Prediction of functional siRNA structure

Zuzana Ho ejší

1.4.2005

Ústav Molekulární Genetiky
odd. Molekulární Virologie

Previous and current work

Institute of Molecular Genetics, dep. Of Molecular Virology
- projects focused on transcription factor c-Myb

2002-2004 Copenhagen Institute of cancer Biology, dep. of Cell Cycle and Cancer
- cell cycle checkpoint pathways (DNA damage checkpoint pathways in G1/S phase)

RNA interference

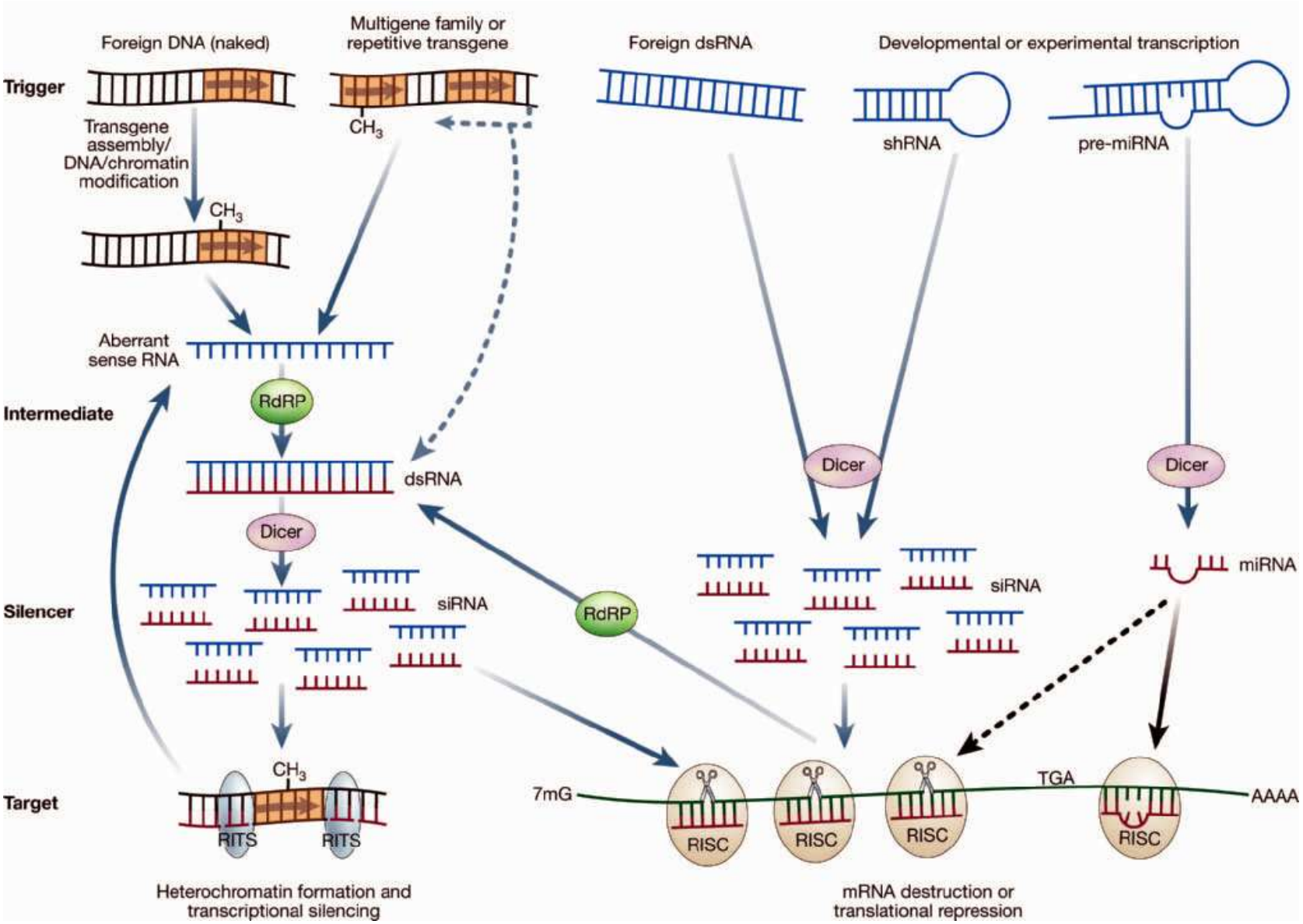
Cleavage of specific mRNA triggered by presence of complementary dsRNA

- first recognized in Caenorhabditis Elegans, evolutionary conserved
- antiviral mechanisms - protecting from RNA viruses
- prevention of random integration of transposable elements

