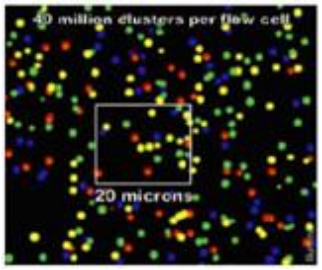
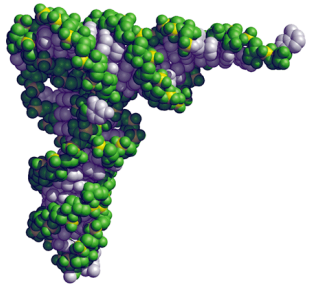


High-throughput methods for studying transcription

Structure versus function of RNA

Michal Koper, IGiB UW

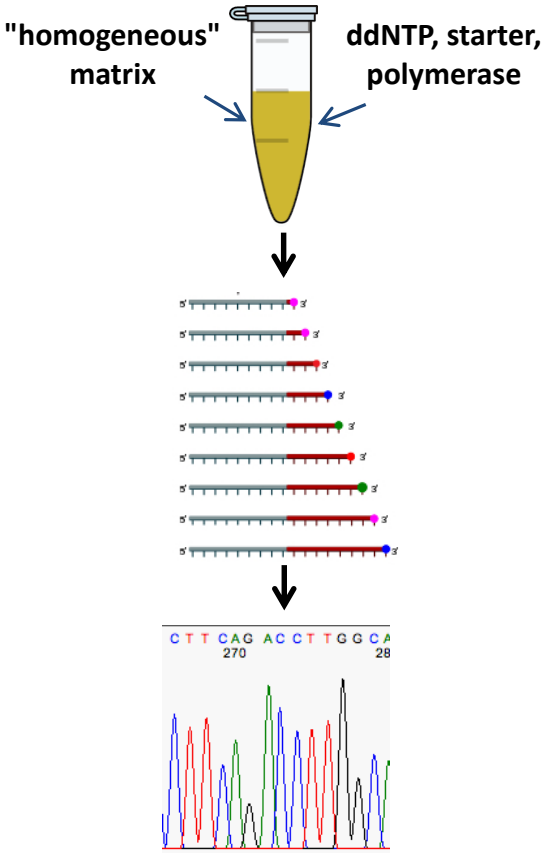


Next Generation Sequencing - NGS

("high-throughput")

Sanger *versus* NGS

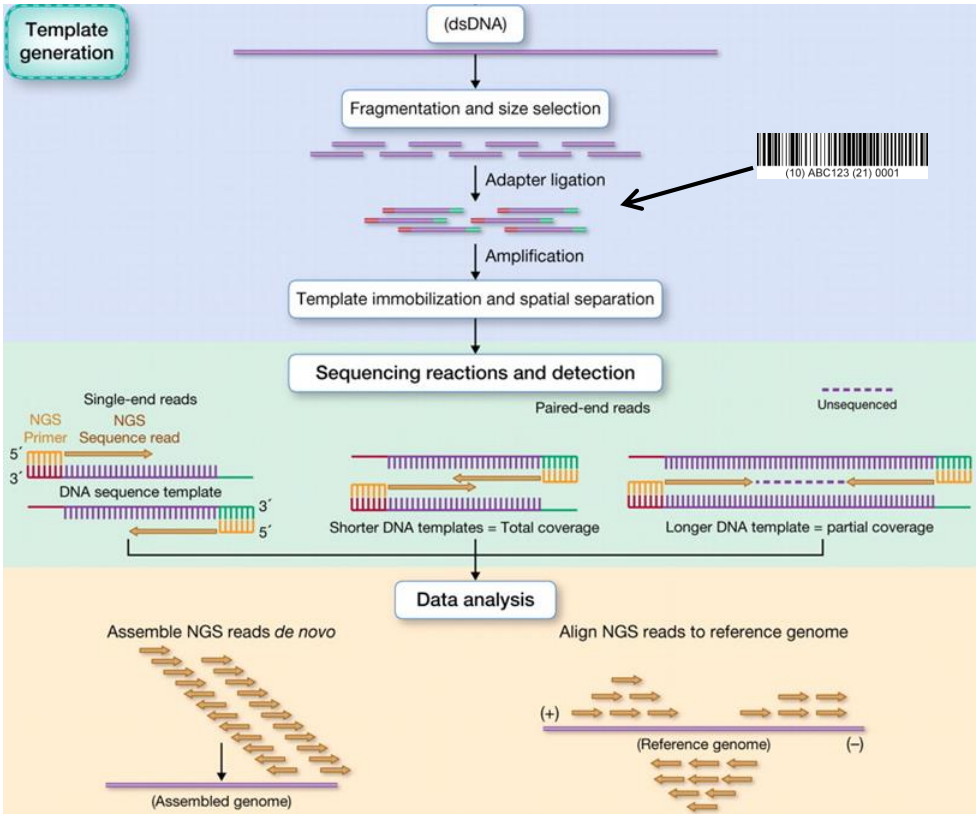
Sanger sequencing



- **Up to 384 reads** (read=sample)
- **600-1200 bp per reading**
- **Up to about 250-400 kbp per "run"**

N G S

(high-throughput; mass parallel sequencing)



Rizzo and Buck; Cancer Prev Resss 2012

- **From hundreds of thousands to over 50 billion reads** (multiplex)
- **typically 2x50 bp to 2x300 bp per read** (up to 2 Mbp)
- **Up to about 16 Tbp per "run"** (Illumina NovaSeq X)

Next-generation sequencing - different technologies

2nd generation sequencing
(now "zombie" platforms)

Pyrosequencing: by synthesis with a cascade of 4 enzymatic reactions (454; Roche) - historically the first commercially available NGS platform (2005-2016); Marquiles et. al., Nature 2005.

SOLID: by ligation of fluorescently labeled short oligonucleotides (Life Technologies)

3rd and 4th generation sequencing (real-time)

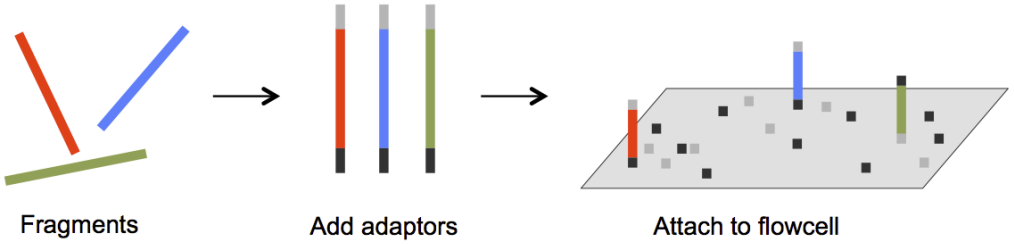
Illumina: by synthesis with reversible termination (SBS - sequencing by synthesis) - **currently the most widely used and efficient technology** (up to 16 Tbp (trillion base pairs) / 2 days of device operation)

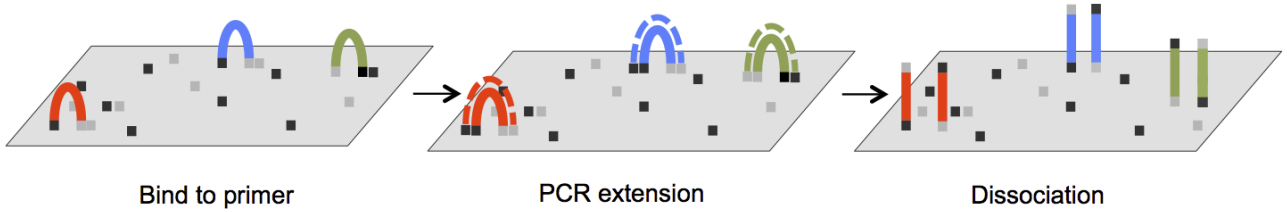
Ion Torrent: by synthesis with on-chip proton detection (Life Technologies/ThermoFisher Scientific)

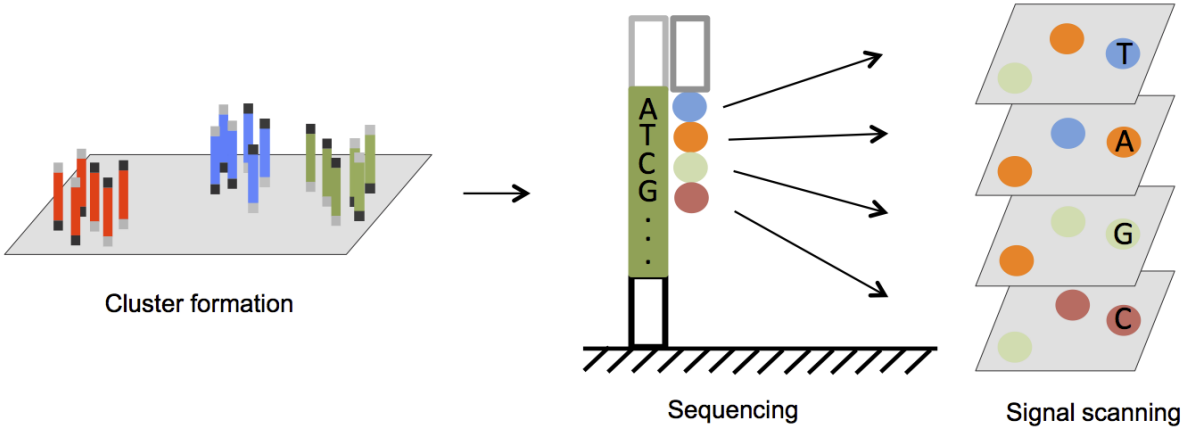
PacBio: SMRT (Single Molecule Real-Time sequencing) technology - long reads (15-25 even up to 50 kb) (Pacfic Biosystems)

Oxford Nanopore: longest readings up to 2 million base pairs (record about 4 million).

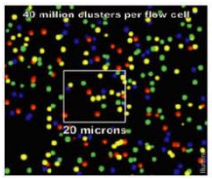
Illumina Technology







Lu et. al, "Next Generation Sequencing in Aquatic Models" in Next Generation Sequencing - Advances, Applications and Challenges, 2016.



Flow cell

Sequencer

http://www.illumina.com/

Illumina sequencing

https://www.youtube.com/watch?v=womKfikWlxM

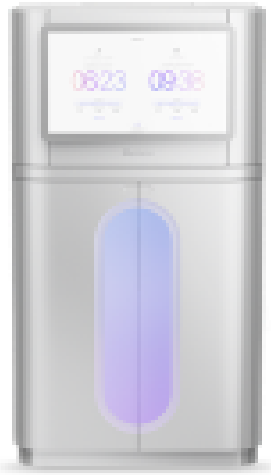
http://www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf

Illumina sequencers

iSeq (1.2 Gb max)

MiSeq (15 Gb max)

NextSeq (120 Gb max)



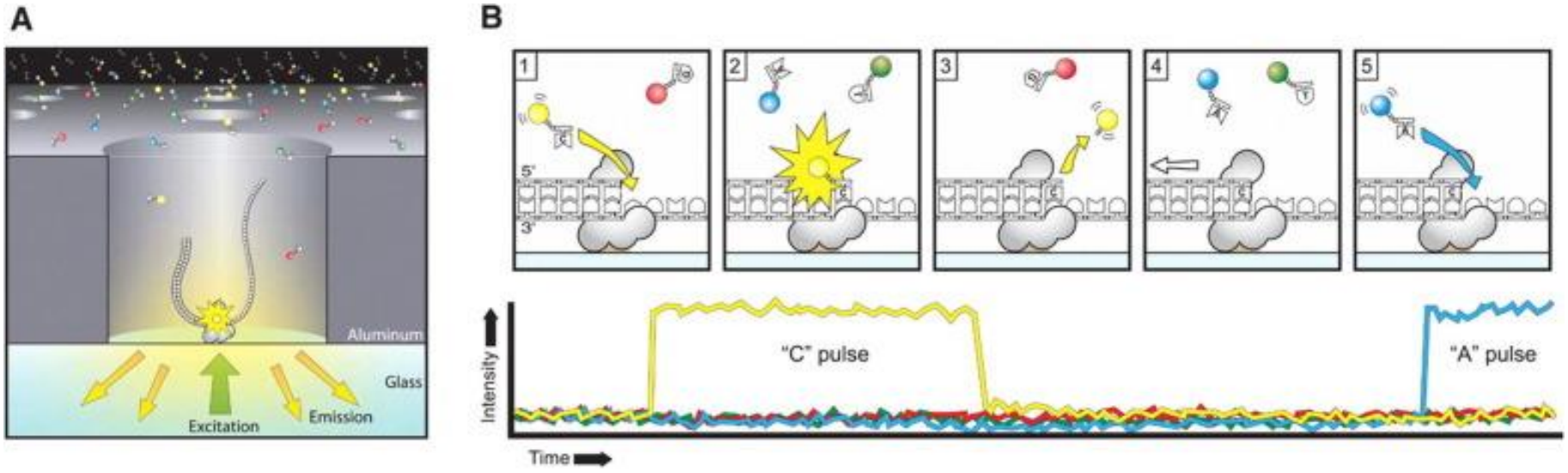
NovaSeq X (16 Tb max)

NGS sequencing - not just short reads

Pacific Biosciences (PacBio)

Single Molecule, Real-Time (SMRT) Sequencing

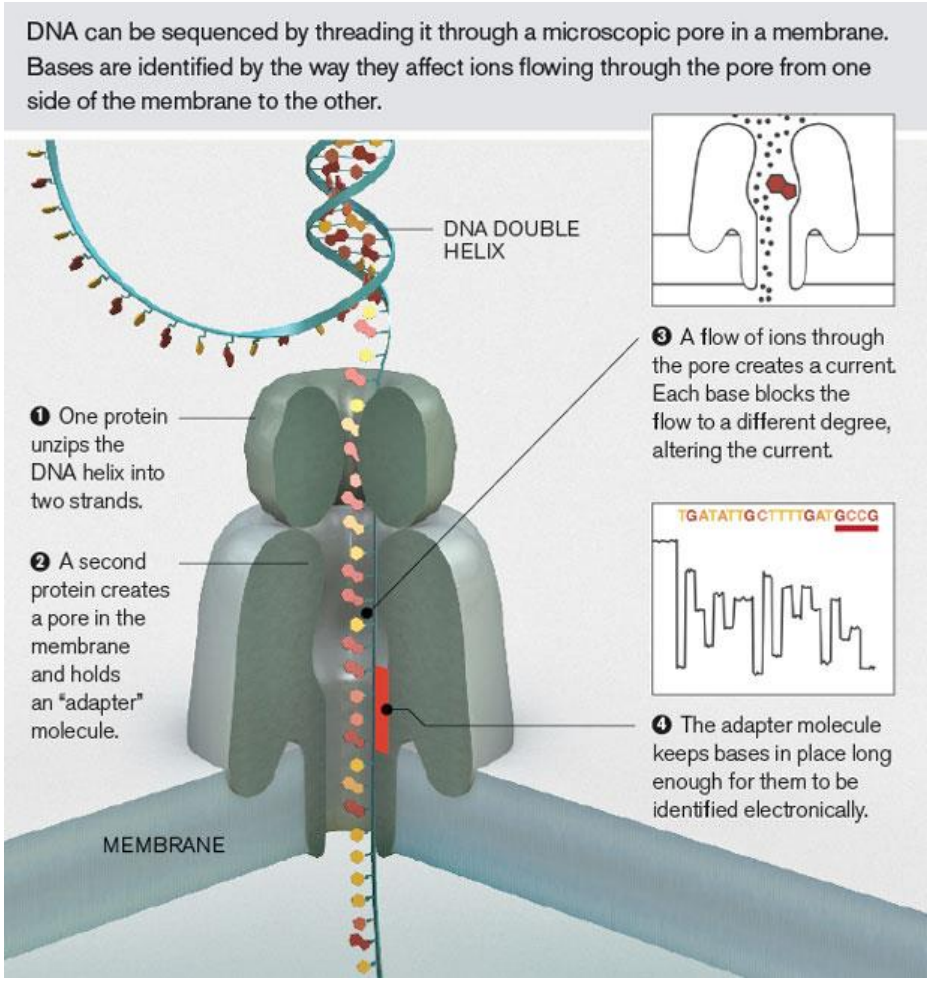
Reads up to 25 kb



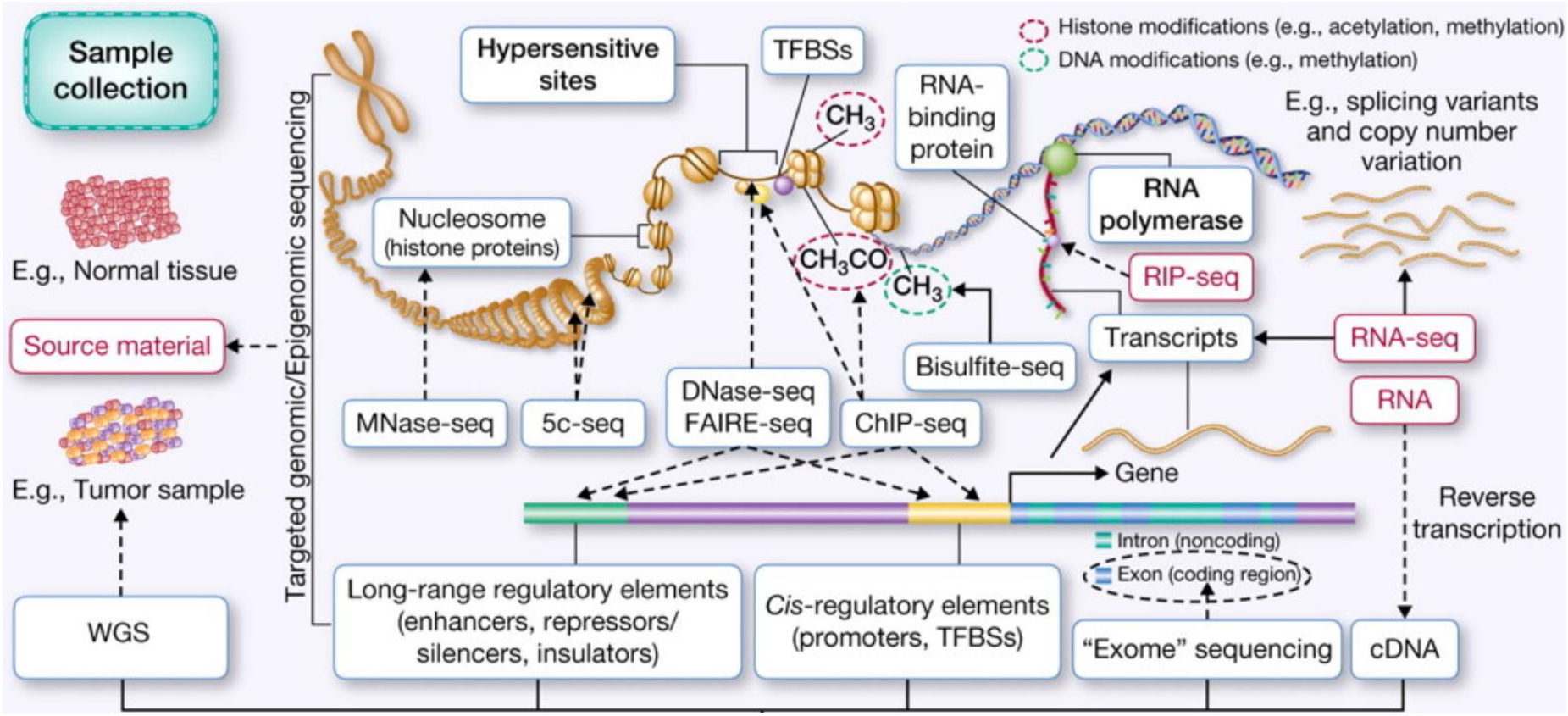
NGS sequencing - not just short reads

Oxford Nanopore - individual reads lenght up to 4 million bp!

