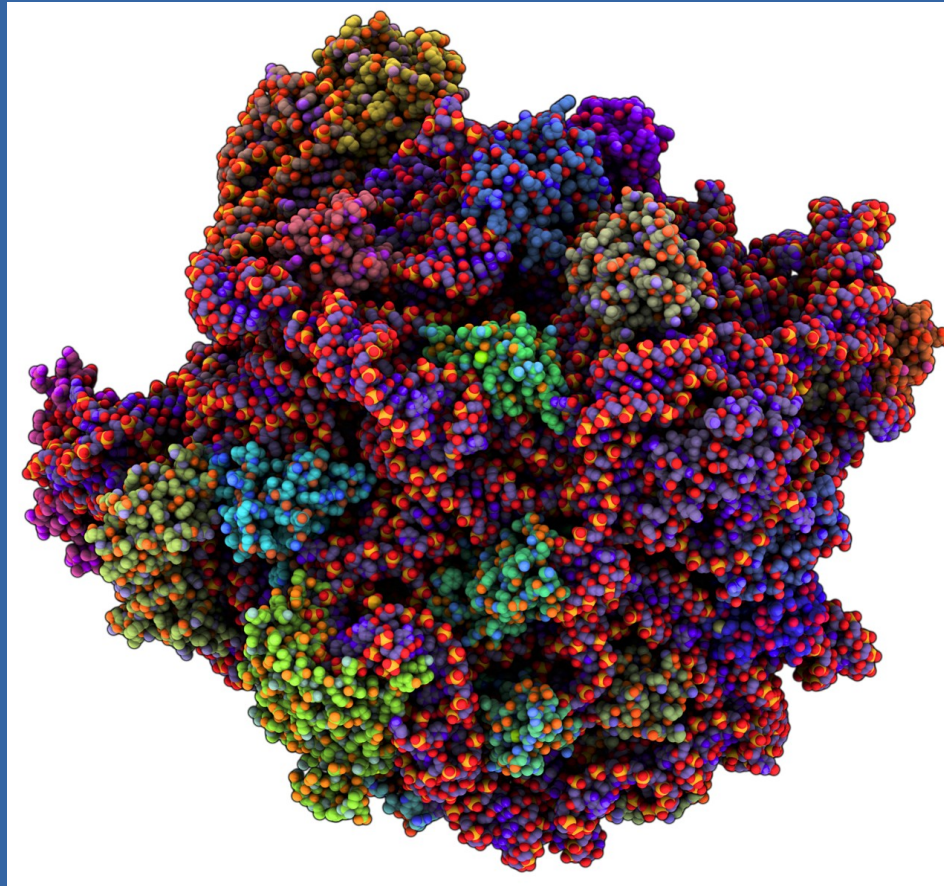


Background



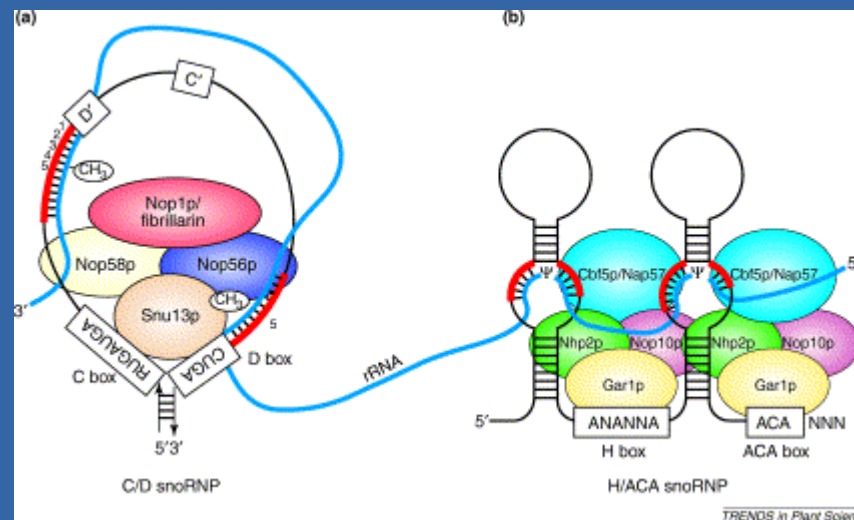
Background



Prader-Willi Syndrome

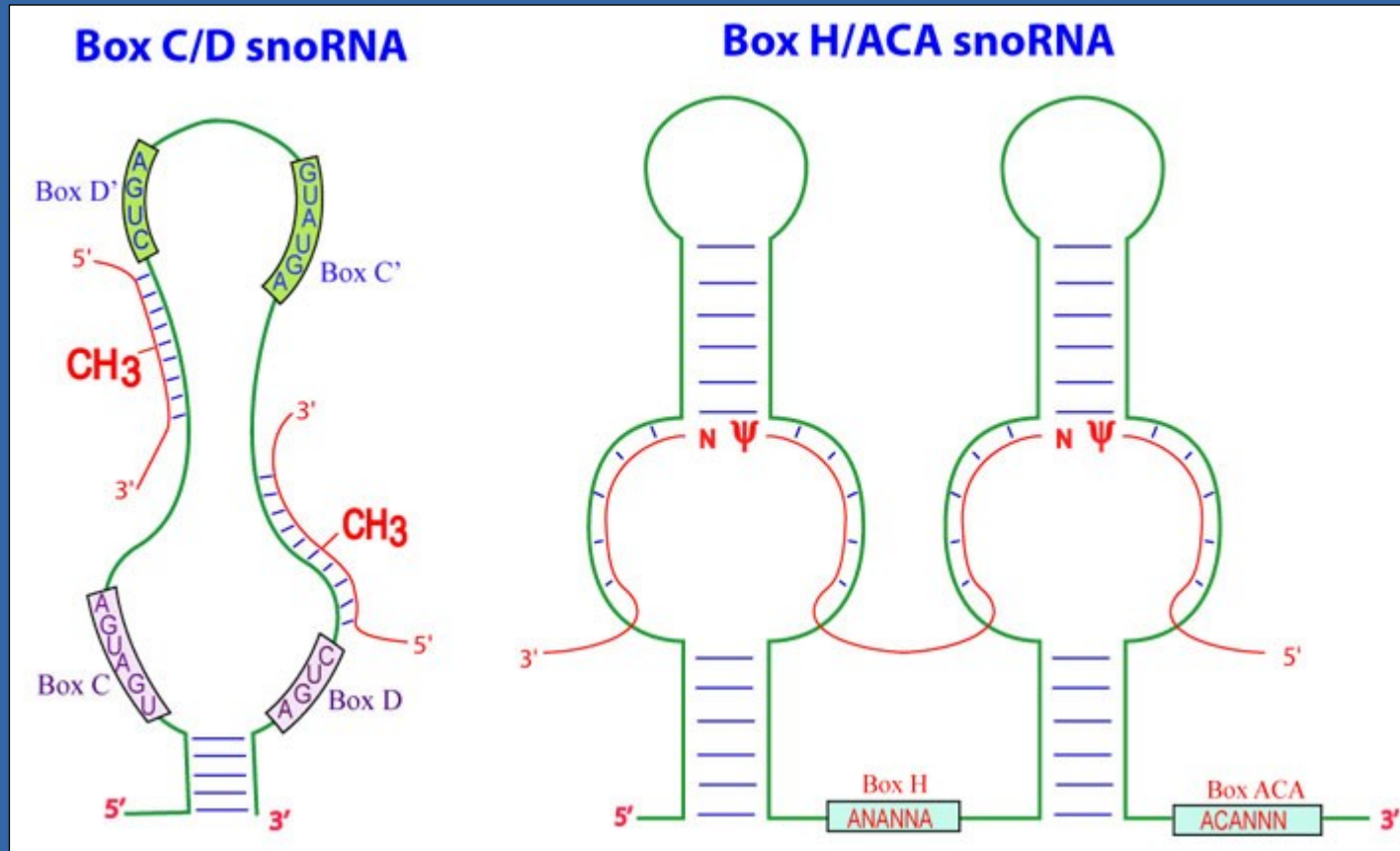
Skryabin BV et al 2007

Background



John W.S Brown et al 2002

Background



Cbox: RUGAUGA
Dbox: CUGA

Hbox: ANANNA
Tail: ACA

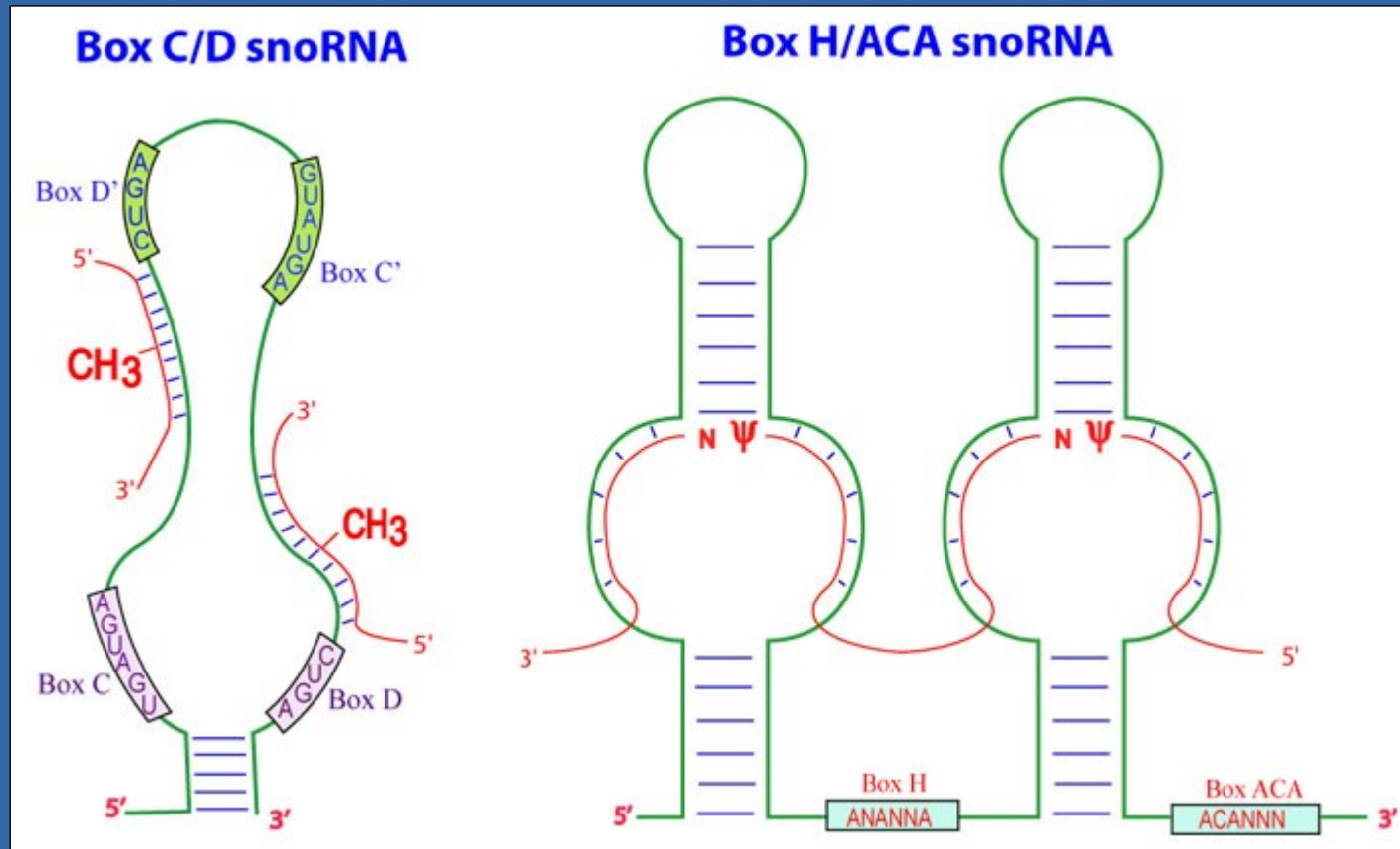
