Поиск возможных сайтов RAN трансляции

No PasaRAN! Team:
Алена Кизенко
Наталья Родина
Лаврентий Данилов
Ярослав Соловьев
Дмитрий Смирнов

Tutor:
Антон Бондаренко

What is RAN translation and why do we do this?

Repeat-associated **non-AUG** (RAN) translation (first time discovered in 2011 *Zu T et al. Proc. Natl Acad. Sci.*)

- **Microsatellite expansion (3-6)**
- Shown in 7 neurological diseases
- Translation from: **introns**, **UTRs**, exons, non-coding RNA (?)
- In **multiple reading frames**
- Products: multiple homopolymeric or dipeptide repeat-containing proteins – involved in protein aggregation
- **G-rich**, role of **G-quadruplexes** unknown

Disease	Repeat	Strand	Number of repeats	RAN protein	Region	Gene
Spinocerebellar ataxia type 8	CAG	antisense	>73	poly-Alanine		overlapping ATXN8OS and ATNX8
	CAG	antisense	>42	polyGln		
	CAG	antisense	>58	polySer		
Myotonic dystrophy type 1	CTG	sense	80-4000		3'UTR	DMPK
	CAG	antisense		poly-Glutamine		
ALS / FTD	GGGGCC	sense	>145	poly-GlycineProline	1st intron	C9orf72
	GGGGCC	sense		poly-GlycineArginine		
	GGGGCC	sense	>38	polyGlycineAlanine		
	GGCCCC	antisense		poly-AlanineProline		
	GGCCCC	antisense		poly-ProlineArginine		
	GGCCCC	antisense		poly-GlycineProline		

Fragile X(FXTAS)	CGG	sense	55–200, (>30)	poly-Glycine	5'UTR	FMR1
	CGG	sense	>88	polyAla		
	CGG	sense		polyArg		
	CCG	antisense		poly-Proline		
	CCG	antisense		poly-Alanine		
Huntington's disease	CAG	sense	>52	HD-poly-Alanine	1st exon	HTT
	CAG	sense	>35	HD-polySerine		
	CTG	antisense		HDpoly-Leucine		
Spinocerebellar ataxia type 31	TGGAA	antisense	22	poly-WNGME, GST-(MEWNG)n	intron	BEAN1
	TTCCA	antisense		RAN транлсяция пока не показана		TK2
Myotonic Dystrophy Type 2	CCTG	sense	80-11000	LPAC	intron 1	CNBP
	CAGG	antisense		QAGR		

Aim:

- Find new potential RAN translation sites
- Create database

Known in literature RAN cases

1

Annotating known cases of RAN translation

Articles → RAN translation cases → Genes

Gene	Strand	Repeat	Repeat length range
FMR1	sense	GCC	55-200
HTT	sense	CAG	>23
HTT	antisense	GCT	>52
HTT	antisense	GCT	>52
HTT	antisense	GCT	>52
HTT	antisense	GCT	>52
HTT	antisense	GCT	>52
HTT	antisense	GCT	>52
HTT	antisense	GCT	>52
HTT	antisense	GCT	>52
C9orf72	sense	GGGGCC	>145
ATNX80S	antisense	CAG	>42
ATNX80S	antisense	CAG	>42
ATNX8OS	antisense	CAG	>58
ATNX8OS	antisense	CAG	>73
DMPK	antisense	CAG	>80

ORF: Consensus + Kozak

1

Annotating known cases of RAN translation

Articles → RAN translation cases → Genes → Search for ORF → Motifs

Kozak consensus sequence

Start **G** — Repeat — Stop

CTG ATT
GTG ATG
ACG GCG

3-6 nucleotides

TAA
TAG
TGA

CANONICAL SEQUENCE START CODON -> REPEATS														
1	2	3	4	5	6	7	8	9	10	11	13	16	18	22
G	C	G	G	C	G	G	C	G	R	Y	G	G	T	G
100%	60%	60%	60%	50%	63%	55%	44%	56%	90%	60%	45%	45%	45%	45%
	A	A	A											
	30%	30%	30%											

CANONICAL SEQUENCE REPEATS -> STOP CODON									
1	2	3	4	5	6	9	10	18	27
Y	A	R	!C	R	G	G	!A	G	G
60%	56%	91%	94%	94%	56%	44%	100%	44%	44%
	G			!T	!T				
	31%			100%	100%				

