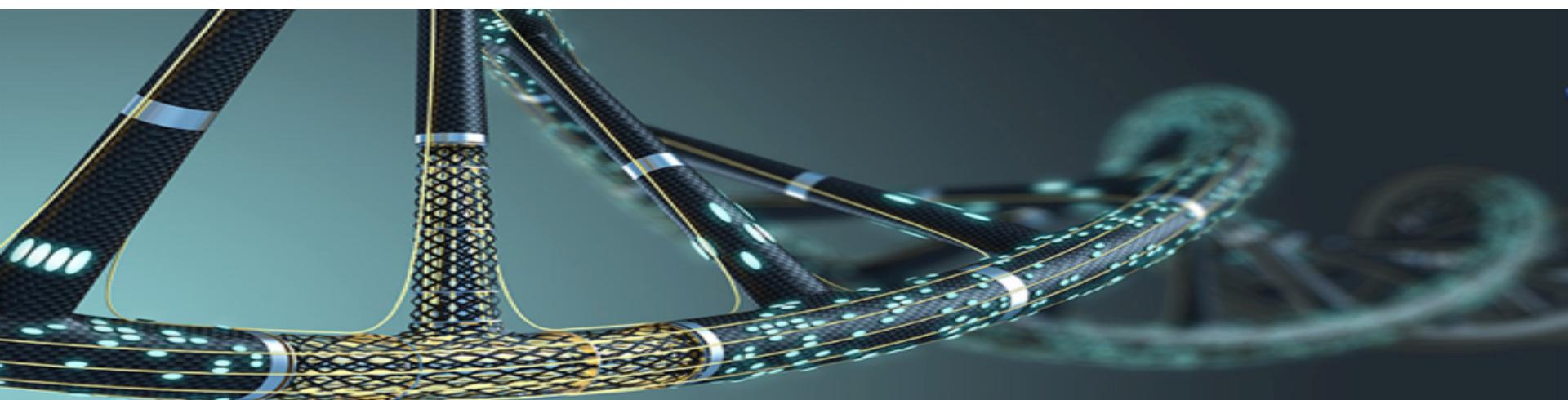




Майнор по биоинформатике

Лекция 15

Мария Попцова



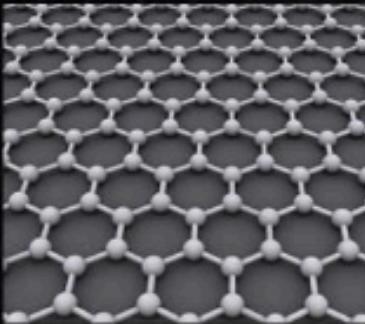
Вторичные структуры ДНК.
Квадруплексы, триплексы,
крестообразные структуры.

the **DAWN** *of the* **SINGULARITY**

The following predictions were made by Ray Kurzweil
in his book The Singularity is Near.

Kurzweil, now the Director of Engineering at Google,
had made 147 predictions since the 1990's and has maintained
an astonishing 86% accuracy rate.

2019-2029



Three-dimensional nanotube lattices are the dominant computing substrate



The digital world makes paper books and documents almost completely obsolete



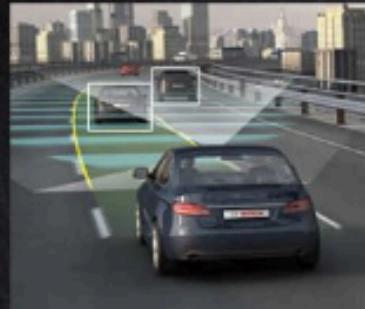
Total power of all computers is comparable to total brainpower of the human race



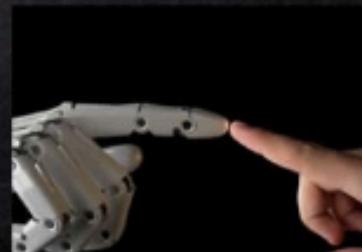
Computers are embedded everywhere in our environment (furniture, jewelry, walls, etc)