Source: <http://biochem.ncsu.edu>

Folding HACA Motif Example

```
Input string (upper or lower case); @ to quit
.....1.....2.....3.....4.....5.....6.....7.....8
AACAAATAATTACATTCGAAGGACCGGTATTATTGTAGGGGATTGTGACTTCAAGGCAACGTCTCTACAACCGAGTTCGAGAATAAGTACAAATGGCTCTTTTGTTATTCGAAAGCTTACAAGTTCGGTATATTCTATATACTCACTTATTACTTTCAAGTACTTCACACGGGCCTGACA
length = 189
AACAAAUAUUAUCAAUCCAAGGACCGGUUAUUUGUAGGGGAUUUGUCUUUCAAGGCAACGUCCUCUCUACAACCGAGUUCGAGAAUUAAGUACAUAUUGGCUCUUUUUGUUAUUCGAAAGCUUACAAGUUCUGGUUAUUAUUAUUAUCACUUAUUAUUAUUAAGUACUUCACACGGGCCUGACA
.....(((((((....(((((((....((((((((((((.....)))))).)))..)))).)))...((((.....((((.....((((.....((((((.....)))))).)))..))))).))).....
minimum free energy = -38.10 kcal/mol
```

VS

Background



Cbox: RUGAUGA
Dbox: CUGA

Hbox: ANANNA
Tail: ACA

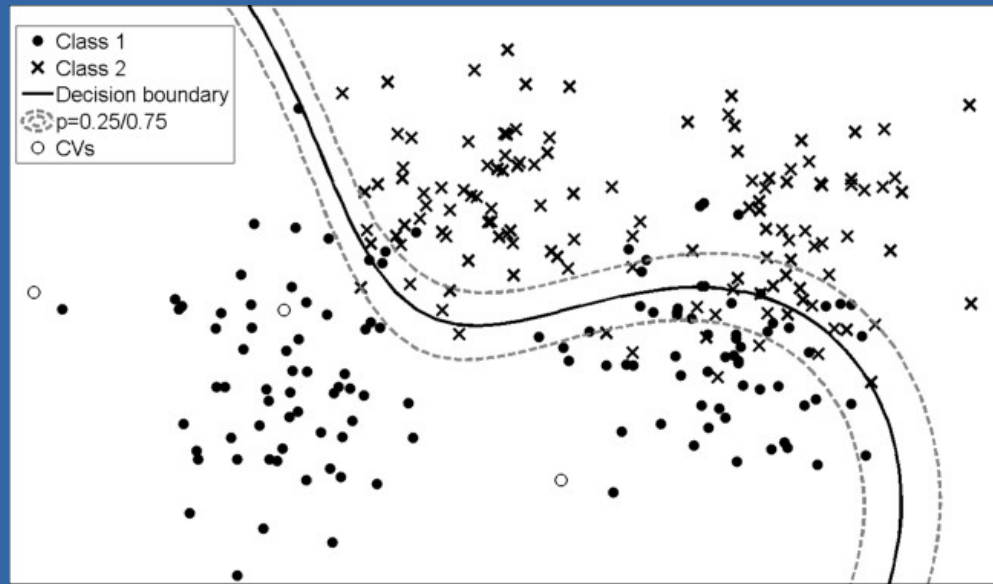
Source: <http://biochem.ncsu.edu>

1D Motifs as a Feature

Hbox	+	-
AGAGAA	76	27
AGAGCA	65	25
AAAGCA	65	21
AGAGGA	63	28
AGAGUA	61	17
AUAGAA	61	17
ACAGCA	58	24
ACAGAA	53	28
AGAUUA	51	28