- Direct sequencing of DNA, RNA
- Ability to recognize chemical modifications of bases

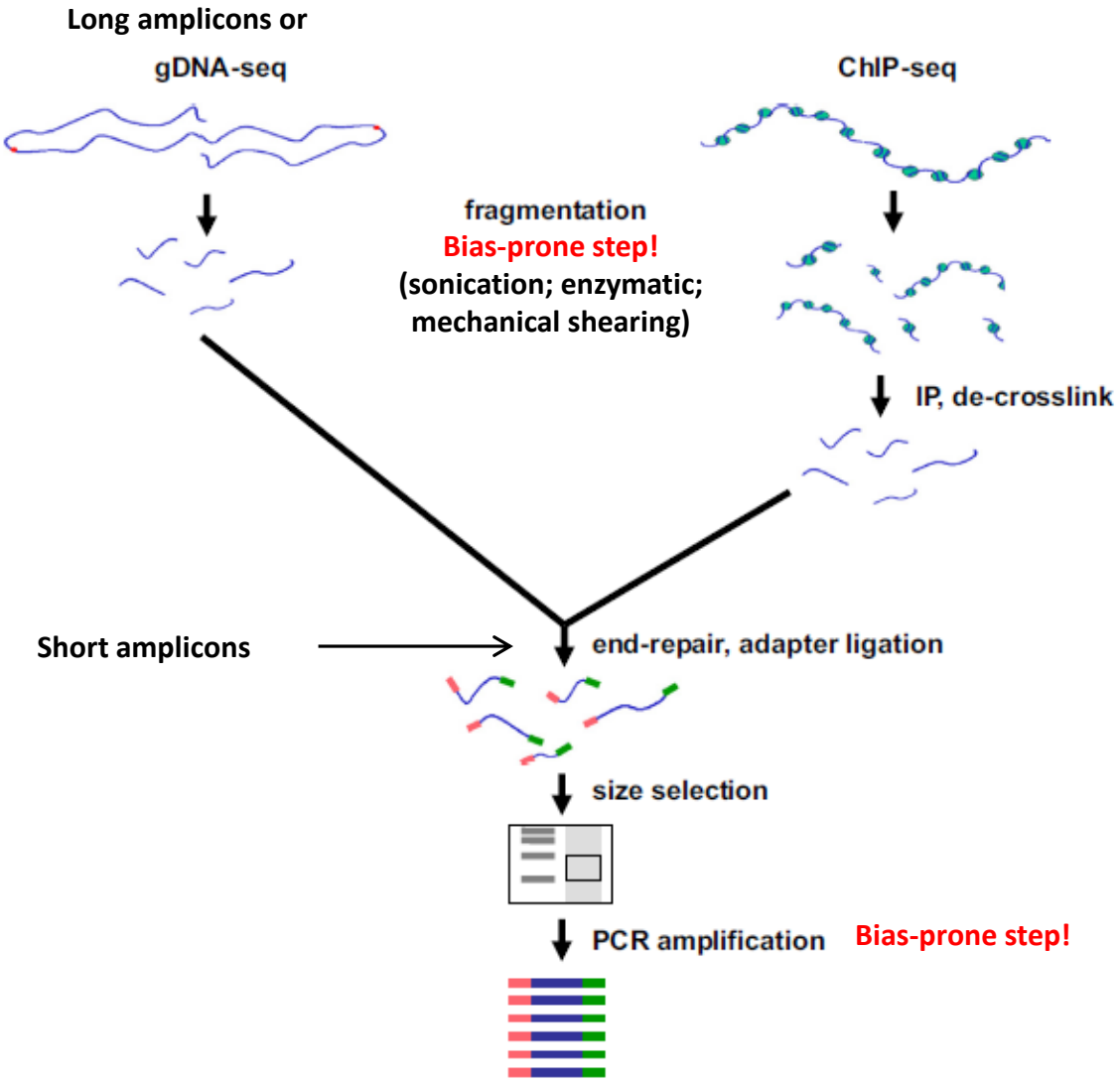


Methods using NGS



Rizzo and Buck; Cancer Prev Res 2012

Preparation of NGS libraries for DNA

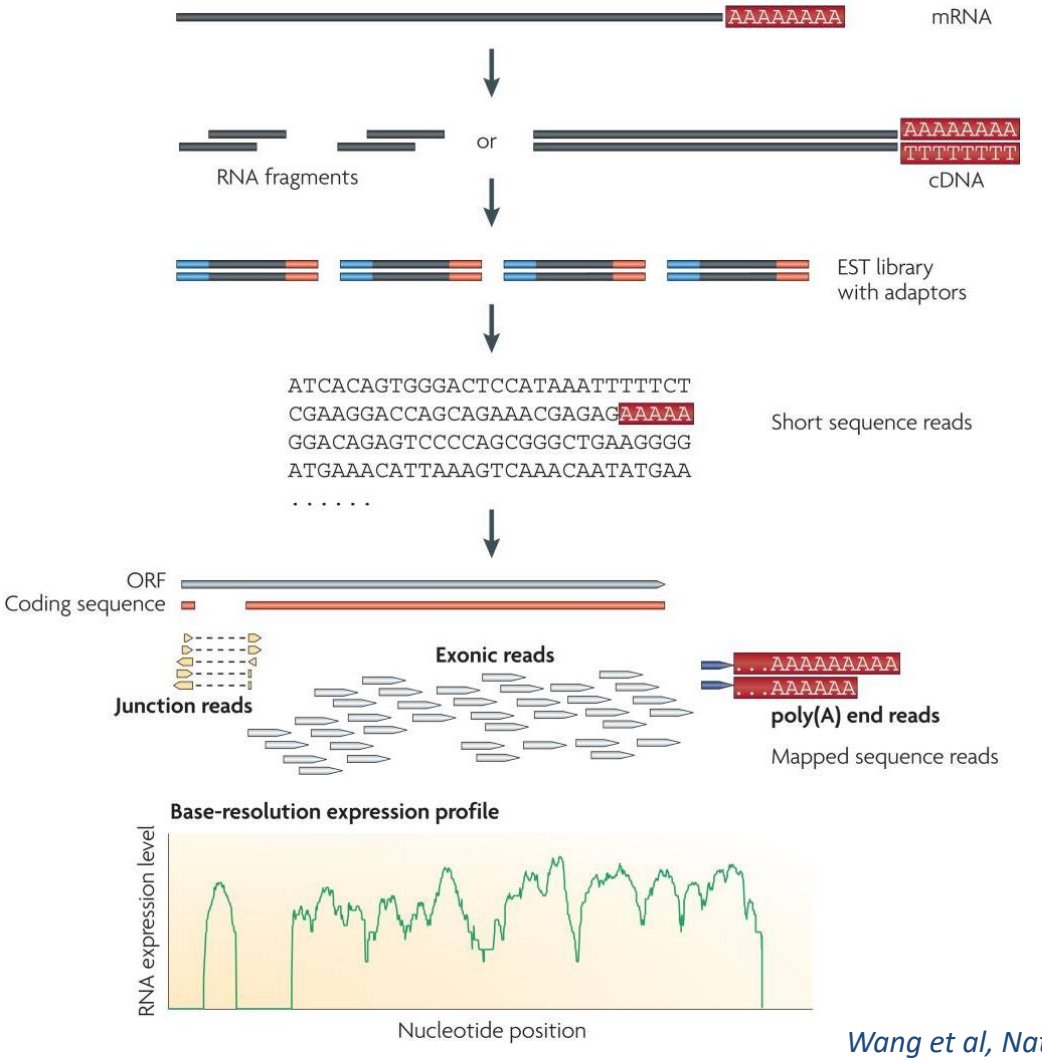


van Dijk et. al, Exp Cell Res. 2014

A typical mRNAseq experiment

First reported by *Mortazavi et. al., Nature 2008*

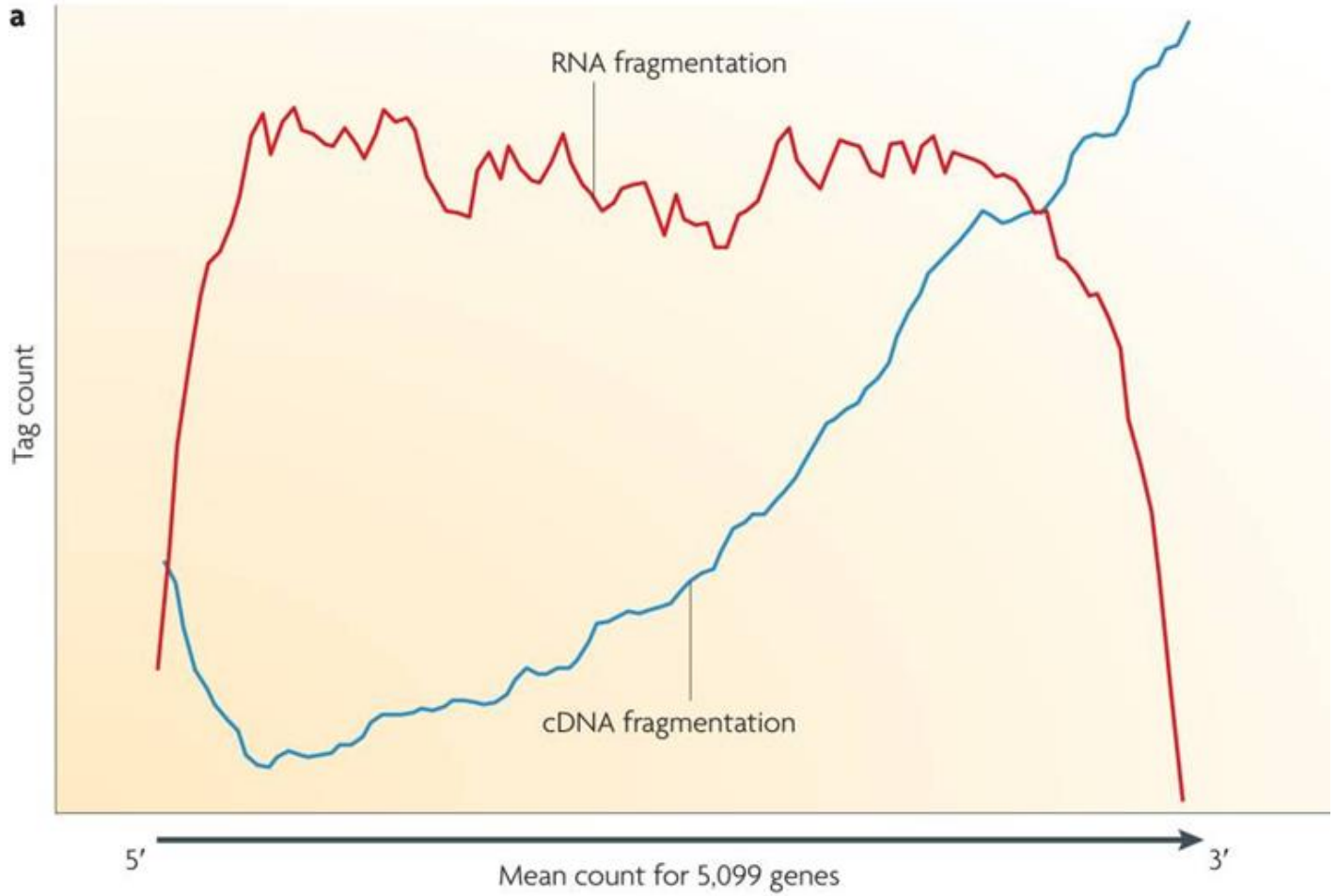
High-quality RNA,
RIN>7.5-8.0



Selection of PolyA+ fraction on oligo-dT bed

Wang et al, Nat Rev Genet. 2009

Why fragment RNA?



Wang et al,
Nat Rev Genet. 2009

Types of libraries for RNA-seq

Total RNA library

- Usually depleted of rRNA.
- Represents whole transcriptome.

PolyA selected library

- oligo(dT) selected RNA's - poly A+ fraction.
- Usually understood as mRNAs but also refers to some ncRNAs with polyA tails.

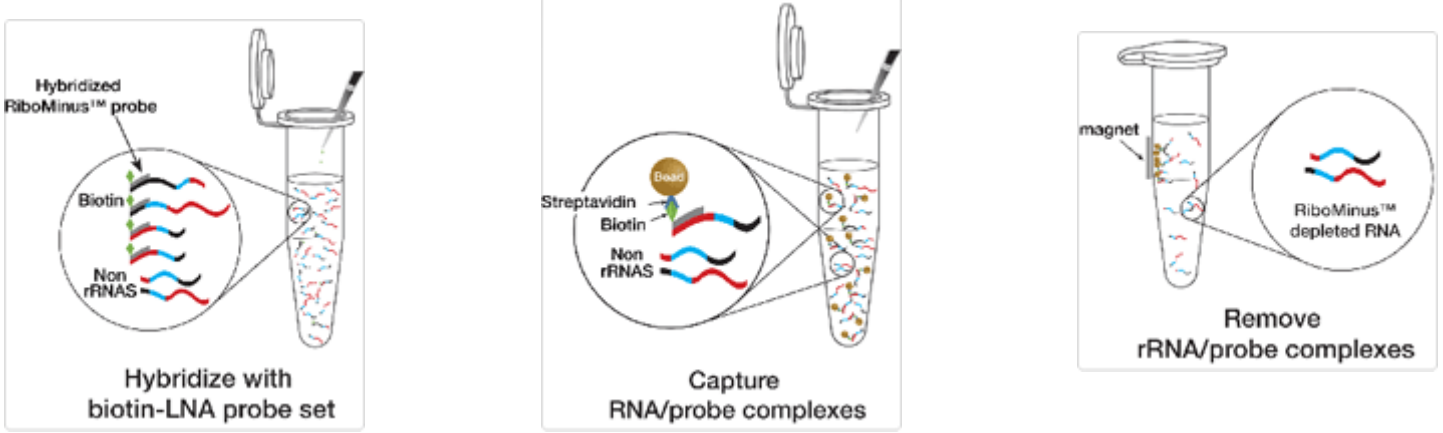
small RNA library

- Usually refers to micro RNAs and other smaller then ~ 200 nt nc RNAs.
- Size selected library.
- Size selection based on magnetics beads.
- Specific RNA isolation protocol which promotes small RNA precipitation.

- **Stranded RNAseq – strand specific**
- **Non-stranded RNAseq – non strand specific**

RNAseq for the entire transcriptome

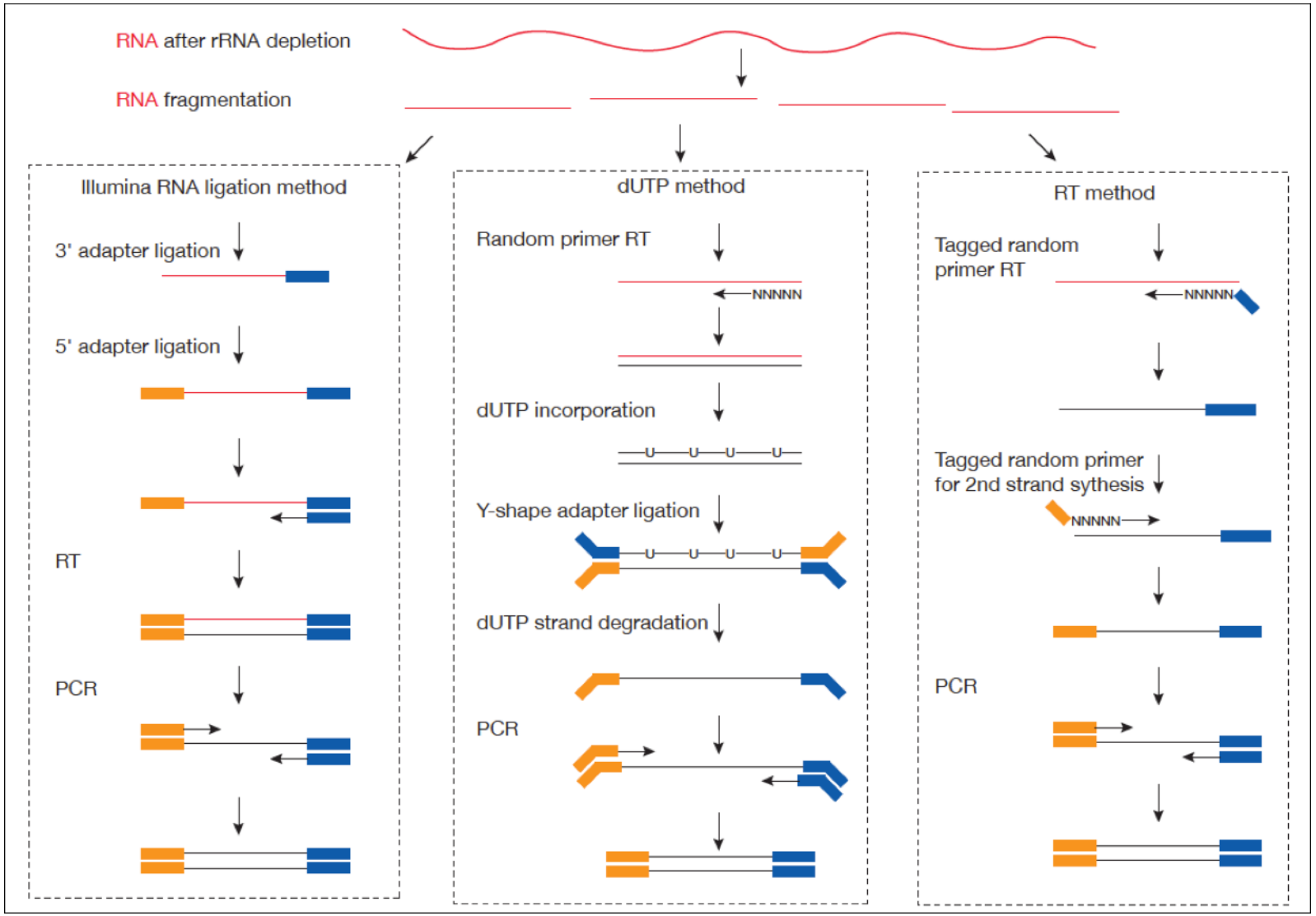
Removal of rRNA from total RNA (*rRNA depletion*)



RiboMinus™ Technology by Life Technologies (ThermoFisher)

Ribo-Zero by Epicenter (Illumina) works similarly.