Final product involved in recognition and cleavage is short (19-23 bp) dsRNA – short interfering RNA (**siRNA**)

Other related short RNAs:

- repeat associated short interfering RNAs (**rasiRNAs**) – chromatin changes
- microRNAs (**miRNAs**) - natural hairpin dsRNAs, regulation of protein translation



Mello and Conte. Nature. 2004 Sep 16;431(7006):338-42

Mechanism of recognition and cleavage

RISC = RNA-induced silencing complex

Argonaute protein family (Ago proteins)

-components of RISCs
-tightly bind ssRNAs
-contain PAZ and PIWI domains

PIWI domain – binds Dicer

- similar to RNase H family domain (cleaves the RNA strand of RNA/DNA duplexes)

PAZ domain – recognizes the 2-nucleotide 3' overhang of siRNA

- prevents processing of unrelated RNA or RNA-turnover products

Features important for siRNA function

dsRNA should form -helix

duplex-end stability - whichever strand is most easily unwound in a 5'-3' direction will be preferentially assembled within RISC

GC contents between 30-60%

Avoid GGGG, CCCC, UUUU, AAAA

particular nucleotides at specific positions

mRNA conformation

Differences between the siRNA design algorithms

Algorithm	Description
GPboost	Weighted sum of sequence motifs/patterns
Ui-Tei	Sequence features
Amarzguioui	Sequence features
Hsieh	Sequence features
Takasaki	Sequence features
Reynolds 1	Hairpin potential, sequence features
Reynolds 2	Sequence features
Schwarz	Difference between 3' and 5' stability
Khvorova	Duplex stability profile
Stockholm 1	Energy features
Stockholm 2	Energy features
Tree	Sequence features in decision tree
Luo	mRNA secondary structure features

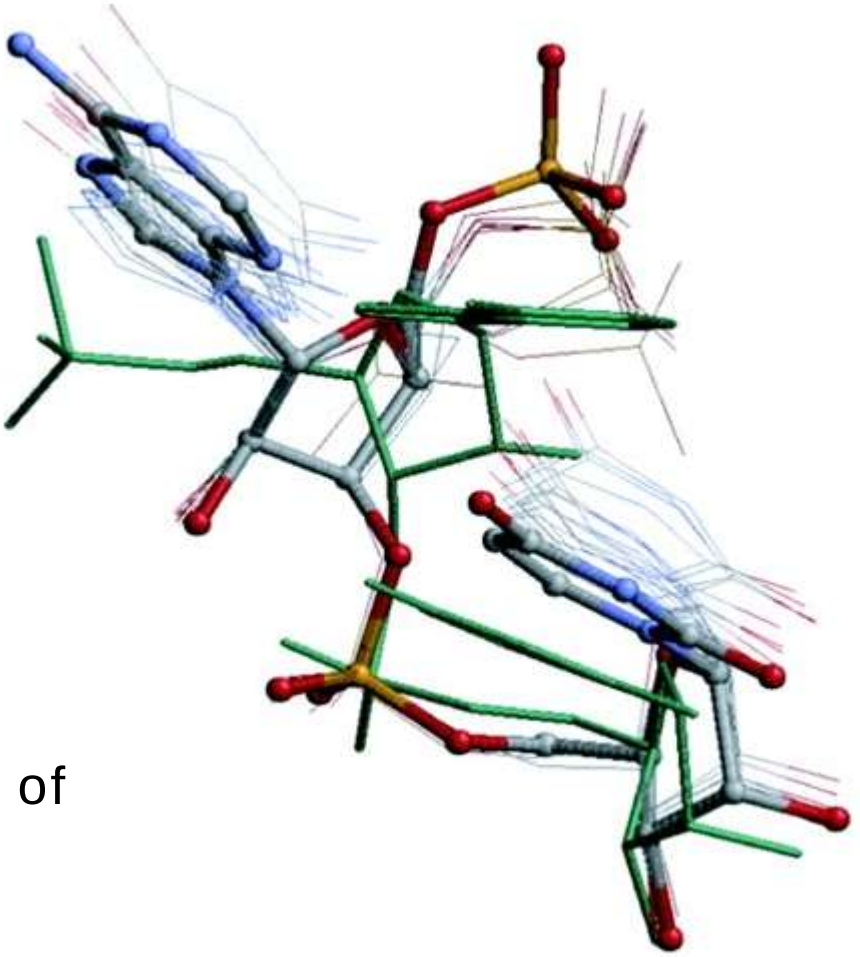
Adapted from Sætrom and Snøve. Biochemical and Biophysical Research Communications 321 (2004