Cbox	+	-
AUGAUGA	366	10
GUGAUGA	238	37
CUGAUGA	92	24
UUGAUGA	84	28
CUGAAGA	70	23
AUGAUGU	57	7
CUGAUGC	57	12
AUGAAGA	45	10
AUGAUGG	43	19

Methods



Kernel Methods

- SVM
- CRUM

Objectives

- Crossfold validation of current snoRNA CRUM
- Add folding data as a feature for crum
- Rewrite crum training program
- Compare crum-snoRNA classifier with existing methods
- Create web interface for easy DNA/RNA feature selection so users can generate their own models

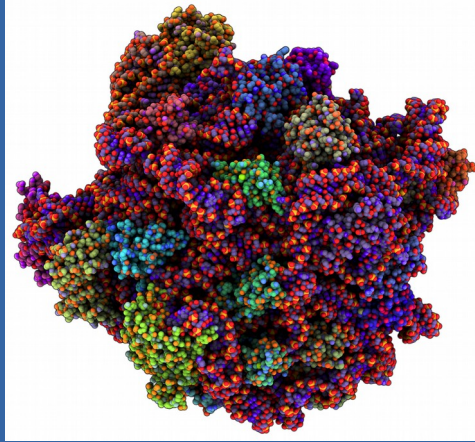
Questions?



The secret of all victory lies in the organization of the non-obvious.

-Marcus Aurelius

Background



- Project title: Classification of snoRNA Types Using Classification Relevance Unit Machines
- pic of ribosome, primary use of the rRNA with modifications performed by the snoRNP snoRNA associates with

Background



Prader-Willi Syndrome

Skryabin BV et al 2007

Importance of snoRNA?

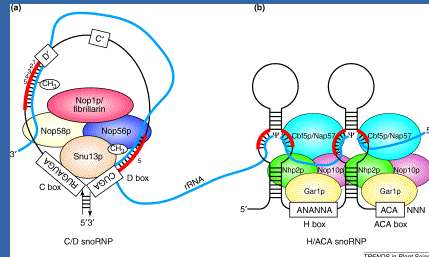
Loss of snoRNA:

- Prader willi: 7 gene deletion on the chr14 in humans
- retardation, infertility exaggerated brow features

Duplication of snoRNA genes:

- Possibly a cofactor in autism

Background



John W.S Brown et al 2002

- Cartoon of the protein/rna complexes formed with snoRNA when in snoRNP
- 4 protein complex associates, 2x on HACA 1 on CD
- NOTE BINDING SITES AT MOTIFS! It's why they are so well conserved across species

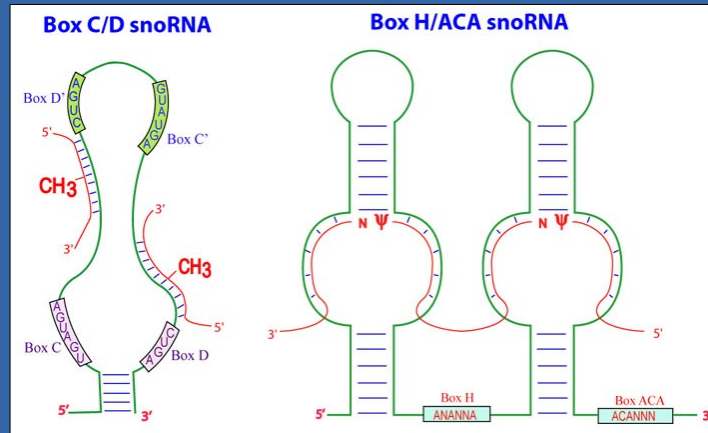
Modifications

HACA rnp = pseudouridylation

C/D rnp = methylation

Post transcriptional modification of rRNA can increase its longevity in the cell against endogenous rnases