Constructing libraries for RNAseq



dUTP-marked strand is selectively degraded by Uracil-DNA-Glycosylase (UDG)

http://onetipperday.blogspot.com/2012/07/how-to-tell-which-library-type-to-use.html

ChIP-chip ("ChIP on chip") and ChiP-seq

- **Chromatin immunoprecipitation**
- **Detection of immunoprecipitated DNA on microarrays or via next-generation sequencing**
- **Allows the study of protein-DNA interactions on a genomic scale : distribution of transcription factor binding or distribution of RNA polymerase II**

Methods 48 (2009) 398–408

Contents lists available at ScienceDirect

Methods

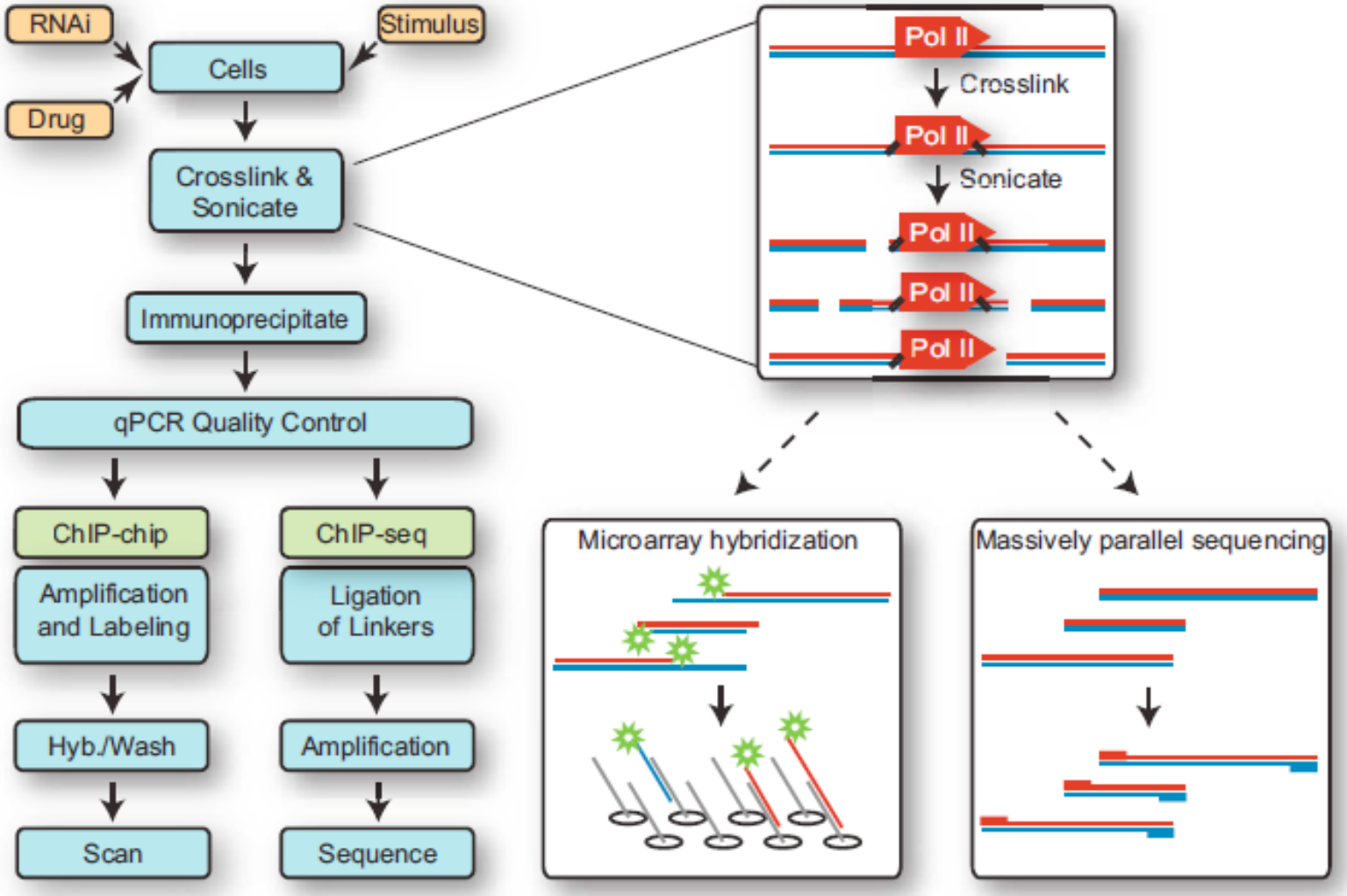
journal homepage: www.elsevier.com/locate/ymeth

Review Article

Using ChIP-chip and ChIP-seq to study the regulation of gene expression: Genome-wide localization studies reveal widespread regulation of transcription elongation

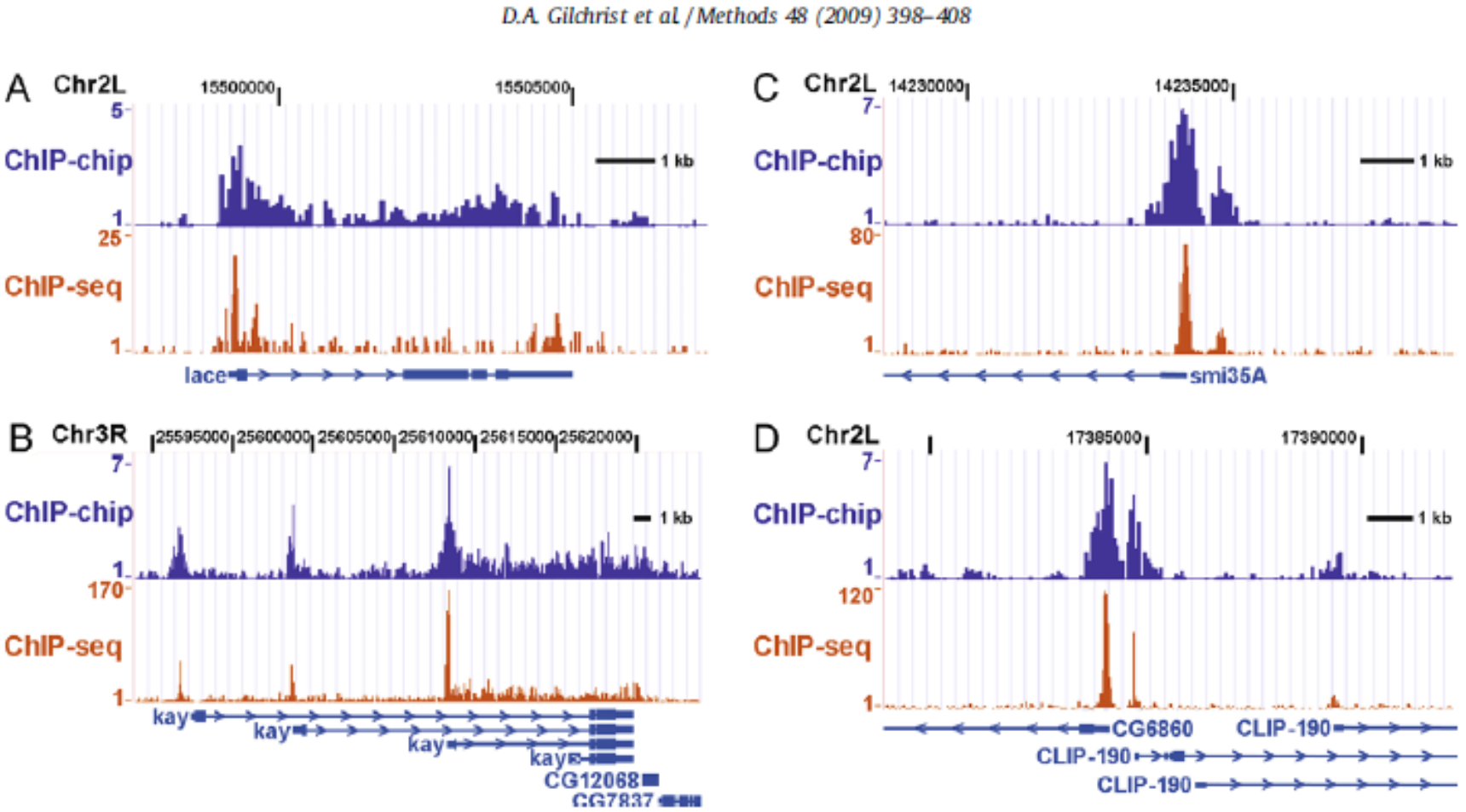
Daniel A. Gilchrist [a], David C. Fargo [b], Karen Adelman [a,*]

ChIP-chip ("ChIP on chip") and ChiP-seq

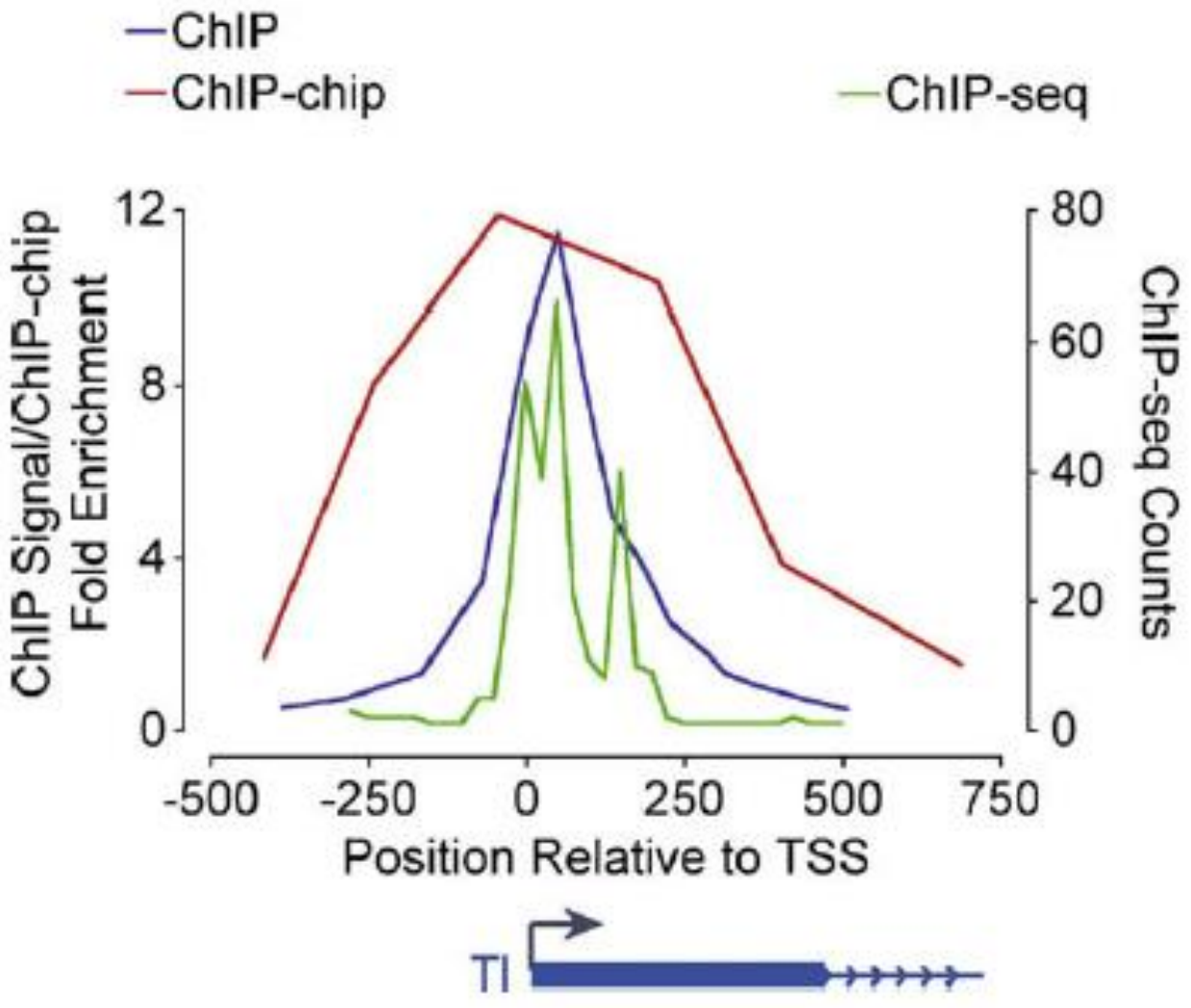


Gilchrist et al, 2009

ChIP-chip ("ChIP on chip") and ChiP-seq



ChIP-chip ("ChIP on chip") and ChiP-seq



Gilchrist et al, 2009

NGS sequencers - capabilities vs. costs

Instrument	Amplification	Run time	Millions of Reads/run	Bases / read	Reagent Cost/run	Reagent Cost/Gb	bp/run	Gbp/run	cost/Gb
Applied Biosystems 3730 (capillary)	PCR, cloning	2 hrs.	0.000096	650	$144	$2,307,692.31	62,400	0	$2,307,692.31
Illumina MiSeq v3	bridgePCR	56 hrs.	25	600	$1,530	$102.00	15,000,000,000	15	$102.00
Illumina NextSeq 500 - High v2	BridgePCR	29 hrs.	400	300	$4,240	$35.33	120,000,000,000	120	$35.33
Illumina HiSeq 2500 - high output v3	BridgePCR	11 days	1500	200	$15,860	$52.87	300,000,000,000	300	$52.87
Illumina HiSeq 2500 - high output v4	BridgePCR	40 hrs.	2000	50	$6,240	$62.40	100,000,000,000	100	$62.40
Illumina HiSeq 4000	BridgePCR	3.5 days	2500	300	$15,400	$20.53	750,000,000,000	750	$20.53
Illumina HiSeq X - Ten	BridgePCR	< 3 days	3000	300	$6,375	$7.08	900,000,000,000	900	$7.08
Ion Torrent – PGM 314 chip v2	emPCR	4 hrs.	0.55	400	$474	$2,154.55	220,000,000	0.22	$2,154.55
Ion Torrent - S5 540 chip	emPCR	2.5 hrs.	80	200	$1,275	$79.69	16,000,000,000	16	$79.69
Pacific Biosciences Sequel	None - SMS	≤6 hrs.	0.385	10000	$700	$181.82	3,850,000,000	3.85	$181.82

Illumina NovaSeq 6000 (2017)
- Up to 20 billion reads, 6 Tb

Illumina NovaSeq X (2022)
- Up to 52 billion reads, 16 Tb

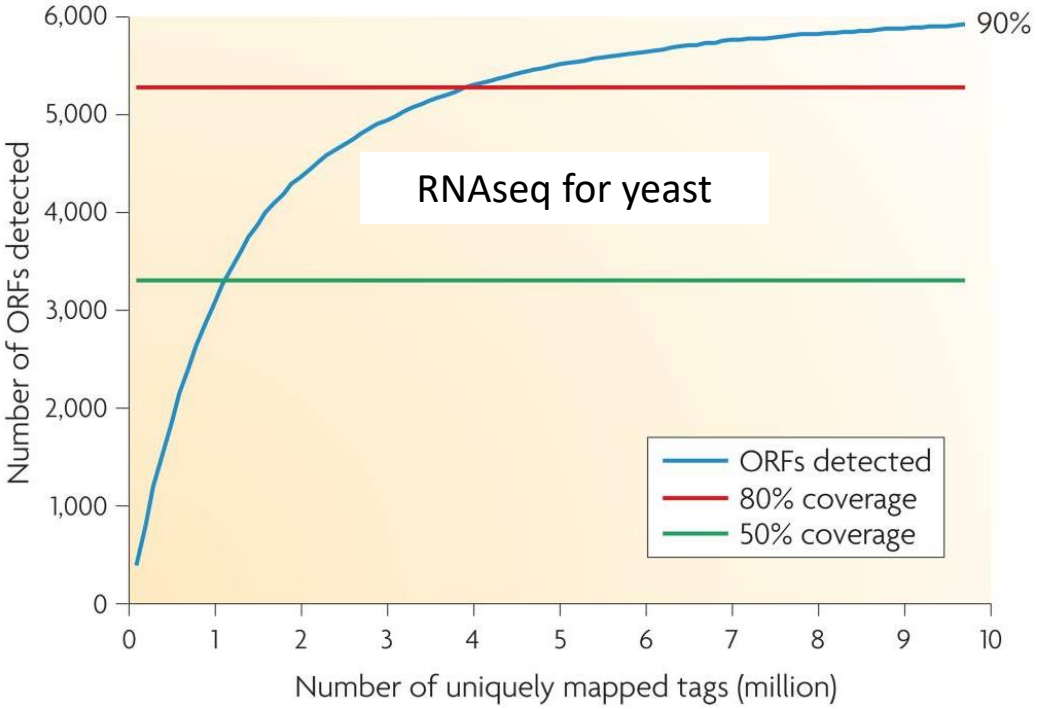
Glenn, TC (2011) Field Guide to Next Generation DNA Sequencers. Molecular Ecology Resources. 2016 Update.

http://www.molecularecologist.com/next-gen-fieldguide-2016/

NGS platform and throughput selection

Genome resequencing - coverage required ~ **30X**
De novo genome sequencing - ~ **100X** coverage required

Required Coverage in RNAseq depends on the complexity of the genome and the intended "**sensitivity**" of the experiment



Wang et al,
Nat Rev Genet. 2009

Yeast transcriptome ~ 5-10 million reads
Human transcriptome ~ 20-30 million for mRNA, 50-200 million for total RNA

NGS "in house" *versus* "outsourcing"

"Must have a box" ... not just the sequencers themselves!

NGS laboratory at the UW's Institute of Public Health

Illumina MiSeq

http://www.illumina.com/

Ion Torrent

PGM Proton

http://www.lifetechnologies.com/pl/en/home/brands/ion-torrent.html

Covaris S220

http://covaris.com

Agilent Bioanalyzer 2100

https://www.genomics.agilent.com

Computing server + workstations + CLC

http://dell.com

Costs of RNAseq experiments

The cheapest option:

mRNAseq 2x150 bp, 20 million reads from $200/sample with bioinformatics analysis

Typical sequencing of total RNA:

Sample service in Poland at a scale of 20 samples (gross):

~ PLN 1,000/sample - library preparation (with "ribodepletion") **~ 500** (mRNA selection)

+

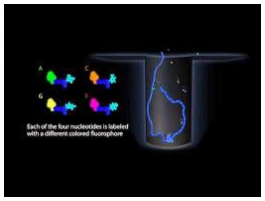
~$900/sample - Illumina 2x100 bp sequencing, 40 million reads/sample

= ~ 1,500 - 2,000/sample (data only for FASTQ format)

NGS development prospects

1. **Simplifying the design of libraries**

PacBio Systems *Sequel*



http://www.pacb.com

2. **Extension of readings without a decrease in their quality**

Oxford Nanopore Technologies
***MinION* (Plug and Play sequencer)**

3. **Real "personal" sequencers**

http://www.sciencemag.org

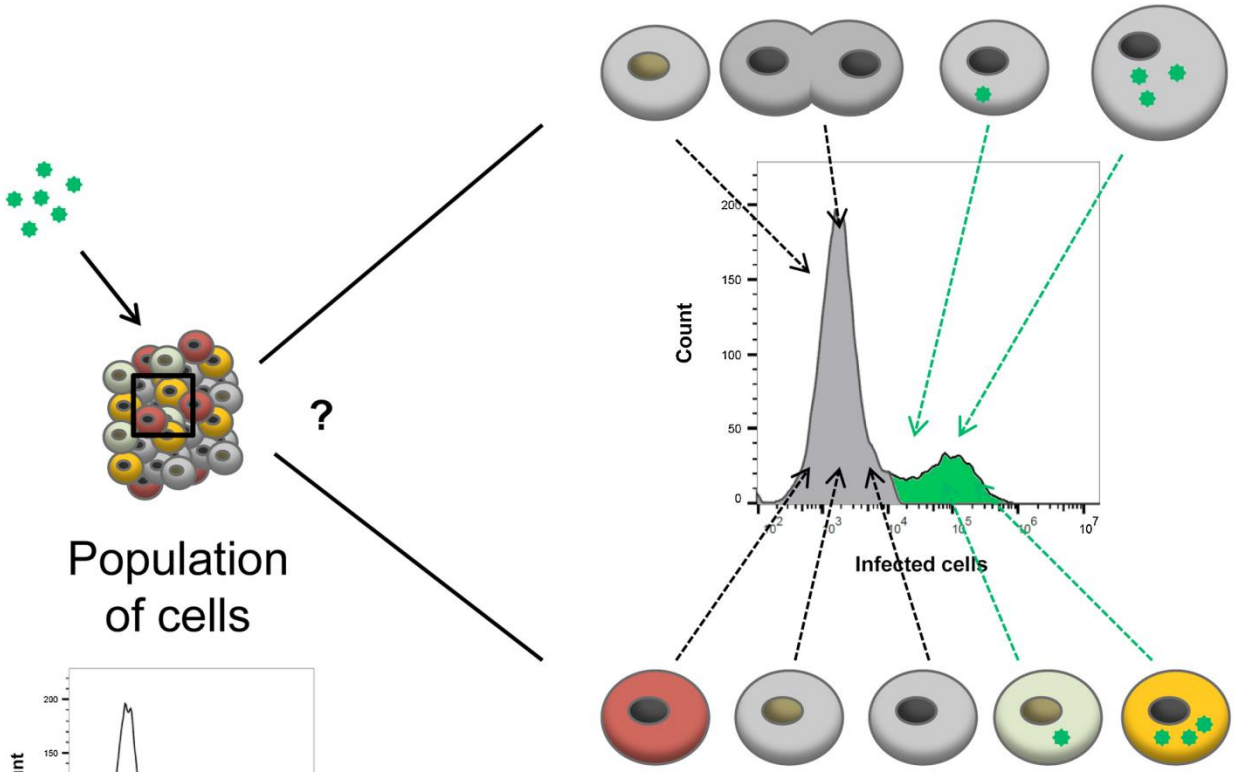
4. **Simplify, automate bioinformatics analysis + educate bioinformaticians**

5. **New applications -> e.g., sequencing from a single cell**

6. **New technologies -> e.g. the return of direct RNA sequencing (DRS; Direct RNA Sequencing).**

