

Noncoding-RNA

ncRNA
identification
And
functionality

What is RNA?

Ribonucleic acid (RNA) is one of the three major components in the cell macromolecules
(along with DNA and Proteins)

During the transcription process, a RNA copy of a DNA segment is made
(mRNA)

RNA plays a central role in the pathway from DNA to proteins
(known as the Central Dogma)

This copy will then be read by a ribosome to form a protein
(RNA strand)

What is RNA? (III)

Initially we thought only proteins were
capable of catalyze chemical reactions
(enzymes)

In 1980 we found that RNA
can act as catalysts too
(ribozymes)

This finding supported the RNA World Hypothesis
**(earlier forms of life used only RNA in order to
store information and catalyze chemical reactions)**

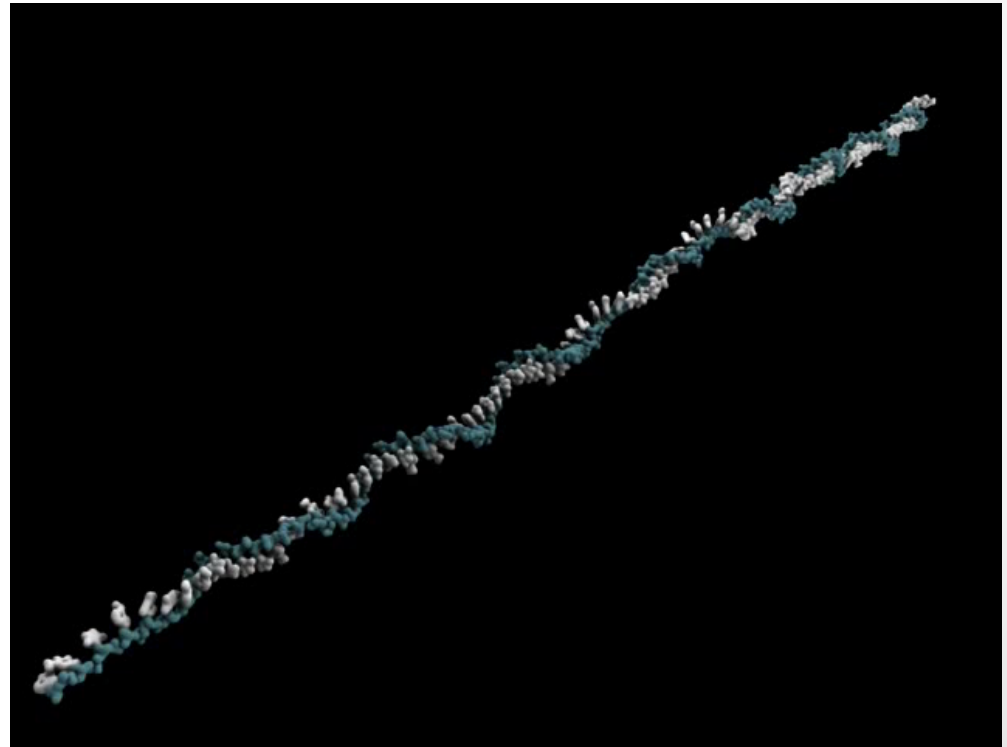
What is RNA? (II) – Why bother?

Typically found in a single-stranded form
(allowing complex folded 3D structures)

