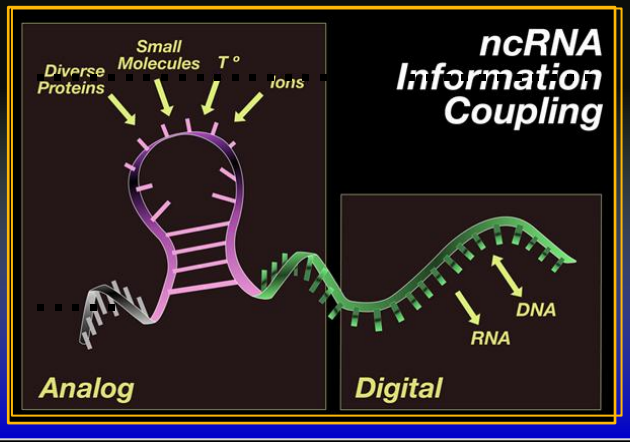
Computational mechanisms and information coding by the non-coding transcriptome



Georges St. Laurent

Brown University and St. Laurent Institute

Providence, Rhode Island

Georges_st_Laurent@brown.edu

IN DEO SPERAMUS

SLI
st. laurent
institute

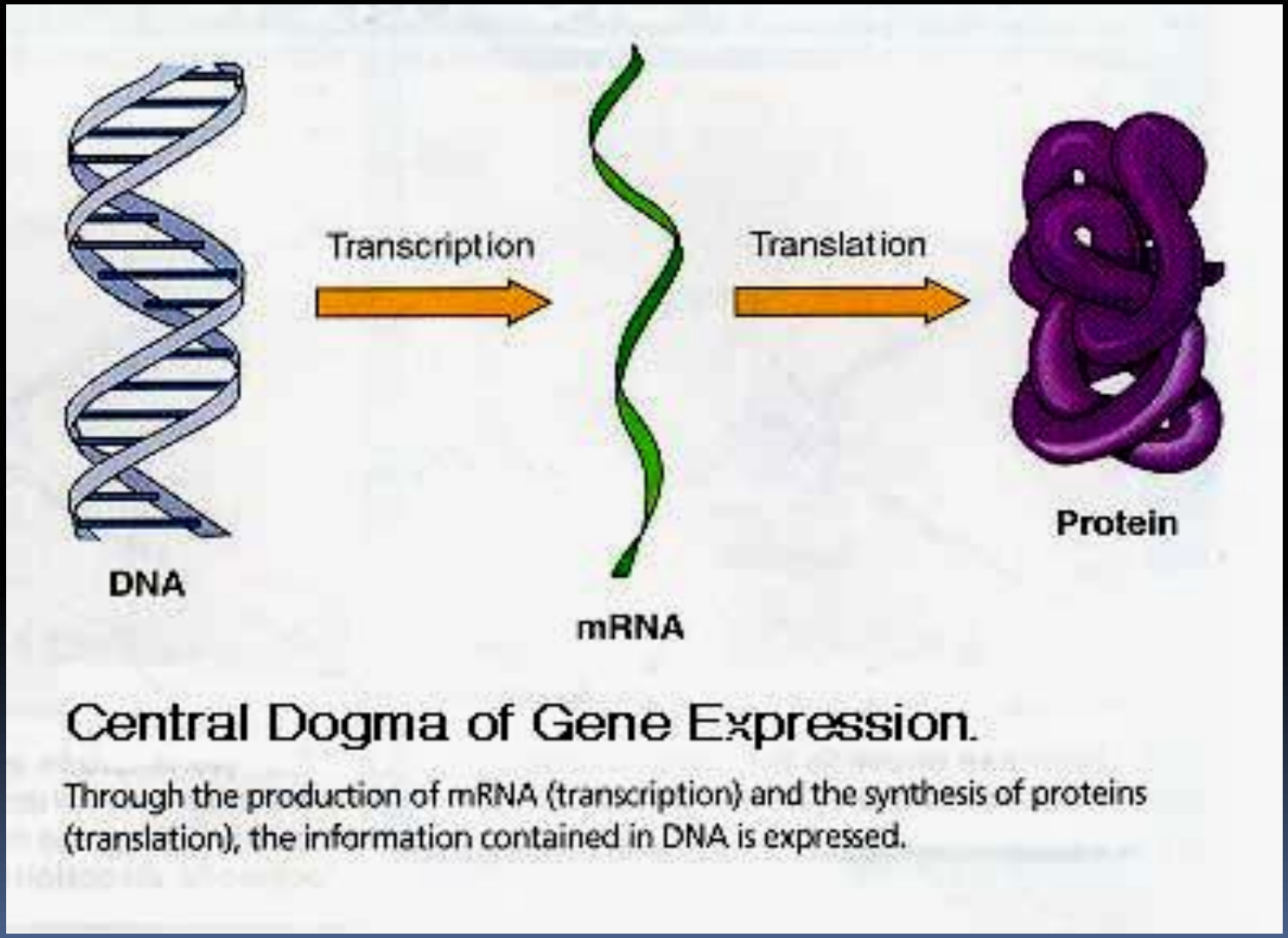
Three Goals of today's talk.....

.....The non-coding transcriptome plays pervasive roles throughout the functional systems of the nervous system.

.....catalyzes a new paradigm of computational complexity and information processing in the nervous system.

....May solve the dilemma of Genomics and Neuroscience.

Francis Crick's Central Dogma of biological information flow

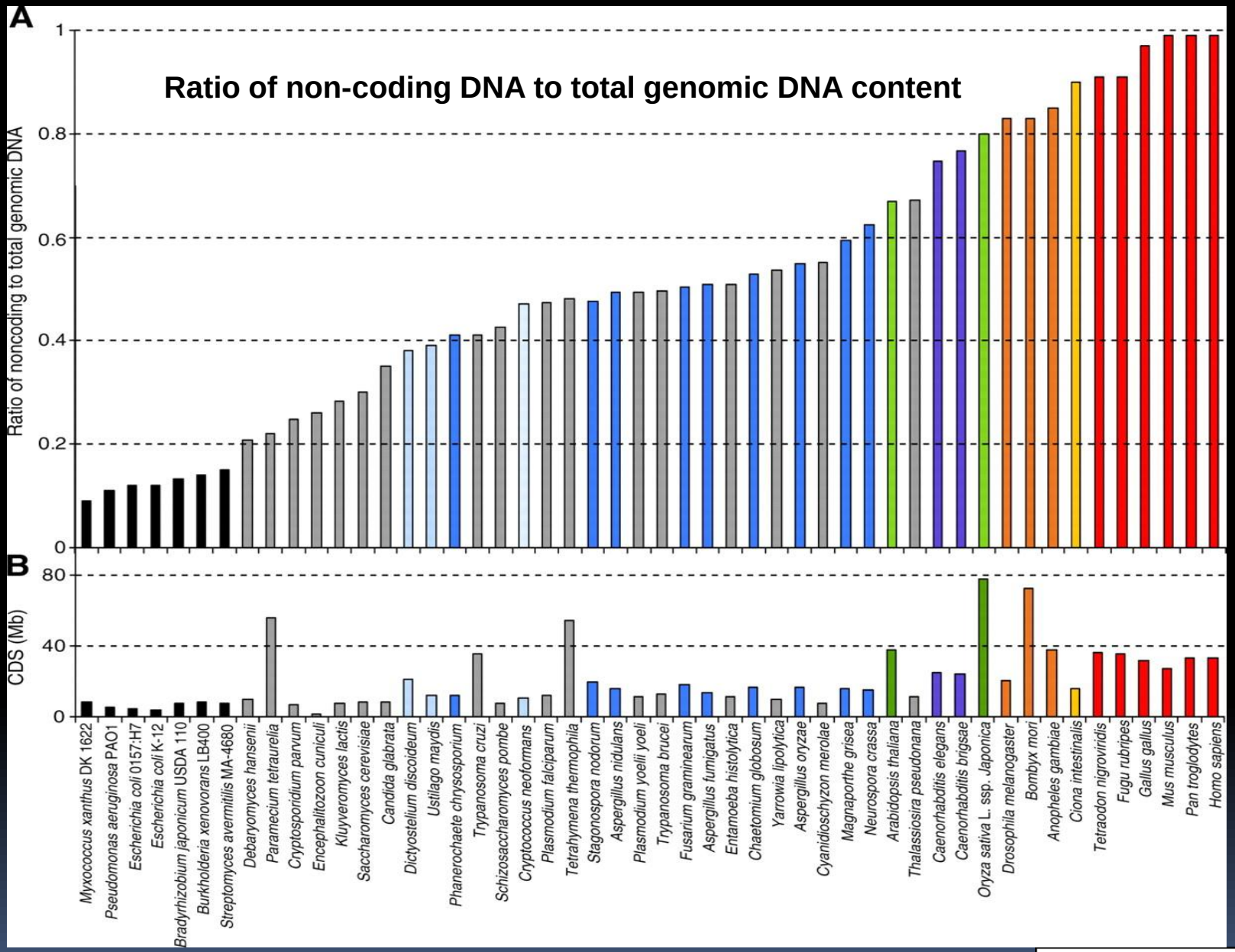


Problems with the Central Dogma.

1) **Where's the complexity?**
   - How many genes?

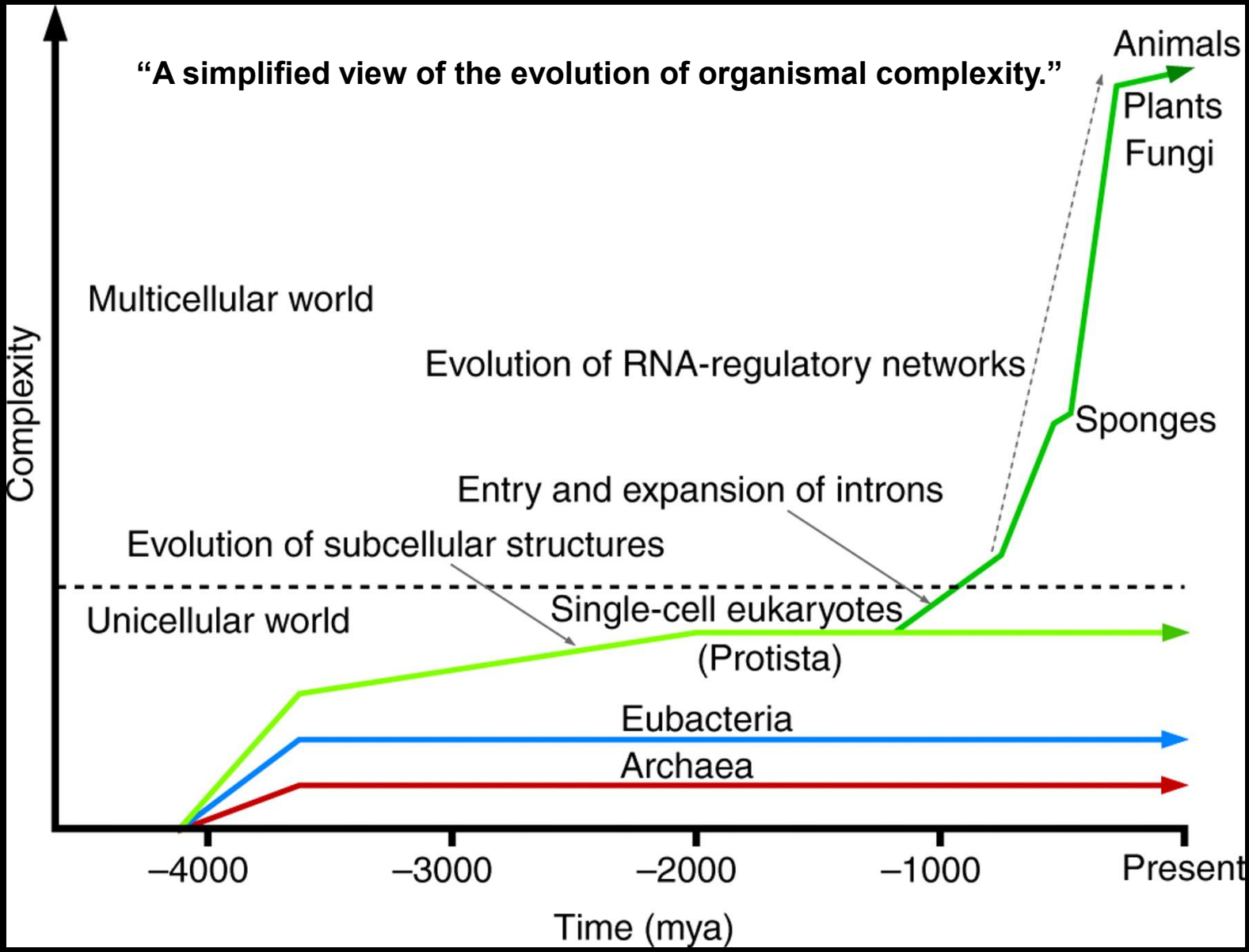
2) **What's in the DNA? … junk?**
   - Why keep it around?

3) **Evolutionof regulatory networks?**
   - Regulator requirements in protein networks increase quadratically.



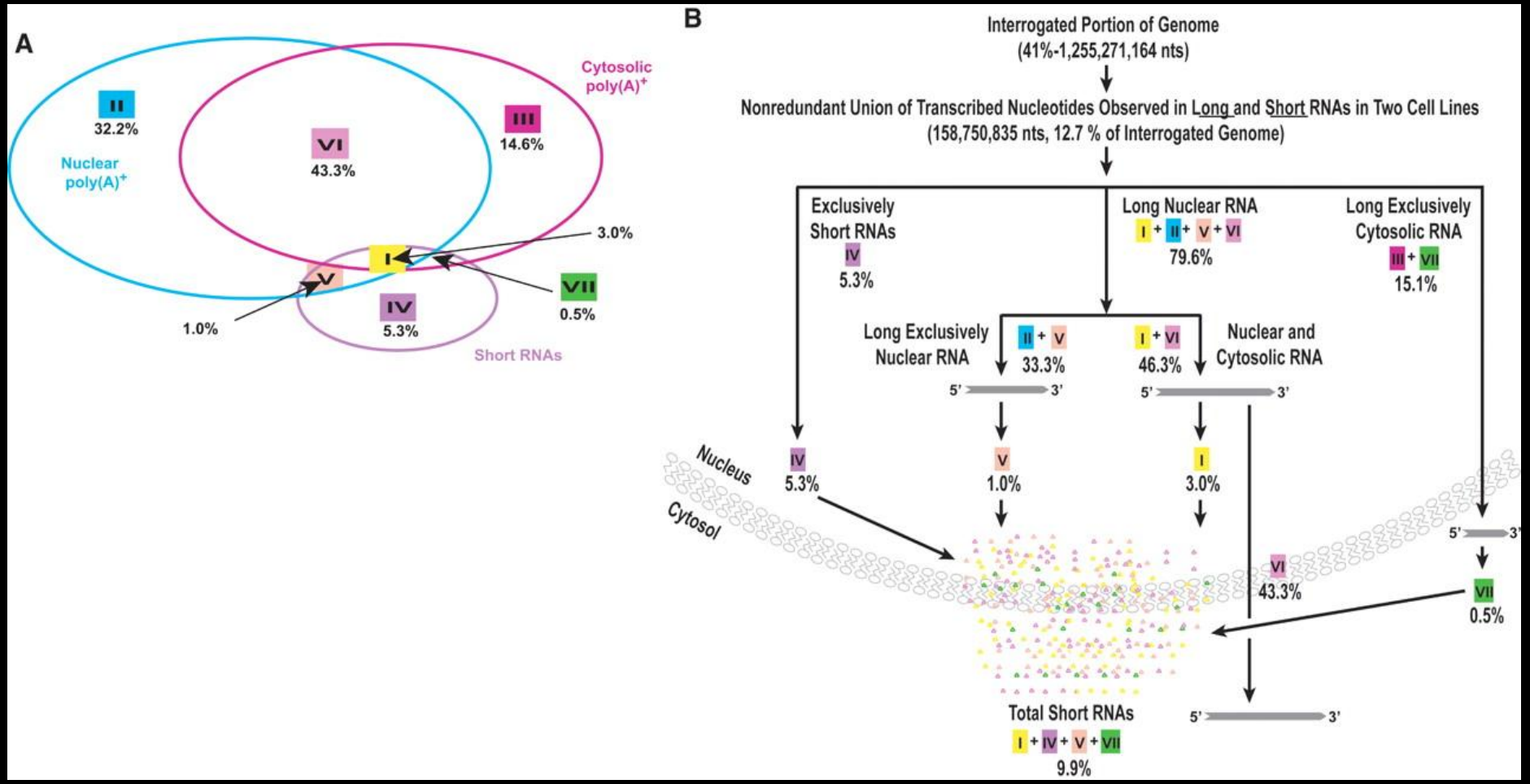
Taft and Mattick, Exp Biol (2007) 210:1526-1547

The Journal of Experimental Biology

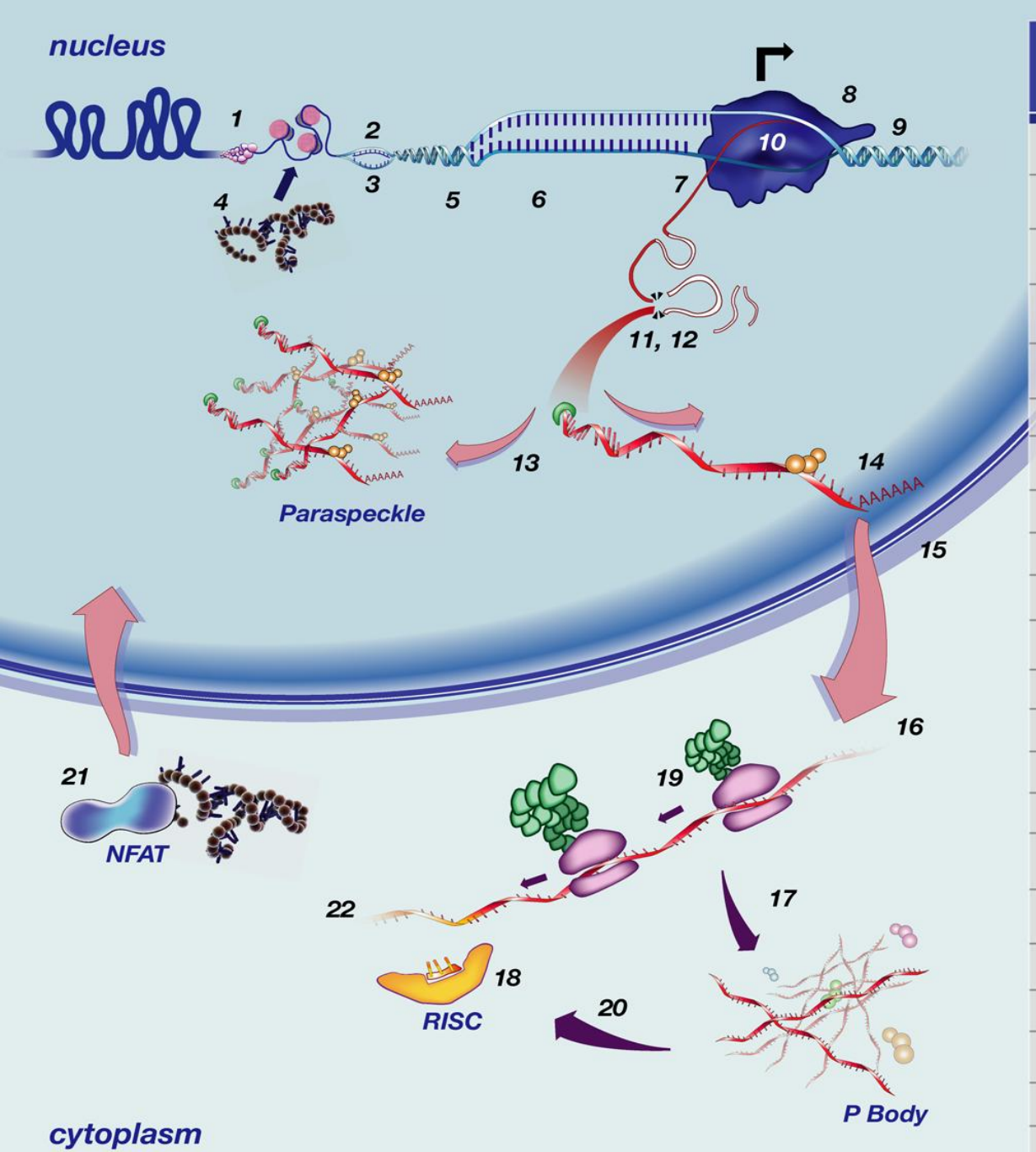


Mattick, J. S. J Exp Biol 2007;210:1526-1547

Relations among the transcribed bases in the nonrepeat portions of the human genome