=> Выравнивание на консенсус и введение системы штрафов

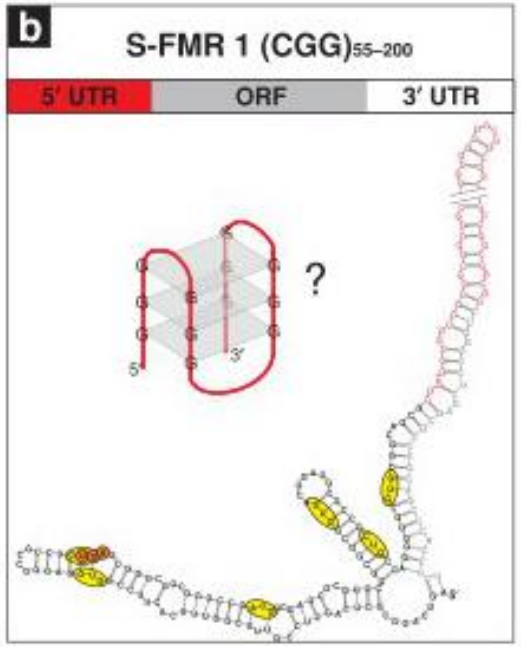
Testing for G-quadruplex

1

Annotating known cases of RAN translation

Articles → RAN translation cases → Genes → Search for ORF → Motifs

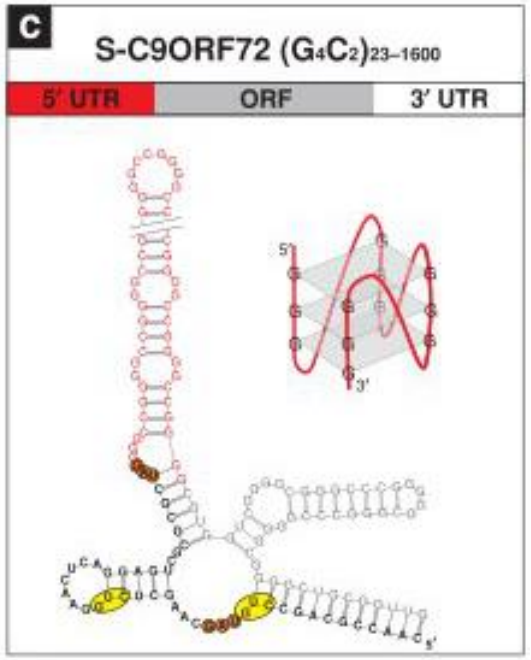
Genes → mRNA structure → G quadruplex or not?



G-quadruplex

(G{3,4}[AUGC]{1,12}){4,4}

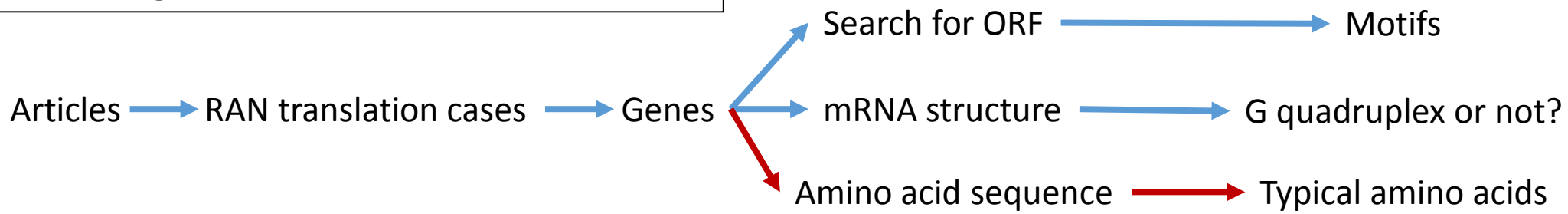
Kim et al. Nucleic Acids Research, 2016
https://doi.org/10.1093/nar/gkw272



Typical amino acid content

1

Annotating known cases of RAN translation



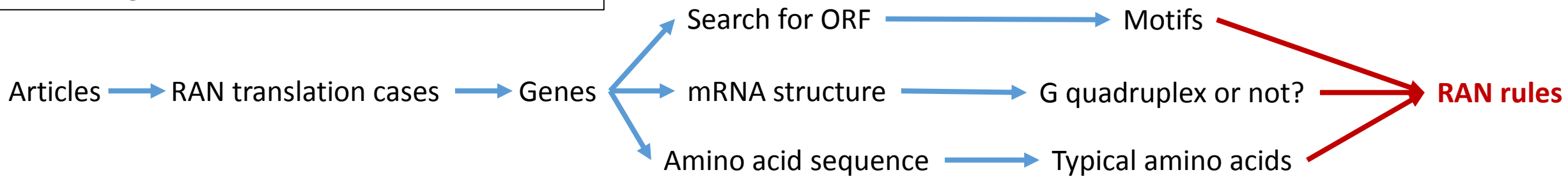
Typical amino acid content:

Ala, Gln, Ser, Gly, Pro, Leu
GlyPro, GlyArg, GlyAla, GlyPro
ProAla, ProArg,
MEWNG, LPAC, QAGR

RAN rules

1

Annotating known cases of RAN translation



Find potential RNA translation sites using RAN rules

2

Annotating known cases of RAN translation

Articles → RAN translation cases → Genes

- Search for ORF → Motifs
- mRNA structure → G quadruplex or not?
- Amino acid sequence → Typical amino acids

→ **RAN rules**

Find potential RNA translation sites using **RAN rules**

Tandem repeats database → **RAN rules** → **sRAN Base: 1136**

https://tandem.bu.edu

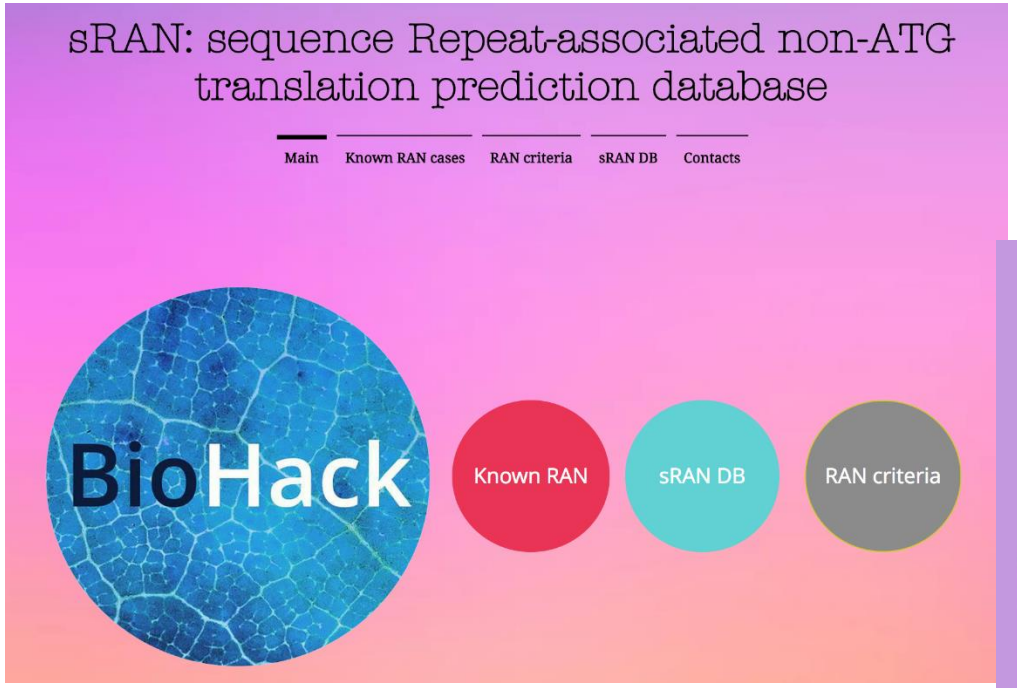
Literature → **RAN rules**

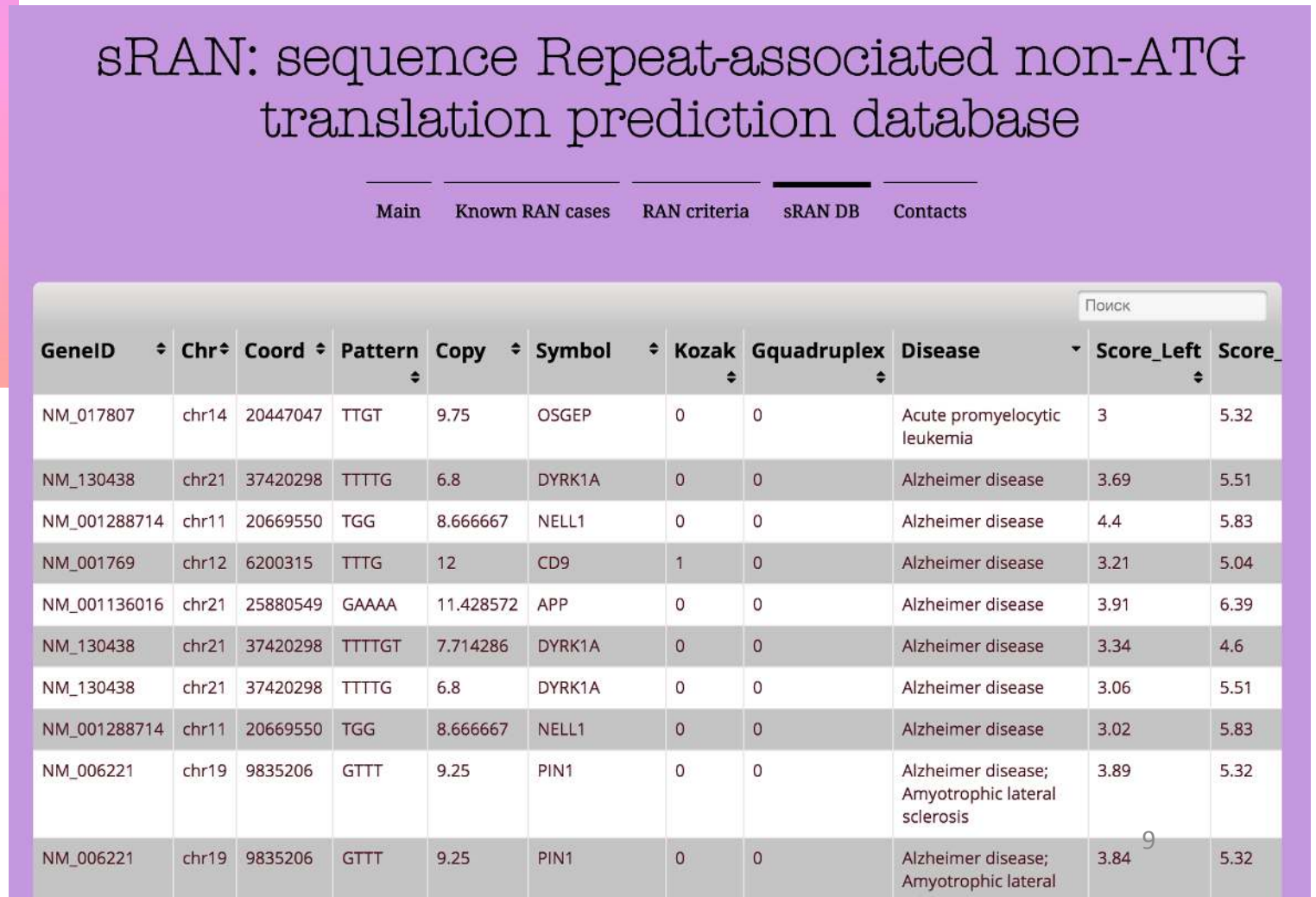
Annotate using **Cider, ClinVar**
For cases of disease → **RAN rules**

sRAN Base: 1136 → Selected **21** cases associated with diseases for analysis

→ Domain search,
- Beta-arch search

Our result – the sRAN online webpage and database!





GeneID	Chr	Coord	Pattern	Copy	Symbol	Kozak	Gquadruplex	Disease	Score_Left	Score_