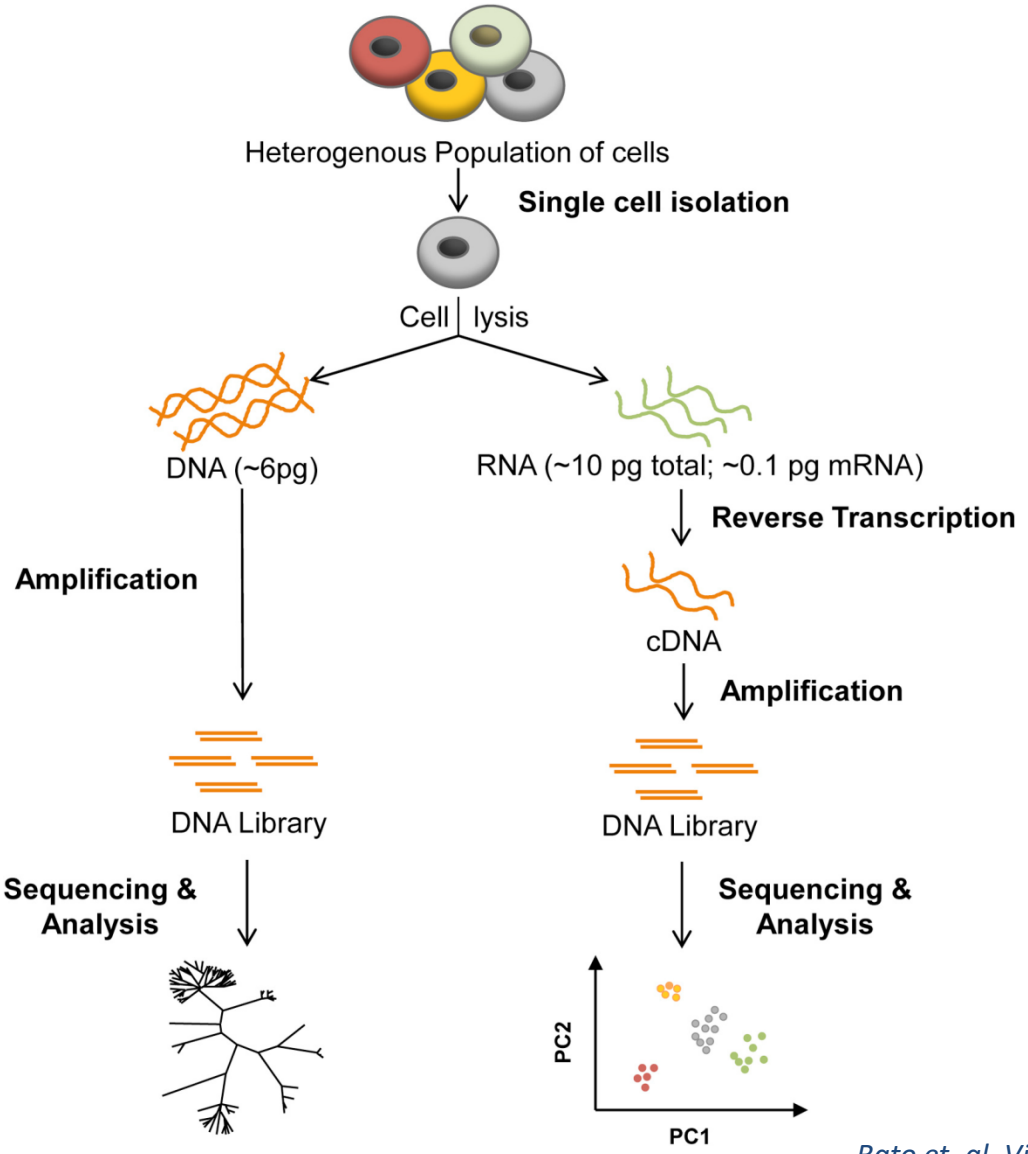
Sequencing from a single cell

Single-cell sequencing (SCS)



Rato et. al, Virus Res 2017

Sequencing from a single cell



Transcriptomic profiles for individual cells differing in physiological state

Ability to study the development of infection dependent on host cell heterogeneity

Infection dynamics at the single cell level

Rato et. al, Virus Res 2017

Structure versus function of RNA

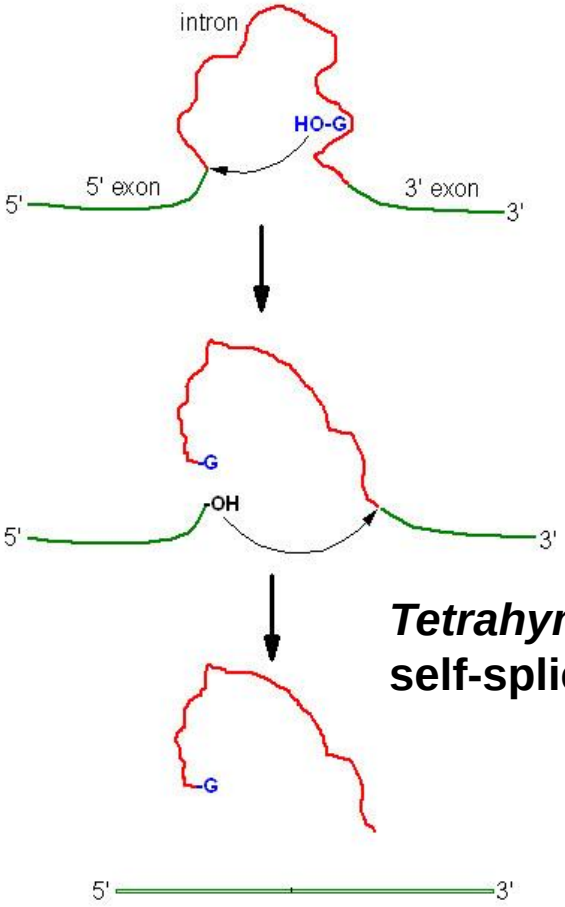
RNA capacity - CATALYTIC RNAs

Nobel 1989

RNA enzymes **– RIBOZYMES**

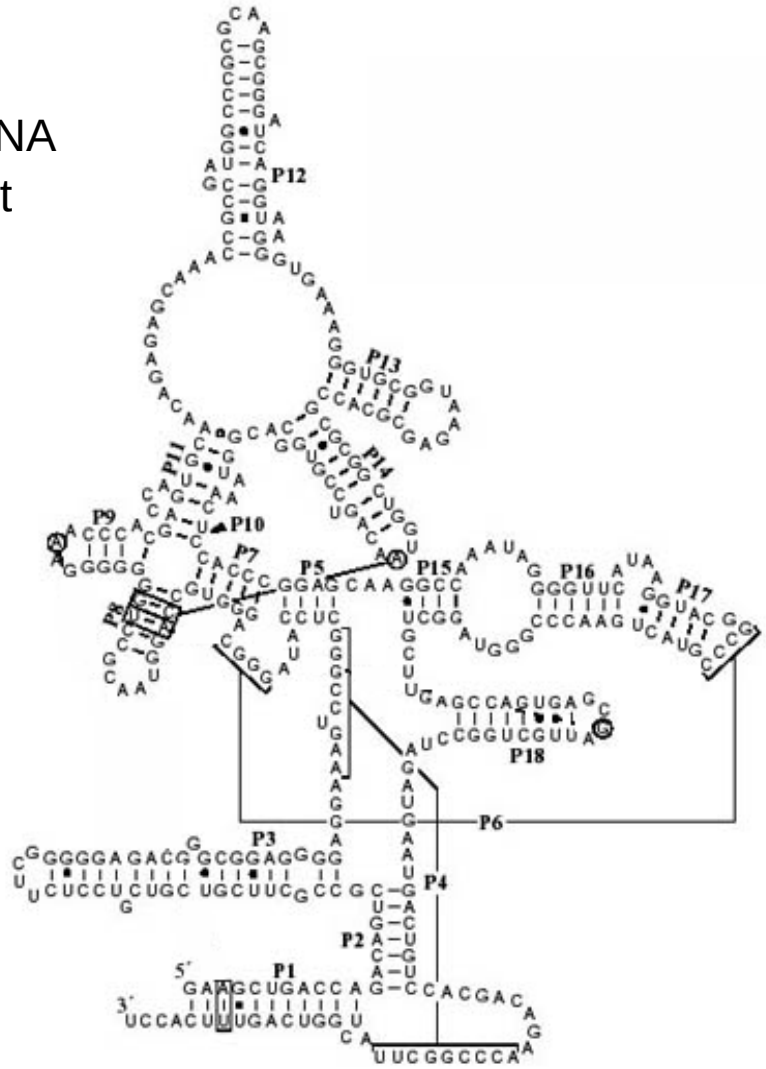
-1981/82 Tom Cech - self-splicing in *Tetrahymena* rRNA

-1982 Sidney Altman - bacterial RNaseP RNA subunit



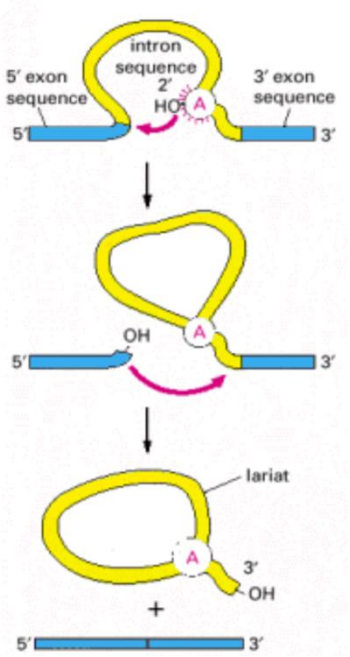
Thomas Cech
Sidney Altman

***Tetrahymena* group I self-splicing intron**

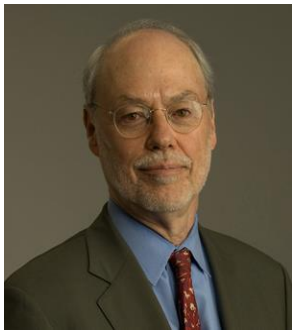


***Escherichia coli* RNaseP RNA**

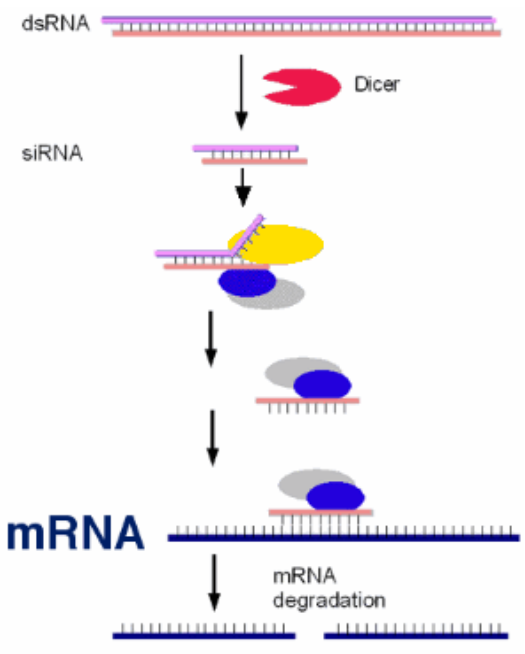
Za J. Kufel



mRNA SPLICING Nobels 1993



Phil Sharp
Richard Roberts



RNAi Nobels 2006

Andrew Fire
Craig Mello

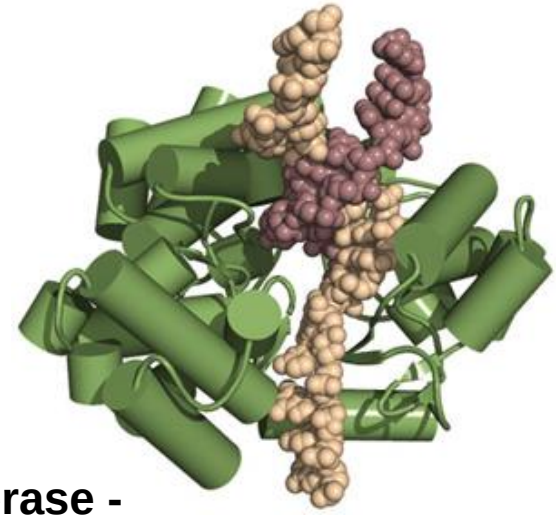
Za J. Kufel

RNAs – STRUCTURE AND FUNCTION

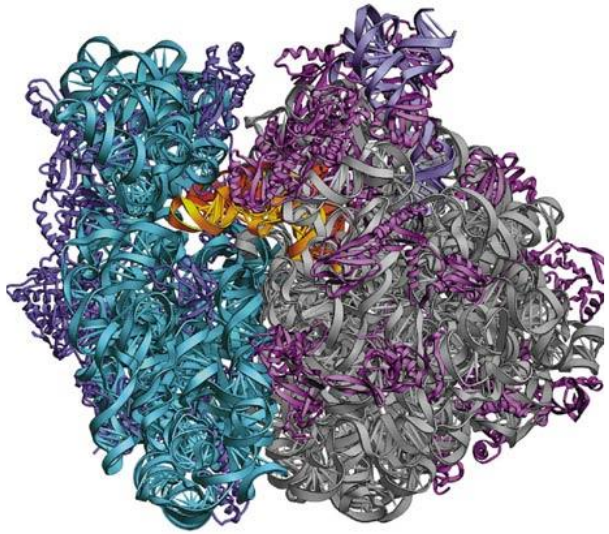
Nobels 2009

Elizabeth Blackburn
Jack Szostak
Carol Greider

Venkatraman Ramakrishnan
Ada Yonath
Thomas Steitz



Telomerase -
maintaing chromosome ends

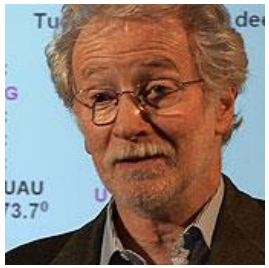


Crystal structure of the ribosome

SELEX = Systematic Evolution of Ligands by EXponential enrichment

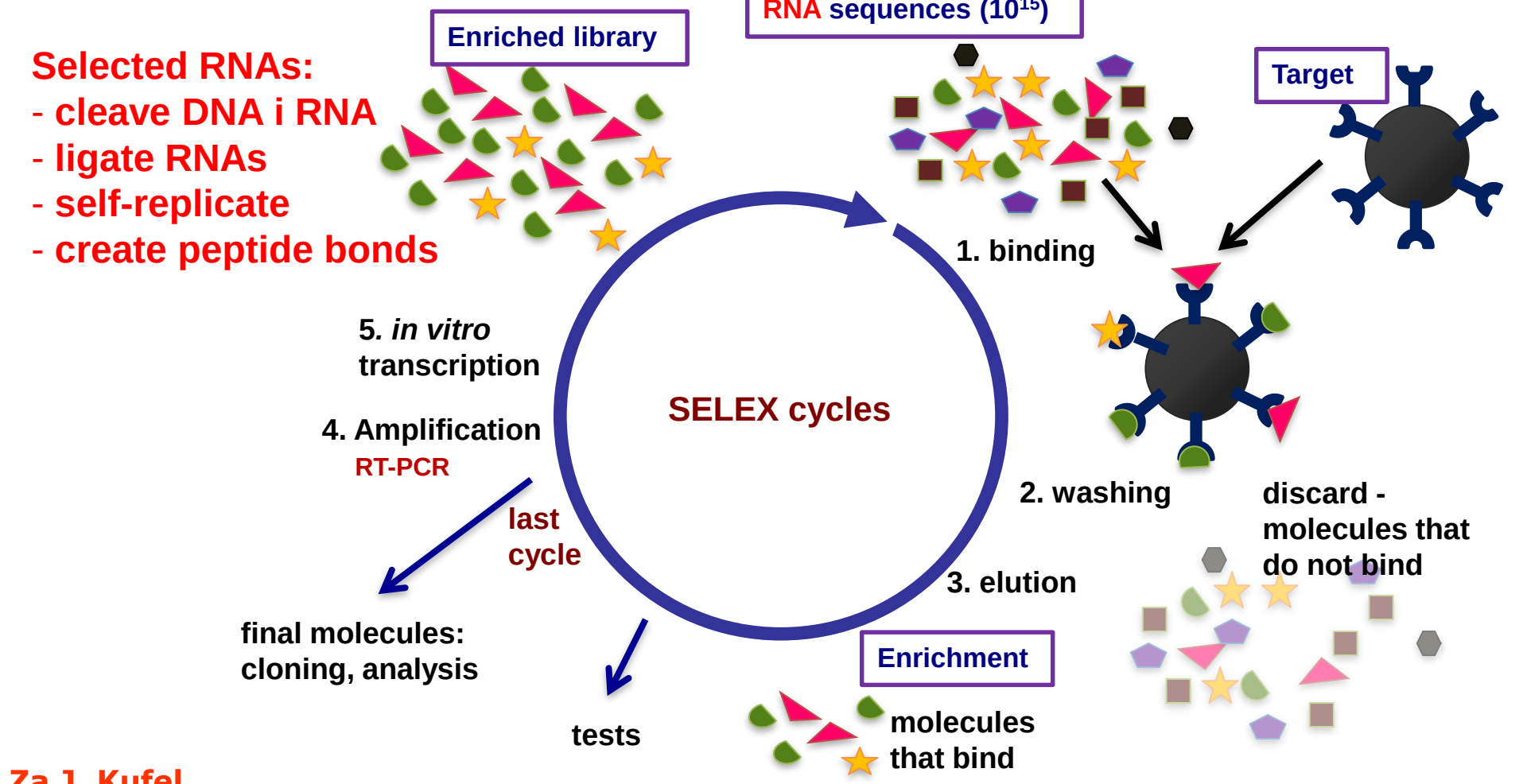
1990


Szostak


Gold

Method of selecting RNA/DNA molecules with desired properties (aptamers, ribozymes) based on cycles of amplification

Selected RNAs:
- cleave DNA i RNA
- ligate RNAs
- self-replicate
- create peptide bonds



Za J. Kufel

RNA folding

Organization of RNA structure – 3 levels

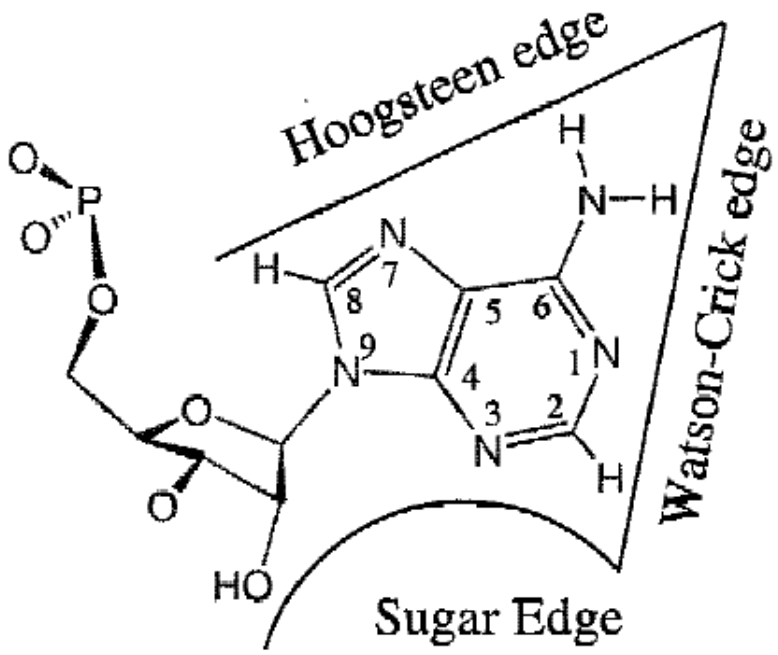
- **Primary structure:** nucleotide sequence
- **Secondary structure:** nucleotide pairings by Watson-Crick rules
- **Tertiary structure:** interactions between distant secondary structures within a molecule

Primary structure - nucleotide sequence

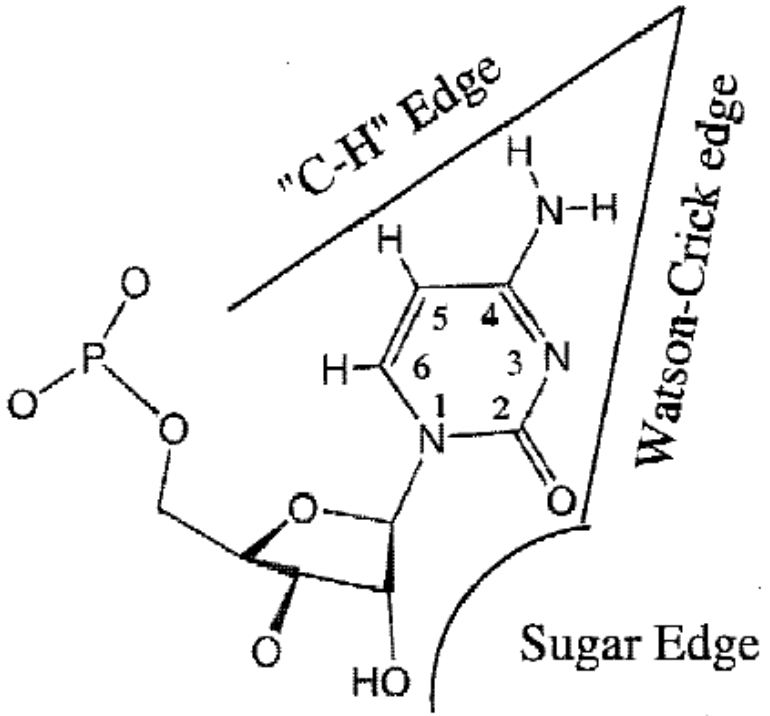
- DNA-dependent sequence but...
- Intron excision, insertions, deletions and...
- Post-transcriptional modifications (Ψ - pseudouracil, D - dihydrouridine)....
- so its exact determination is possible by cDNA sequencing, sometimes necessarily in combination with mass spectrometry