Ribosomes are made up of both RNA and proteins components

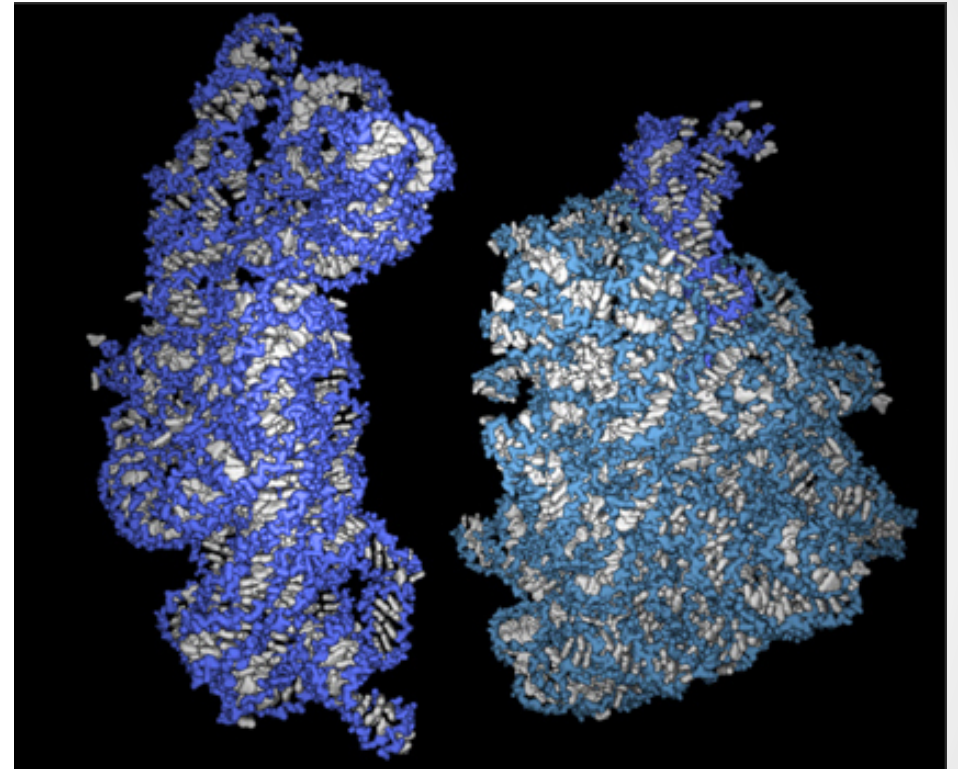
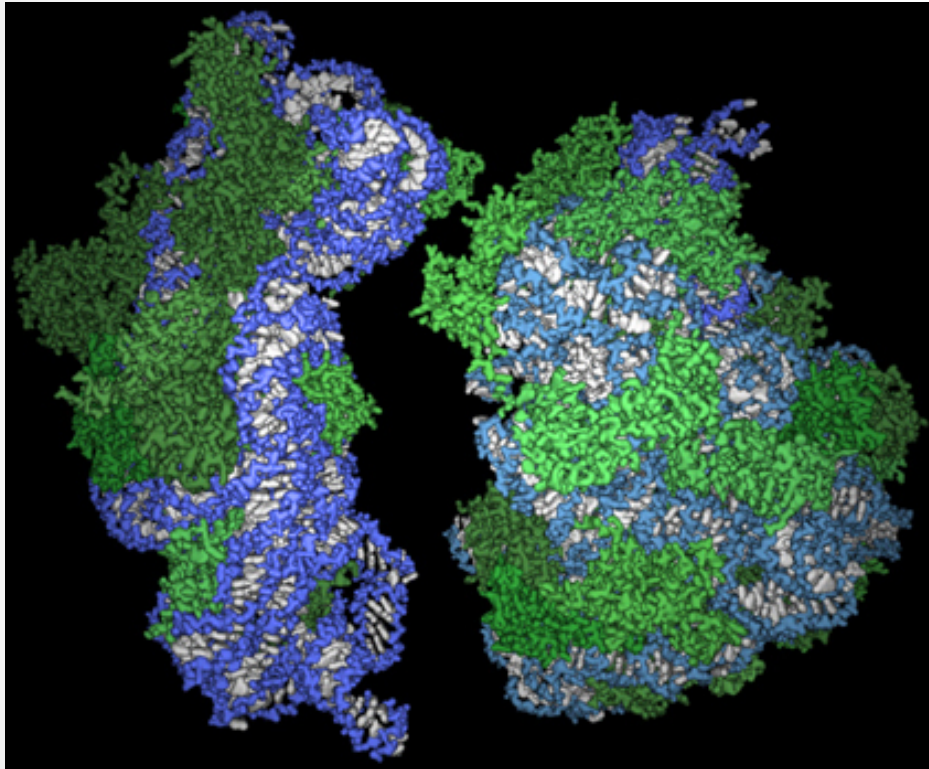
Structural and biochemical analyses revealed that the mechanisms central for translation are catalyzed by RNA, not protein
(allowing complex folded 3D structures)

Suggesting that the use of RNA by early life forms to carry out chemical reactions preceded the use of proteins!