NM_017807	chr14	20447047	TTGT	9.75	OSGEP	0	0	Acute promyelocytic leukemia	3	5.32
NM_130438	chr21	37420298	TTTTG	6.8	DYRK1A	0	0	Alzheimer disease	3.69	5.51
NM_001288714	chr11	20669550	TGG	8.666667	NELL1	0	0	Alzheimer disease	4.4	5.83
NM_001769	chr12	6200315	TTTG	12	CD9	1	0	Alzheimer disease	3.21	5.04
NM_001136016	chr21	25880549	GAAAA	11.428572	APP	0	0	Alzheimer disease	3.91	6.39
NM_130438	chr21	37420298	TTTTGT	7.714286	DYRK1A	0	0	Alzheimer disease	3.34	4.6
NM_130438	chr21	37420298	TTTTG	6.8	DYRK1A	0	0	Alzheimer disease	3.06	5.51
NM_001288714	chr11	20669550	TGG	8.666667	NELL1	0	0	Alzheimer disease	3.02	5.83
NM_006221	chr19	9835206	GTTT	9.25	PIN1	0	0	Alzheimer disease; Amyotrophic lateral sclerosis	3.89	5.32
NM_006221	chr19	9835206	GTTT	9.25	PIN1	0	0	Alzheimer disease; Amyotrophic lateral	3.84	5.32

Links

Webpage:
mityasmirnov.wixsite.com/sran

Github:
https://github.com/Borschevik/RAN-Translation

mityasmirnov.wixsite.com/sran