Background



Cbox: RUGAUGA
Dbox: CUGA

Hbox: ANANNA
Tail: ACA

Source: <http://biochem.ncsu.edu>

Talk about folding of 1D strands
stem v loop v dangling

-base ambiguity: R = purines A or G

Modifications

HACA rnp = pseudouridylation

C/D rnp = methylation

Folding HACA Motif Example

[illegible]

VS

[illegible]

Folding of 1D strand

Top: HACSA

-2 loops

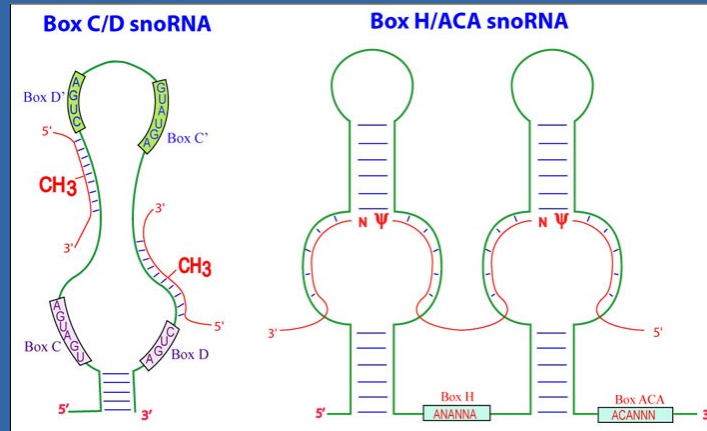
Bottom: random DNA seq

-5 loops

Talk about dot paren notation for folding

- Pre supplied structure motifs for folding
- Results ideal, not typical. Needs work.

Background



Cbox: RUGAUGA
Dbox: CUGA

Hbox: ANANNA
Tail: ACA

Source: <http://biochem.ncsu.edu>

Talk about basic 1D motif frequency features

-base ambiguity: R = purines A or G

Modifications

HACA rnp = pseudouridylation

C/D rnp = methylation

1D Motifs as a Feature

Hbox	+	-	Cbox	+	-
AGAGAA	76	27	AUGAUGA	366	10
AGAGCA	65	25	GUGAUGA	238	37
AAAGCA	65	21	CUGAUGA	92	24
AGAGGA	63	28	UUGAUGA	84	28
AGAGUA	61	17	CUGAAGA	70	23
AUAGAA	61	17	AUGAUGU	57	7
ACAGCA	58	24	CUGAUGC	57	12
ACAGAA	53	28	AUGAAGA	45	10
AGAUUA	51	28	AUGAUGG	43	19

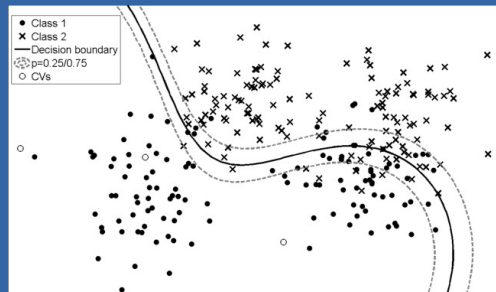
Top 10 frequency of motifs between classes

Neg class is the opposing class, not random dna

Old feature generation method for demonstration

Talk about kmers, sliding window frequencies etc.

Methods



Kernel Methods

- SVM
- CRUM

CRUM v SVM

Crum:

- Probabilistic (not hard decisions)
- sparser kernels than SVM (faster classification)

SVM:

- need to compare the 2 for project
- faster training but generates more support vectors
- (6x more)
- requires cross validation of decisions because of nonprobabilistic nature, computationally slower

Objectives

- Crossfold validation of current snoRNA CRUM
- Add folding data as a feature for crum
- Rewrite crum training program
- Compare crum-snoRNA classifier with existing methods
- Create web interface for easy DNA/RNA feature selection so users can generate their own models

10 fold cross validation

Haca:

TrainErr =%6.6

TestErr = %8.69

CD:

TrainErr =%10.47

TestErr =%13.58

For k=5

Questions?



The secret of all victory lies in the organization of the non-obvious.

-Marcus Aurelius