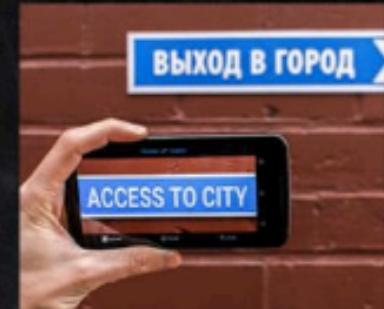
Creative AI is now capable of making complex art and music



Autonomous vehicles now dominate our roads



Humans begin to develop deep relationships with AI

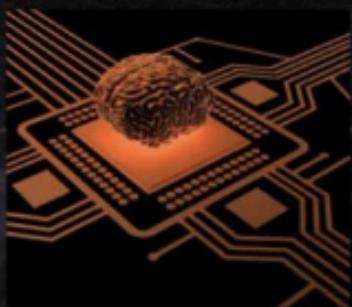


Language translation machines are now routinely used in conversation

2029-2039



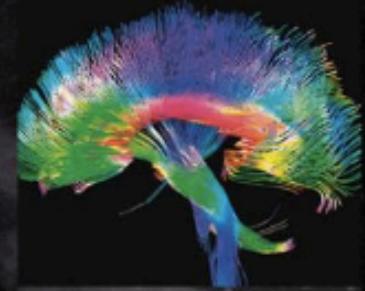
The manufacturing, agricultural and transportation sectors of the economy are almost entirely automated



Computers are now capable of autonomously learning and creating new knowledge



A \$1,000 USD personal computer is now 1,000x more powerful than the human brain



Advanced brain mapping leads to hundreds of distinct subregions in the brain being identified



Artificial intelligences claim to be conscious and openly petition for recognition of this fact



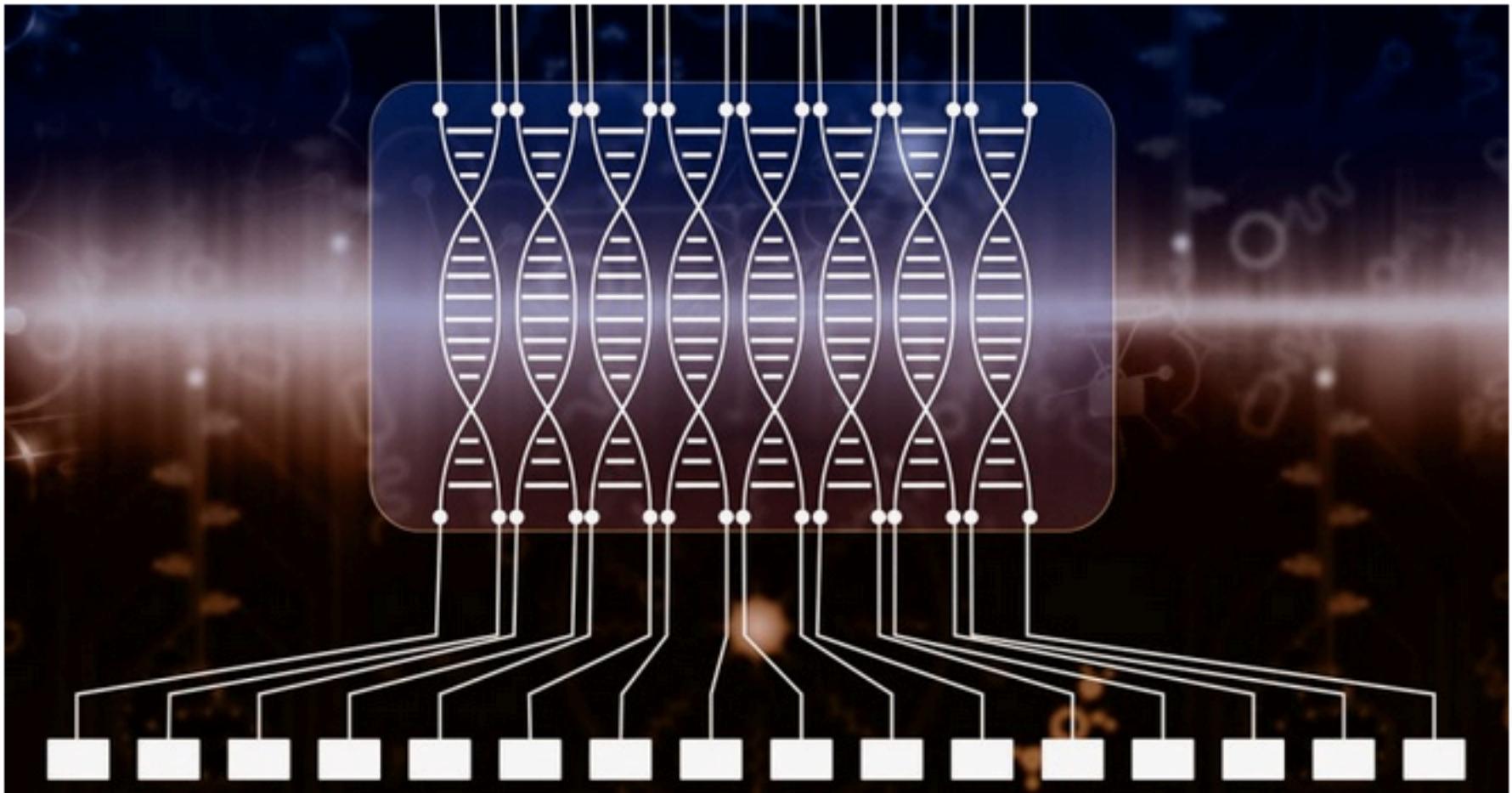
VR eyeglasses and headphones are replaced with computer implants