P. Kapranov et al., Science 316, 1484 -1488 (2007)



Mechanism of Regulation	Examples
1. Epigenetic Guidance (Imprinting, Methylation, Demethylation)	*IGF2r; KvLQT1* (100,101)
2. RNA – DNA Triplex Inhibition of Promoter Region	*DHFR locus, upstream transcripts* (39)
3. RNA based enhancement of transcription	*Air; EVF-2* (108)
4. RNA Induced Transcriptional Gene Silencing (TGS)	*RITS* (102)
5. Modifications to Functional RNAs	*snoRNAs* (51)
6. X Inactivation	*XIST; Tsix* (113)
7. ncRNAs targeting activators and repressors of transcription	*NRSE;SRA;HSR1* (58,100,101)
8. Stimulation of Initiation	*U1 to TFIIH* (46)
9. Transcription Elongation	*7SK to PtefB* (42)
10. ncRNA targeting Pol II	*B2 RNA* (44,45)
11. Splicing	*HBII52; SC35 complex, TTP riboswitch, snRNA; N-myc* (43,72)
12. Transcript Editing	*HBII-52; TS* (72)
13. Nuclear Sequestration	*CTN – paraspeckles* (102)
14. Polyadenylation	*IDE* (106)
15. Transport to the cytoplasm	*NF90 - Tau mRNA* (107)
16. mRNA Stabilization/ Destabilization	*Brl-2; HIF-∝1* (103)
17. Sequestration to P Body by RISC Complex	*miRNA* (83,84,89,90)
18. Transcript cleavage by RISC	*siRNA* (104)
19. Translation Initiation Arrest	*B1; B200; BCMA* (92,91)
20. Release of Translation Inhibition	*CAT mRNA in response to stress* (83)
21. Nuclear Localization of TFs.	*NRON* (13)
22. Co-translational protein targeting	*7SL RNA* (112)

PNAS PNAS PNAS PNAS

Specific expression of long noncoding RNAs in the mouse brain

Tim R. Mercer*, Marcel E. Dinger*, Susan M. Sunkin†, Mark F. Mehler‡, and John S. Mattick*§

*Australian Research Council (ARC) Special Research Centre for Functional and Applied Genomics, Institute for Molecular Bioscience, University of Queensland, St. Lucia, QLD 4072, Australia; †Allen Institute for Brain Science, Seattle, WA 98103; and ‡Institute for Brain Disorders and Neural Regeneration, Departments of Neurology, Neuroscience and Psychiatry, and Behavioral Sciences, Einstein Cancer Center and Rose F. Kennedy Center for Research in Mental Retardation and Developmental Disabilities, Albert Einstein College of Medicine, Bronx, New York, NY 10461

Edited by Huda Y. Zoghbi, Baylor College of Medicine, Houston, TX, and approved November 21, 2007 (received for review July 17, 2007)

A major proportion of the mammalian transcriptome comprises long RNAs that have little or no protein-coding capacity (ncRNAs). Only a handful of such transcripts have been examined in detail, and it is unknown whether this class of transcript is generally functional or merely artifact. Using *in situ* hybridization data from the Allen Brain Atlas, we identified 849 ncRNAs (of 1,328 examined) that are expressed in the adult mouse brain and found that the majority were associated with specific neuroanatomical regions, cell types, or subcellular compartments. Examination of their genomic context revealed that the ncRNAs were expressed from diverse places including intergenic, intronic, and imprinted loci and that many overlap with, or are transcribed antisense to, protein-coding genes of neurological importance. Comparisons between the expression profiles of ncRNAs and their associated protein-coding genes revealed complex relationships that, in combination with the specific expression profiles exhibited at both regional and subcellular levels, are inconsistent with the notion that they are transcriptional noise or artifacts of chromatin remodeling. Our results show that the majority of ncRNAs are expressed in the brain and provide strong evidence that the majority of processed transcripts with no protein-coding capacity function intrinsically as RNAs.

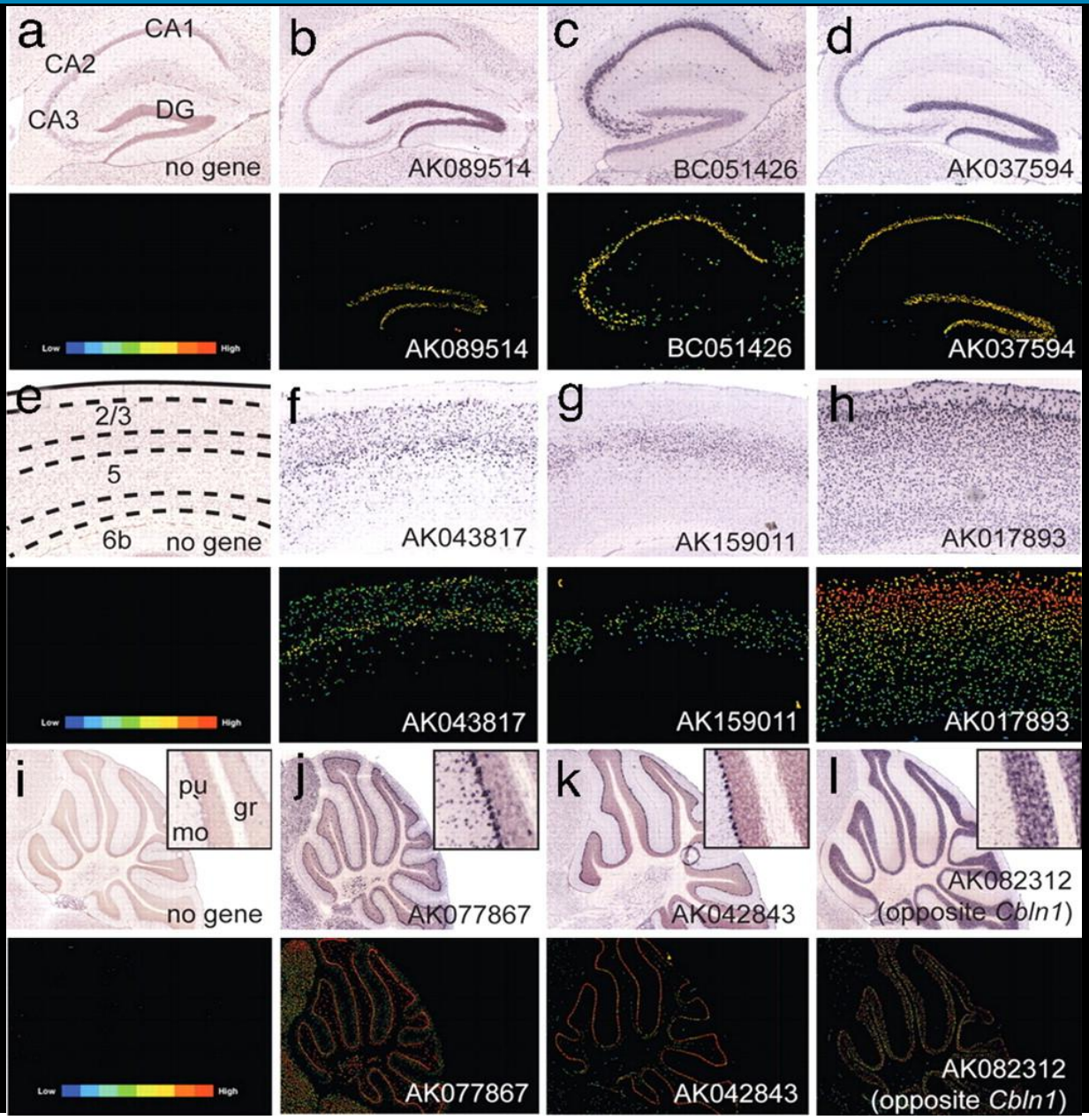


Fig. 1

Mercer T. R. et.al. PNAS 2008;105:716-721

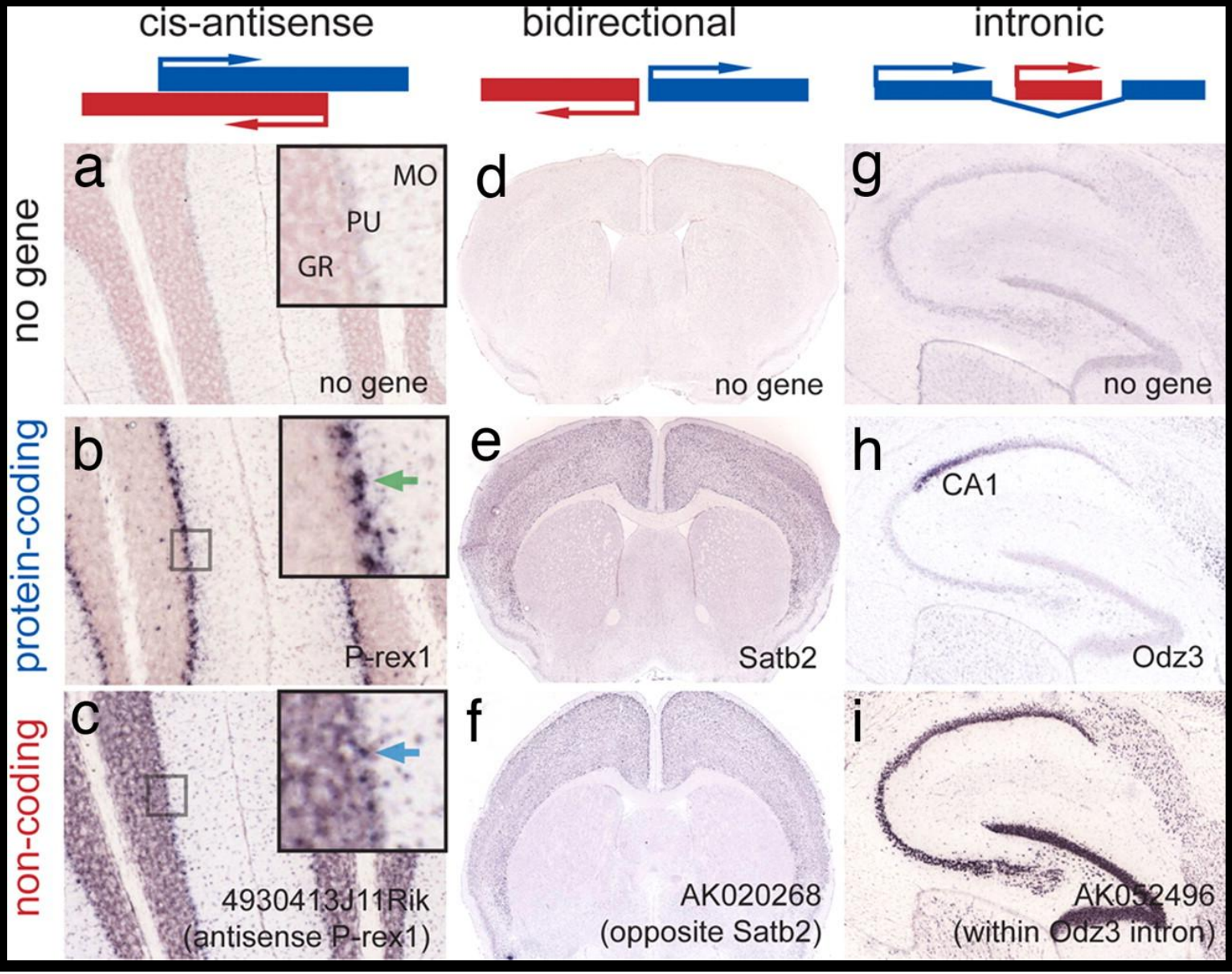


Fig. 2

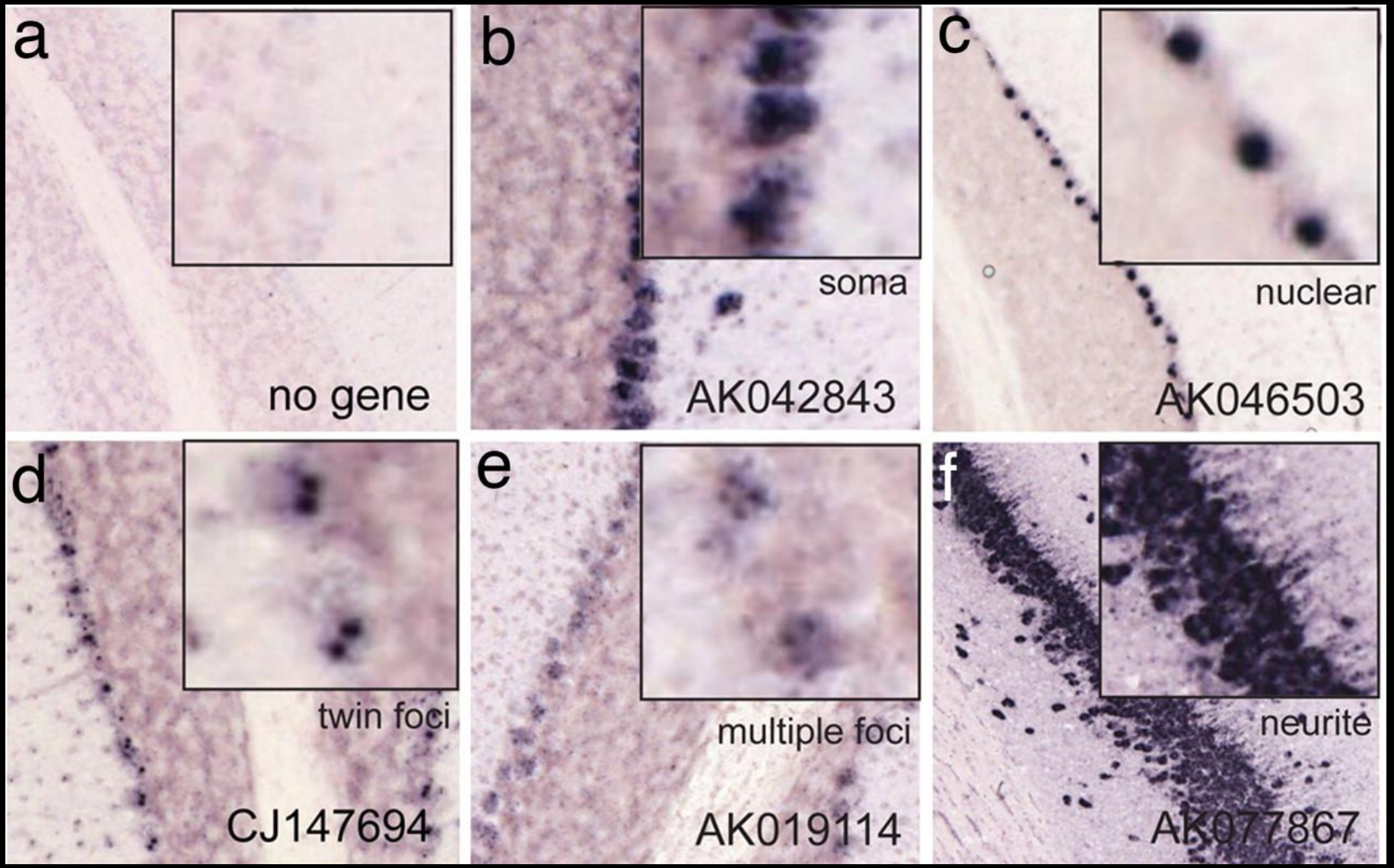
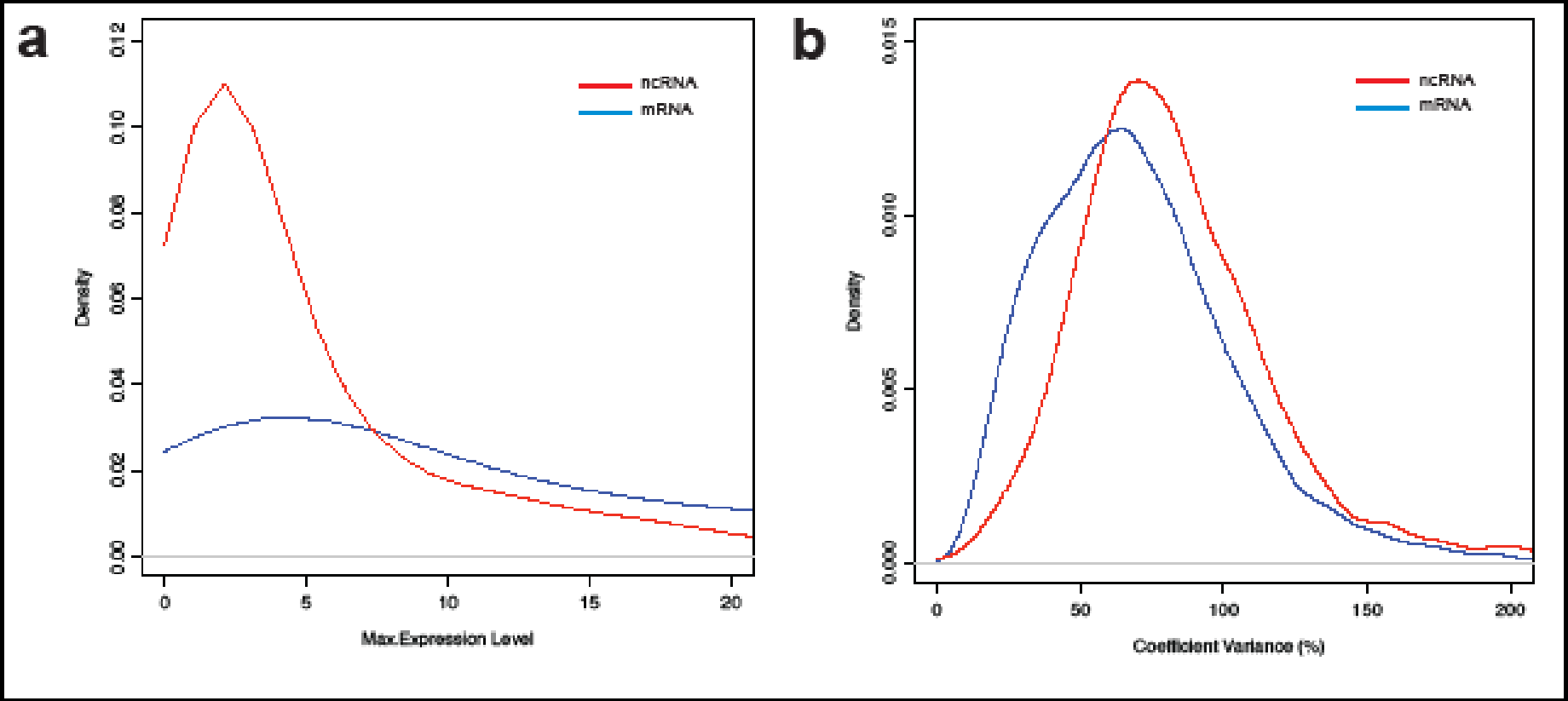


Fig. 4

Mercer T. R. et.al. PNAS 2008;105:716-721

PNAS

Long ncRNAs: lower expression levels but higher spatial variation



Conclusion: ~20K long ncRNAs expressed in human brain.

Fig. S5

ncRNAs comprise a Computational Matrix

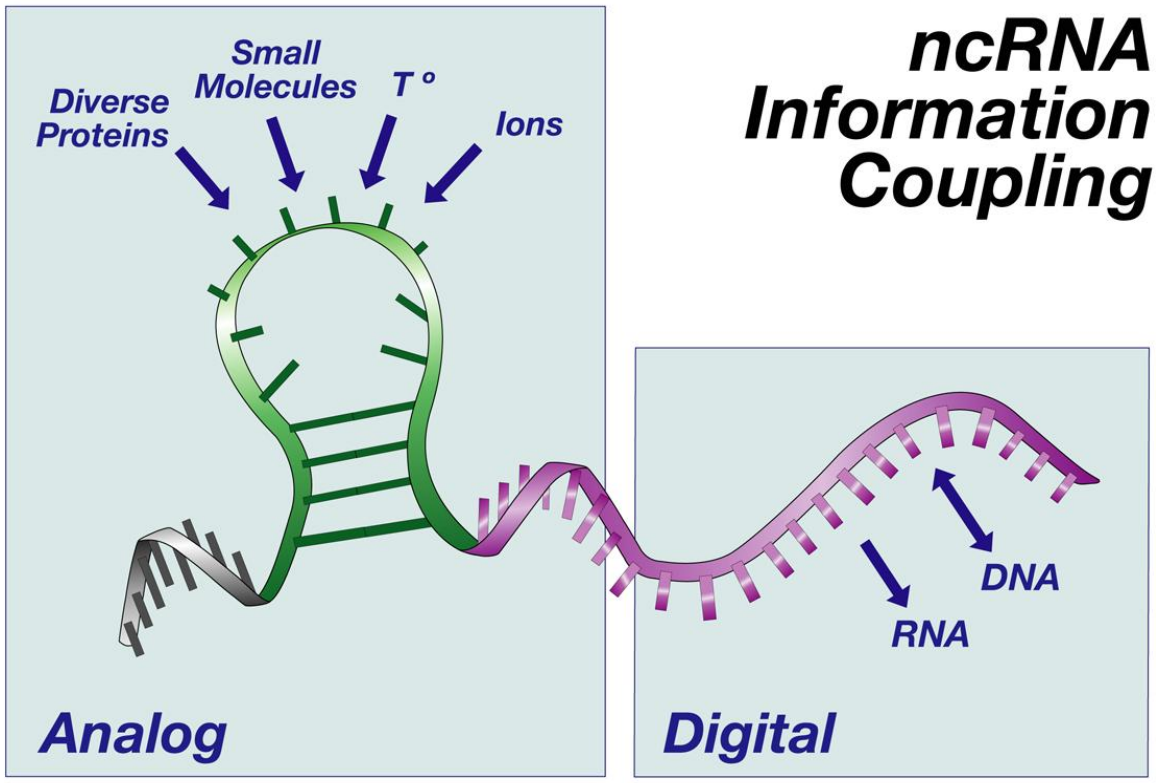
Opinion TRENDS in Neurosciences Vol.30 No.12

Noncoding RNAs: couplers of analog and digital information in nervous system function?