Edges of bases involved in nucleotide pairing

Purines (G or A)



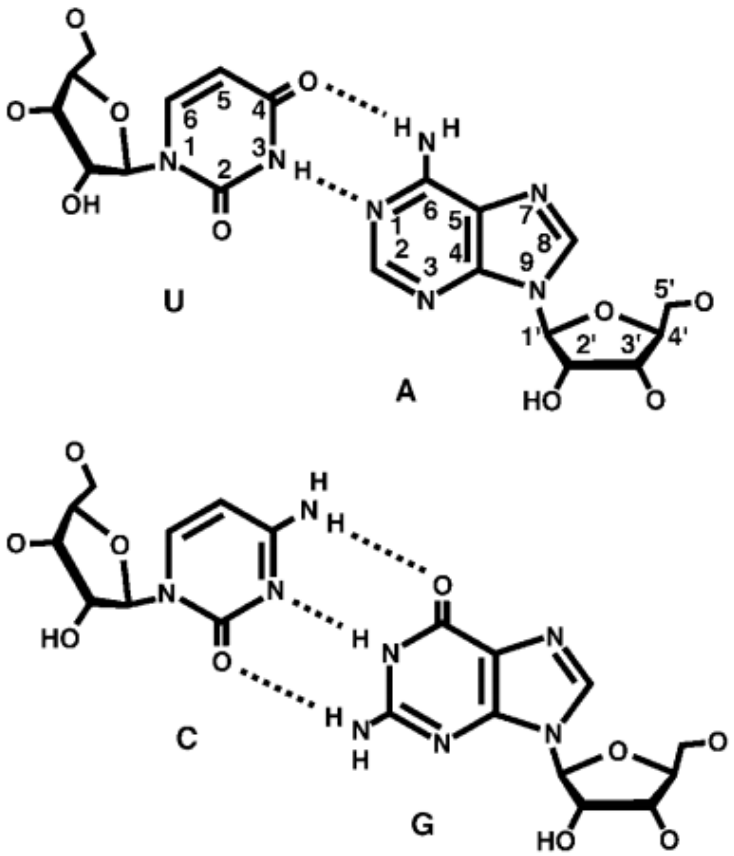
Pyrimidines (T, C or U)



Geometric nomenclature and classification of RNA base pairs. NB Leontis and E Westhof. RNA 2001. 7: 499-512

Secondary structure: canonical and non-canonical nucleotide pairings

- **Canonical pairing:** nucleotide pairings according to **Watson-Crick** rules (G:C, A:T, A:U)
- **Non-canonical pairings**: ***wobble-type*** (G:U) and those resulting from interactions between the atoms of the other edges of the base molecules - many possibilities.
- **The pairings are** formed by **double-stranded segments:** and **single-stranded** segments that separate them



RNA canonical and non-canonical base pairing types: a recognition method and complete repertoire

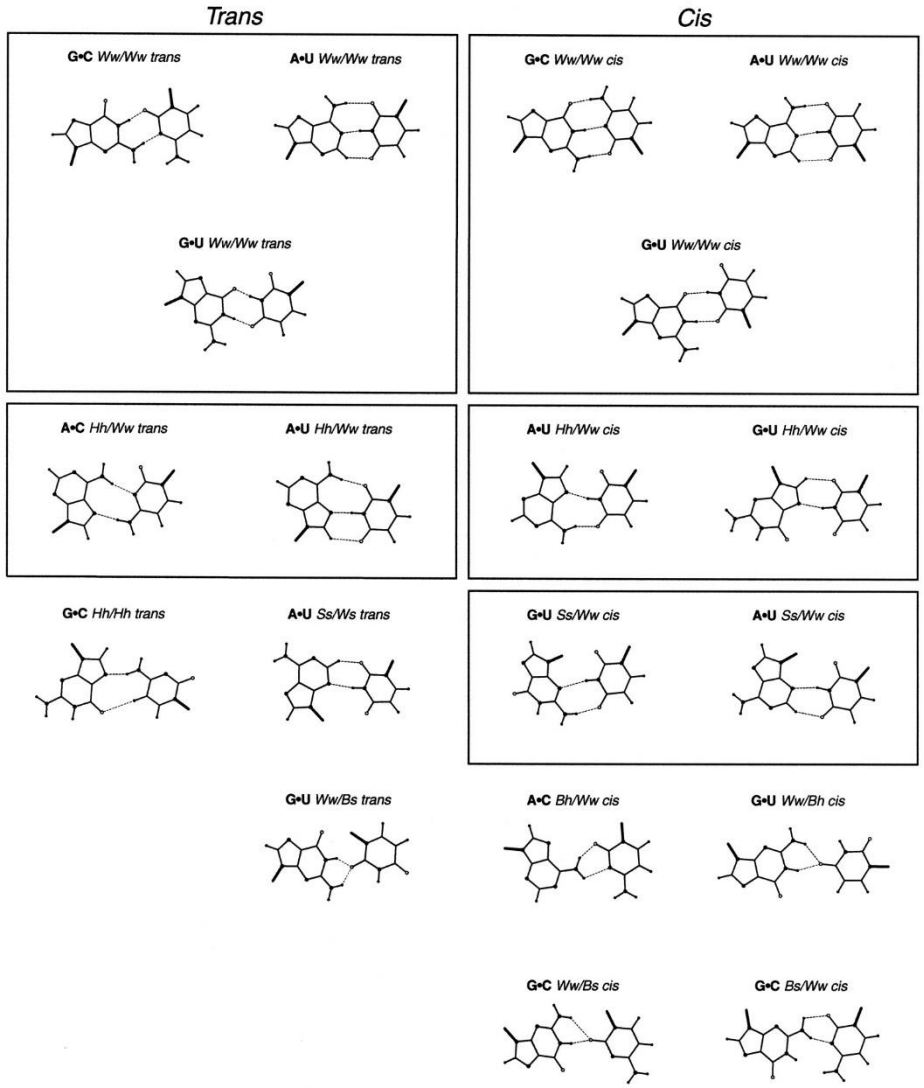


Figure 8.(Previous page and above) Two H-bond base pairing types found in HR-RNA-SET.

Lemieux S , Major F Nucl. Acids Res. 2002;30:4250-4263

Nucleic Acids Research

RNA canonical and non-canonical base pairing types: a recognition method and complete repertoire

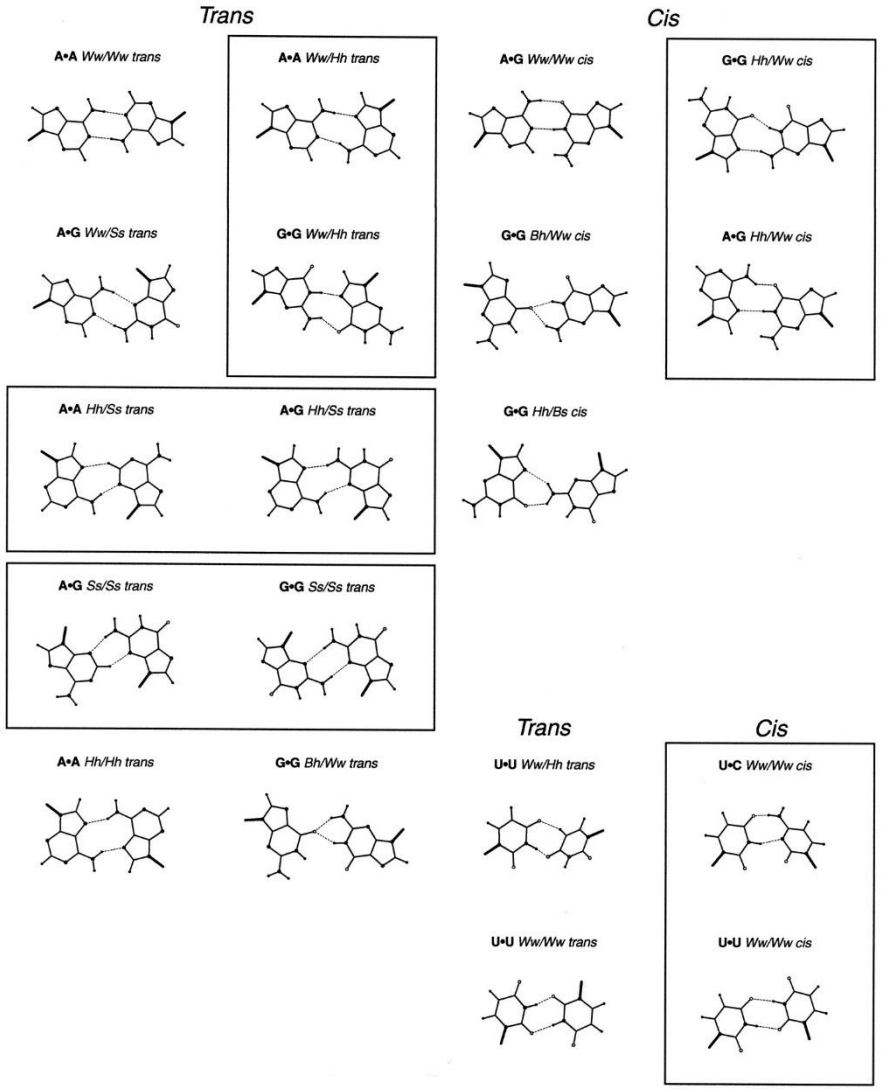
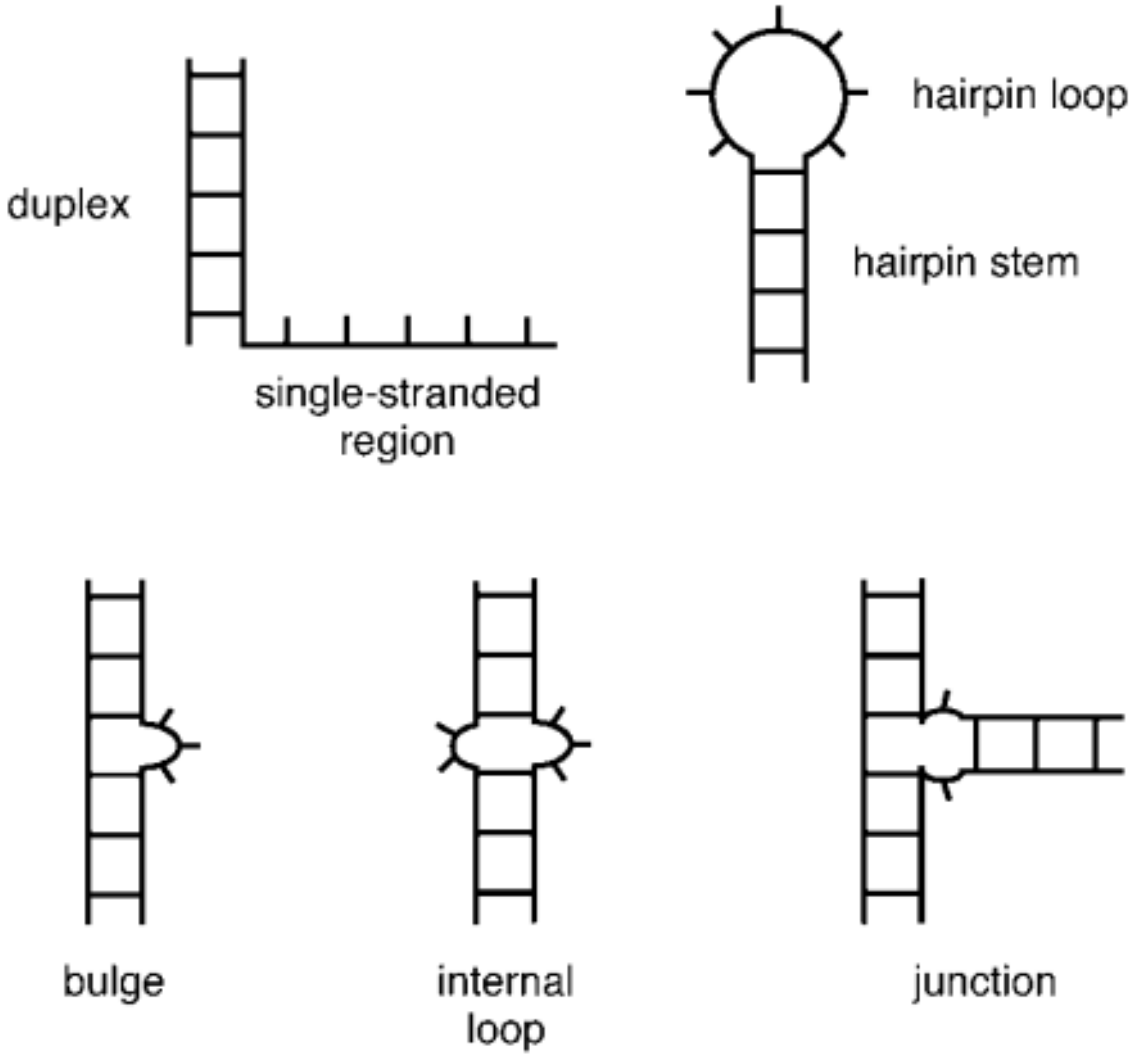


Figure 8.(Previous page and above) Two H-bond base pairing types found in HR-RNA-SET.

Lemieux S , Major F Nucl. Acids Res. 2002;30:4250-4263

Nucleic Acids Research

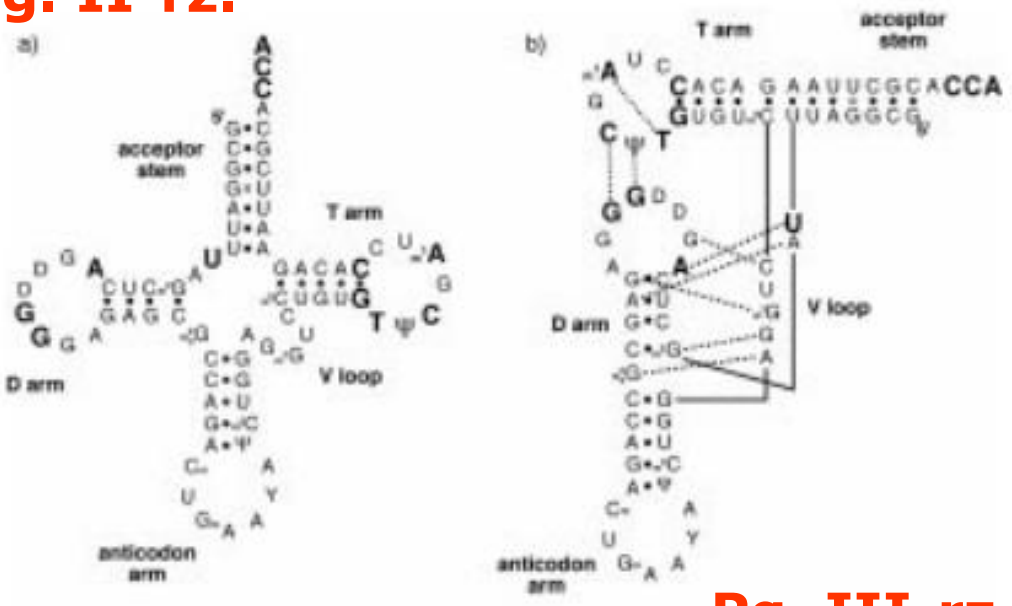
The most common elements of the secondary structures with Watson-Crick pairings



Robert T. Batey, Robert P. Rambo, and Jennifer A. Doudna Angew. Chem. Int. Ed. 1999

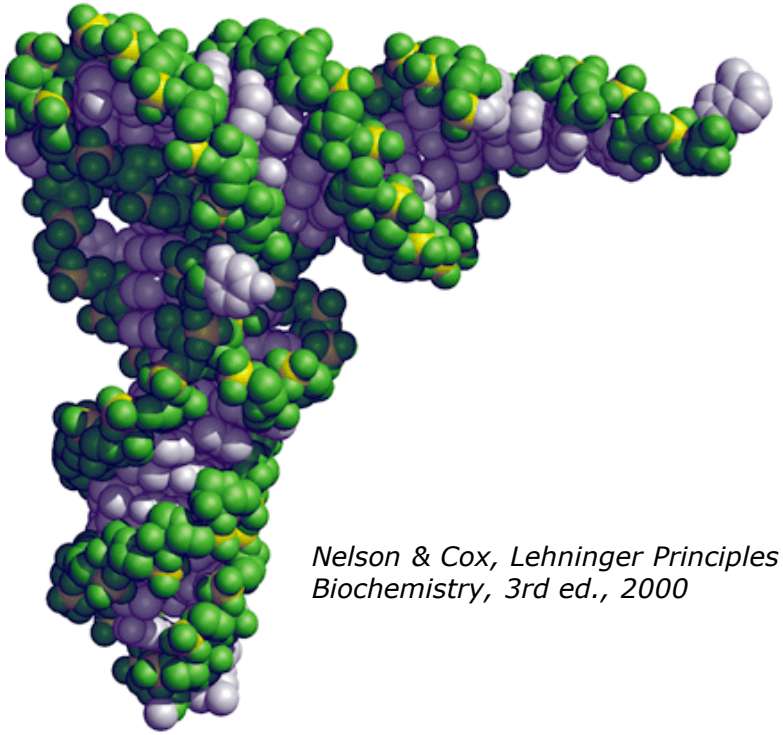
Example of structure: tRNA^{Phe}

Pg. II-rz.

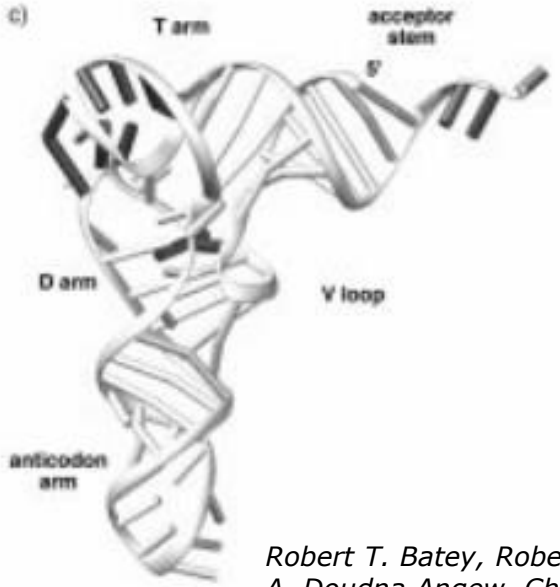


Fixed (invariant) nucleotide positions for class I tRNAs are in bold. They take part in biological functions or are responsible for organizing the architecture of the molecule

Pg. III-rz.