Sequence characteristics used by different algorithms

Algorithm	siRNA sense strand position																																	
	1				2		3		6		7		8		9		10	11		13		15		16		17		18		19					
	A	C	G	U	A	U	A	U	A	C	G	U	A	G	G	U	U	C	G	A	G	A	U	A	G	U	A	U	A	U	A	C	G	U
Reynolds 1 and 2							1										1				−1	1	1	1	1	1	1	1	1	2	−1	−1	1	
Ui-Tei	−1	1	1	−1																										1	−1	−1	1	
Amarzguioui	−1	1	1	−2	−1	−1	−1	−1	1																1	1	1	1	2		−1	2		
Hsieh										−1								1	1	1				1							−1	1		
Takasaki	−3.97		7.4	−3.75					2.33		2.4	−2.59	3.02	−2.35	−2.35	2.3						2.7									−2			

Adapted from Sætrom and Snøve. Biochemical and Biophysical Research Communications 321 (2004

Our approach



Find any rules for prediction of functional

siRNAs based on their 2/3 D structure

Adapted from Schneider B, Moravek Z, Berman HM. Nucleic Acids Res. 2004 Mar 11;32(5):1666

Available data sets

Author	Amarzguioui	Harborth	Hsieh	Khvorova	Reynolds	Ui-Tei	Vickers
Count	46	44	108	14	240	53	76

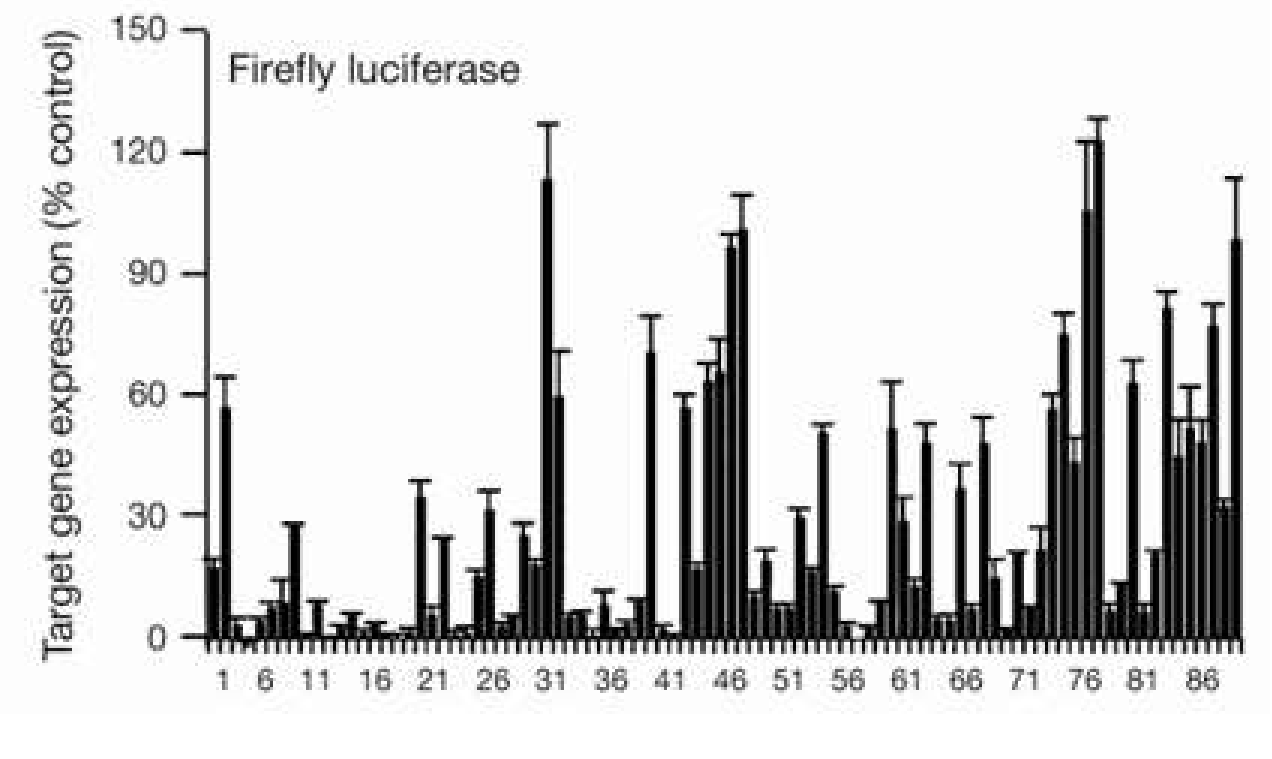
Data from Reynolds

90 siRNAs each targeting every other position of the regions

Human cyclophilin B (M60857) 193-390 [90]
efficacy >= 80 [9]
efficacy <= 20 [41]