Georges St. Laurent III[1,2] and Claes Wahlestedt[3]

[1] The George Washington University Medical Center, Department of Biochemistry and Molecular Biology, 2300 I Street, NW, Ross Hall 232, Washington, D.C. 20037, USA
[2] Immunovirology – Biogenesis Group, University of Antioquia, A.A. 1226, Medellín, Colombia
[3] The Scripps Research Institute, Molecular and Integrative Neurosciences Department (MIND), 5353 Parkside Drive, Jupiter, FL 33458, USA

ncRNAs are unique as information processors



-Fine Tuning

-Wide Array of Interactions

-Environmental Sensor

-High Density

-Noise Filtering

-Precision

Capacity to couple the digital Information dimension of sequence homology to the analog information dimension of macromolecular shape.

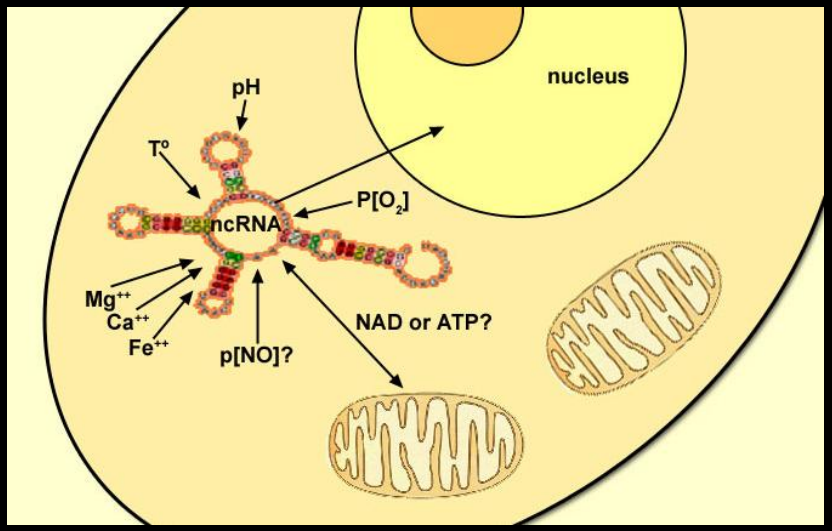
Sensory signaling changes ncRNA secondary structure *in-vivo*

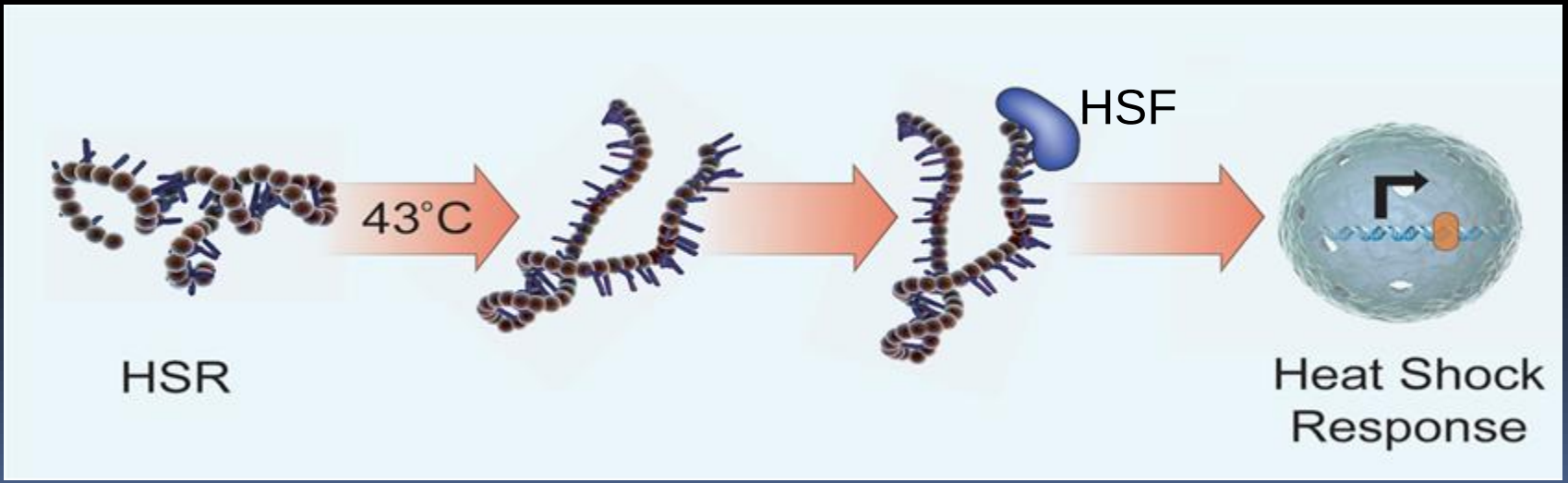
Unique sensory features:

→ Reversibility

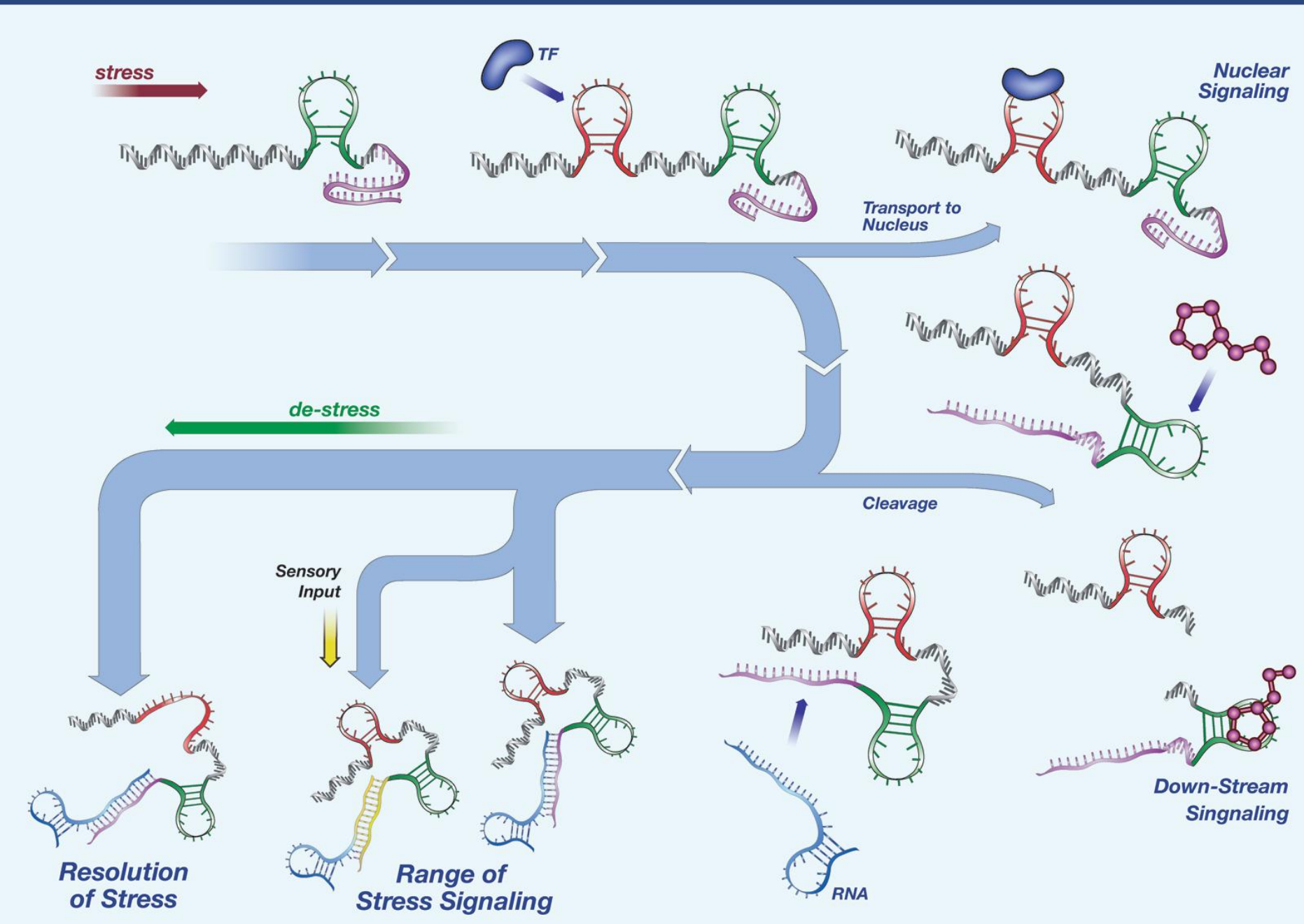
→ Sensitivity

→ Plasticity





Molecular Digital Analog Computation with ncRNA



ncRNA Information Theory and Thermodynamics: more information for less thermal energy

$$\frac{\text{Information Content}}{\text{Thermal Dissipation}} = \frac{\text{Codable Degrees of Freedom}}{\text{Thermal Degrees of Freedom}}$$

$$= \frac{\sum p_i \ln p_i}{\sum \frac{1}{w} \ln \frac{1}{w}}$$

ncRNA features an enhanced Shannon Entropy to Thermal entropy ratio.

$$\frac{\text{Information Content}}{\text{Thermal Dissipation}} = \frac{\sum p_i \ln p_i}{\sum \frac{1}{w} \ln \frac{1}{w}}$$

ncRNAs can regulate the early pathogenisis of complex disease

nature medicine

ARTICLES

Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of β-secretase

Mohammad Ali Faghihi[1,2], Farzaneh Modarresi[1], Ahmad M Khalil[1], Douglas E Wood[3], Barbara G Sahagan[3], Todd E Morgan[4], Caleb E Finch[4], Georges St. Laurent III[5,6], Paul J Kenny[7] & Claes Wahlestedt[1]

ß-secretase = BACE-1 (ß-site APP cleaving enzyme)

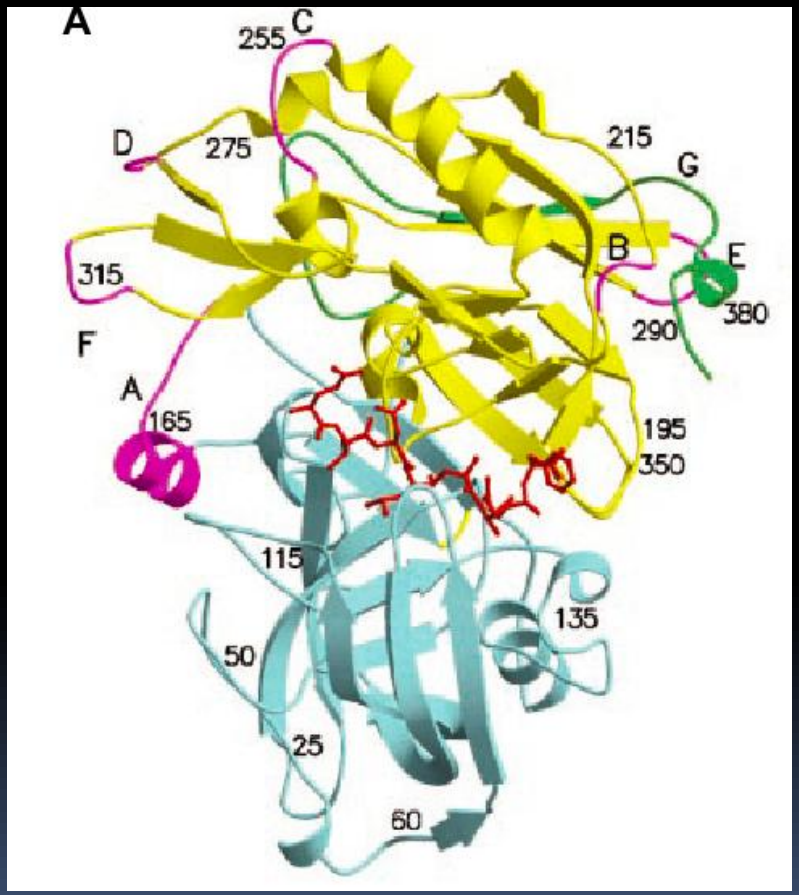
Rate limiting for Aß 1-42 generation

Finely tuned

Stress responsive

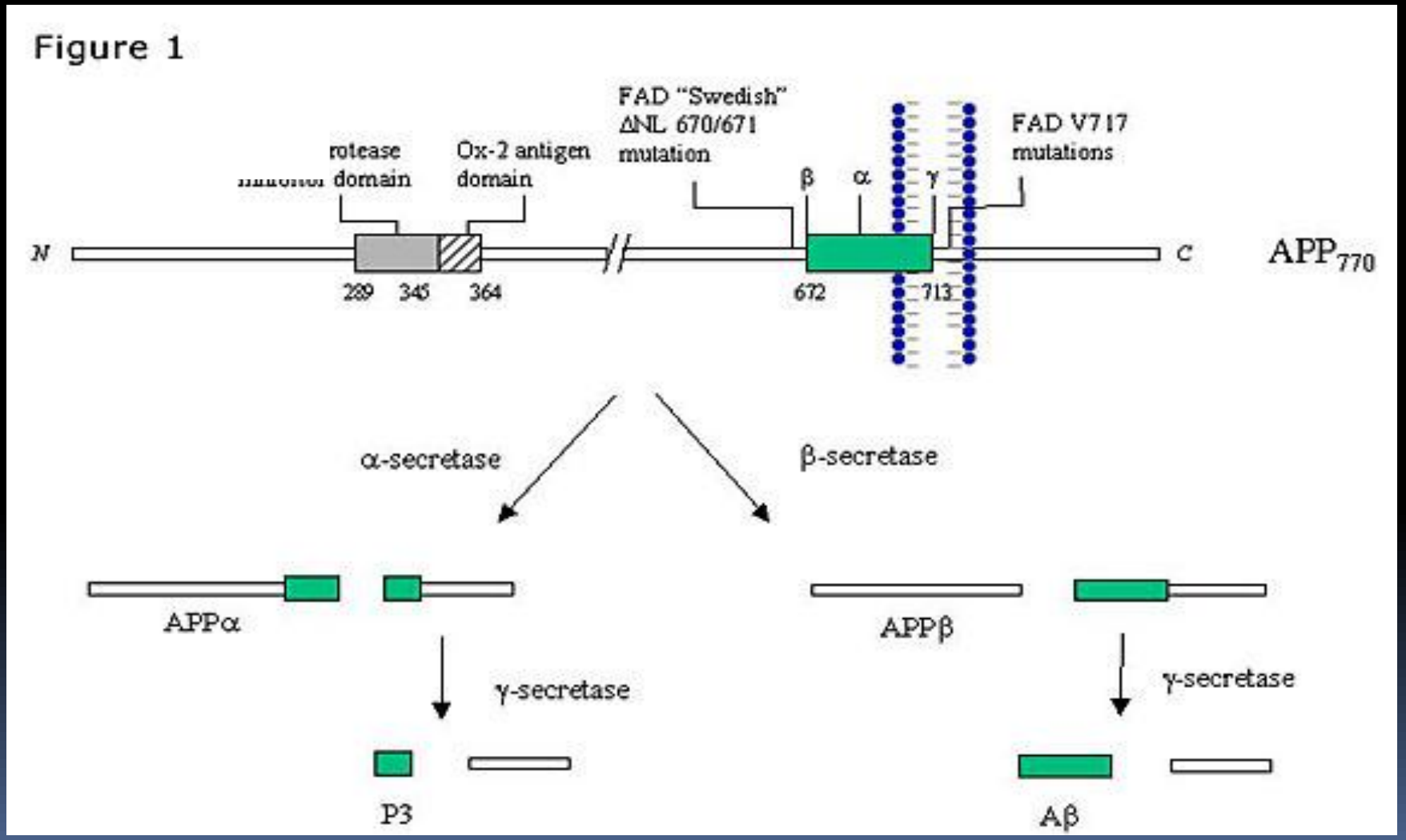
Up-regulated in Alzheimer's disease

Inhibitor: disease modifying therapy

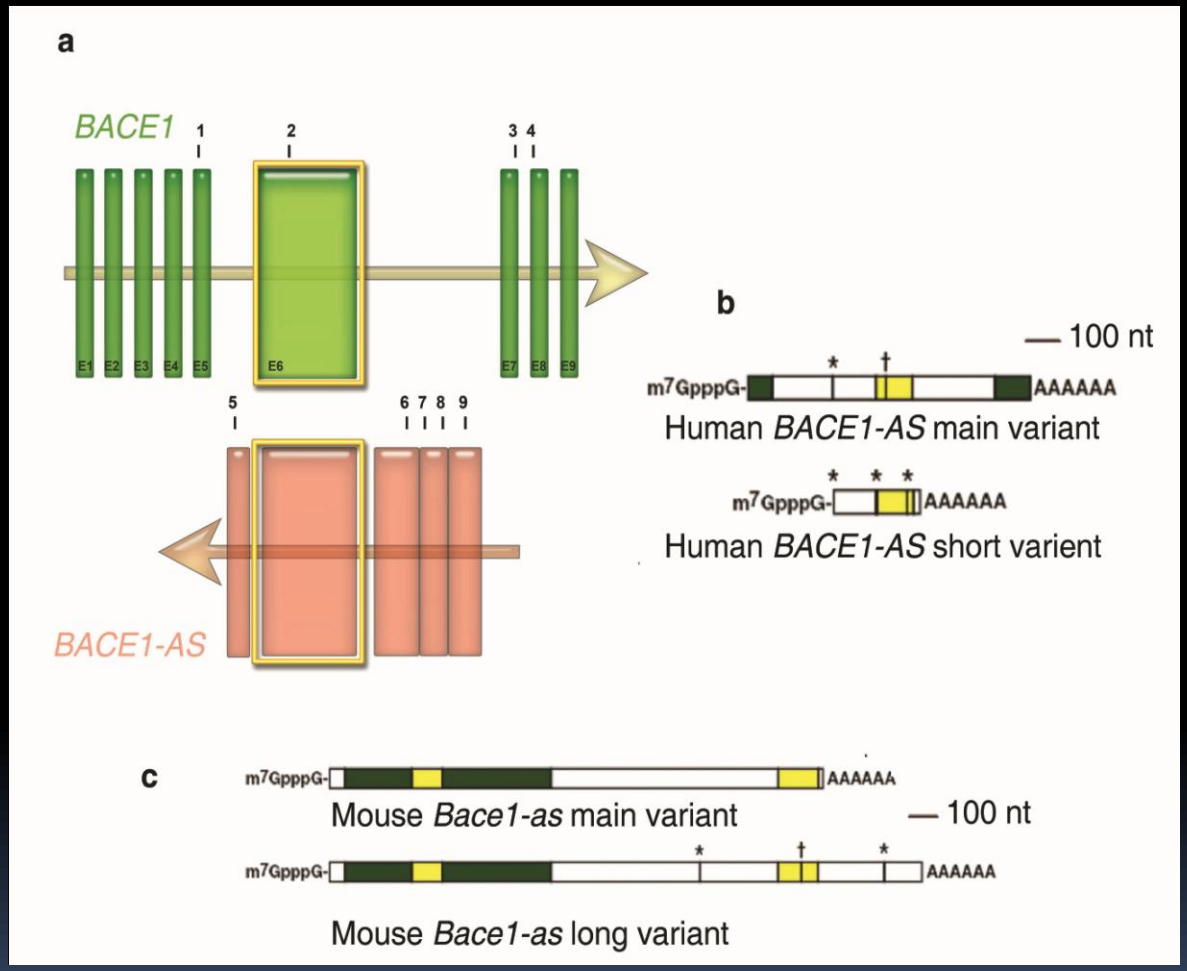


Gurney *et al.*, *Nature* **402**:533, 1999

BACE1 (beta-site APP cleaving enzyme)