- 2036 – Using programming as an approach to biology, humans will program cells to cure diseases.
- 2037 – Gigantic leap in understanding mysteries of human brain. Some of the algorithms **will be deciphered and included into the neural networks of computers**.

My additions: Nobel prize in physiology and medicine

- 203X – for discovery of algorithms of genome functioning
- 203X – for cell programing

Algorithms are encoded in the genome

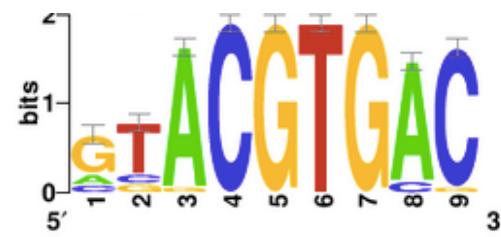
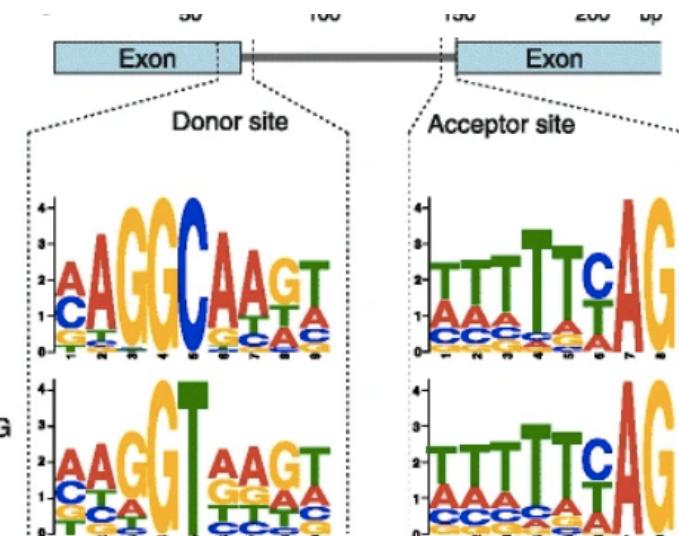


Layers of genomic encoding

- DNA sequence

DNA-seq, RNA-seq,
ChIP-seq

- Encoding into sequence: genes, exons, introns, promoters (some), enhancers (unclear), transcription factor binding sites