What is RNA? (III) – Why bother?

PDB #[2HGR](#) and #[2HGU](#)



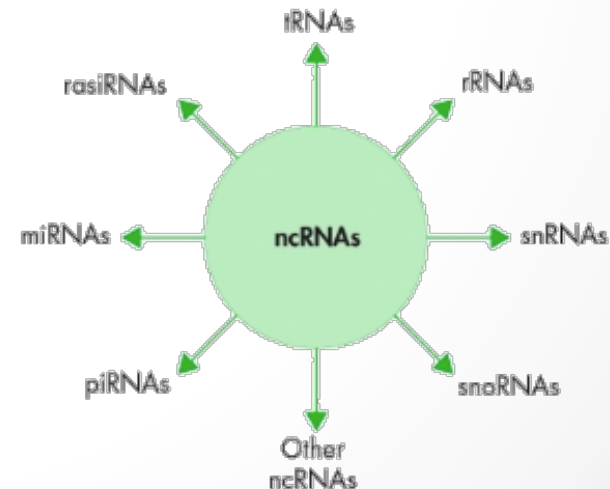
Protein structure shown in green, RNA shown in blue and white

What is RNA? (IV) – Types of RNA

RNA is produced by a process called RNA polymerase (RNAP or RNAPol)
(RNAP is an enzyme needed for constructing RNA chains from DNA - transcription)

- mRNA
 - messenger RNA – carries information from DNA to the ribosome in order to begin translation
 - Boring!! We all know it!
- ncRNA
 - Every other function RNA performs..
 - Here's a small list:

ra*si*RNAs, tRNAs, rRNAs, snRNAs, snoRNAs,
piRNAs, miRNAs, other ncRNAs (yes there's
more!)



What is RNA? (IV) – Types of RNA (II)

Involved in Protein Synthesis...

Type (RNA)	Abbreviation	Function
Messenger	mRNA	Codes for proteins
Ribossomal	rRNA	Translation
Signal Recognition particle	SRP RNA	Membrane Integration
Transfer RNA	tRNA	Translation
Transfer-messenger	tmRNA	Rescuing stalled ribossomes

What is RNA – Types of RNA (III)

Involved in post-transcriptional modification
or DNA replication...

Type (RNA)	Abbreviation	Function
Small nuclear	snRNA	Splicing and other functions
Small nucleolar	snoRNA	Nucleotide modification of RNAs
SmY	SmY	mRNA trans-splicing
Small Cajal body-specific	scaRNA	Type of snoRNA; Nucleotide modification of RNAs
Guide	gRNA	mRNA nucleotide modification
Ribonuclease P	Rnase P	tRNA maturation
Ribonuclease MRP	Rname MRP	rRNA maturation, DNA replication
Y	?	RNA processing, DNA replication
Telomerase	?	Telomere synthesis

What is RNA – Types of RNA (IV)

Involved in Regulation...

Type (RNA)	Abbreviation	Function
Antisense RNA	aRNA	Transcriptional attenuation; mRNA degradation; mRNA stabilisation; Translation block
Cis-natural antisense transcript	?	Gene regulation
CRISPR	crRNA	Resistance to parasites, probably by targeting their DNA
Long noncoding	Long ncRNA	Various
MicroRNA	miRNA	Gene regulation
Piwi-interacting	piRNA	Transposon defense, maybe other functions
Small interfering	siRNA	Gene regulation
Trans-acting	tasiRNA	Gene regulation
Repeat associated siRNA	rasiRNA	Type of piRNA; transposon defense
7SK RNA	7SK	negatively regulating CDK9/cyclin T complex

What is RNA? (IV) – Types of RNA (V)

Parasitic RNAs...