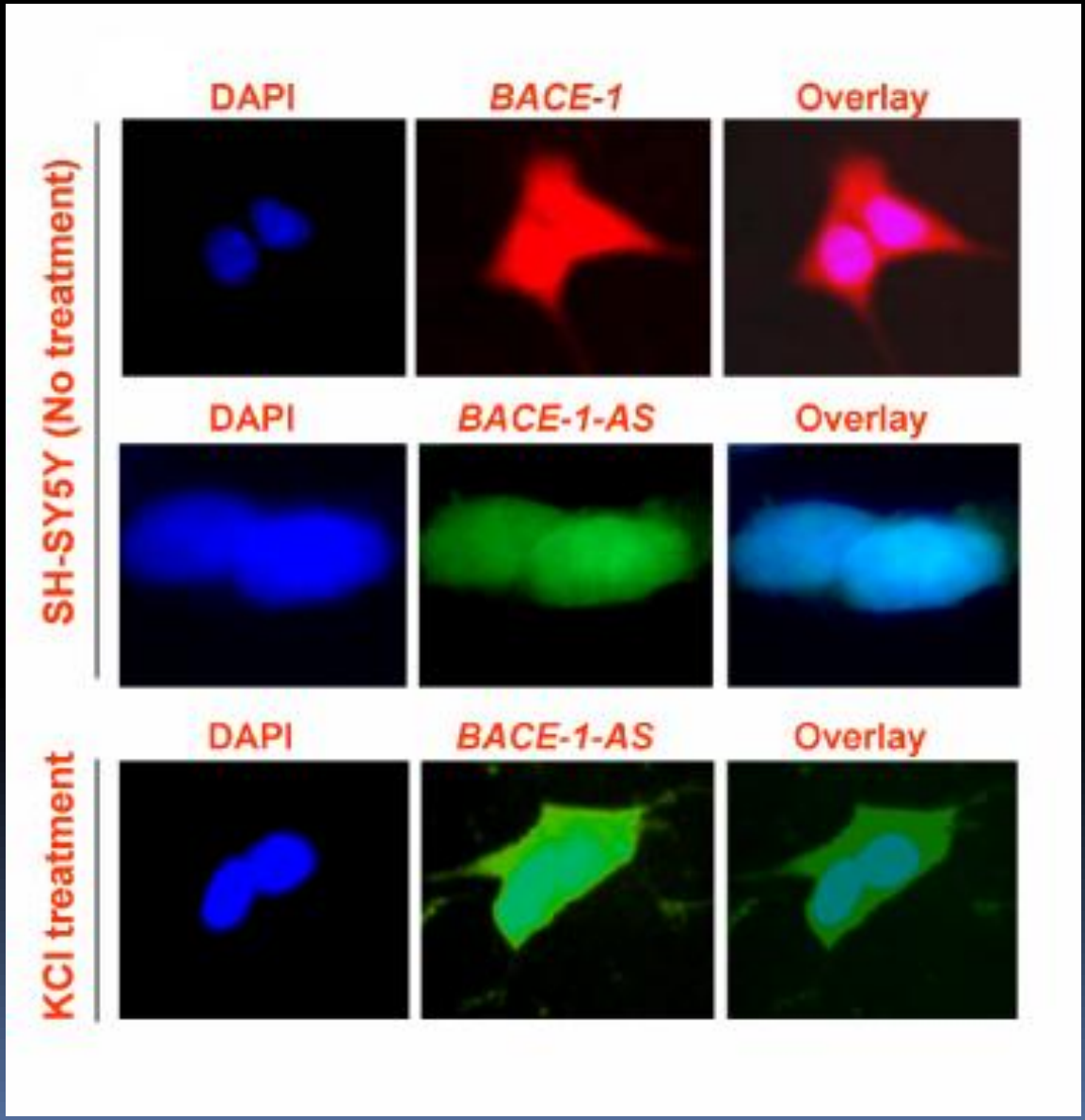


BACE1 Genomic Locus

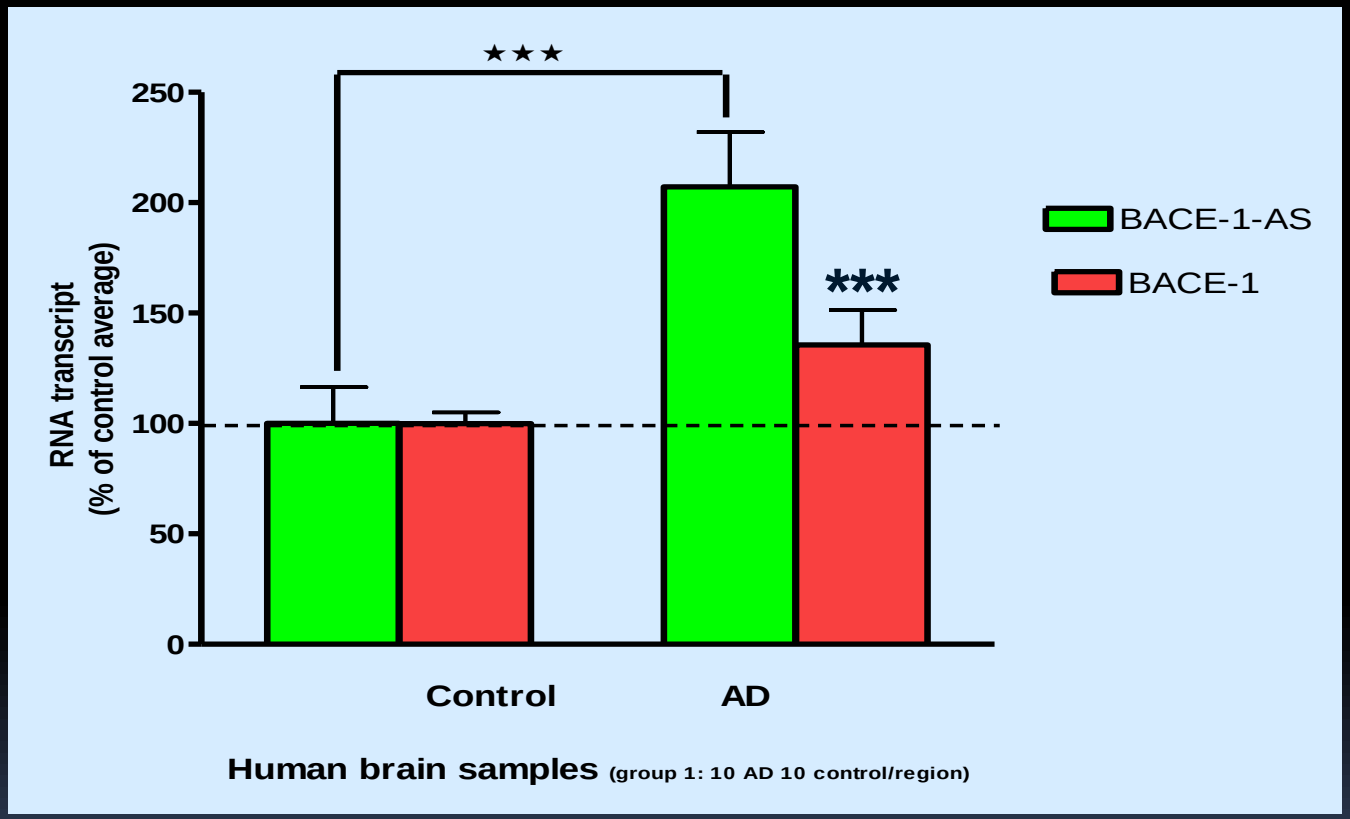


miR-485-5p binding site is located in the overlapping region of *BACE1* and *BACE1*-AS transcripts.

FISH images show a nuclear enrichment pattern



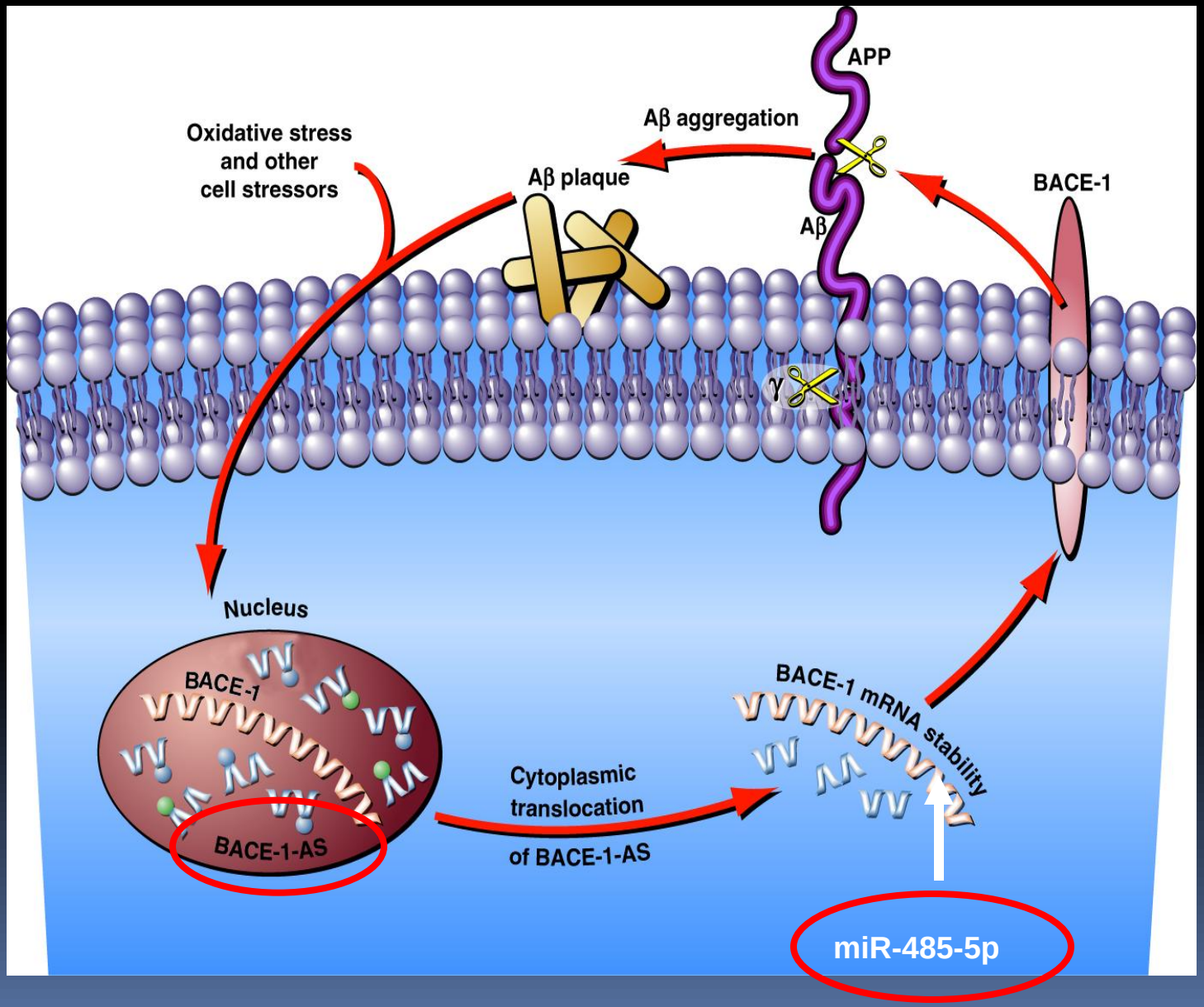
BACE1-AS is elevated in Alzheimer's disease brain (as is *BACE1* itself)



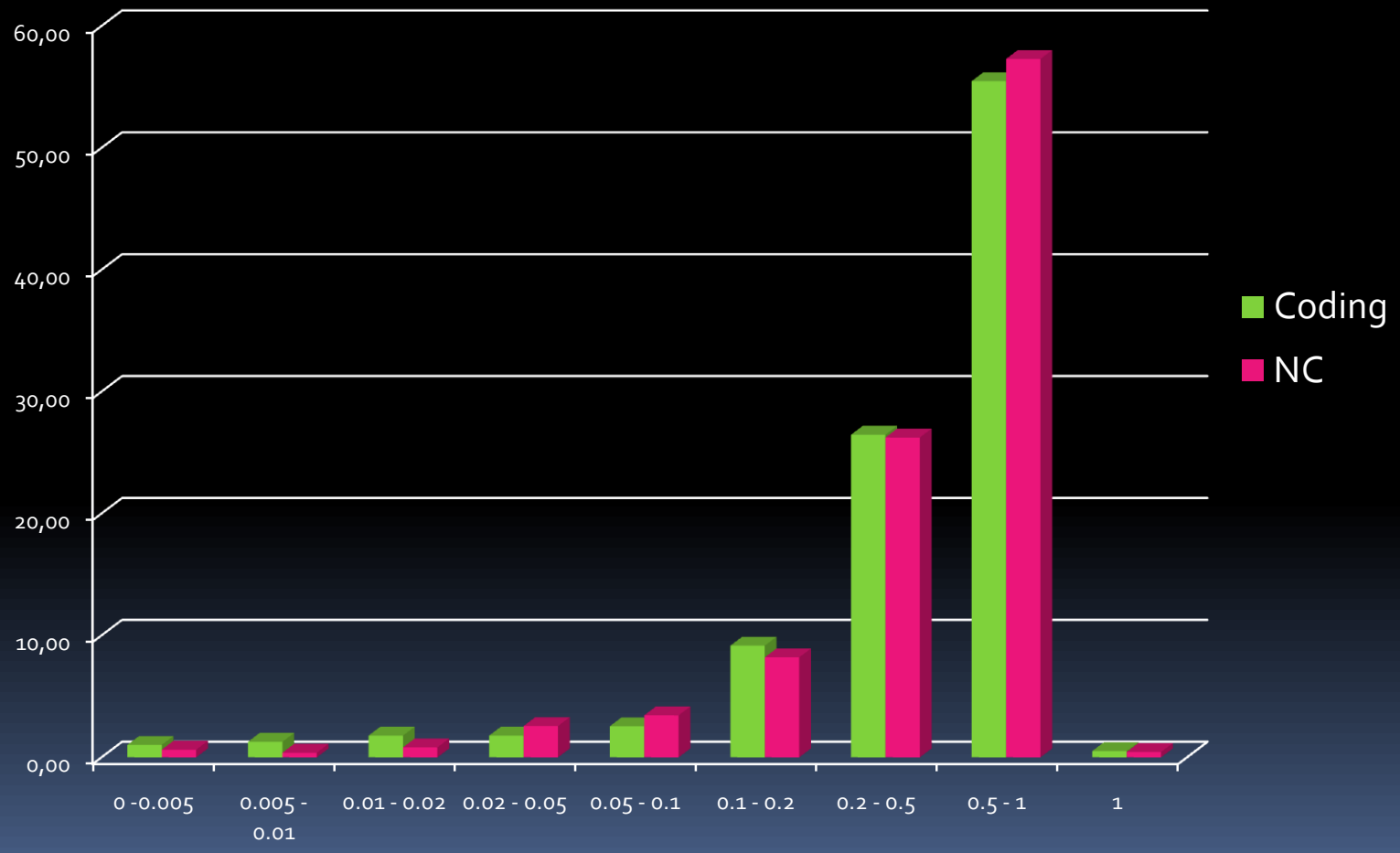
Hippocampus (n=40 each)

Faghihi *et al. Nature Medicine* 14:723, 2008

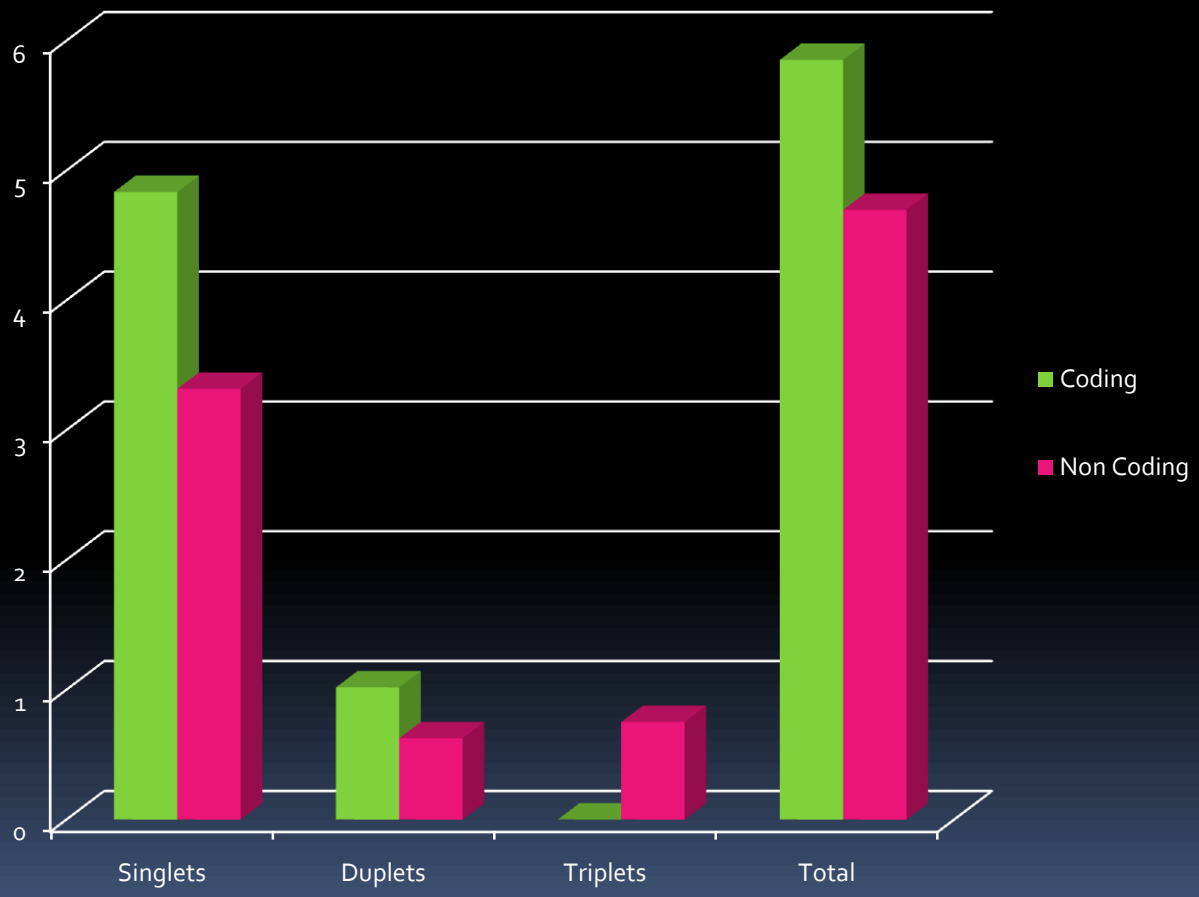
Dual and synergistic BACE1 regulation by BACE1-AS



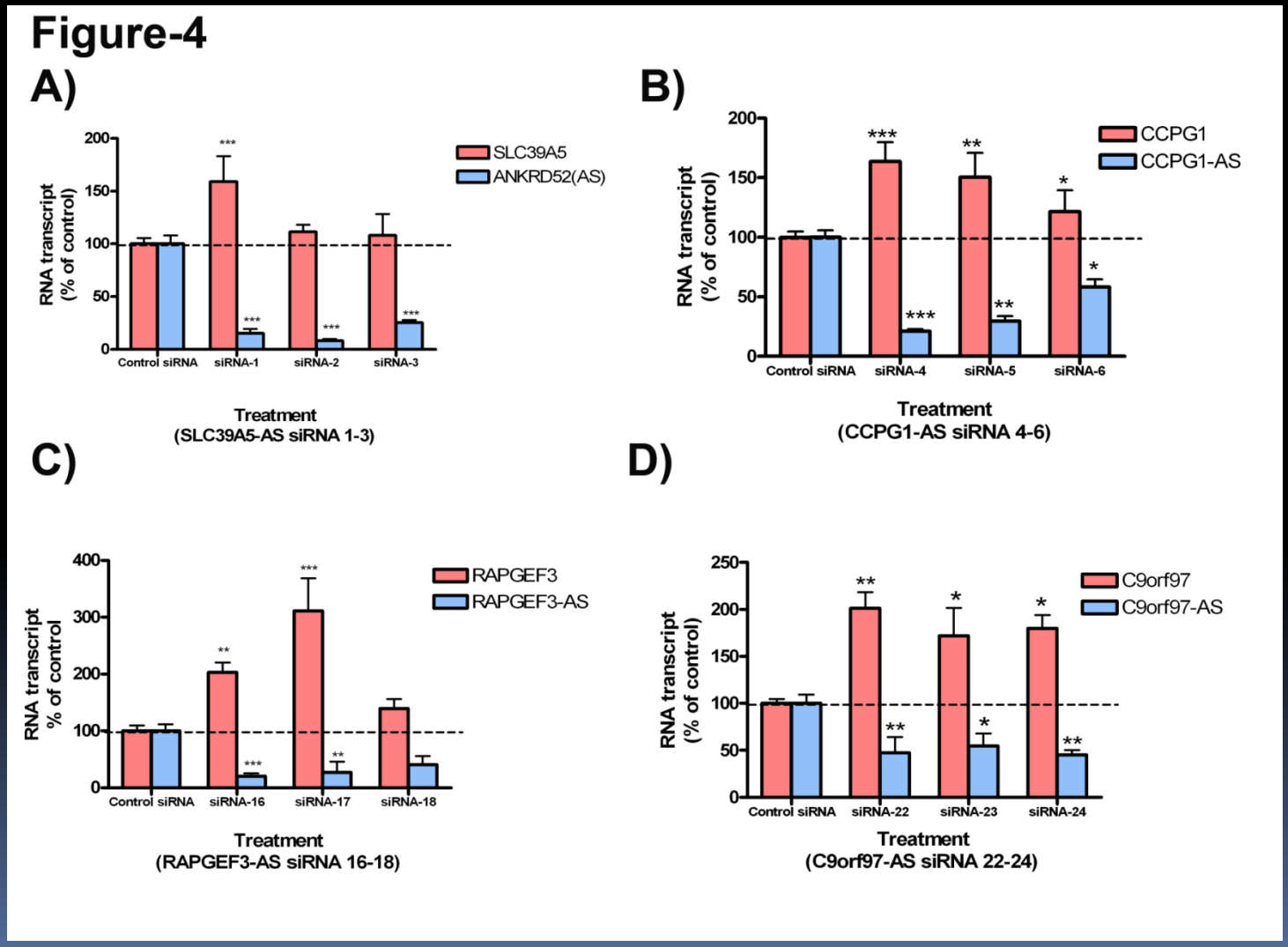
siRNA Screen for NAT modulation of cell viability
700 NATs – 2000 siRNAs