	U	C	A	G	
U	UUU Phe UUC Phe	UCU Ser UCC Ser	UAU Tyr UAC Tyr	UGU Cys UGC Cys	U
C	UUA Leu UUG Leu	UCA Ser UCG Ser	UAA TER UAG TER	UGA TER UGG Trp	C
A	CUU Leu CUC Leu CUA Leu CUG Leu	CCU Pro CCC Pro CCA Pro CCG Pro	CAU His CAC His CAA Gln CAG Gln	CGU Arg CGC Arg CGA Arg CGG Arg	A
G	AUU Ile AUC Ile AUA Ile AUG Met	ACU Thr ACC Thr ACA Thr ACG Thr	AAU Asn AAC Asn AAA Lys AAG Lys	AGU Ser AGC Ser AGA Arg AGG Arg	G
	GUU Val GUC Val GUA Val GUG Val	GCU Ala GCC Ala GCA Ala GCG Ala	GAU Asp GAC Asp GAA Glu GAG Glu	GGU Gly GGC Gly GGA Gly GGG Gly	U C A G

Legend:

- Hidrophobic - Imino (purple)
- Hidrophobic - Aromatic (red)
- Polar - Acid (blue)
- Polar - Basic (green)
- Polar - Neutral (yellow)

First Layer of Genome Annotation

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

chr11:128,805,270-128,807,789 2,520 bp.

enter position, gene symbol, HGVS or search terms

go

See us FREE @ ASHG Wed 7



Haplotypes to GRCh37 Reference Sequence
Patches to GRCh37 Reference Sequence
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)

RefSeq gene predictions from NCBI

Human RNA Editing from the Database of RNA Editing

Gene Expression in 53 tissues from GTEx RNA-seq of 8555 samples (578 donors)

H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE

DNASEI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)

Transcription Factor ChIP-seq (161 factors) from ENCODE with Factorbook Motifs

100 vertebrates Basewise Conservation by PhyloP

Multiz Alignments of 100 Vertebrates

Simple Nucleotide Polymorphisms (dbSNP 150) Found in >= 1% of Samples

Database of Genomic Variants: Structural Variation (CNV, Inversion, In/dele)