Nelson & Cox, Lehninger Principles of Biochemistry, 3rd ed., 2000



Robert T. Batey, Robert P. Rambo, and Jennifer A. Doudna Angew. Chem. Int. Ed. 1999

Tertiary structure: 3D interactions between distant elements of the structure

Interactions between helical motifs

- Coaxial stacking
- Adenosine platform
- Helical interactions through the 2'-hydroxyl group

Interactions between helical and unpaired motifs

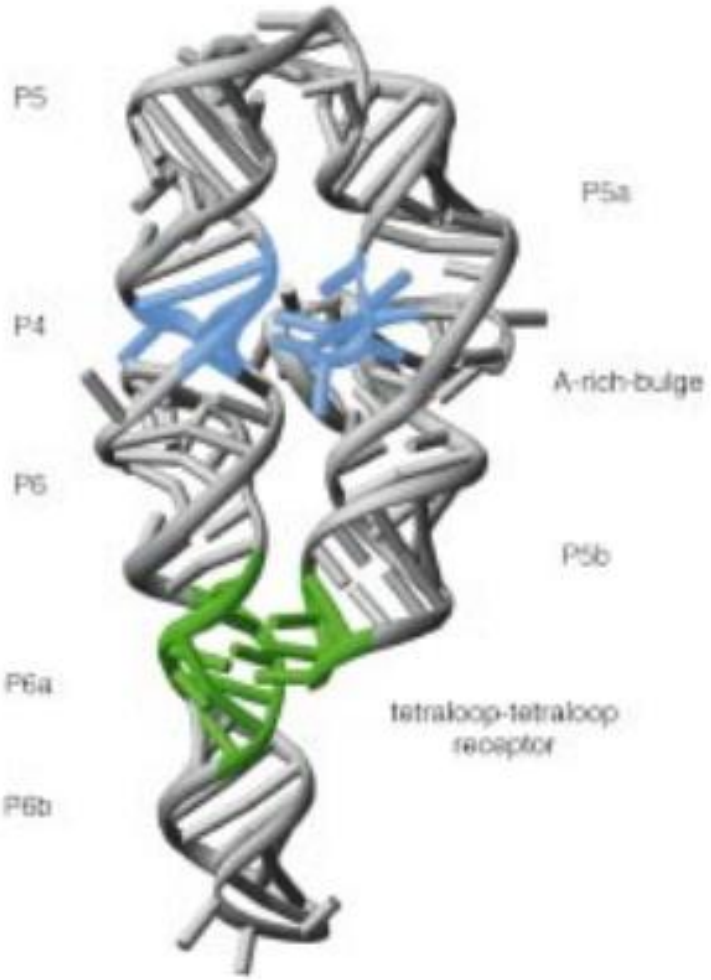
- Triplets and Triplexes of Bases
- "tetraloop" motif
- Core metal motif
- Ribose zipper motif

Tertiary interactions between unpaired bases

- Loop-to-loop interactions
- pseudoknots

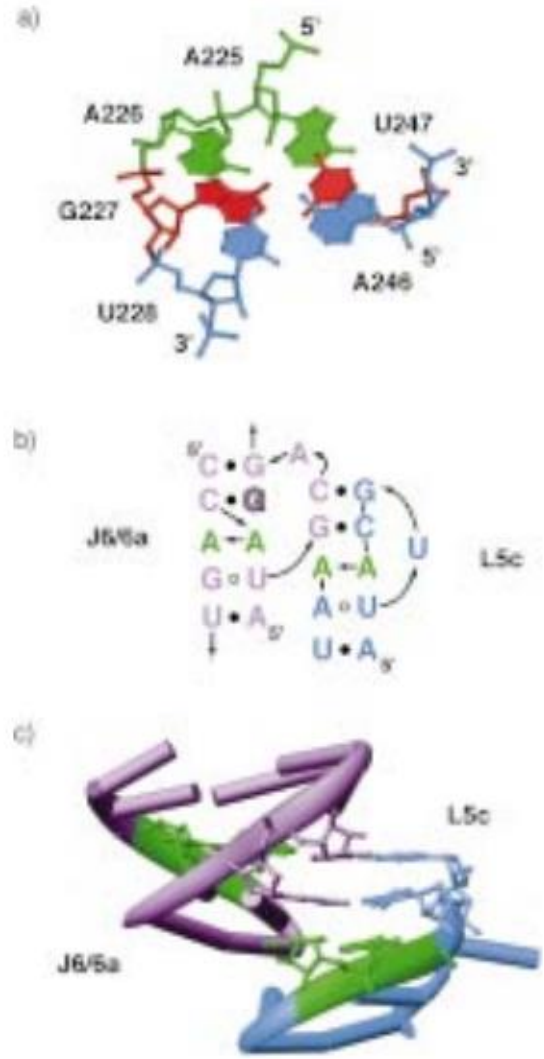
Interactions between helical motifs

Model of the P4-P6 domains of the Tetrachymena intron
(PDB accession number 1gid)

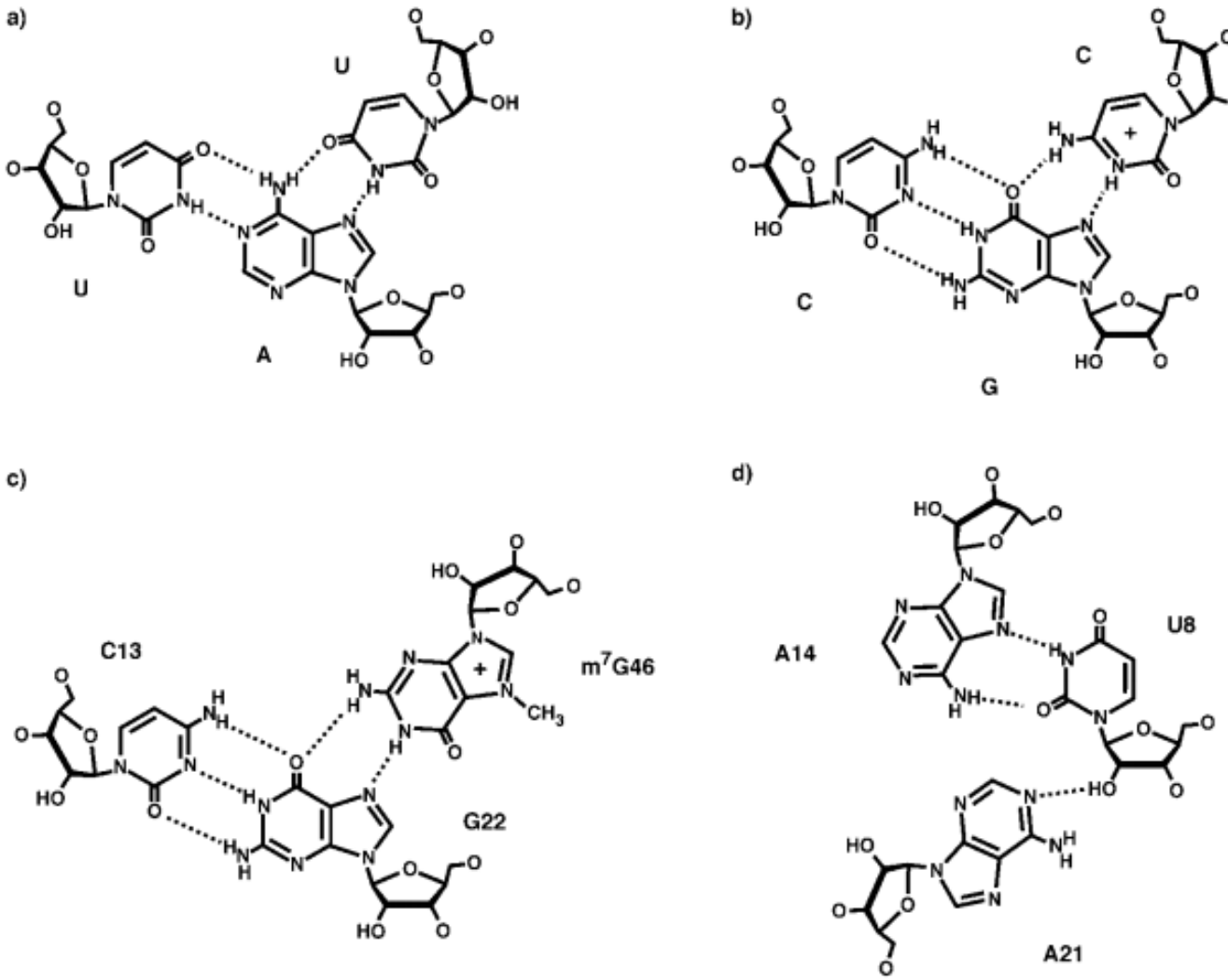


. Batey, Robert P.
and Jennifer A. Doudna
Chem. Int. Ed. 1999

Adenosine platform *Tetrachymena* intron

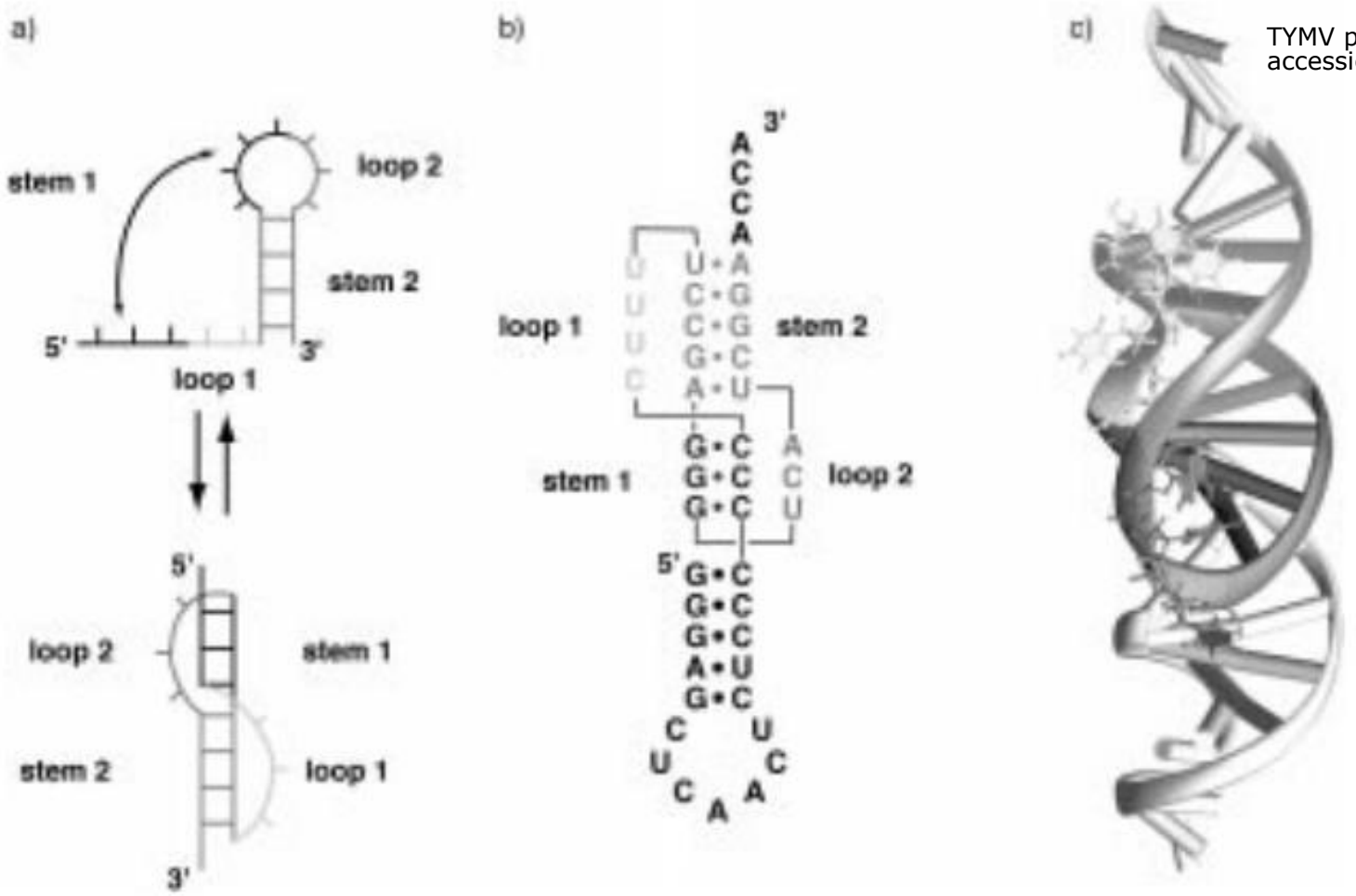


Triple base pairings in RNA



Robert T. Batey, Robert P. Rambo, and Jennifer A. Doudna Angew. Chem. Int. Ed. 1999

Pseudoknots

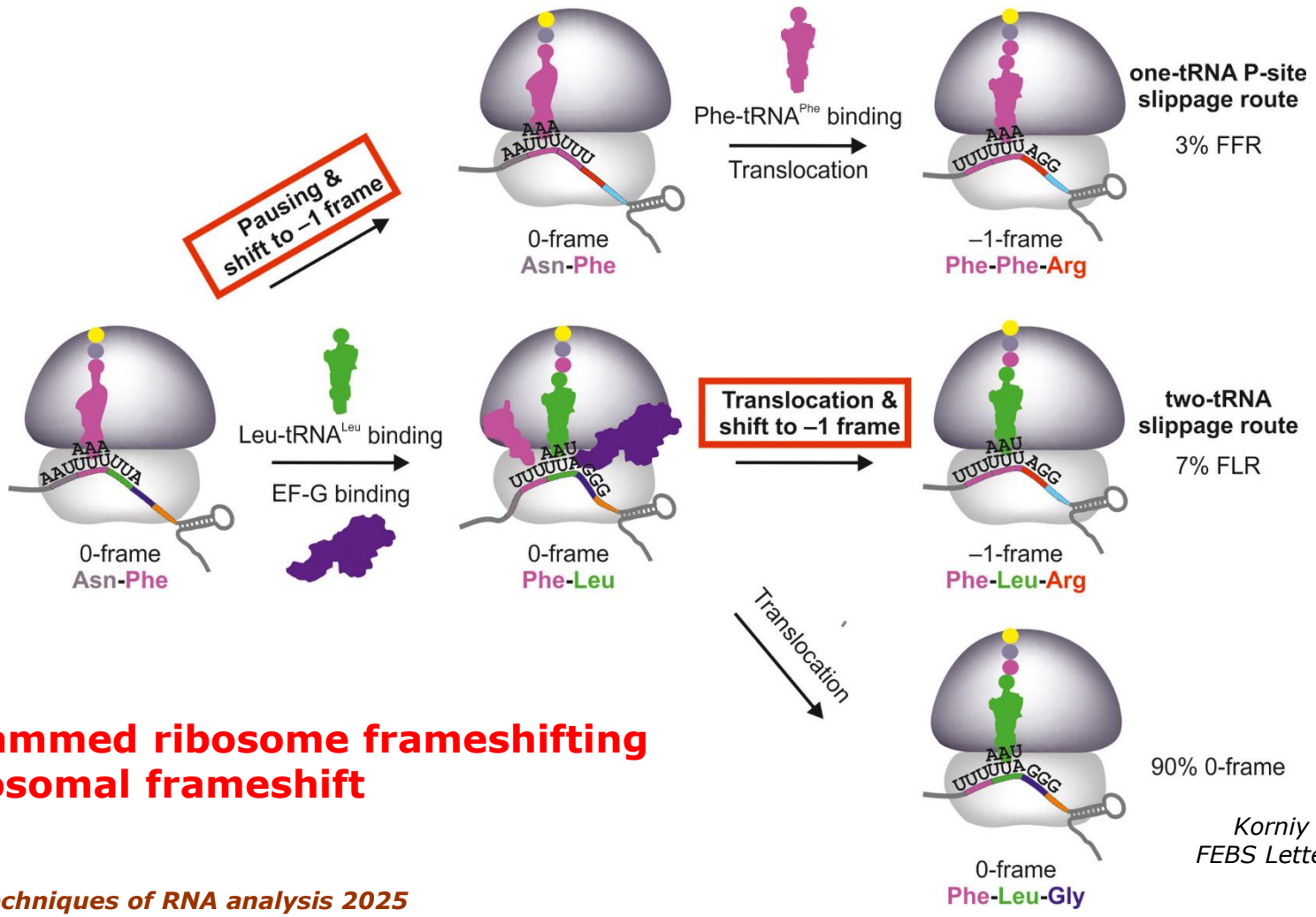


TYMV pseudoknot (PDB accession number 1a60)

Robert T. Batey, Robert P. Rambo, and Jennifer A. Doudna Angew. Chem. Int. Ed. 1999

Programmed shift in the reading frame of viral protein translation

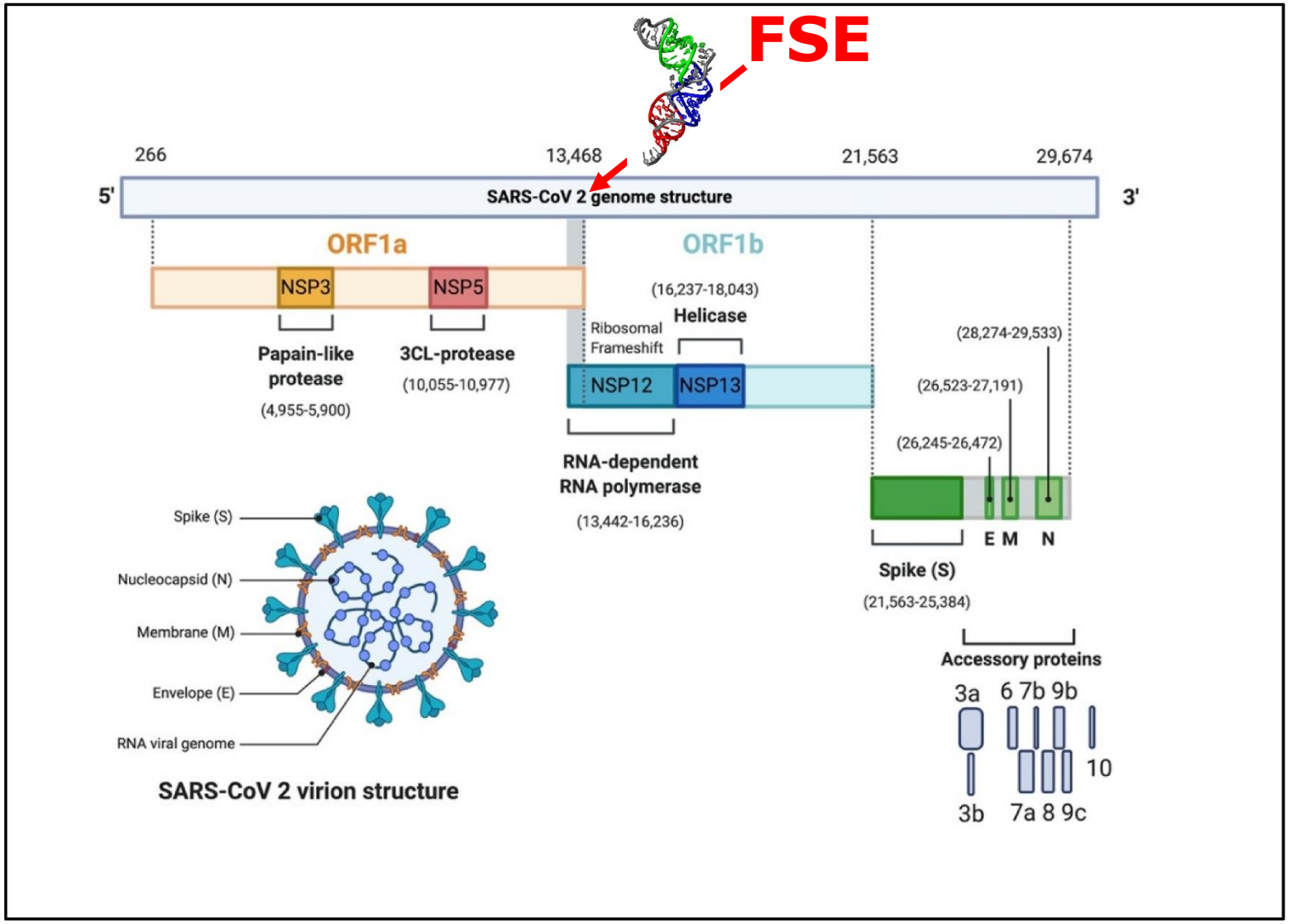
Kinetic mechanisms of FFR (upper) and FLR (lower) -1PRF pathways on the gag-pol mRNA of HIV-1



Programmed ribosome frameshifting = Ribosomal frameshift

Korniy et al, FEBS Letters 2019

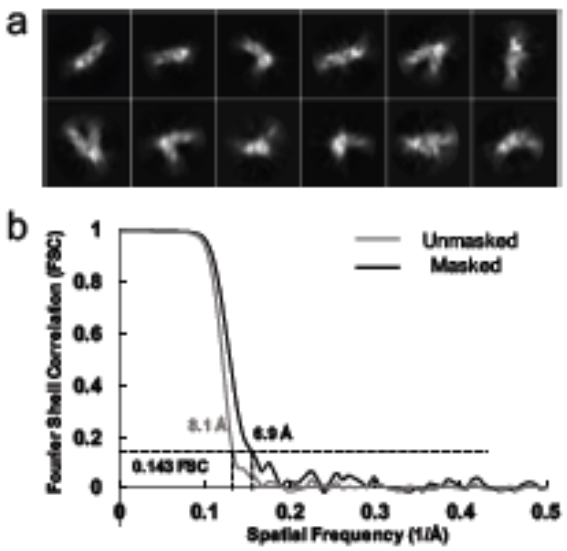
Translation reading frame shift element (FSE) - pseudoknot in RNA SARS_CoV2