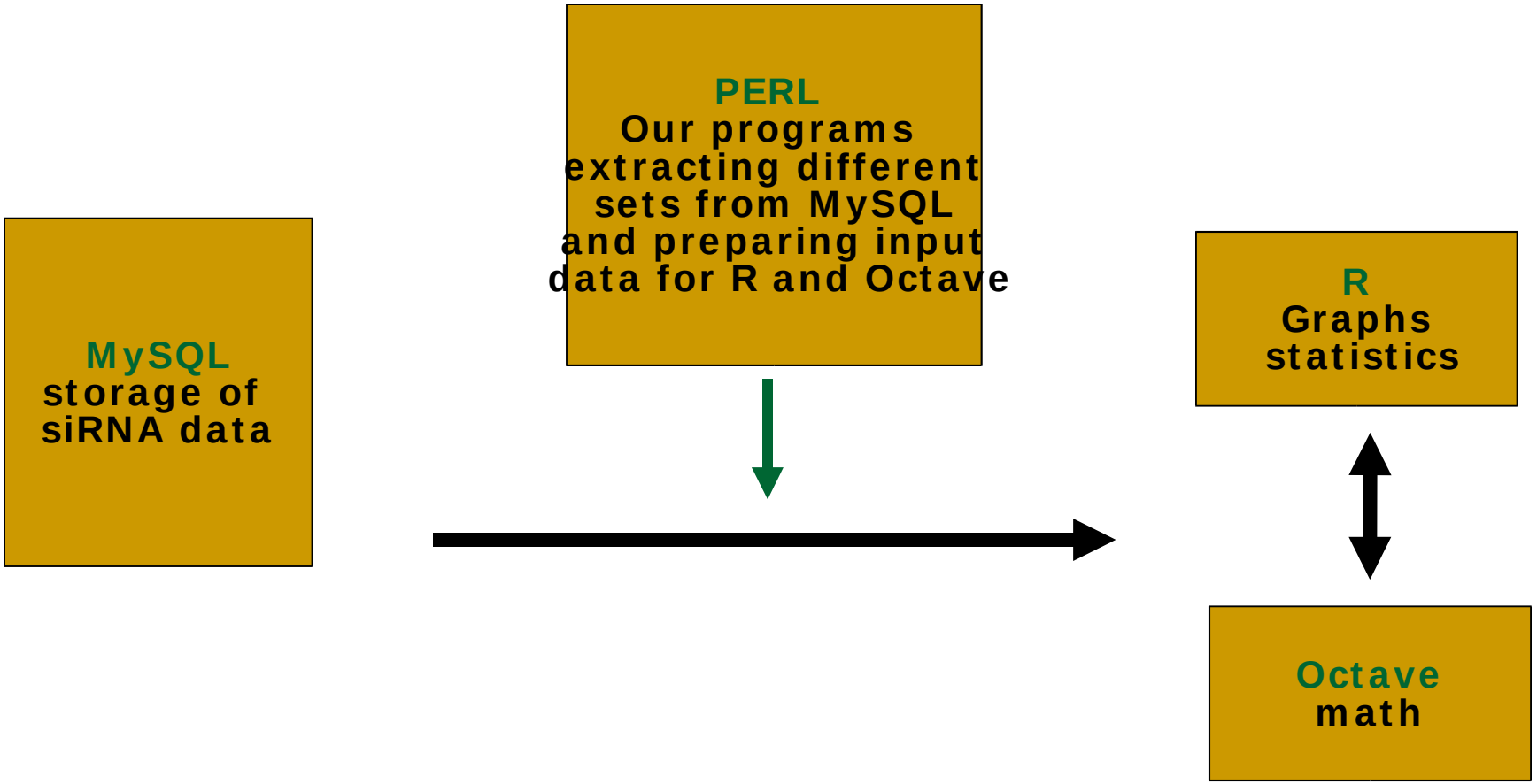
Firefly luciferase (U47298) 1434-1631 [90]
efficacy >= 80 [8]
efficacy <= 20 [49]