Type (RNA)	Function	Found in
Retrotransposon	Self-propagating	Eukaryotes and some bacteria
Viral Genome	Information carrier	Double-stranded RNA viruses, positive and negative sense RNA viruses, satellite viruses and reverse transcribing viruses
Viriod	Self-propagating	Infected plants
Satellite	Self-propagating	Infected cells

...and others

Retrotransposon	Self-propagating	Eukaryotes and some bacteria
-----------------	------------------	------------------------------

Ok.. Maybe there's more to it than just
translation

**Besides translation many RNAs do not code
proteins..**

About 97% of the transcriptional output is non-
protein-coding in eukaryotes

97% !!!

Suspicion arises

- **Mattick JS(2001): The evolution of controlled multitasked gene networks: the roles of introns and other noncoding RNAs in the development of complex organisms**

“ We suggest that these RNAs (eRNAs) have evolved to function as endogenous network control molecules which enable direct gene-gene communication and multitasking of eukaryotic genomes “

Suspicion arises (II)

- **Mattick JS (2001): Noncoding RNAs: the architects of eukaryotic complexity:**

“...Analysis of a range of complex genetic phenomena in which RNA is involved or implicated, including co-suppression, transgene silencing, RNA interference, imprinting, methylation, and transvection, suggests that a higher-order regulatory system based on RNA signals operates in the higher eukaryotes and involves chromatin remodeling as well as other RNA-DNA, RNA-RNA, and RNA-protein interactions.”

Suspicion arises (III)

- **Mattick JS (2003): "Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms":**

"... A number of startling observations about the extent of non-protein-coding RNA (ncRNA) transcription in the higher eukaryotes and the range of genetic and epigenetic phenomena that are RNA-directed suggests that the traditional view of the structure of genetic regulatory systems in animals and plants may be incorrect."

Suspicion arises (IV)



- **The Hidden Genetic Program of Complex Organisms:**

“ We may be witnessing such a turning point in our understanding of genetic information. “

“Biologists assumed that proteins alone regulate the genes of humans and other complex organisms. But an overlooked regulatory system based on RNA may hold the keys to development and evolution “

Sense a trend yet?!

Got motivated?! Ok.. We'll dig a little deeper..

Motivation

- **Most of ncRNAs are involved in translation:**
 - the Ribosome consists of more than 60% of ribosomal RNA;
 - tRNA form an “adaptor molecule” between mRNA and proteins;
 - snoRNAs are involved in rRNA maturation and drive covalent modifications of rRNA, tRNA and snRNAs
 - Rnase P and, evolutionary from Rnase MRP matures tRNA
 - SRP RNA recognizes and transports specific proteins

- **ncRNAs in RNA splicing:**
 - Remove introns sequences (required for mature mRNA)
 - Another RNP RNA class (snRNP or tri-snRNP)
 - Self-splicing introns catalyze their own removal from host transcripts

Motivation (II)

- **ncRNAs in gene regulation:**
 - Thousands of genes are regulated by ncRNAs
- **Trans-acting ncRNAs:**
 - Regulate gene expression (miRNA) – down silencing
 - Rnase P influences gene expression
 - In the human nucleus Rnase P is required for normal and efficient transcription of various ncRNAs (including tRNA, rRNA, SRP RNA and snRNA)

Motivation (III)

- **Cis-acting ncRNAs:**
 - Involved in protein coding influencing their expression
 - Regulate operons by forming terminator structures when an excess of the regulatory amino acid of the operon is detected
 - They also take action when a deficiency of charged tRNA and stall the ribosome forming an ant terminator structure
 - Solve the lack of Iron (Iron response elements or IRPs)

- **ncRNAs in genome defense:**
 - Piwi-interacting RNAs form RNA-protein complexes with Piwi proteins silencing retrotransposons(?)
 - Form spaces helping the cell to fight infection

Motivation (IV)

- **ncRNAs and chromosome structure:**
 - Telomerase is an RNP enzyme that adds specific repeats (“TTAGGG”) to telomeric regions – which contain condensed DNA material and give stability to the chromosomes
 - Xist (X-inactive-specific transcript) is a long ncRNA gene on the X chromosome in placental mammals.
 - In drosophilids are involved in dosage compensation (helping determining the sex of the animal)

Motivation (last one.. promise)

Ok.. I'll stop.. But did you know ncRNA is responsible for some diseases too?