Repeating Elements by RepeatMasker

Epigenetics

Levels of DNA Packaging in Eukaryotes

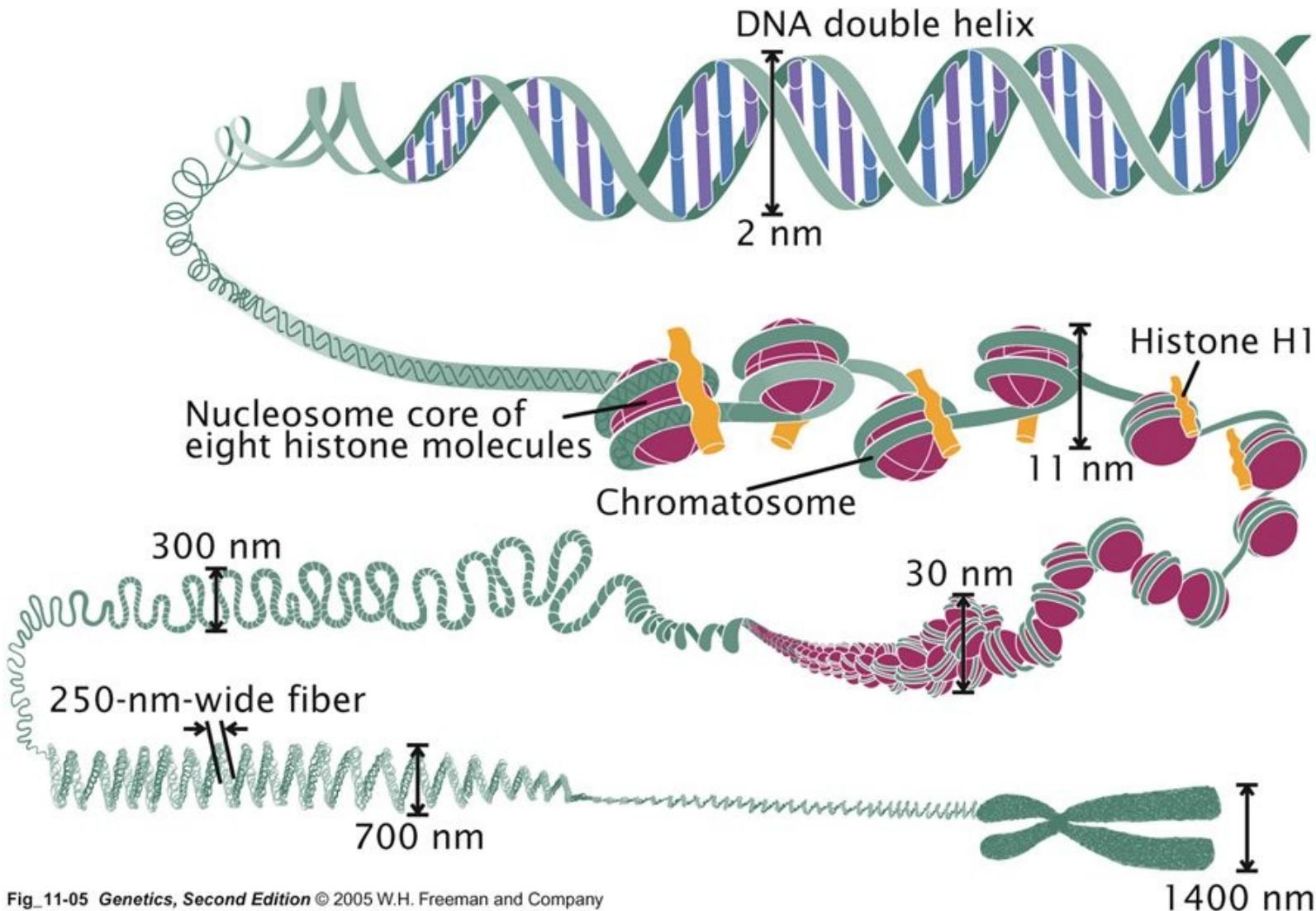
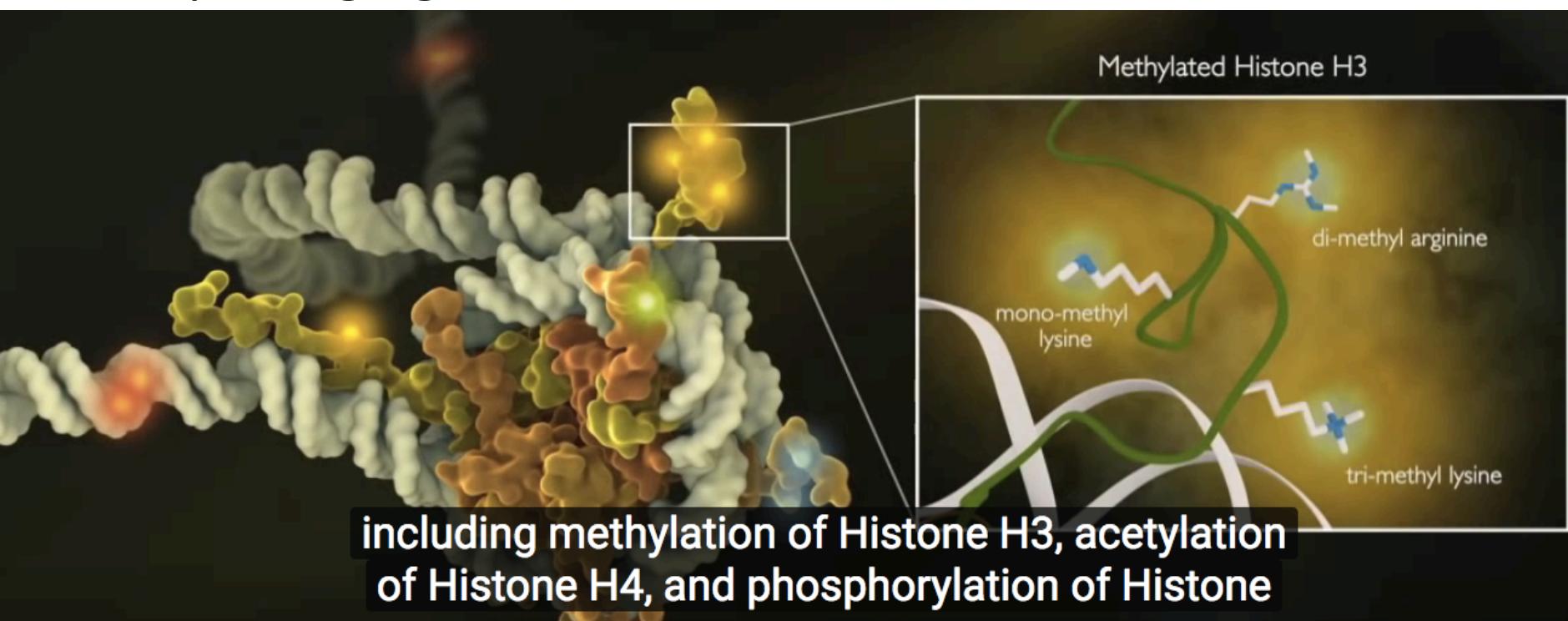