Reynolds cyclophylin B

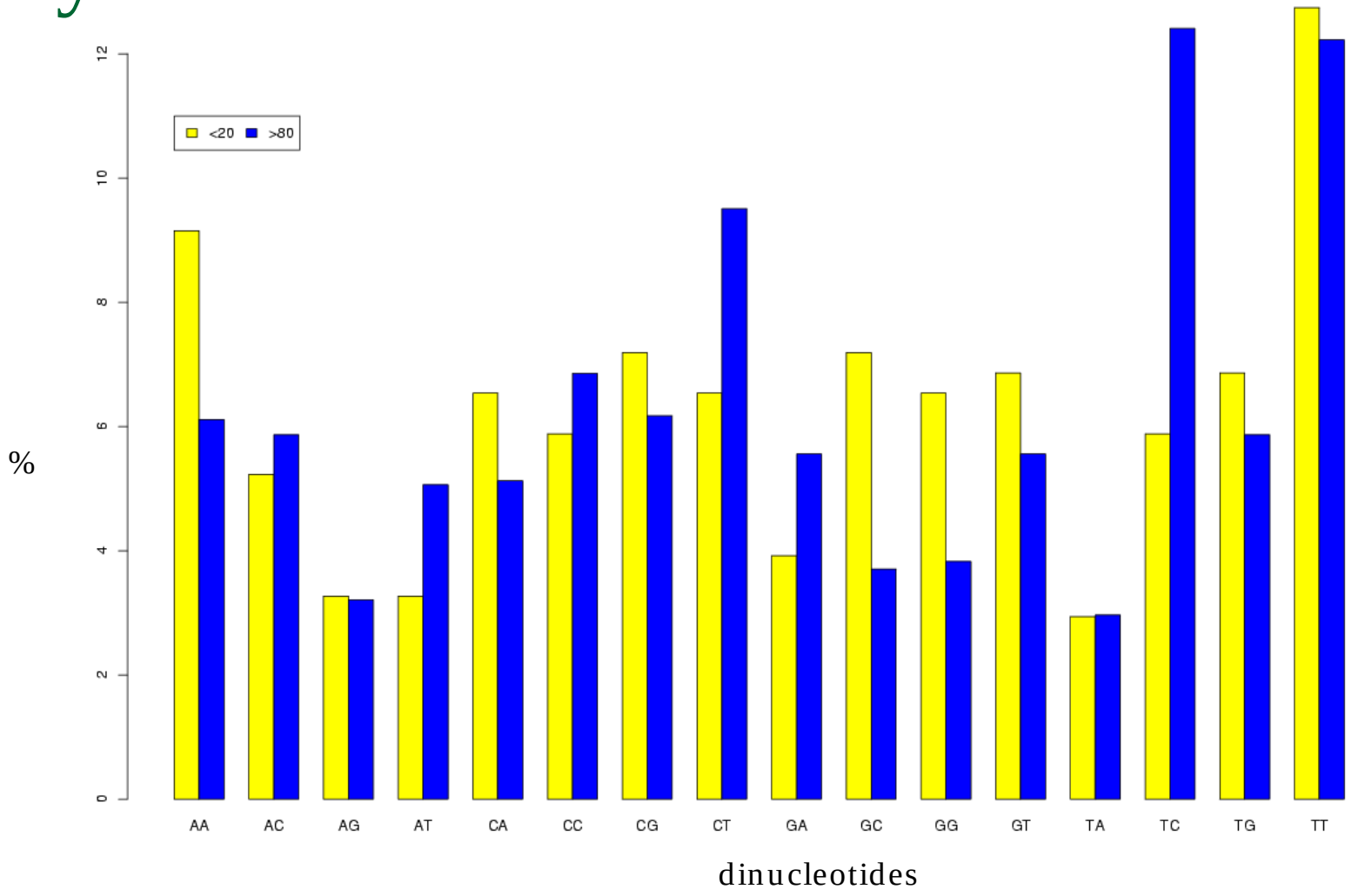


Reynolds A, Leake D, Boese Q, Scaringe S, Marshall WS, Khvorova A. Nat Biotechnol. 2004 Mar;22(3):