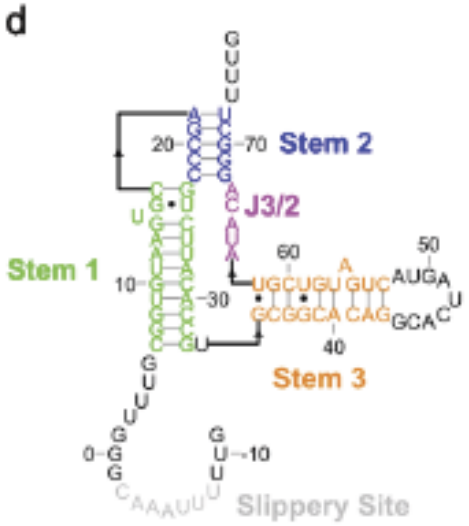


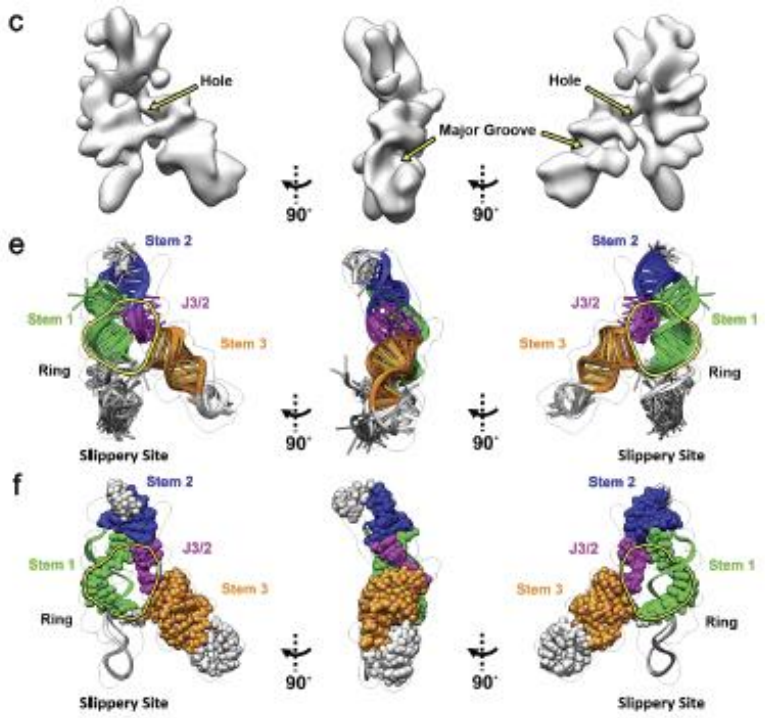
Lo'ai Alanagreh et al, Pathogens 2020

Translation reading frame shift element (FSE) - pseudoknot in RNA of SARS_CoV2 virus

Cryo-electron Microscopy and Exploratory Antisense Targeting of the 28-kDa Frameshift Stimulation Element from the SARS-CoV-2 RNA Genome







Zhang K. et al, (R. Das Lab), bioRxiv 2020, https://doi.org/10.1101/2020.07.18.209270

Potential target for drugs - inhibitors of FS: low-molecular-weight compounds or oligonucleotides!

RNA structure modeling

Predicting RNA structure:

- **Predicting the secondary structure: algorithms that calculate the map of base pairings for structures with the lowest free energy, such as Zucker's algorithm (mfold) - a *physical approach*.**

- **Aligment with a homologous sequence of known structure - *an evolutionary approach*.**

Studying the actual structure of RNA:

- **Biochemical analysis of the pairing map: nuclease digestion (degradation map) of single-stranded strucutres**

- **Structural studies, e.g., X-ray, Cryo-EM**

The best results yields data integration

Ribo-switches and RNA aptamers

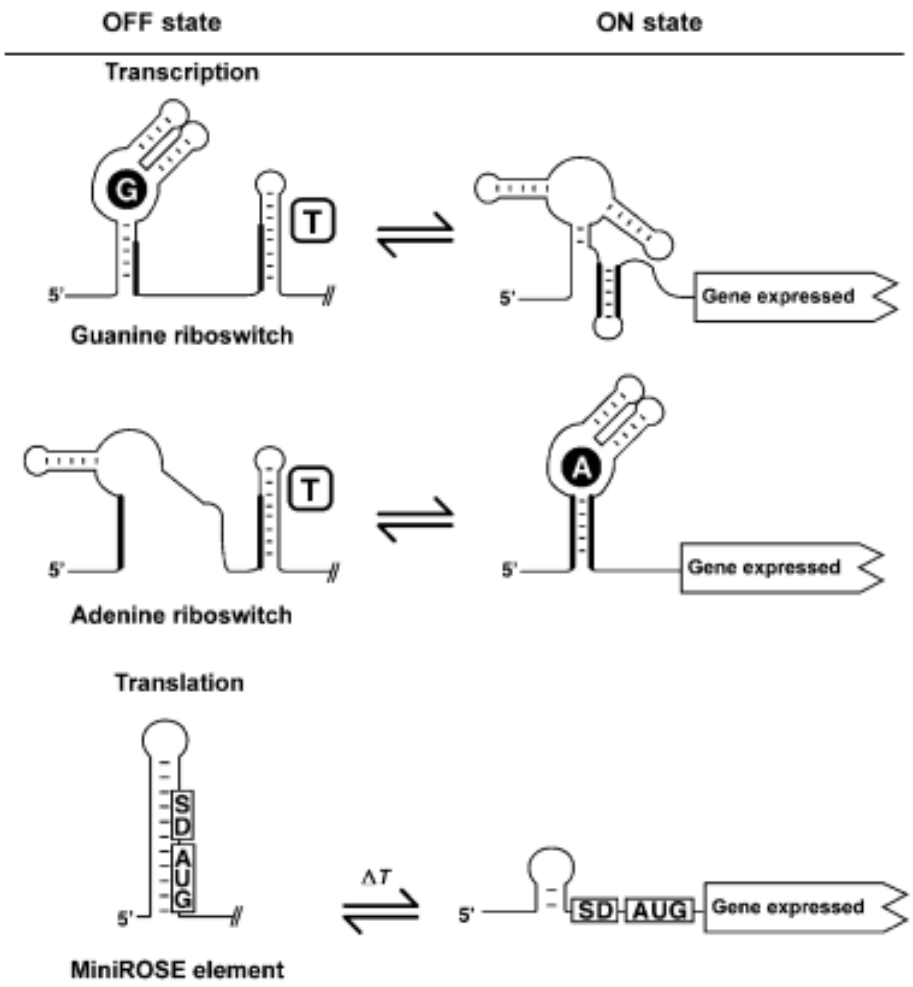
RNA aptamers:
structural elements that bind
low molecular weight ligands or proteins

Riboswitches

Sensor RNAs capable of binding low molecular weight ligands

- **Binding of low molecular weight ligand changes RNA structure**
- **Binding elements (aptamers) are most often found in UTRs**
- **Regulation of transcription**
- **Regulation of translation**
- **Response to environmental stimuli**
- **Originally detected in *Procaryota* and treated as a remnant of the "RNA world"**
- **Currently known examples in *Eucaryota***

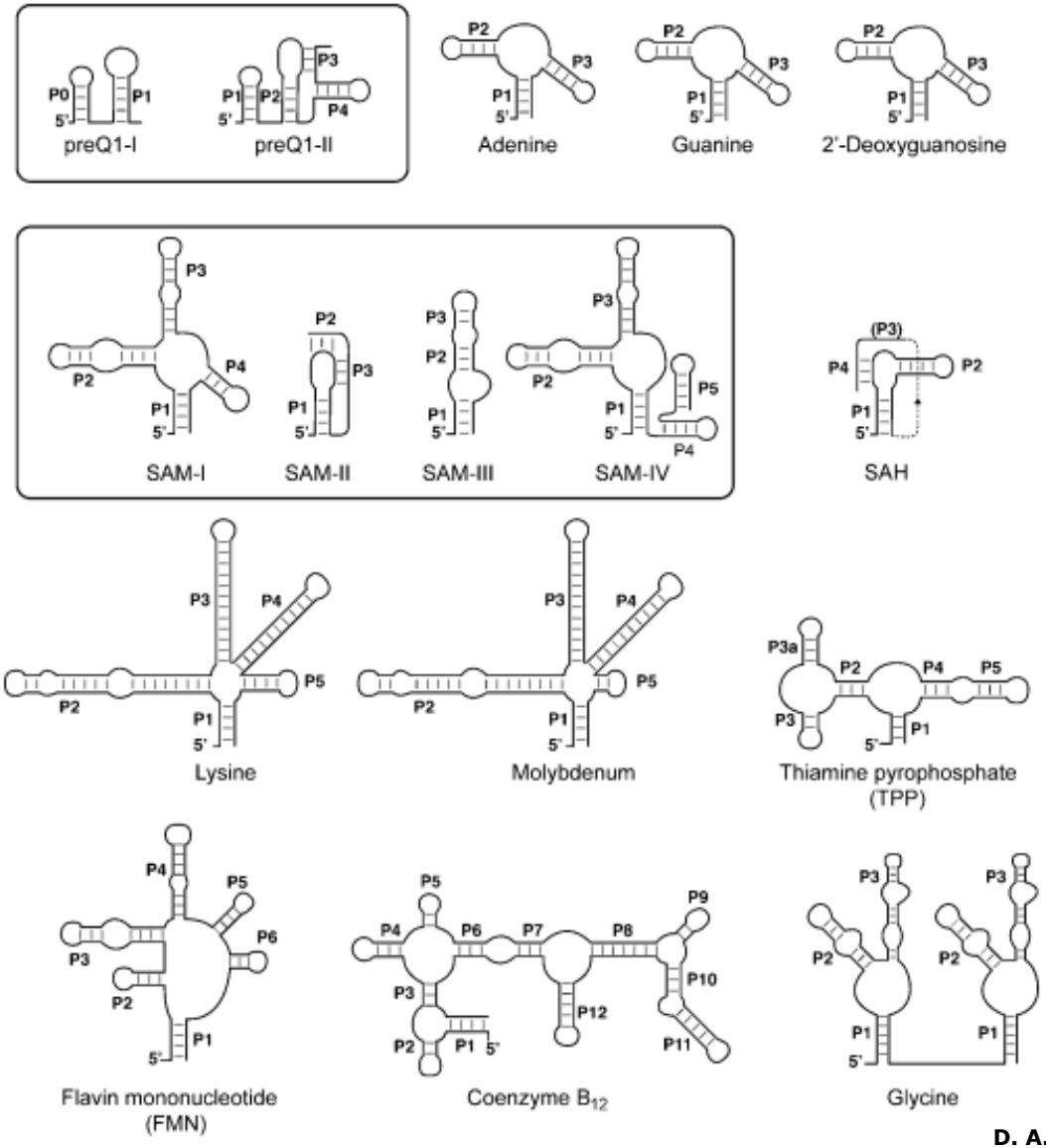
Examples of riboswitches



"RNA thermometer"

D. A. Lafontaine et al, Riboswitches as Promising Regulator, ChemBioChem 2009, 10, 400 - 41

Examples of aptamers that bind low molecular weight ligands



D. A. Lafontaine et al, Riboswitches as Promising Regulator, ChemBioChem 2009, 10, 400 - 41

Application in practice

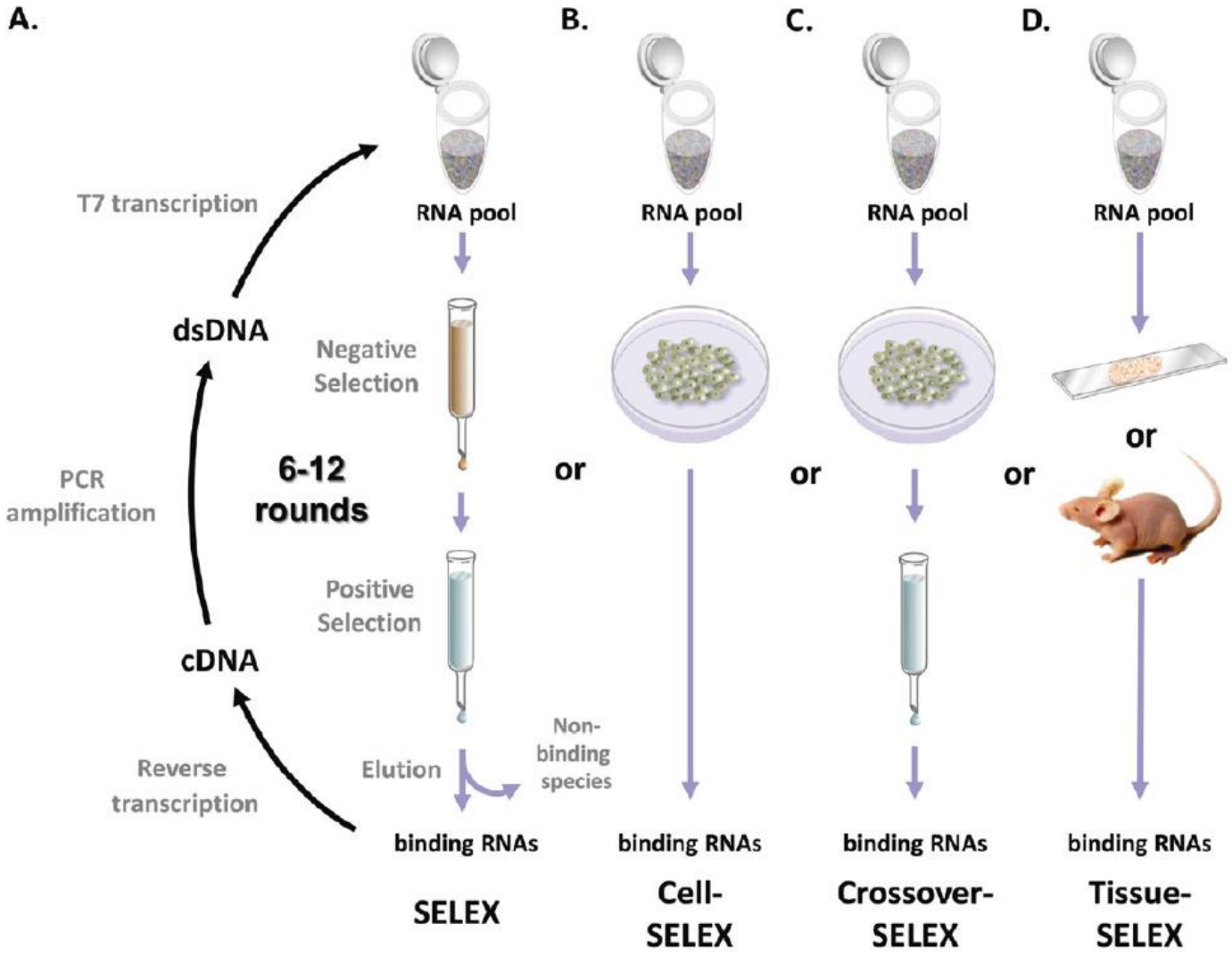
Medicine - instead of antibodies

- **Kd even an order of magnitude lower!**
- **Stability problem (sometimes an advantage)**
- **Mostly against cell receptors (e.g., anti-VEGF in AMD) or transcription factors (e.g., anti-AML1 in acute myeloid leukemia).**
- **Modifications: pegylation, substitution of modified bases**
- **First drugs available!**
- **Platelet tests similar to ELISA but with RNA aptamers instead of antibodies, e.g., virus diagnostics.**

Biosensors

- ***In vivo* systems with reporter genes**
- ***In vitro* translation**

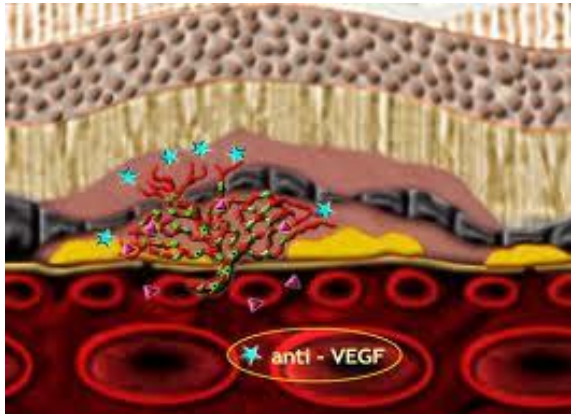
New variations of the SELEX technique



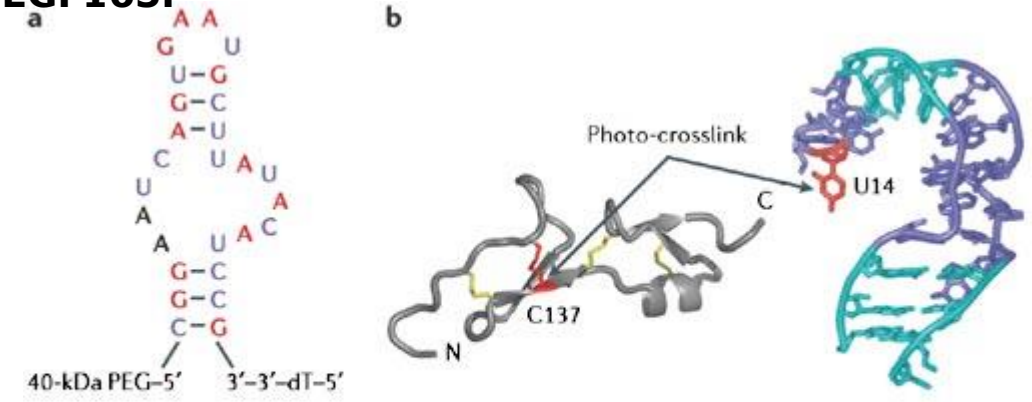
Germer et al, Int J Biochem Mol Biol (2013)

Pegaptanib - the first drug based on an RNA aptamer

Age-related macular degeneration (AMD) - proliferation and leakage of blood vessels stimulated by VEGF (vascular endothelial growth factor).



First aptamer approved as a drug in the US (2004): Pegaptanib (Macugen). A pegylated anti-VEGF aptamer, inhibits angiogenesis by binding the extracellular form of the factor VEGF165.



Eugene et al, Nature Reviews Drug Discovery (2006)

RNA structure mapping using nucleases

Mapping RNA structure with RNases

- **Factors that cut RNA within single-stranded regions - double-stranded structurses are protected**

- **Enzymatic or chemical RNases:**

- U2, T1, V1, RnI, ChSI nucleases
- Heavy metal ions: Pb^{2+}
- alkalain degradation

- **Specificity to sequences (to bonds between specific bases)**

Mapping RNA structure with RNases

- ***In vitro* experiments**

-Cons: non-physiological RNA concentrations, not always physiological salt concentrations in buffers

- Pros: simple to make, possible mapping of large molecules, easy addition of low molecular weight ligands

- **Schematic of the experiment:**

1. *In vitro* transcription in the presence of e.g.^{32} P-UTP, or labeling of cold transcripts by kinase (5'end) or ligation (T4 RNA ligase).
2. Purification of labeled RNA.
3. Incubation in buffer with Rnases.
4. Polyacrylamide gel separation and autoradiography.
5. Analysis of results, coupled with RNA structure modeling algorithms (mfold).