Cancer, Prader-Willi syndrome, Autism, Cartilage-hair hypoplasia, Alzheimer's, miR-96, hearing loss..

So.. Why isn't all RNA functional?!!

Several publications have started using the term **functional RNA (fRNA)**, as oposed to ncRNA to describe regions functional at the RNA level..

But if most of the fRNAs interact in so many ways within the cell..

The distinction is fading.. Thought mRNA is still the main player in the RNA World Hypothesis

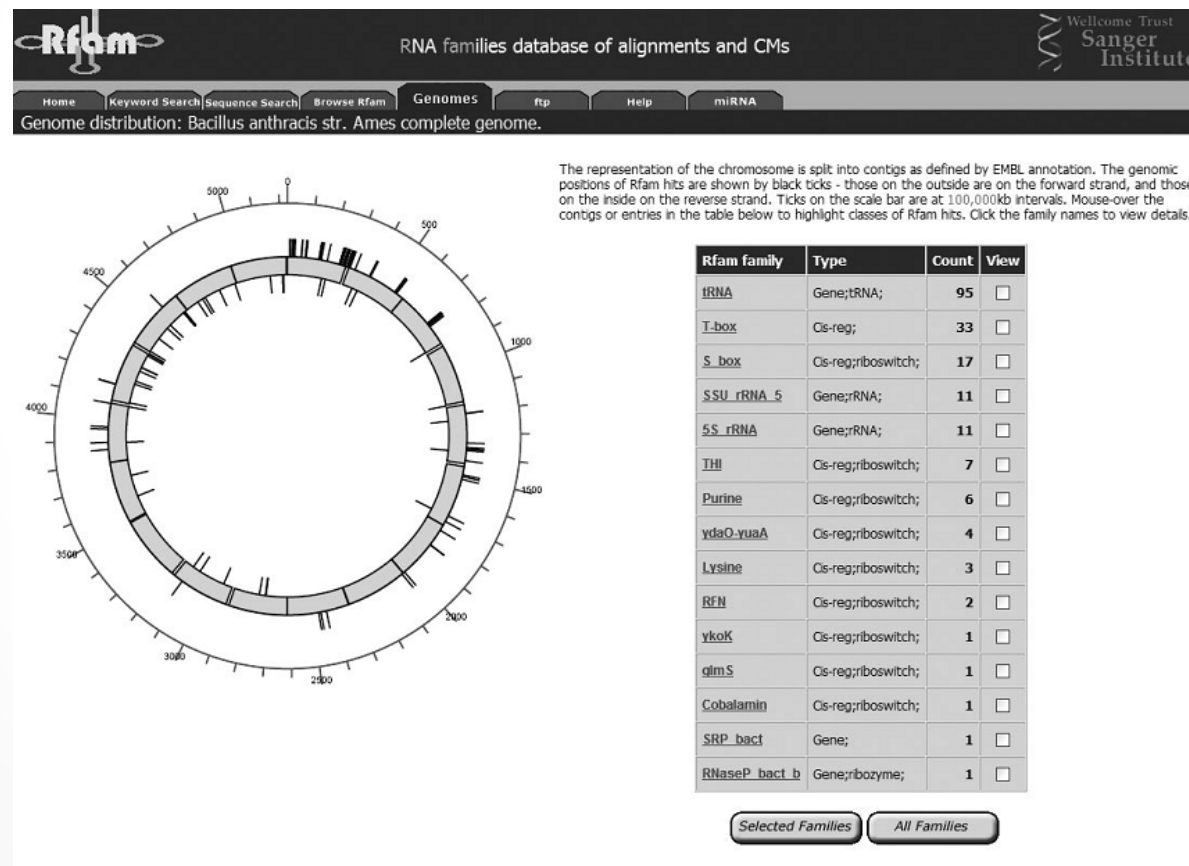
Break

15 minutes for cookies and coffee

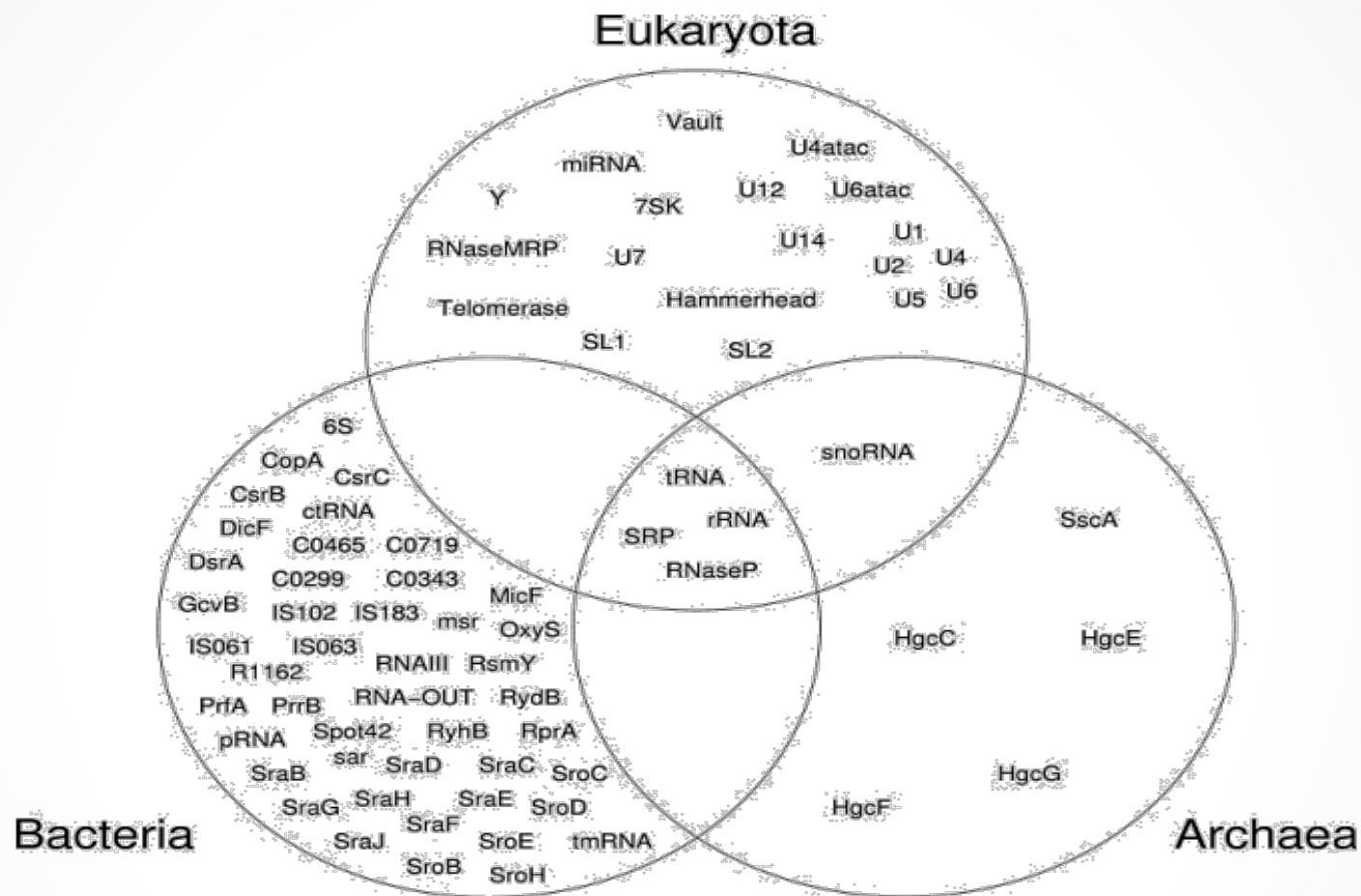
In part II we'll talk what we, computer scientists, can do to help biologists identifying RNAs and label their functions (interactions)