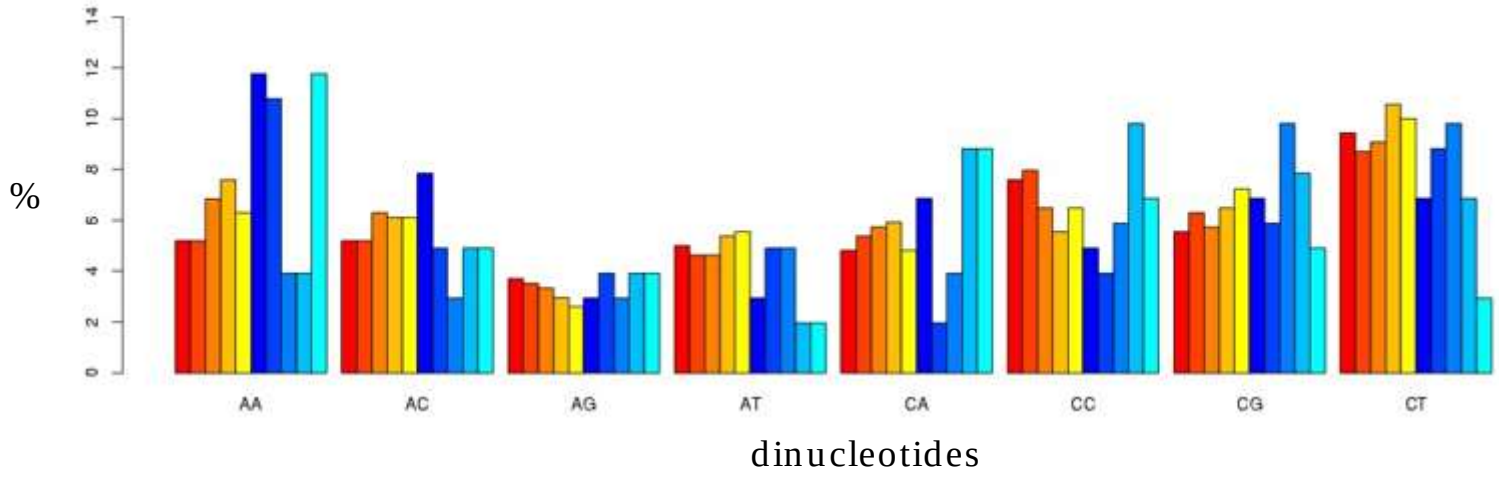
Bioinformatic workflow

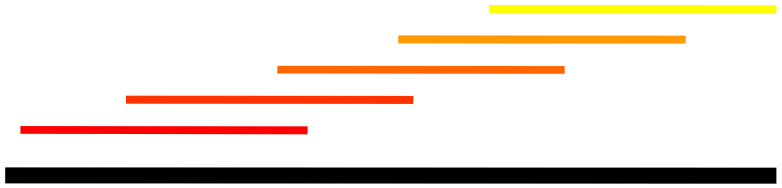


Number of dinucleotides in both Reynolds sets

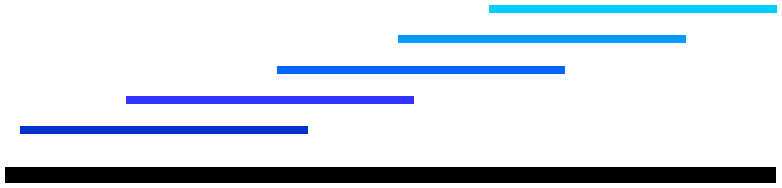


Number of dinucleotides in both Reynolds sets with 3 nucleotide shift



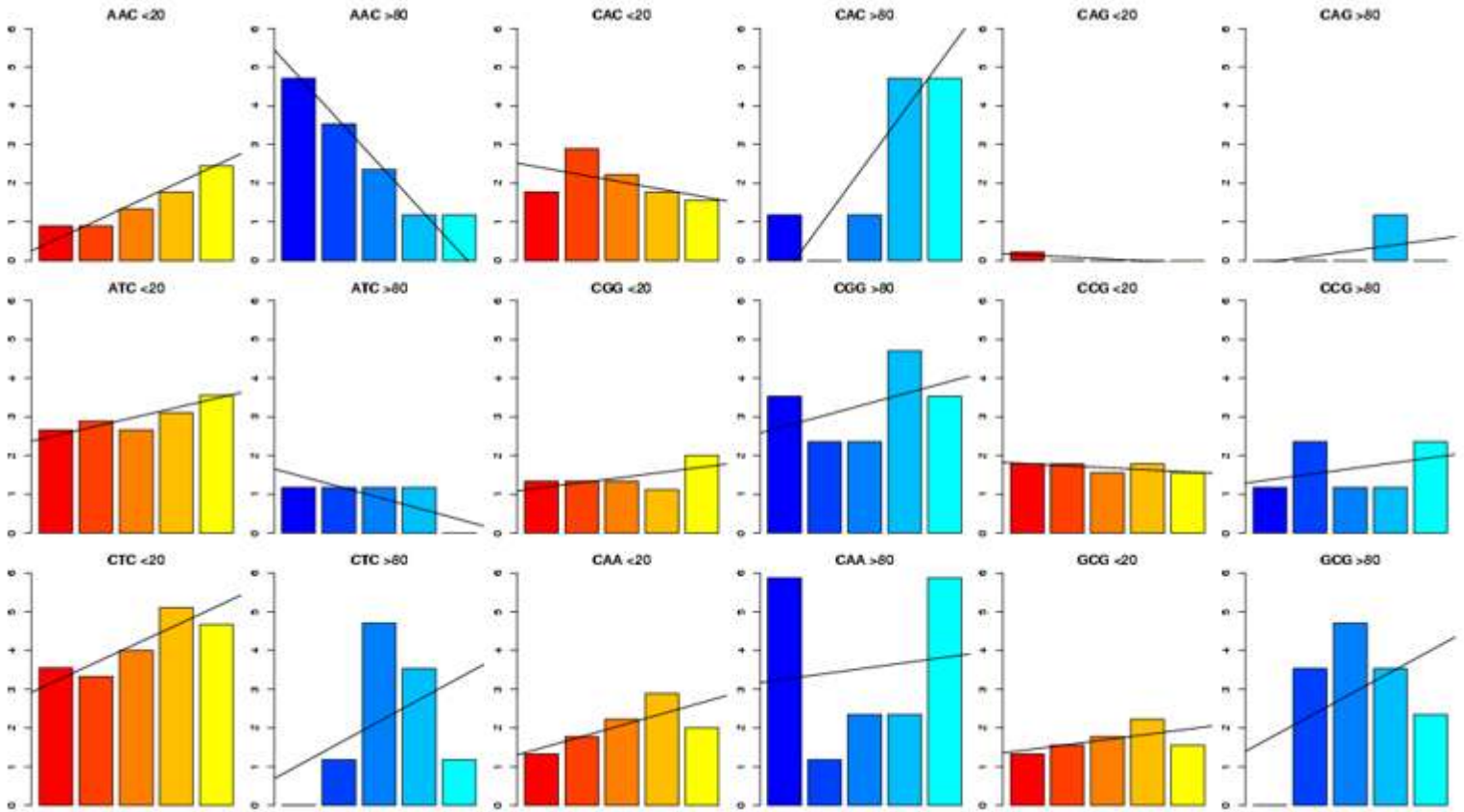


> 80



< 20

Number of trinucleotides in both Reynolds sets with 3 nucleotide shift