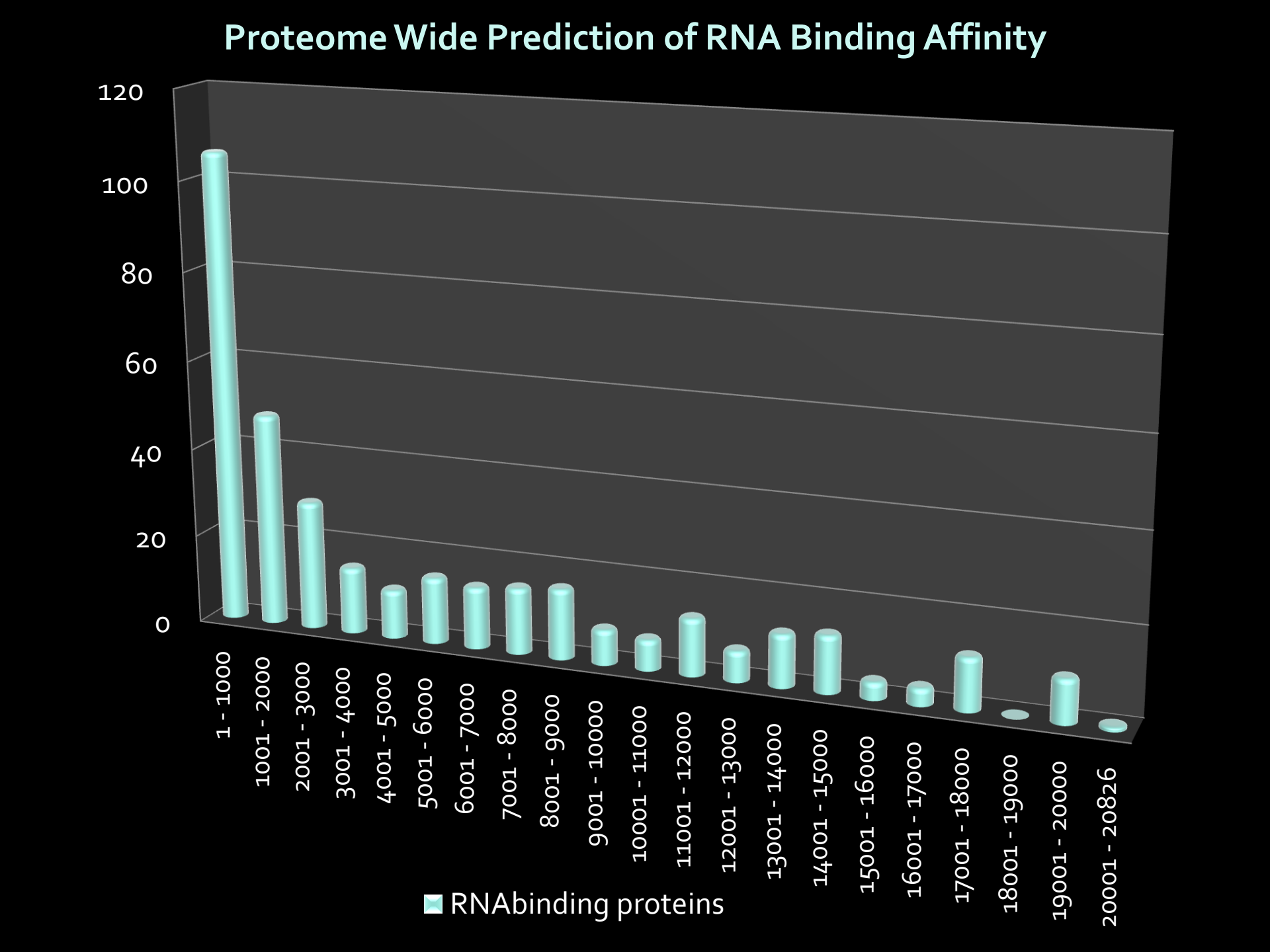


ncNATs score almost as high as coding NATs in cell viability screen.



Validation in approx. 60% of hits




Proteome Wide Prediction of RNA Binding Affinity
120
100
80
60
40
20
0
1 - 1000
1001 - 2000
2001 - 3000
3001 - 4000
4001 - 5000
5001 - 6000
6001 - 7000
7001 - 8000
8001 - 9000
9001 - 10000
10001 - 11000
11001 - 12000
12001 - 13000
13001 - 14000
14001 - 15000
15001 - 16000
16001 - 17000
17001 - 18000
18001 - 19000
19001 - 20000
20001 - 20826
RNAbinding proteins

RNA Protein Complex Hi Throughput Pipeline

- Proteome wide Prediction of RNA Binding regions
- Endogenous Flag Tag of Predicted RNA Binding proteins
- Cryogenic Flag IP of RNA – Protein Complexes
- RNA Deep Sequencing and Peptide Mass Spec
- Bioinformatics Identification of RNAs and proteins
- Systems Biology Analysis of Datasets → Network construction

Cryogenic ImmunoprecipitationTechnique

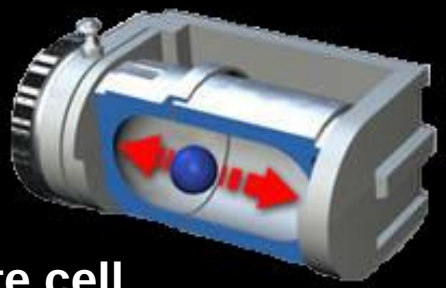
Treatments
Depolarization
Stress
Inflammation
Cytokines
Drugs

Cells

- PC12 cells
- primary cells
- neuronal progenitors

Freezing

ENTROPY



Create cell ‘grindate’

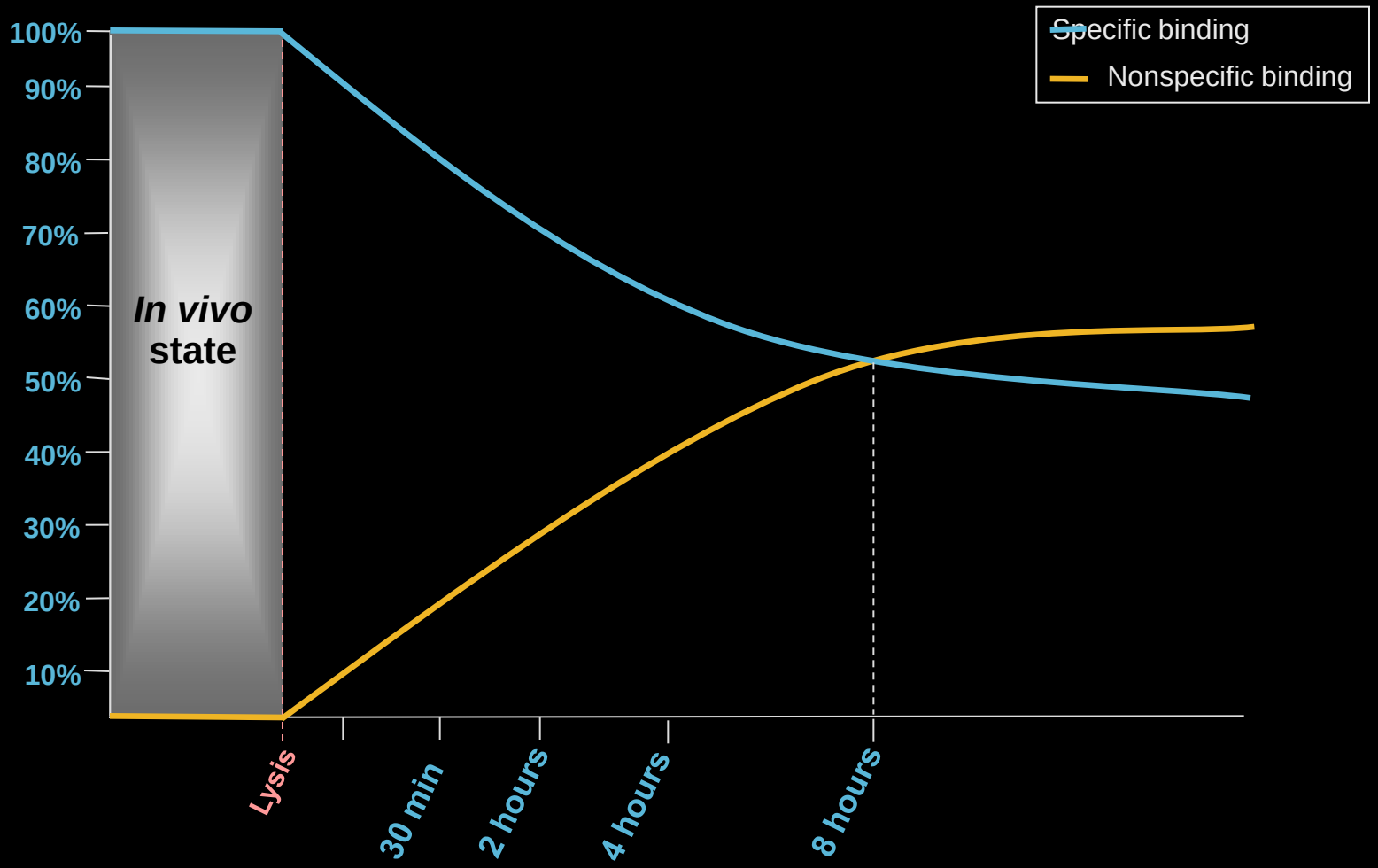
ENTROPY

Immunoprecipitation

30 min

Proteomics, RNA-Seq, or Chip-SeqAnalysis

Difficult Timescales for RNP Immunoprecipitation



Features of Cryogenic IP for RNP Studies

•Rapid technique prevents RNA degradation and loss of transient macromolecular interactions

• Rearrangement is not a significant problem

• Yields of > 90% for bait protein and associated RNA

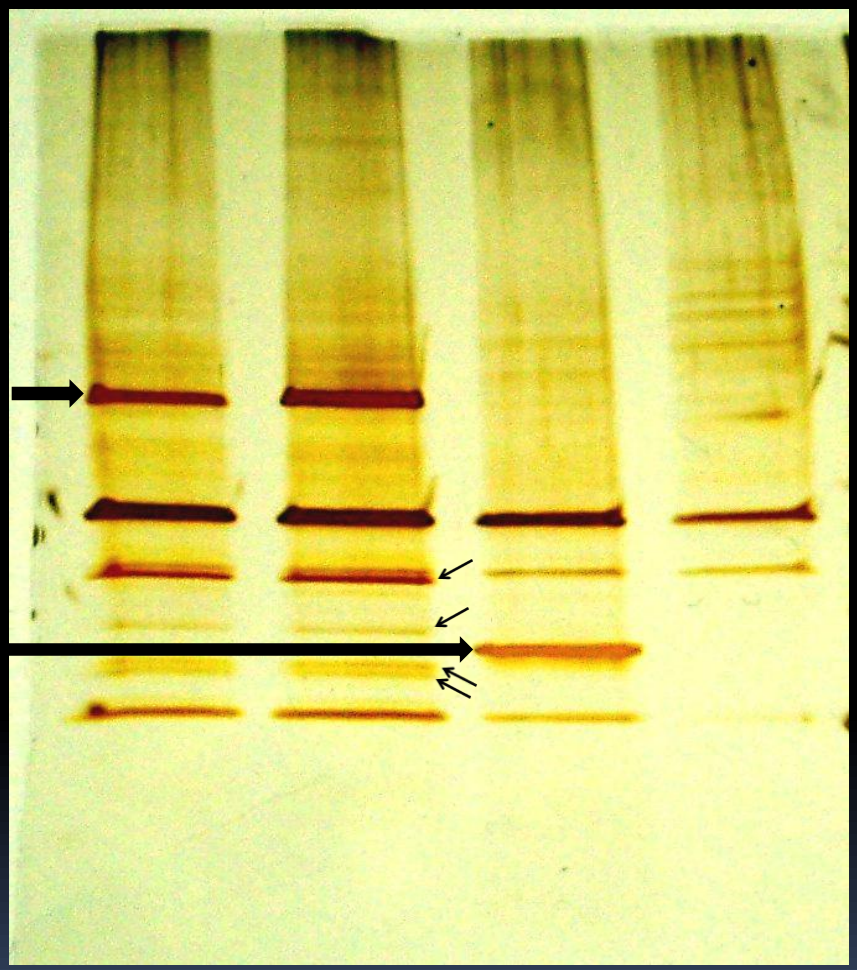
• Does not depend on a particular protein tag

• No cross-linking necessary

• Able to capture weak interactions

An ideal technique for studying maturing RNP complexes

Transfected with:



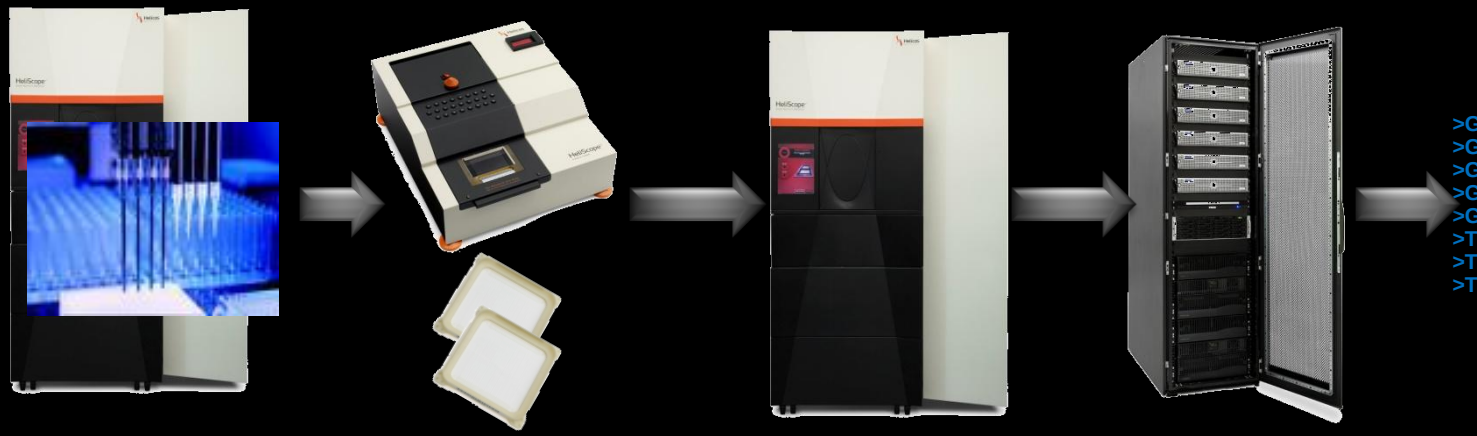
NF90 = 90 kDa

HuR = 37kDa

Cryogenic Immunoprecipitation of RNA – Protein Complexes

High-throughput tools for ncRNA Systems Biology

Helicos single molecule sequencing



>GATAGCTAGCTAGCTACACAGAGAT
>GATAGACACACACACACACAGCGCA
>GTACTACACACAGCGACACAGTCTA
>GTCGAACACACATGAACACATGAGC
>GTGTCACACACGACTACACATGCAT
>TAGTGACACACGTAGACACGACAGT
>TCTCGACACACTATCACACGACTCA
>TGCACACACACTCGTACACGAGACG

Sample Preparation

HeliScope™ Single Molecule Sequencer

Bioinformatic Analysis Engine

Output

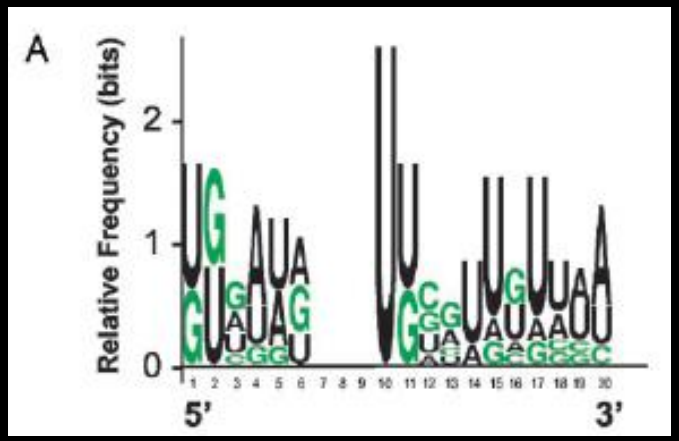
Capacity = 10 billion nucleotides / run

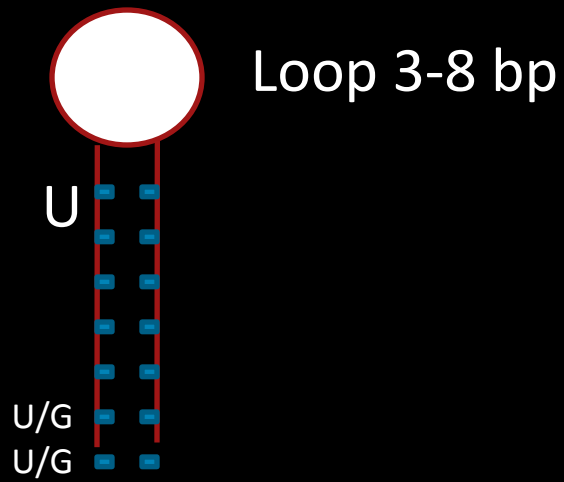
HuR Associated Transcriptome isolated by Cryo-IP:

- Affy All Exon Array = 11,155 called probes
- Illumina Deep Sequencing = 6 million total sequence tags
- Coding and non-coding represented in top 3000
- Natural Antisense RNAs such as HIF1α-AS represented in top 100
- 60% overlap between top 3000 sequence tags and Affy.
- Helicos comparison pending (permits very small sample sizes)

RNA Motif #1 for HuR association

Lopez de Silanes, et. al. (2004) PNAS. "**Identification of a target RNA motif for RNA-binding protein HuR**" [MyriamGorospe's Lab]





→Found the Gorospe motif in 4536 of 11150 sequences

… … (versus 3521 in a mononucleotide shuffled control)

→Z-score =~ 20.69

RNA Motif #2 for HuR association

Ma, et. al. (1996) JBC **"Cloning and Characterization of HuR, a Ubiquitously Expressed Elav-like Protein"**

A U UUUU A

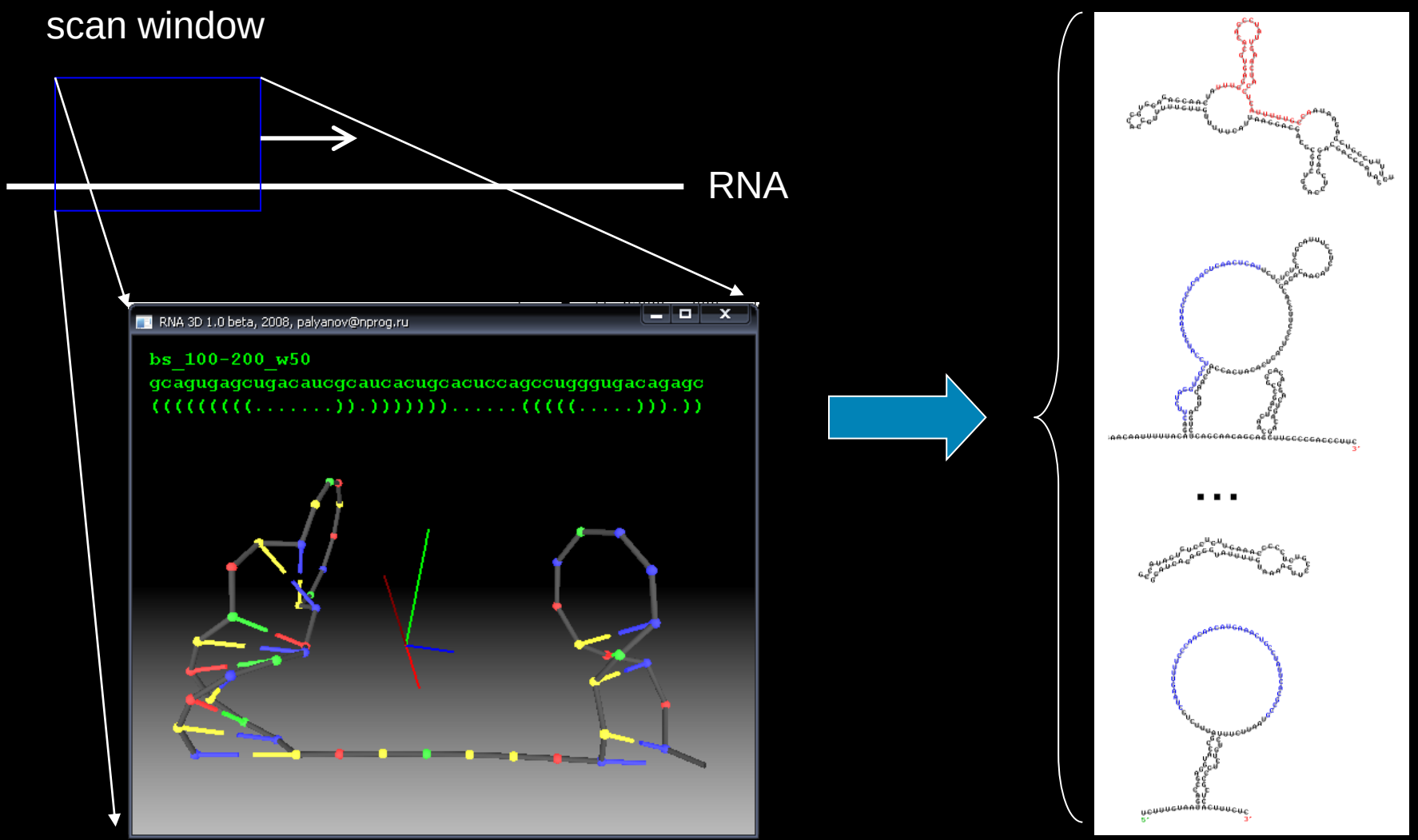
→Found the Ma motif in 2230 of 11150 sequences