Identification (Rfam)

- Rfam: a comprehensive collection of non-coding RNA families
 - Families are represented by multiple sequence alignments and profiled stochastically (non-deterministic way) in context-free grammars



Identification (Rfam II)



Distribution of the Rfam family members in the three kingdoms of life

Identification (fRNAdb)

- fRNAdb: a platform for mining/annotating functional RNA candidates from non-coding RNA sequences

S. no.	Description	Number of transcripts	Min/max
1	Length of the sequence (nt)	13 693	15/107 797
2	Number of exons	7166	0/60
3	Number of overlapping ESTs	4184	0/6490
4	Number of mapped positions	7158	0/892
5	GC-content (%)	13 693	4/87
6	Maximum length of potential ORF (amino acids)	12 655	0/1664
7	Percentage of bases that is covered with repeat elements	6460	0/100
8	Repeat elements reside proximal upstream/downstream	2219	
9	Known gene that is a potential sense/antisense of this transcript (exon overlapping required)	936	
10	Number of protein homologs (GenBank NR)	5811	0/250
11	Known gene that includes this transcript within its intron	951	
12	Known gene region that overlaps with the mapping extent of this transcript (strand not considered)	4245	
13	Known gene that overlaps with this transcript within its intron in different strand	965	
14	Known gene where this transcript is possibly a part of its 3'-UTR	757	
15	Known gene where this transcript is possibly a part of its 5'-UTR	77	
16	Known gene within upstream 5 kb	1011	
17	Known gene within downstream 5 kb	402	
18	Average conservation score over the mapped exonic region	6184	0/93
19	Maximum conservation score over the mapped exonic region	5741	0/98
20	Maximum conservation score within 500 base upstream from the mapped 5' terminal	6878	0/255
21	Overlapping UCSC ultra conserved region	24	0/4
22	Number of canonical splice signals in this transcript	751	0/30
23	Number of poly(A) signals in this transcript	8081	0/199
24	Number of CpG island	1353	0/4
25	Associated transposon free region	1137	
26	Number of RFAM known RNA motifs in this transcript	5511	0/12
27	Number of RNAz predictive RNA motifs in this transcript	1185	0/24
28	Number of EvoFold predictive RNA motifs in this transcript	888	0/7
29	Maximum Z-score of RNA secondary structure over this transcript. Scores lower than -6 are significant. Higher scores are considered insignificant. Stored scores= raw score \times -10	252	0.0/121.0
30	Number of cell lines responding to Affy probes in exon regions of this transcript (Affymetrix Transcriptome Phase 2 Tiling Array Analyses)	1593	0/11

Identification (fRNAdb II)

Basic DB/ID Expert Sort Column **A**

Basic Filters:

This is a collection of short-cuts which allows you a quick selection of frequently used filters. Please check items that suit your own criteria. Details of these selection items are described [here](#).

- Mapped
- Well conserved on average (Avg.>50%)
- Multi-exon
- No overlapping known gene
- Tiny ORF (<40aa)
- No protein homolog
- No 5'UTR host
- No up/downstream gene within 5kbp
- Highly Probable Secondary Structure (Rfam and RNAz and EvoFold)
- Has CpG island
- Well conserved at best (Max.>50%)
- EST-supported
- Antisense
- Low Repeat Coverage (<30%)
- No proximal repeats
- No 3'UTR host
- Potential Secondary Structure (Rfam or RNAz or EvoFold)
- Affy-Txn-Ph2 supported

Total hits: 13693

Clicking on each column label let you sort table by that column instantly but limited within the showing table. For sorting entire result entries, please Legend: [DETAIL](#) opens a detail information page, [SEQ](#) opens a sequence page, and [GB](#) opens a Genome Browser window.