Mapping the 5'UTR structure of the *A. nidulans* arginase mRNA

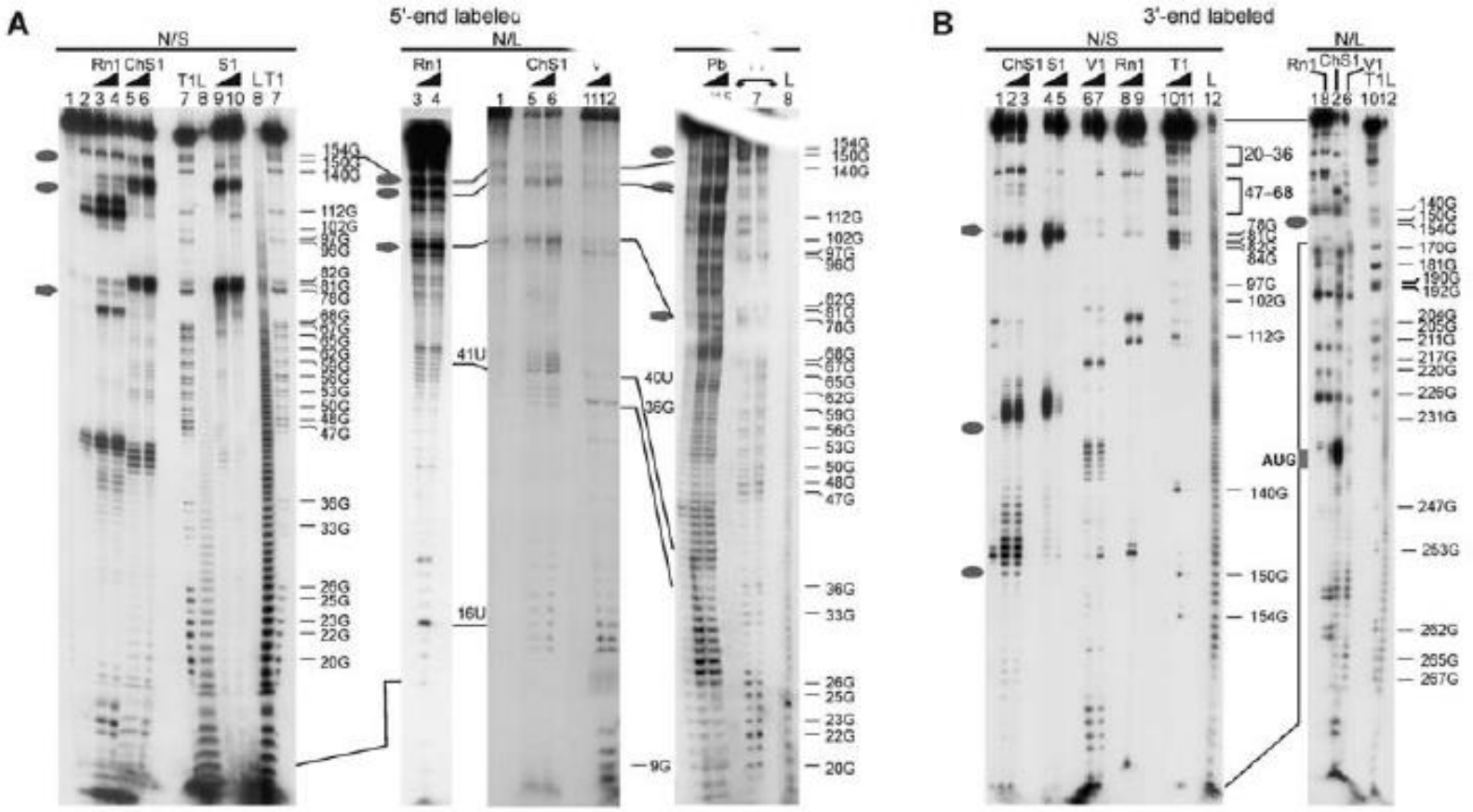
Biol. Chem., Vol. 388, pp. 135–144, February 2007 • Copyright © by Walter de Gruyter • Berlin • New York. DOI 10.1515/BC.2007.015

L-Arginine influences the structure and function of arginase mRNA in *Aspergillus nidulans*

Piotr Borsuk[1,2], Anna Przykorska[2], Karina Blachnio[2], Michal Koper[1,2], Jerzy M. Pawlowicz[2], Malgorzata Pekala[1] and Piotr Weglenski[1,2,*]

- **The 5'UTR of *agaA* mRNA has a complex potential secondary structure.**
- **L-arginine binds to the 5'UTR of arginase mRNA.**
- **L-arginine specifically alters the 5'UTR structure of arginase mRNA *in vitro*, D-arginine does not.**
- **There is an intron in the 5'UTR, the position of which suggests the possibility of its alternative splicing: 19 nt downstream of the experimentally known 3' splice site is a second conserved signal sequence for the 3' intron splice site.**

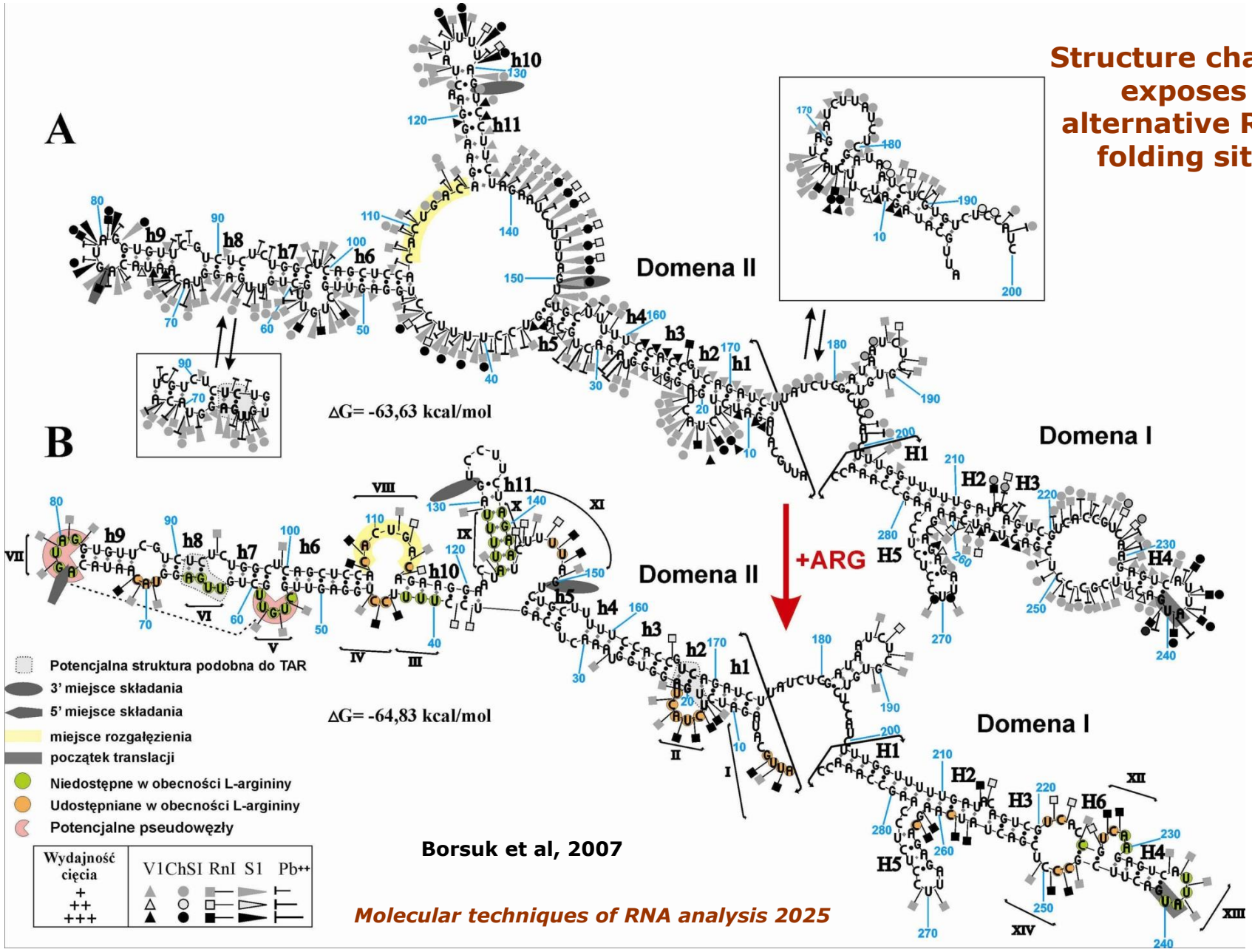
Mapping the 5'UTR structure of the *A. nidulans* arginase mRNA



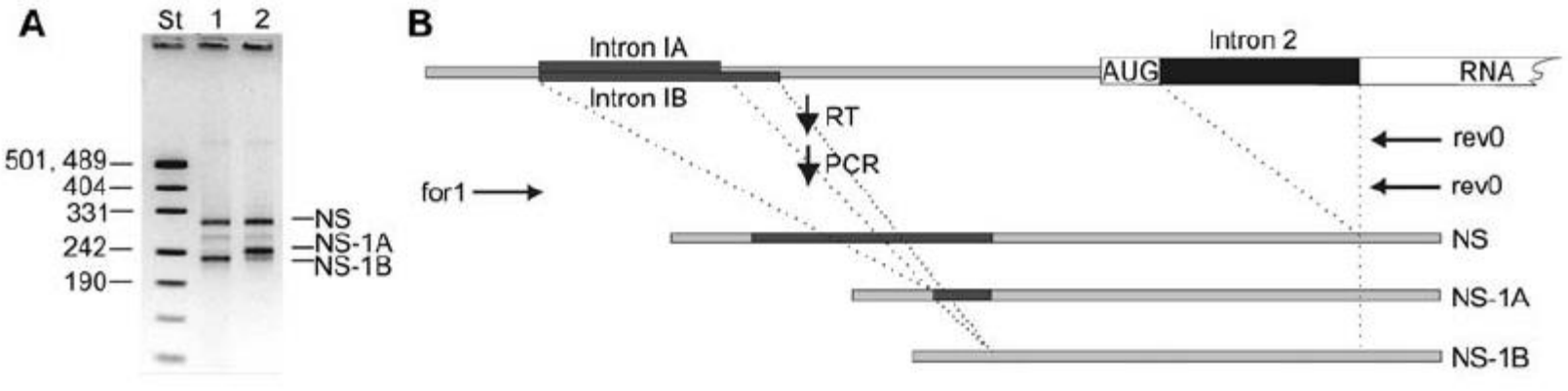
Borsuk et al, 2007

Figure 3 Probing the structure of *agaA* mRNA 5′-UTR fragments.
In vitro transcribed N/L or N/S 5′-UTR fragments were labelled at the 5′- (A) or 3′- (B) ends and subjected to enzymatic and chemical probing. (A) Lane 1, incubation control in TMK buffer containing 1.5 mM $ZnCl_2$. Lane 2, incubation control in TMK buffer. Lanes 3 and 4, 0.2 and 0.3 U of nuclease RnI, respectively. Lanes 5 and 6, 0.2 and 0.3 U of nuclease ChSI, respectively. Lane 7, RNase T1 ladder. Lane 8, alkaline ladder. Lanes 9 and 10, 12.5 and 25 U of nuclease S1, respectively. Lanes 11 and 12, 0.1 and 0.2 U of RNase V1, respectively. Lane 13, incubation control in HEPES buffer. Lanes 14 and 15, 8 and 16 mM lead acetate, respectively. (B) Lane 1, incubation control in TMK buffer. Lanes 2 and 3, 0.2 and 0.3 U of nuclease ChSI, respectively. Lanes 4 and 5, 12.5 and 25 U of nuclease S1, respectively. Lanes 6 and 7, 0.1 and 0.2 U of RNase V1, respectively. Lanes 8 and 9, 0.2 and 0.3 U of nuclease RnI, respectively. Lanes 10 and 11, RNase T1 ladder. Lane 12, alkaline ladder.

Arginase 5'UTR changes structure upon the presence of L-arginine



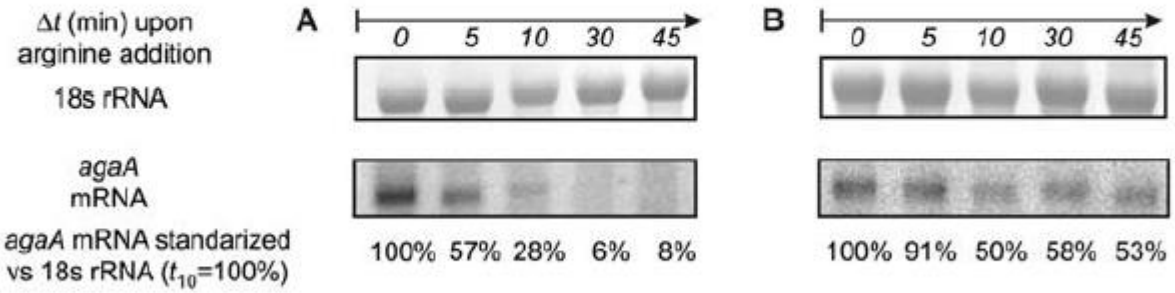
Mapping RNA structure with RNases



RT-PCR of the *agaA* 5′-UTR.

P. Borsuk et al.

Alternative folding of the 5'UTR of *agaA* mRNA affects transcript stability

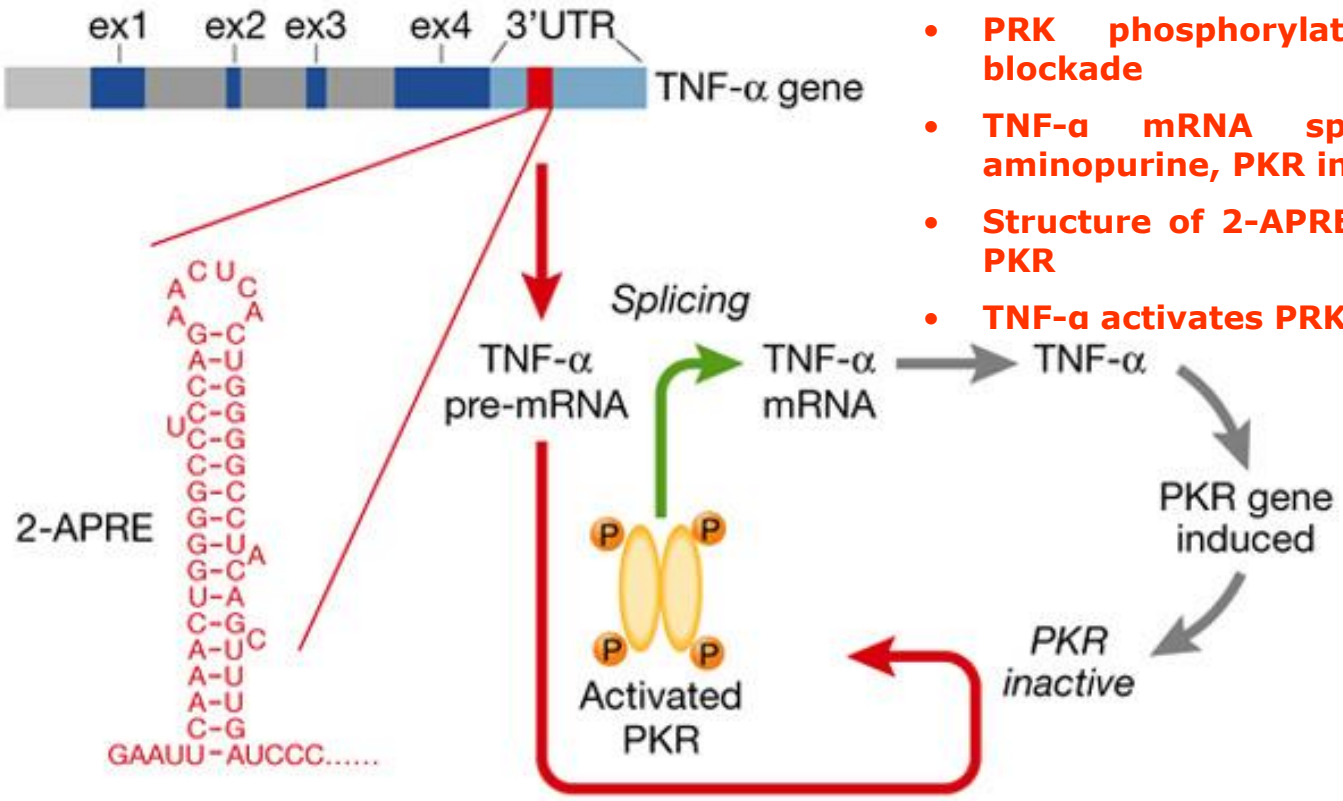


Decay of arginase mRNA in the absence (A) and presence (B) of L-arginine (2 mM) upon transcription inhibition.

Examples of riboswitches in eukaryotes

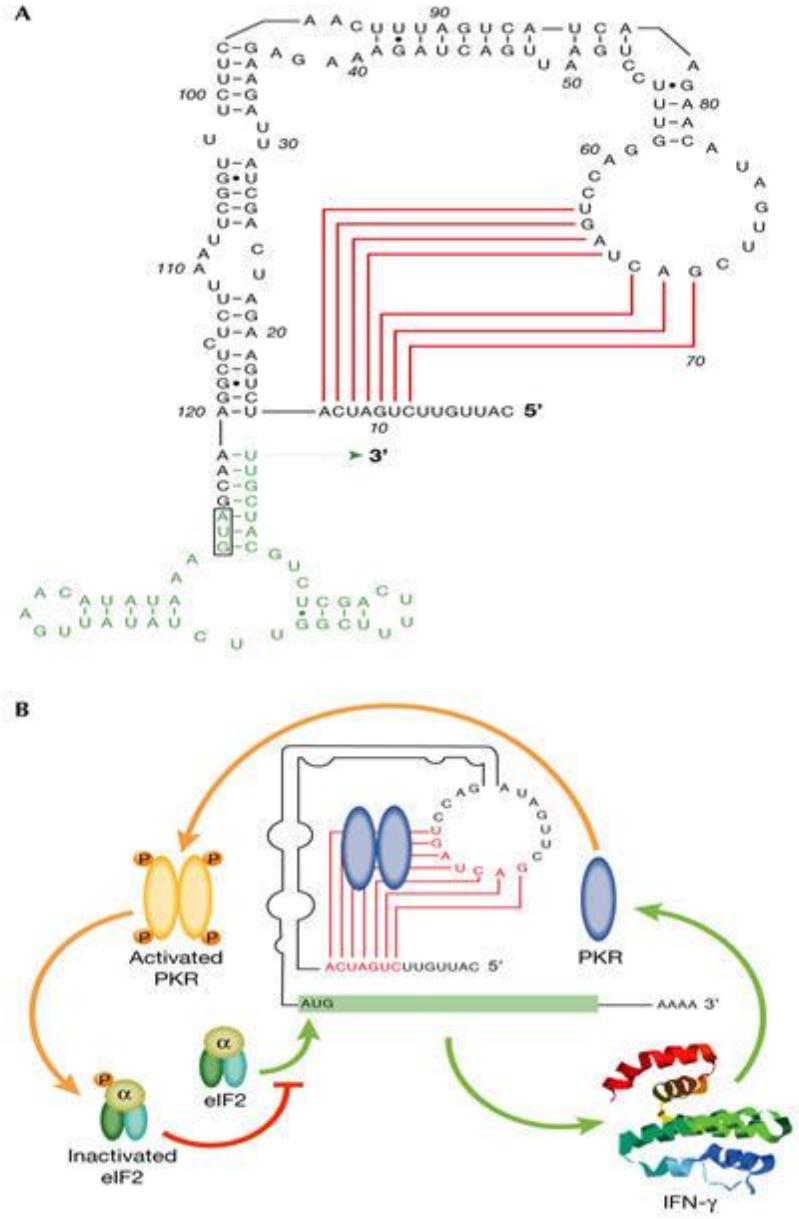
Positive feedback loop in TNF-α mRNA folding

- PKR - RNA-dependent protein kinase, a mediator of the immune response
- Self-activation of PKR after dsRNA binding, e.g., viral
- PRK phosphorylates eIF2 - translation blockade
- TNF-α mRNA splicing sensitive to 2-aminopurine, PKR inhibitor
- Structure of 2-APRE in TNF-α mRNA activates PKR
- TNF-α activates PRK gene transcription



Raymond Kaempfer, RNA sensors: novel regulators of gene expression
EMBO reports VOL 4 | NO 11 | 2003

Negative feedback loop in IFN-γ mRNA translation



- **H-type pseudoknot in the 5'UTR of IFN-γ mRNA activates PKR**
- **PKR through eIF2 phosphorylation inhibits IFN-γ mRNA translation**
- **IFN-γ enhances PKR expression**

Raymond Kaempfer, RNA sensors: novel regulators of gene expression, EMBO reports VOL 4 | NO 11 | 2003

Thank you for your attention