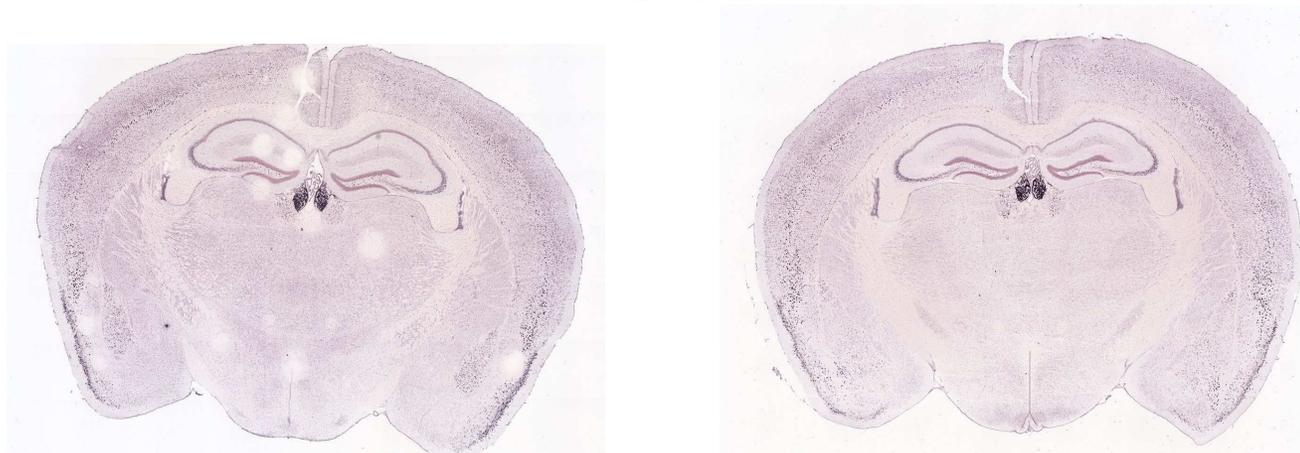
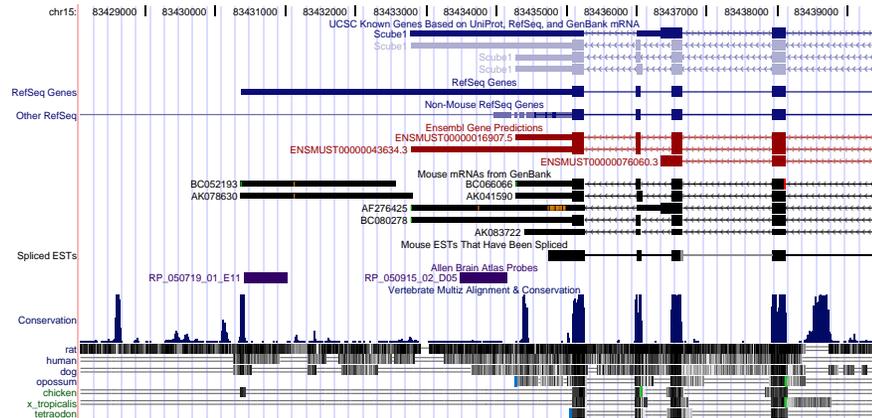
55 ncRNAs involved in 733 correlation pairs ($\rho \geq 0.8$)

Motivation
RNA motif search
Neuronal
▷ expression
RNA co-folding
PETcofold
Summary



Neuronal RNA co-expression

Example - intergenic *mCG1027775.1* and 3' UTR *Scube1*:



High correlation over entire brain ($\rho = 0.84$).

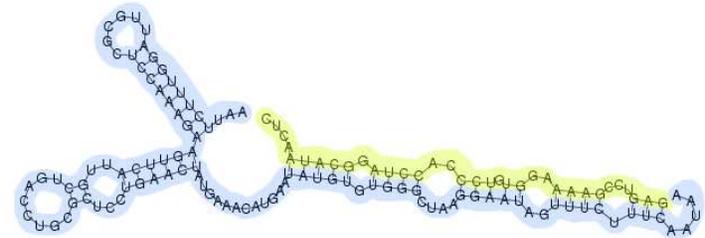
High correlation in Medial habenula (Thalamus), Piriform area (Olfactory bulb) and Cerebral cortex.

Motivation
RNA motif search
Neuronal
▷ expression
RNA co-folding
PETcofold
Summary

RNA co-folding

Motivation
RNA motif search
Neuronal expression
▷ RNA co-folding
PETcofold
Summary

RNA duplex of
intergenic *D630023B12Rik*
and *Cox6c* mRNA



Goal: Search for potential RNA interaction partners of CMfinder motifs in whole human genome and how well can RNA co-folding describe RNA co-expression:

- Seed by reverse complementary BLAST
- Calculate thermodynamic stable RNA duplexes applying RNAduplex
- Calculate E-value that estimates statistical significance of RNA duplex (extreme value distribution)

Summary

Motivation
RNA motif search
Neuronal expression
RNA co-folding
PETcofold
▷ Summary

- CMfinder whole genome scan:
~ 25% of candidates have > 50% of their positions realigned
- RNA expression correlation in ABA:
Structured 3' UTRs significantly over-expressed
⇒ acting as transcription factor or trans-acting ncRNA
- Whole genome scan for significant RNA interaction partners of
CMfinder candidates
- PETcofold:
Detection of kissing loops in RNA interactions

Thank you

Motivation
RNA motif search
Neuronal expression
RNA co-folding
PETcofold
Summary

- Allen Institute for Brain Science, Seattle
Susan Suskin, Michael Hawrylycz
- University of Washington, Seattle
Larry W Ruzzo, Zizhen Yao
- University of Freiburg
Rolf Backofen, Andreas Richter
- University of Copenhagen
Jan Gorodkin, Elfar Torarinsson