… … (versus 1267 in a dinucleotide shuffled control)

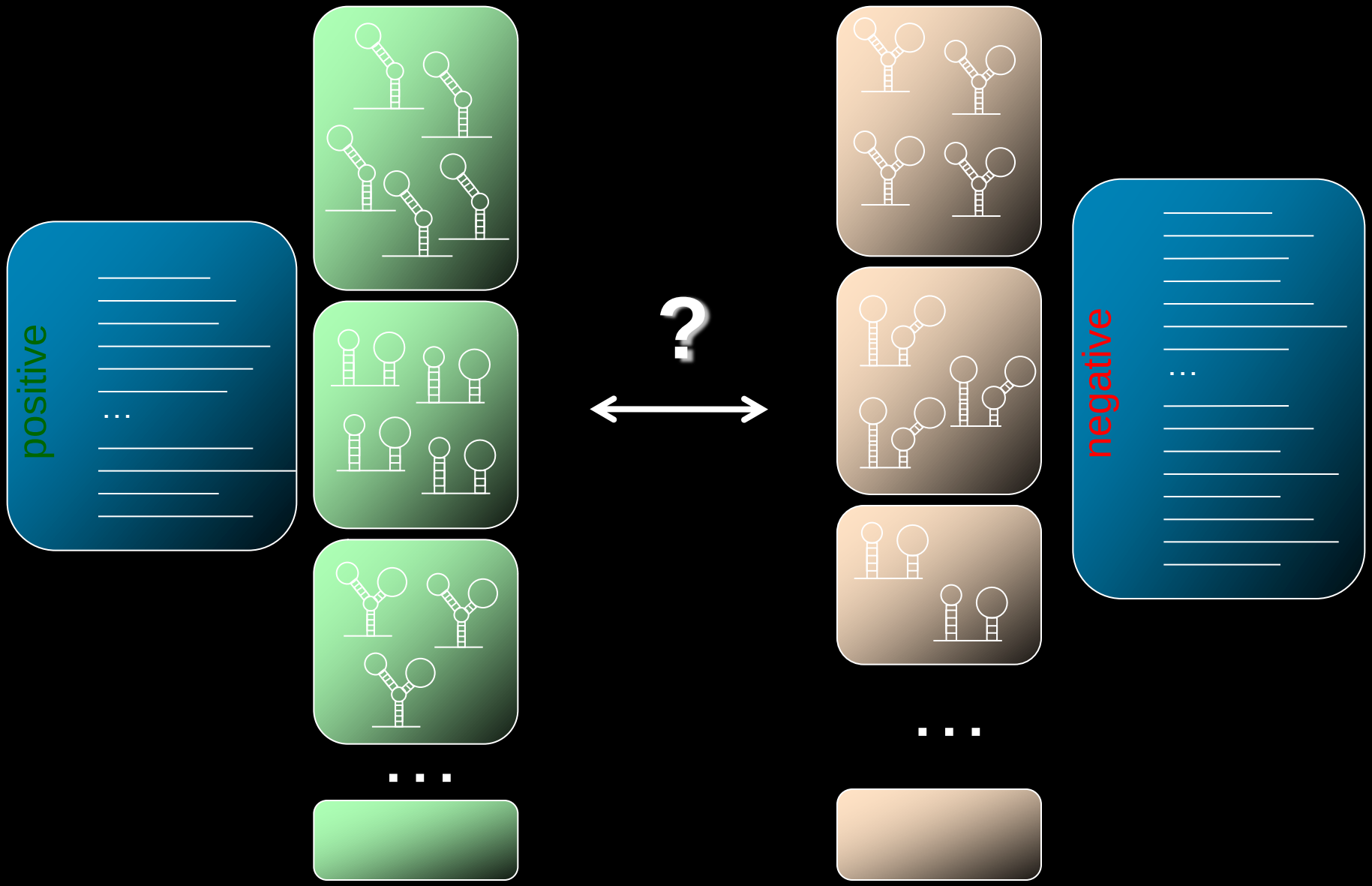
→Z-score =~ 29.72

→ Both motifs informative, but suggests HuR responds to a wider range of information signals.

Deciphering information content defining HuR interactions withthe Transcriptome



Determine clusters of similar structures

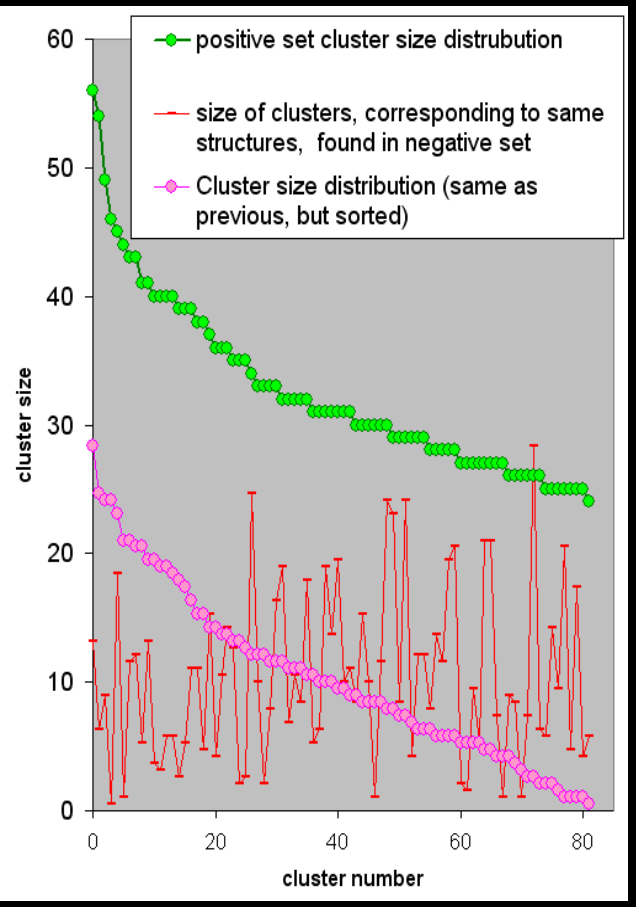


Calculate clusters size distribution (for scanning window length = 50, 45 and 40)

positive set cluster size distrubution

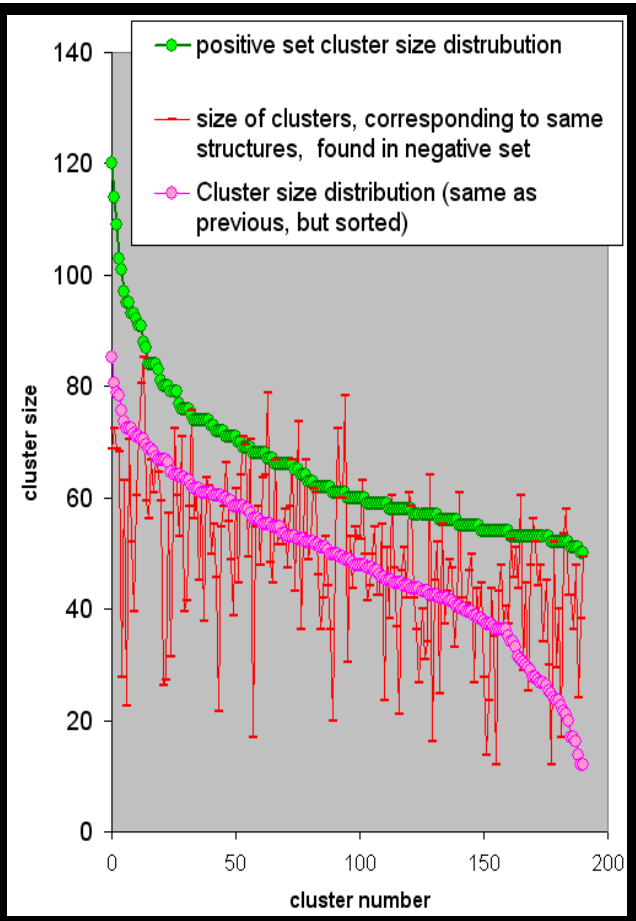
size of clusters, corresponding to same structures, found in negative set

Cluster size distribution (same as previous, but sorted)



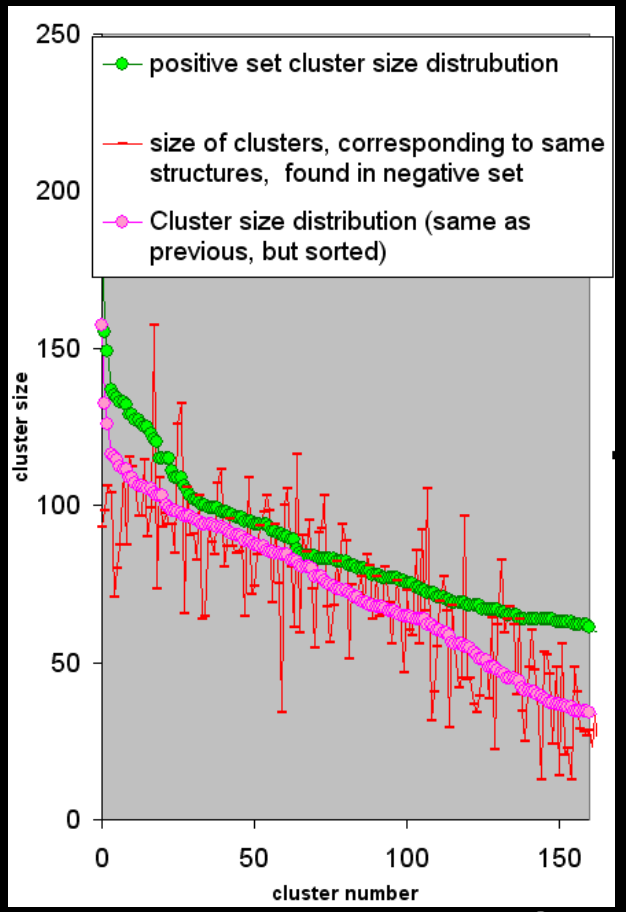
window length = 50

Structures 45-50 bp length



window length = 45

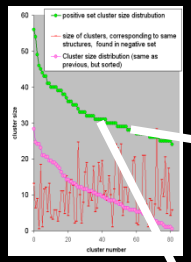
Structures 40-45 bp length

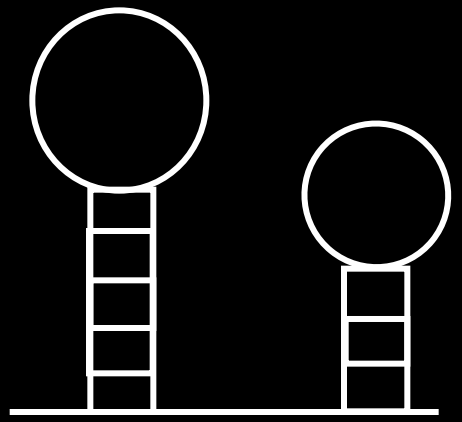


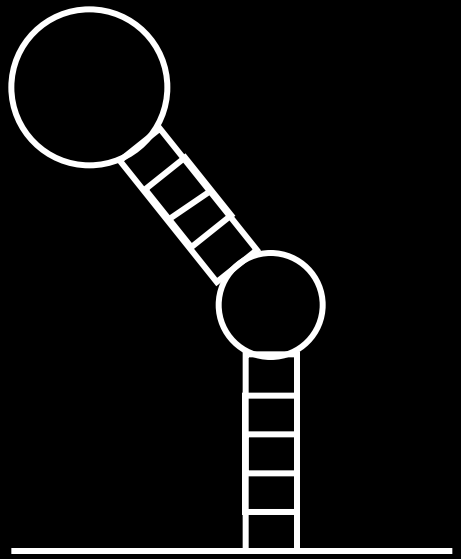
window length = 40

Structures 35-40 bp length

Structures (length 45..50), which constitute biggest clusters of positive set





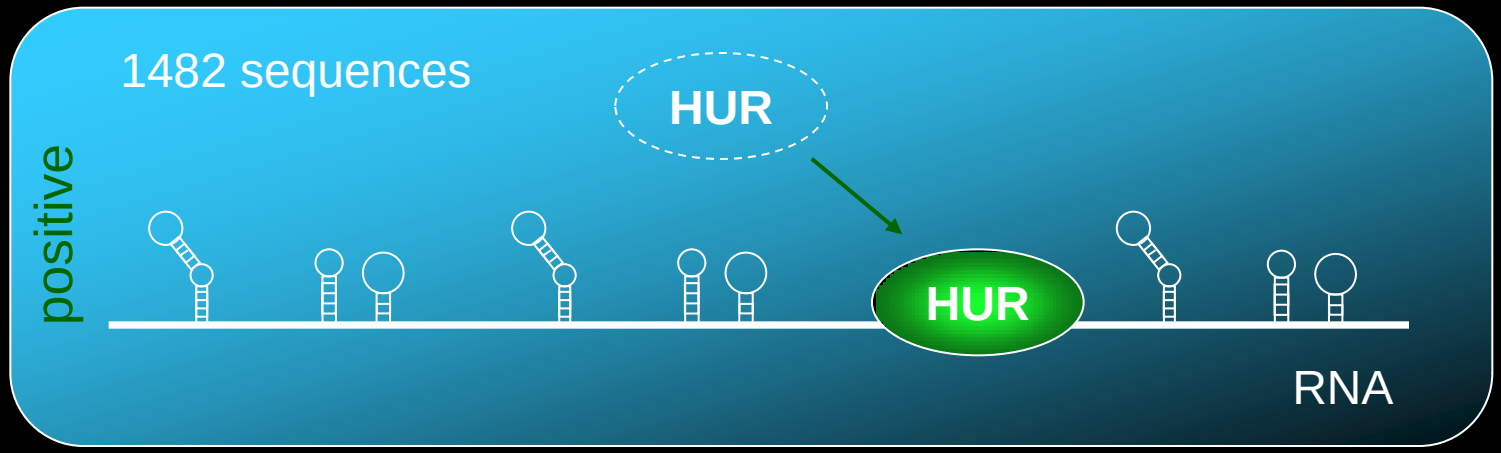


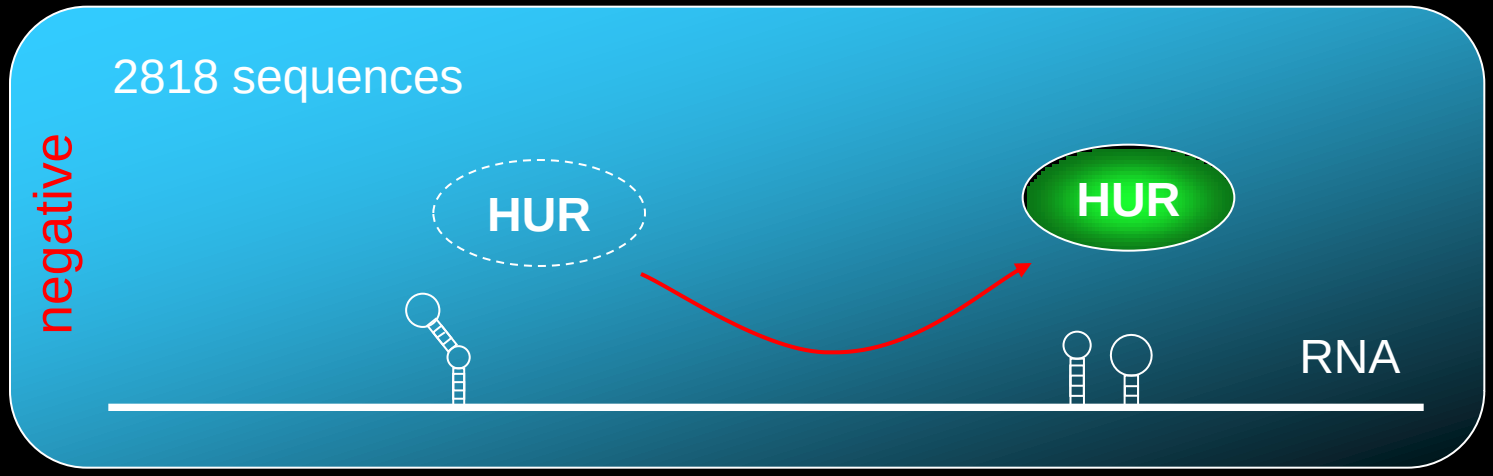
(((((((.....)))))))....(((((((.....)))))))
(((((((.....)))))))....(((((((.....)))))).))
(((((((.....)))))))....(((((((.....)))))).)))
(((((((.....)))))))..(((.(((((.....))))).)))
((((((...((((.....))))......))))))..(((((....)))))
(((((((.....)))))))....(((((((.....)))))).))
(((((((....)).)))))).(((((((.....)))))))
(((((((.....)))))))...((((..(((.....)))))))
((((((.....))))))....(((((((.....)))))))

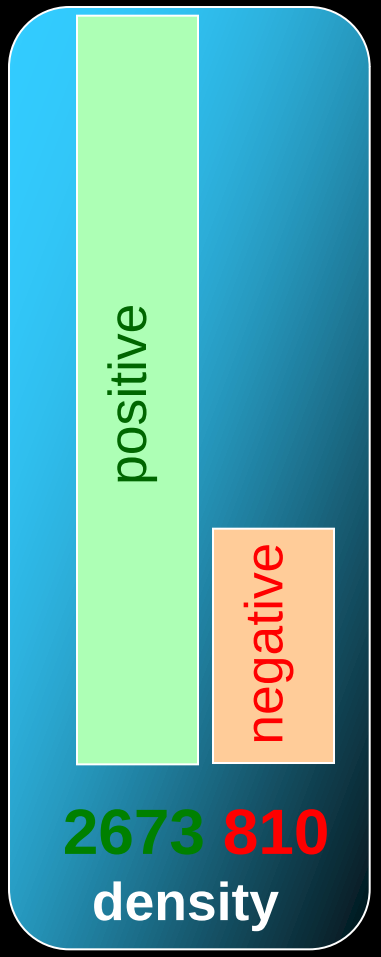
((...(((((..((..(((.....)))))).)))))....))
(((....(((((..((..(((.....)))))).)))))....)))
((...(((((..((((.....))))..)))))......))
(((.....(((((.....)))))).)))
(((...(((((..((((((....))))))......)))))....)))
(((.....(((((((.....)))))))).)))
((..(((((..((..(((.....)))))).)))))......))
((.....(((((..((..(((.....)))))).)))))...))
((.(((((..((((.....))))..)))))......))