Principle of prediction

data sets :

Several different sets of chosen di or trinucleotides

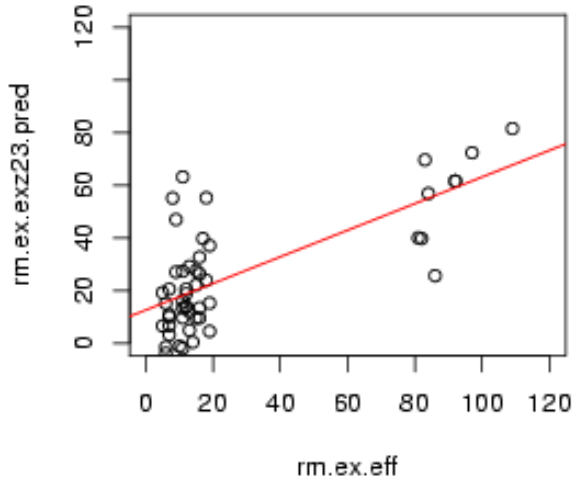
Set of functions that allow obtaining of parameters which show importance of each di/trinucleotide for resulting efficacy of concrete siRNAs

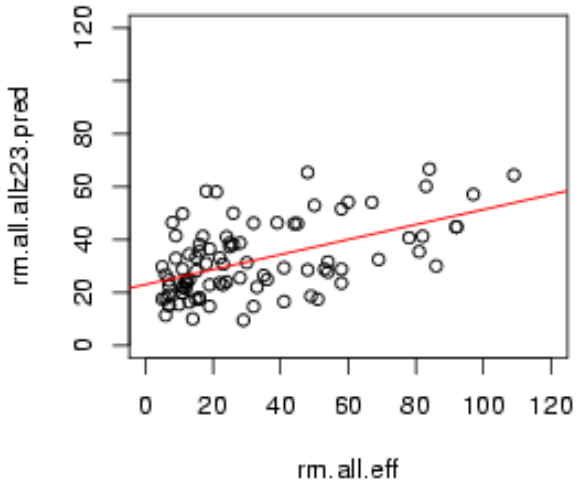
- eg. for sequence ATGTGGCATC with efficacy 82 for ATG [1..10], GGG [1..10], ATC [1..5]:

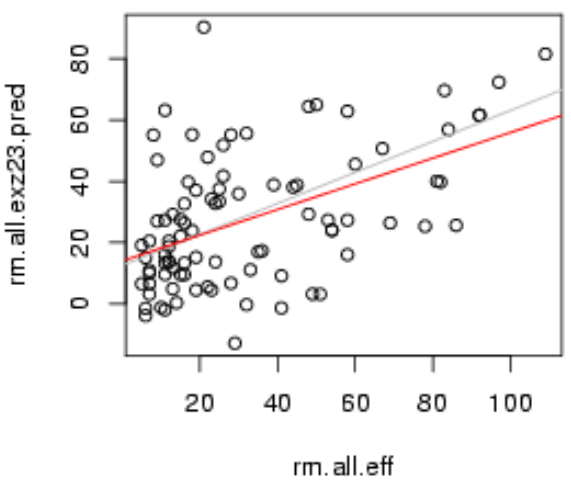
- 1.ATG + 0.GGG + 0.ATC = 82

...system of linear equations of all siRNAs with efficacy < 20 and > 80 from one half of Reynolds data

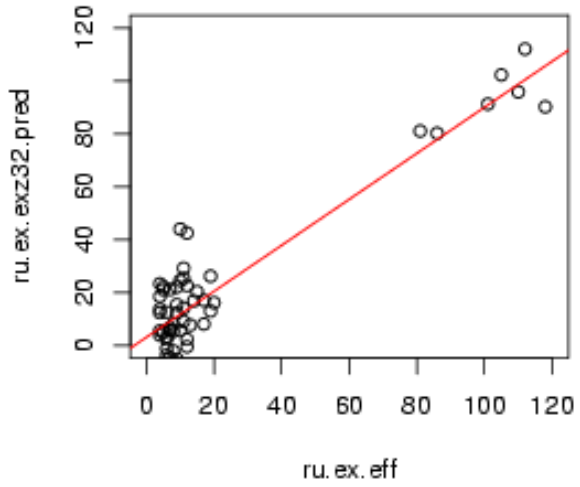
Prediction for a set of dinucleotides

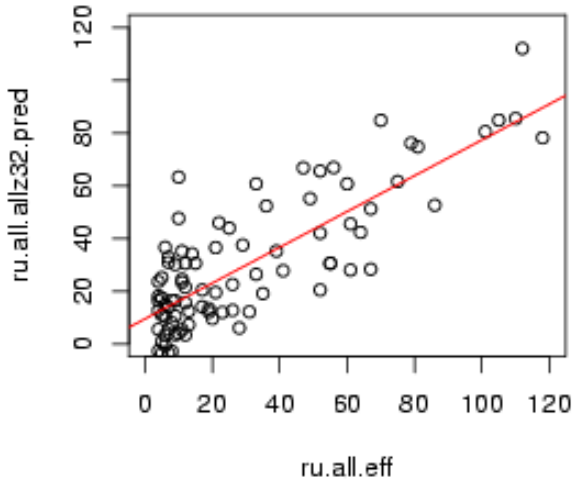


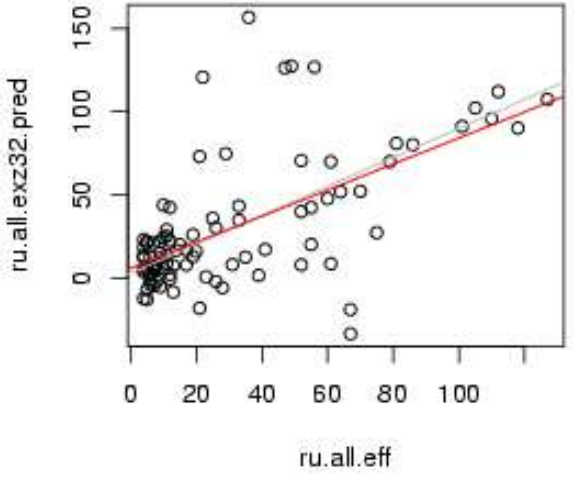




Prediction for a set of trinucleotides







mechanism

dsRNA

- produced by RNA-templated RNA polymerization (viruses) siRNA

- hybridization of overlapping transcripts (repetitive sequences - transgene arrays or transposons) rasiRNA

-endogenous transcripts with complementary 20-50 bp inverted repeats miRNA

dsRNA processed by dsRNA specific RNase-III-type endonucleases - Drosha and Dicer