B

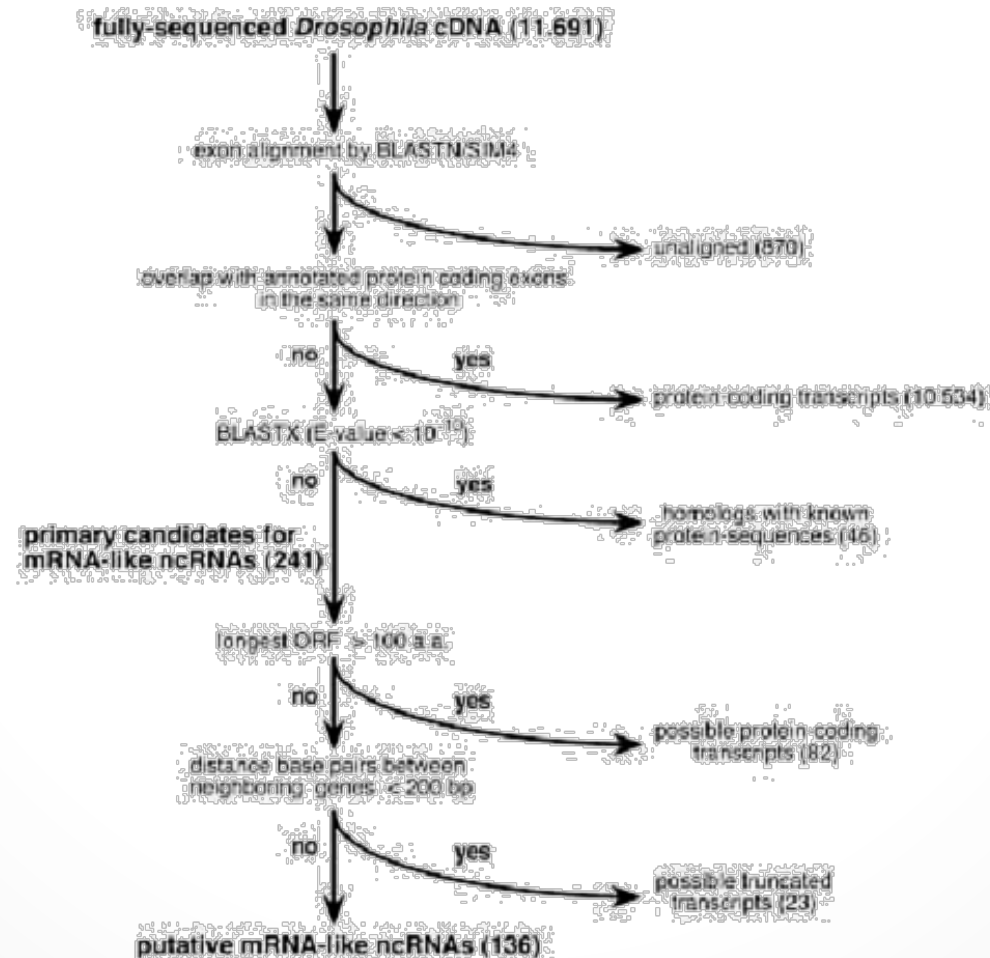
no	ID	Acc.	Original	Common Name	Class	Length (nt)	# of exons	# of ESTs	# of mapping	GC%	M C ler (a
1	FR000001	AB007954	HIT000000201		NPCT	6565	2	7	1	40	
2	FR000002	AB007955	HIT000000202		NPCT	5997	1	0	1	40	
3	FR000003	AB007956	HIT000000203		NPCT	5929	1	4	1	40	
4	FR000004	AB007958	HIT000000215		NPCT	6453	1	0	1	37	
5	FR000005	AB007973	HIT000000220		NPCT	6387	1	2	1	39	
6	FR000006	AB007975	HIT000000222		NPCT	5951	1	1	1	47	
7	FR000007	AB007976	HIT000000223		NPCT	5817	1	7	1	34	

A: frequently used filters

B: transcript found

Function (example..)

Identification and expression analysis of putative mRNA-like non-coding RNA in *Drosophila*



The End

Thank you for watching!