HUR-binding transcripts have ~3 times more *special* local secondary structures than HuR non-binding transcripts

• Z-score ≥ 38



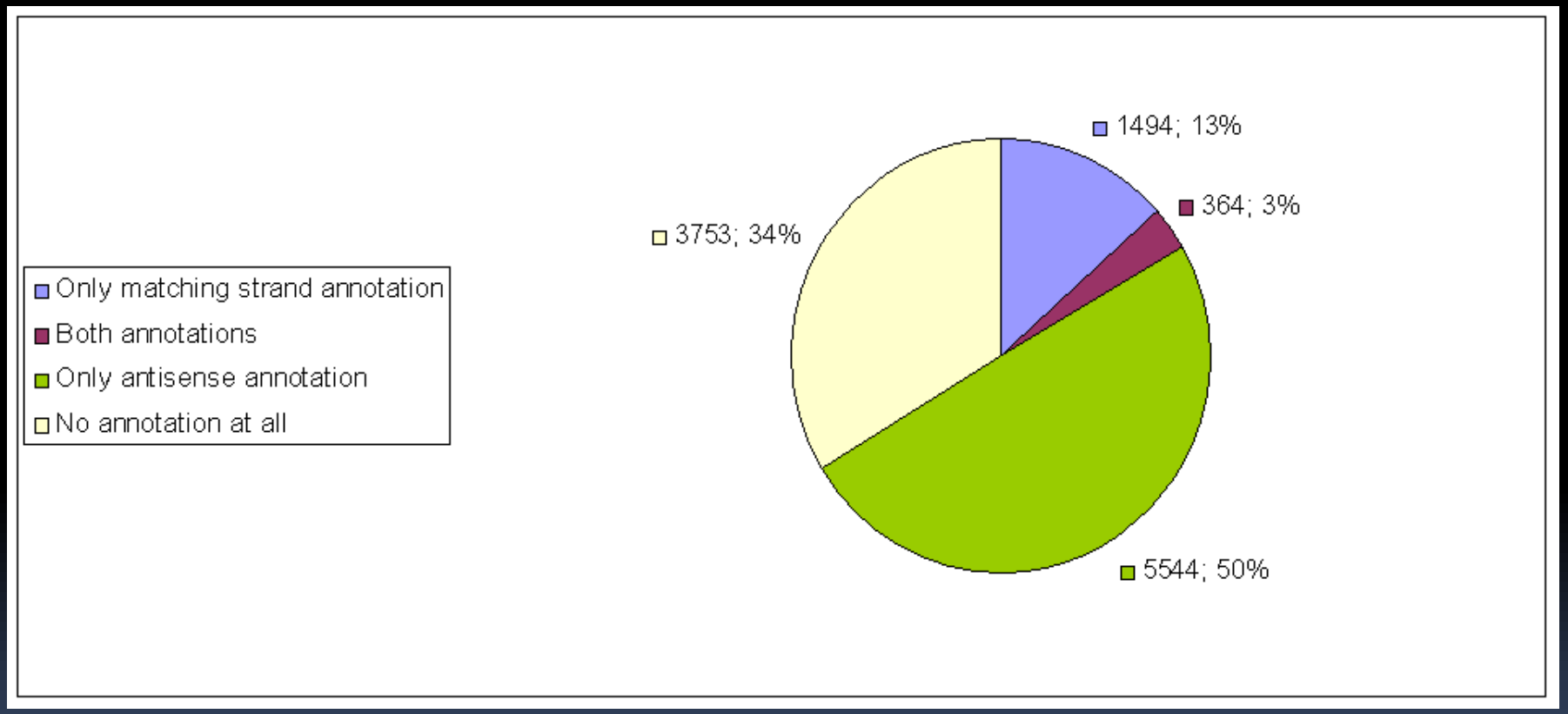




Signal to noise decays with decreasing structure size

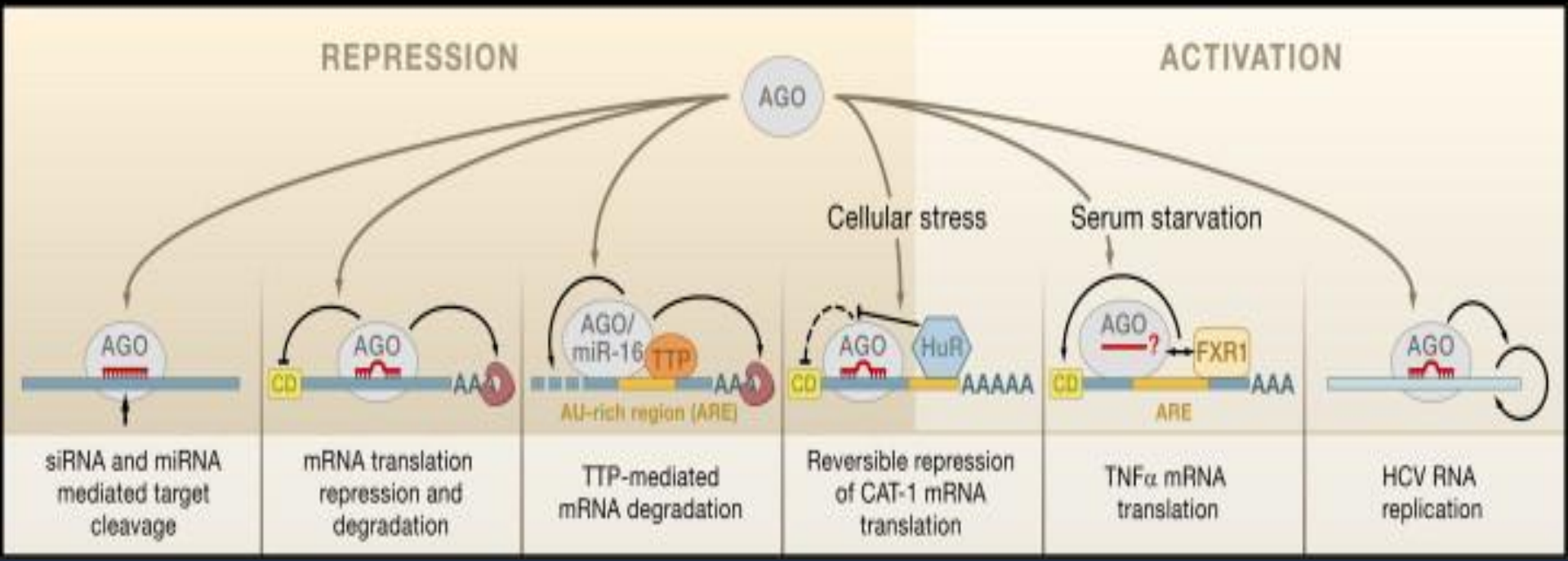
HuR Associated Transcriptome isolated by Cryo-IP

Affy All Exon Array yields 11,155 called probes



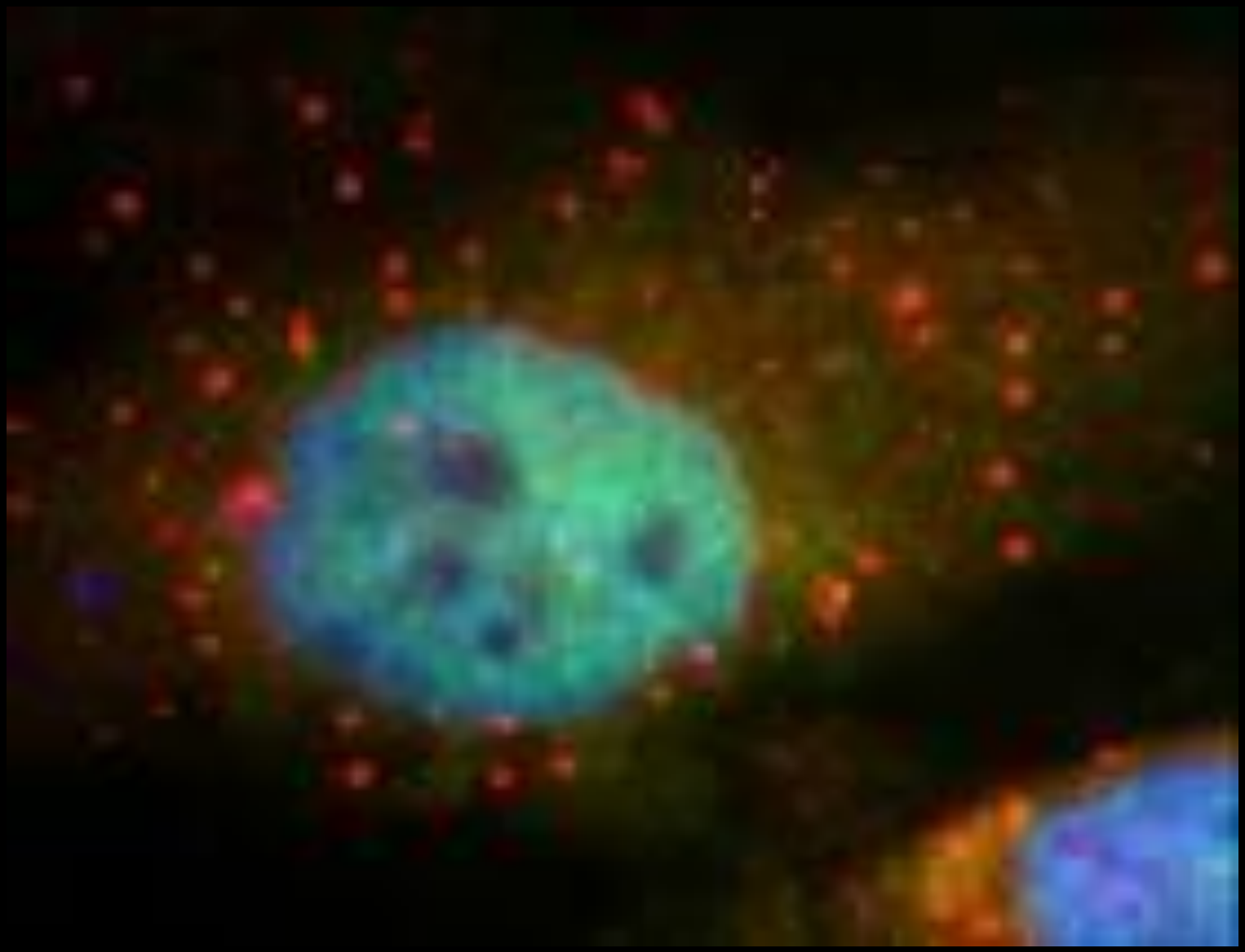
Antisense transcripts comprise 50% of associated RNAs

Multiplexed Computation of Gene Expression



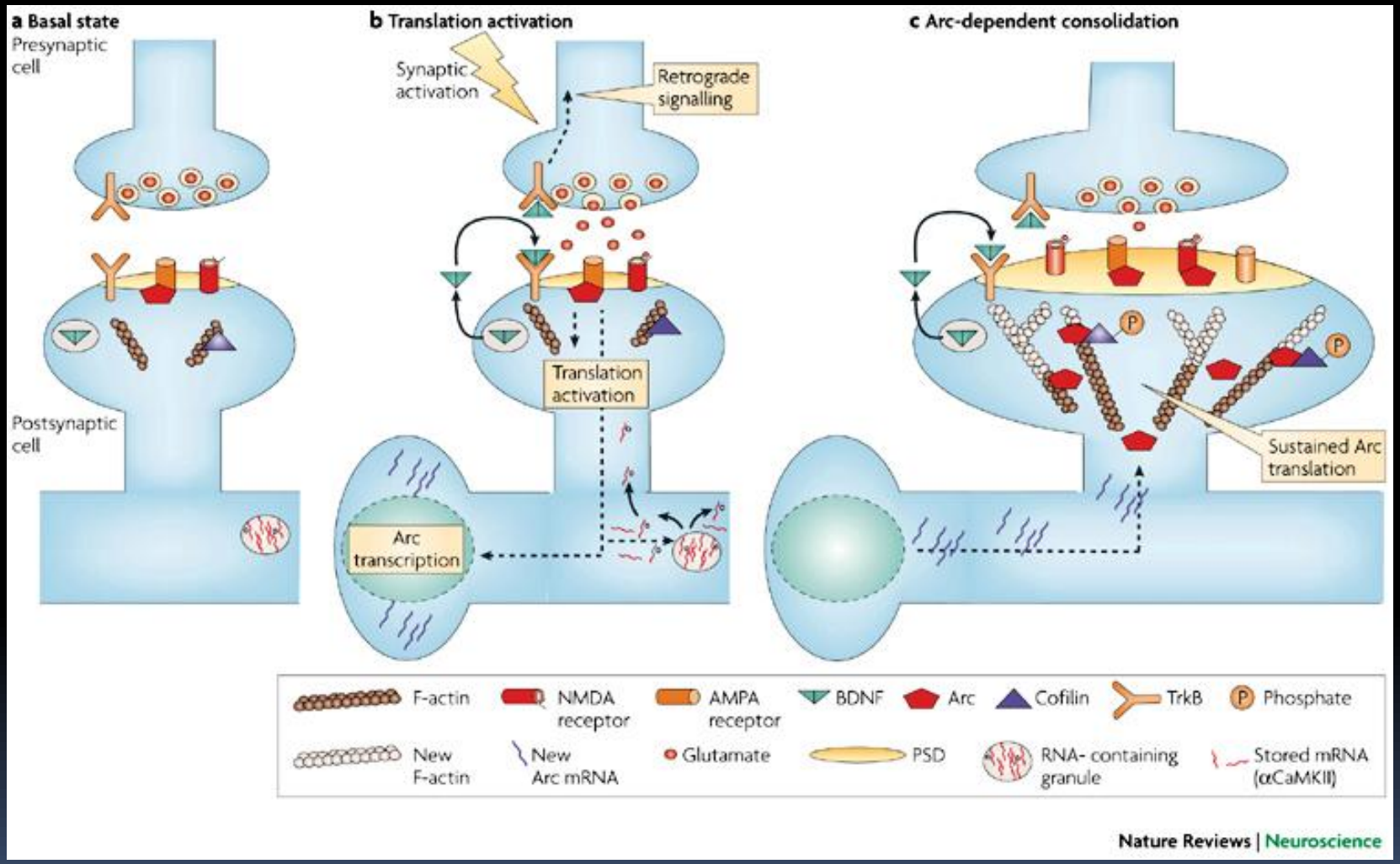
→ Another example: lin 28 - let 7 interactions.

Cytoplasmic P Bodies – Supercomputing Warehouse for RNA.



nature REVIEWS MOLECULAR CELL BIOLOGY

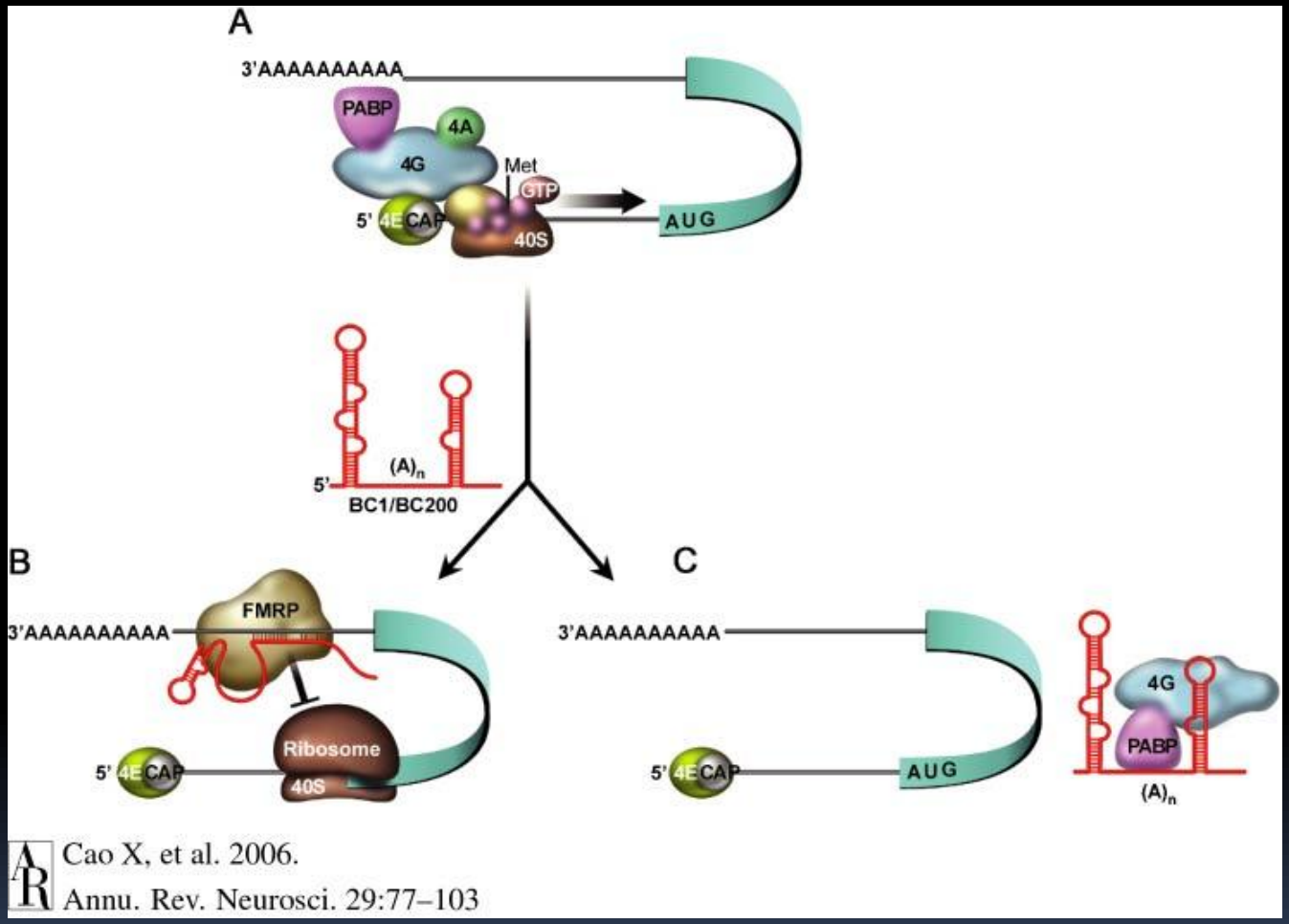
Scaffolding Machineries regulate synaptic translation



Bramham and Wells (2007)

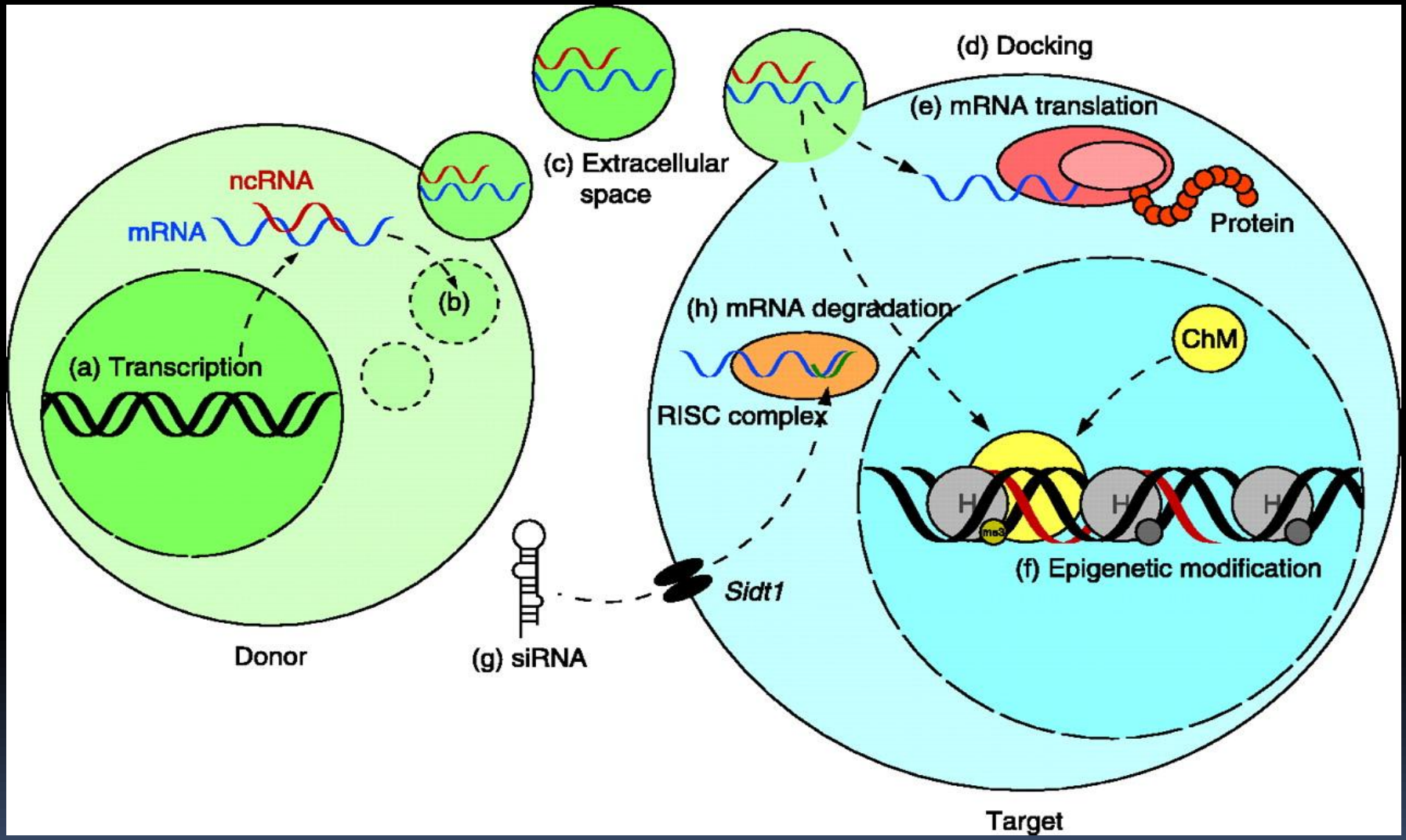
nature REVIEWS NEUROSCIENCE

ncRNAs modulate synaptic translation machineries

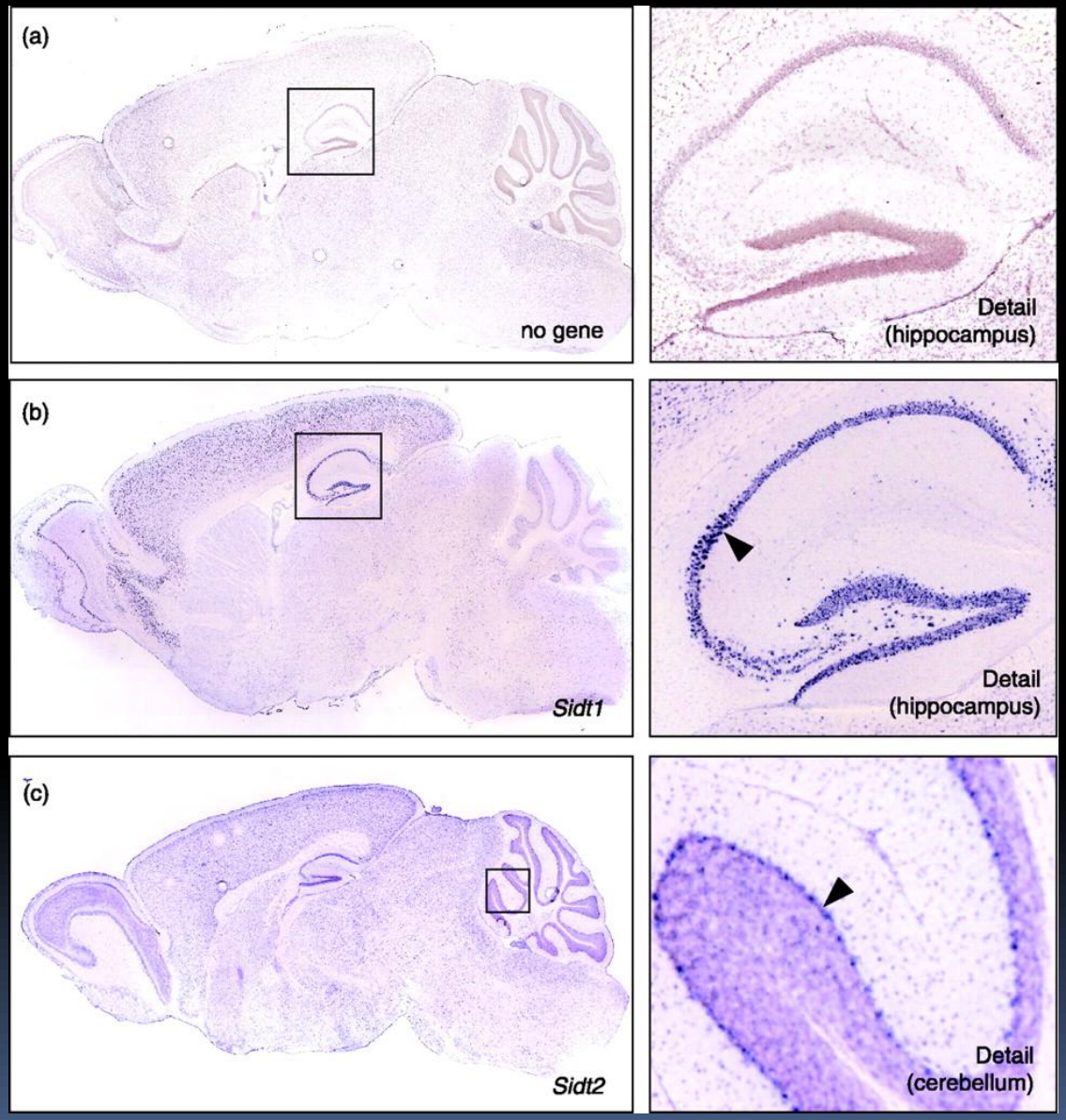


Information content supplied from a range of ncRNAs may modulate these machineries to produce many "Colors and Flavors"of LTP and LTD

RNA as an intercellular communicator



Dinger et. al. (2008)



Sid2 Expression in Mammalian Brain

Dinger et. al. (2008)

Editing may play an active role in the computational matrix.

Neuroscience Letters 466 (2009) 89–98

Contents lists available at ScienceDirect

Neuroscience Letters

journal homepage: www.elsevier.com/locate/neulet

Mini-Review

Enhancing non-coding RNA information content with ADAR editing

Georges St. Laurent III[a,b], Yiannis A. Savva[a], Robert Reenan[a,*]

[a] Department of Molecular Biology, Cell Biology, and Biochemistry, Brown University, Providence, RI 02912, USA
[b] Immunovirology – Biogenisis Group, University of Antioquia, A.A. 1226 Medellin, Colombia

ARTICLE INFO

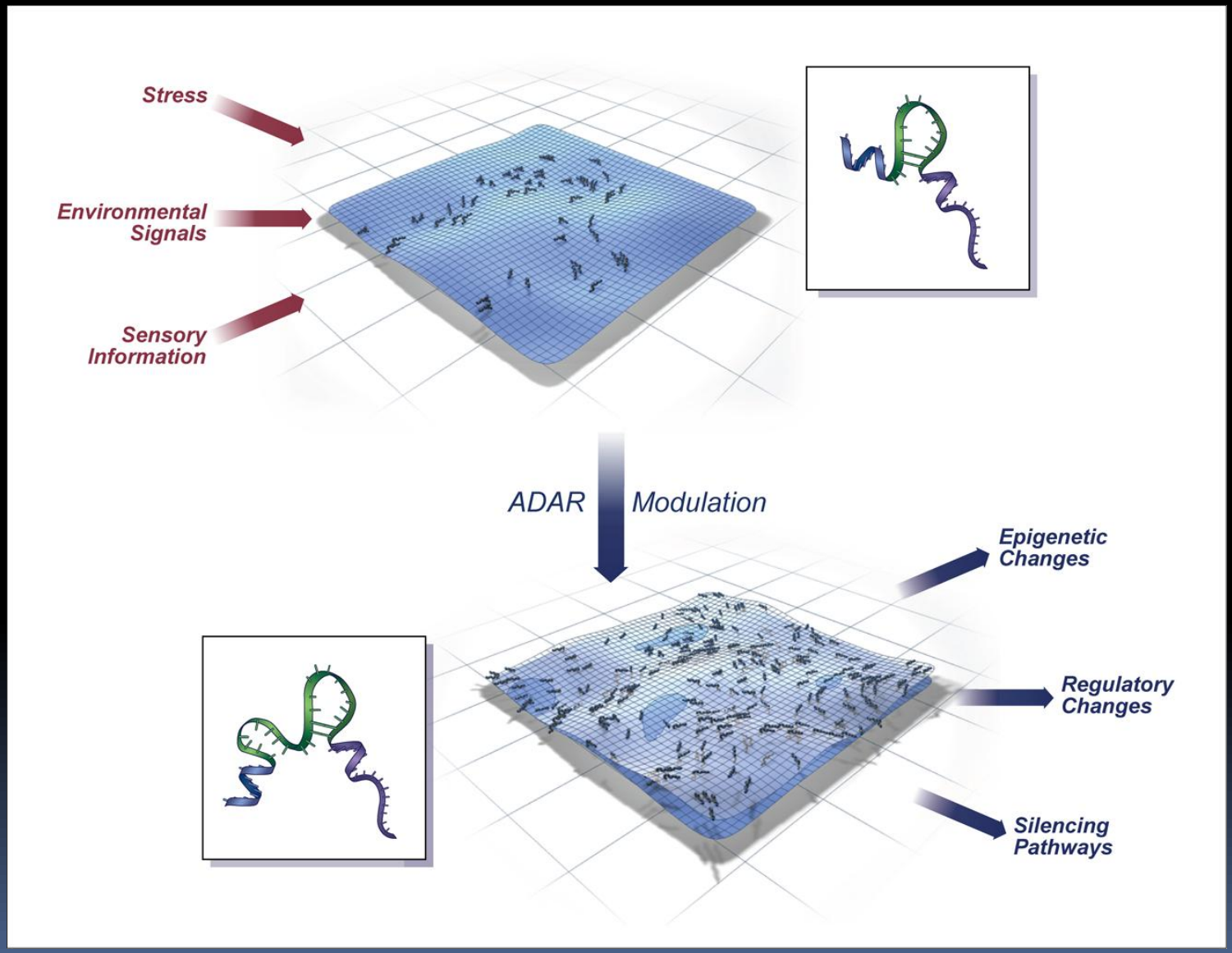
Article history:
Received 1 July 2009
Received in revised form 2 September 2009
Accepted 5 September 2009

Keywords:
RNA editing
ADAR
Nervous system transcriptome
Non-coding RNA (ncRNA)
ADAR activity during stress and inflammation

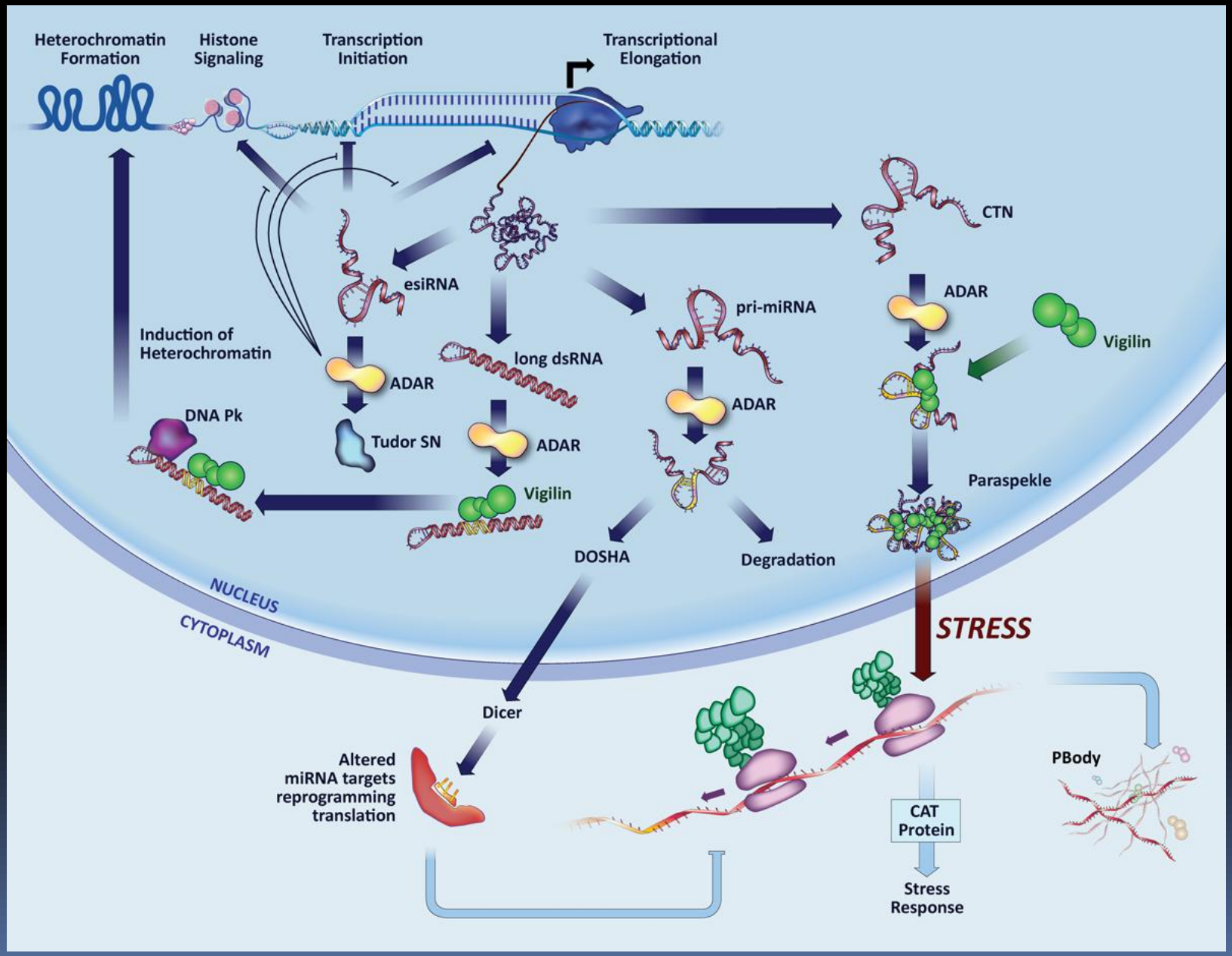
ABSTRACT

The depth and complexity of the non-coding transcriptome in nervous system tissues provides a rich substrate for adenosine de-amination acting on RNA (ADAR). Non-coding RNAs (ncRNAs) serve diverse regulatory and computational functions, coupling signal flow from the environment to evolutionarily coded analog and digital information elements within the transcriptome. We present a perspective of ADARs interaction with the non-coding transcriptome as a computational matrix, enhancing the information processing power of the cell, adding flexibility, rapid response, and fine tuning to critical pathways. Dramatic increases in ADAR activity during stress response and inflammation result in powerful information processing events that change the functional state of the cell. This review examines the pathways and mechanisms of ADAR interaction with the non-coding transcriptome, and their functional consequences for information processing in nervous system tissues.

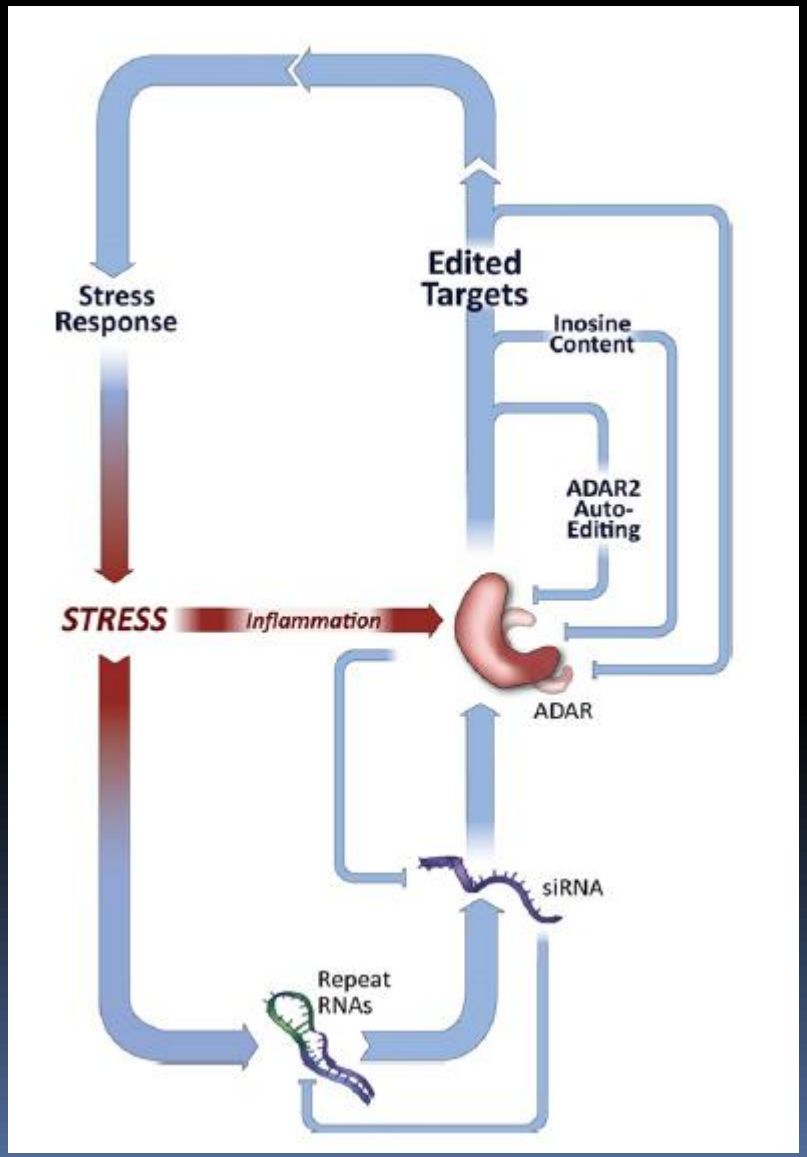
The Transcriptome as a computational Matrix



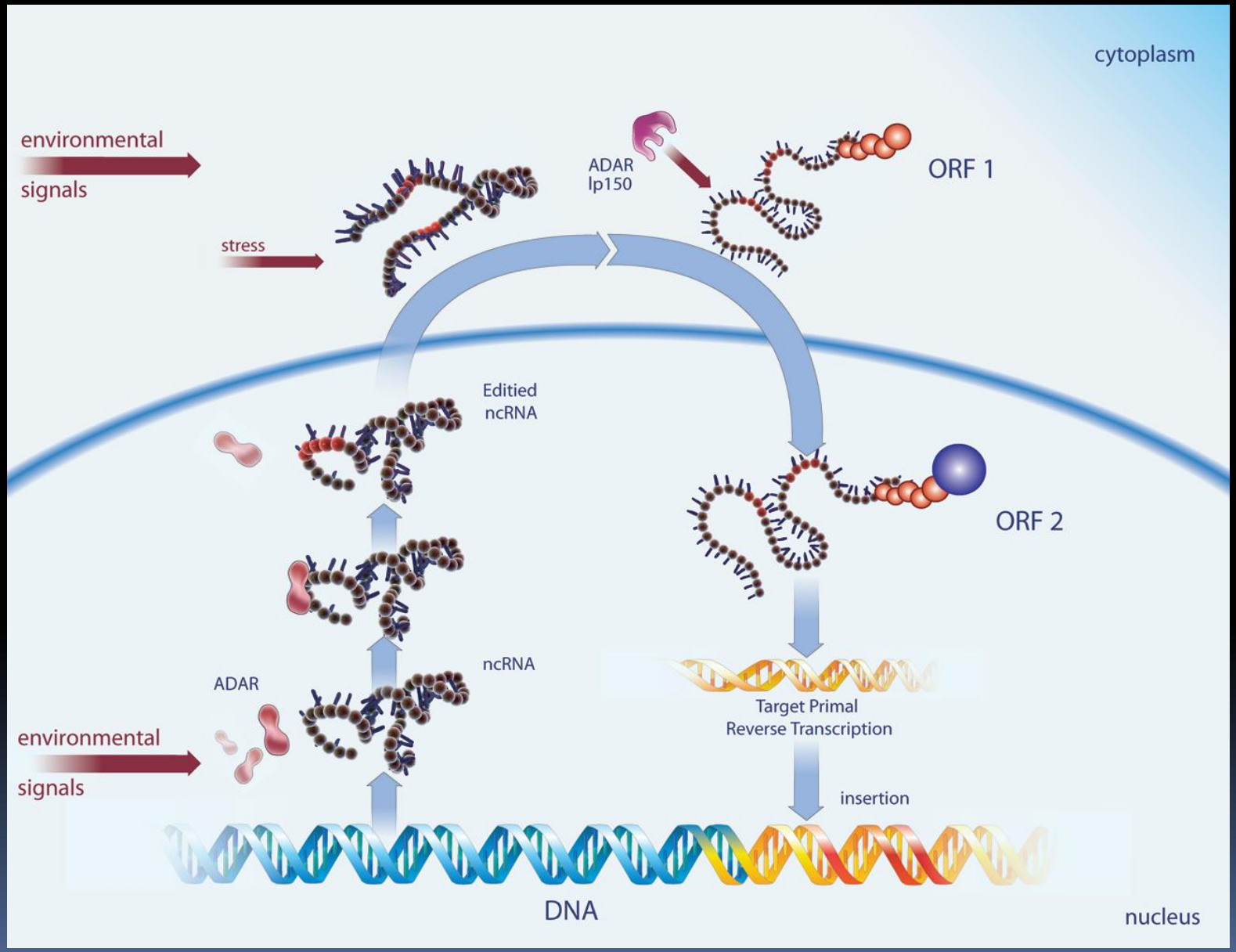
ADAR participates in ncRNA information processing



ADAR participates in Inflammation Cascade Feedback Loops



ncRNA – protein machineries mediate two way information flow



Conclusions

1. Non-coding Regions directly correlate with organismal complexity across evolution.

2. ncRNAs are differentially expressed, processed, and localized in cell types, tissues, and biological processes.

3. ncRNAs play functional roles in processes such as development, stress response, and disease.

4. ncRNAs have unique information coding and processing capabilities: density, range, and flexibility.

5. Therefore, in mammalian cells, the combinatorial space of RNA – protein interactions likely functions as a molecular supercomputer, impacting the great majority of pathways and cellular functions.

Acknowledgements

- Project Team Members
  - Mo Heydarian
  - Dennis Vorobiev
  - Dmitry Schtokalo
  - Sergey Nechkin
  - Andrey Polyanov

  } based in Novosibirsk Russia

- Collaborators
  - Mohammad Faghihi, Scripps Florida
  - Claes Wahlestedt, Scripps Florida
  - Rob Reenan, Brown University
  - Tim McCaffrey, GWU

Q & A ?