Fig. 11-05 *Genetics, Second Edition* © 2005 W.H. Freeman and Company

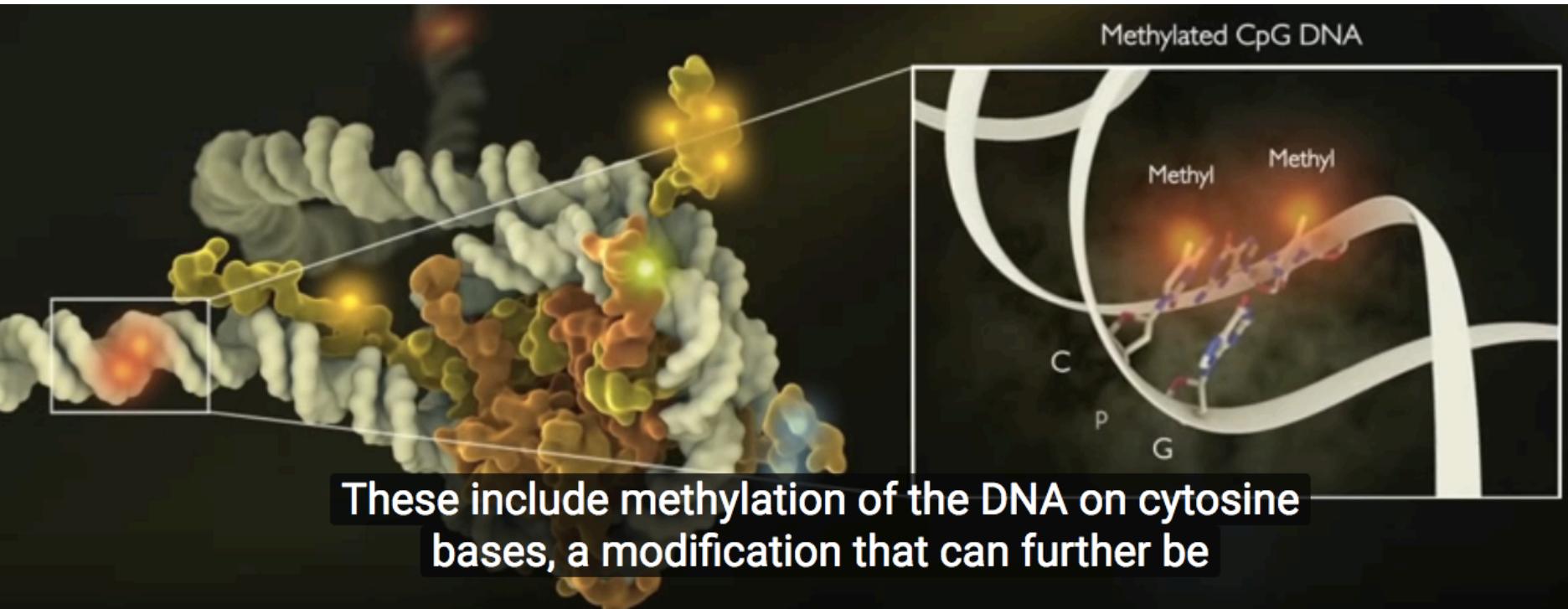
Layers of genomic encoding

- DNA marks: Epigenomics
 - Methylation, histone modifications, chromatine packaging

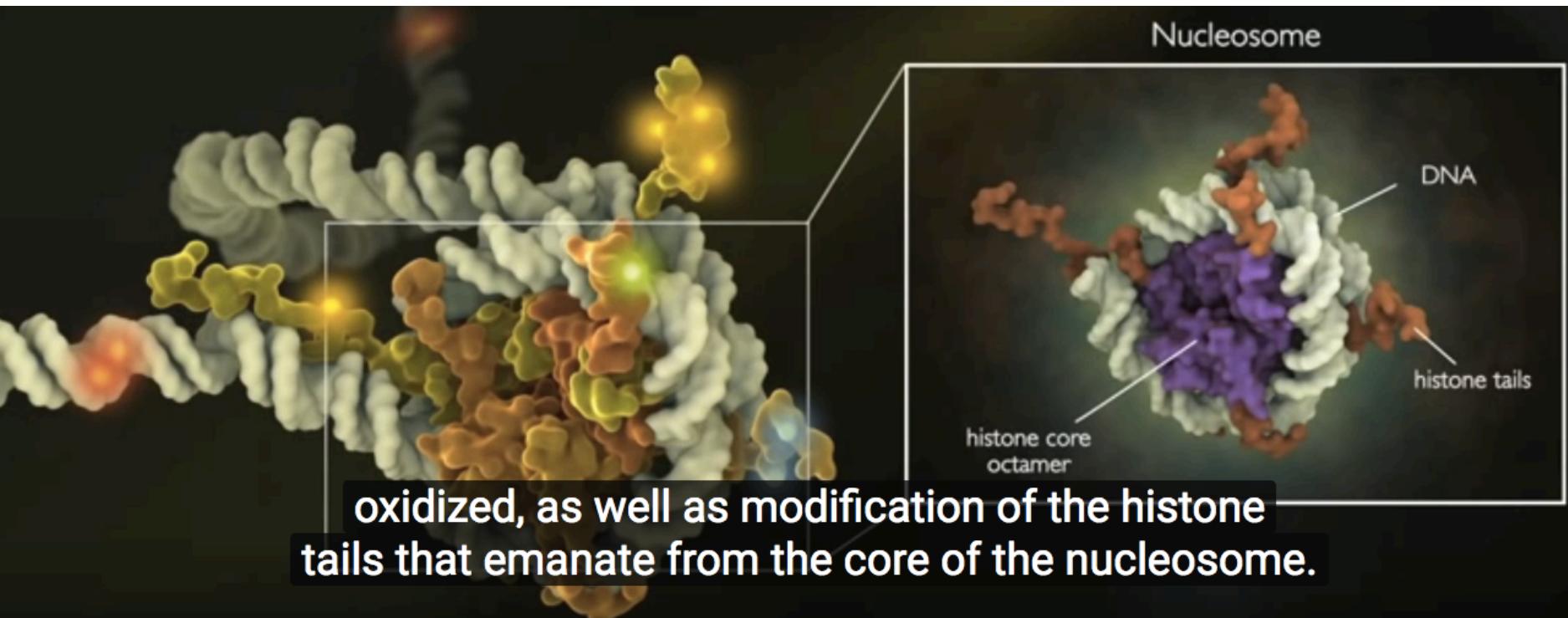
ChIP-seq, Bisulfite, MeDIP-seq, MAD-seq, CLEVER-seq, Mnase-seq, Hi-C, Hi-3C and more....



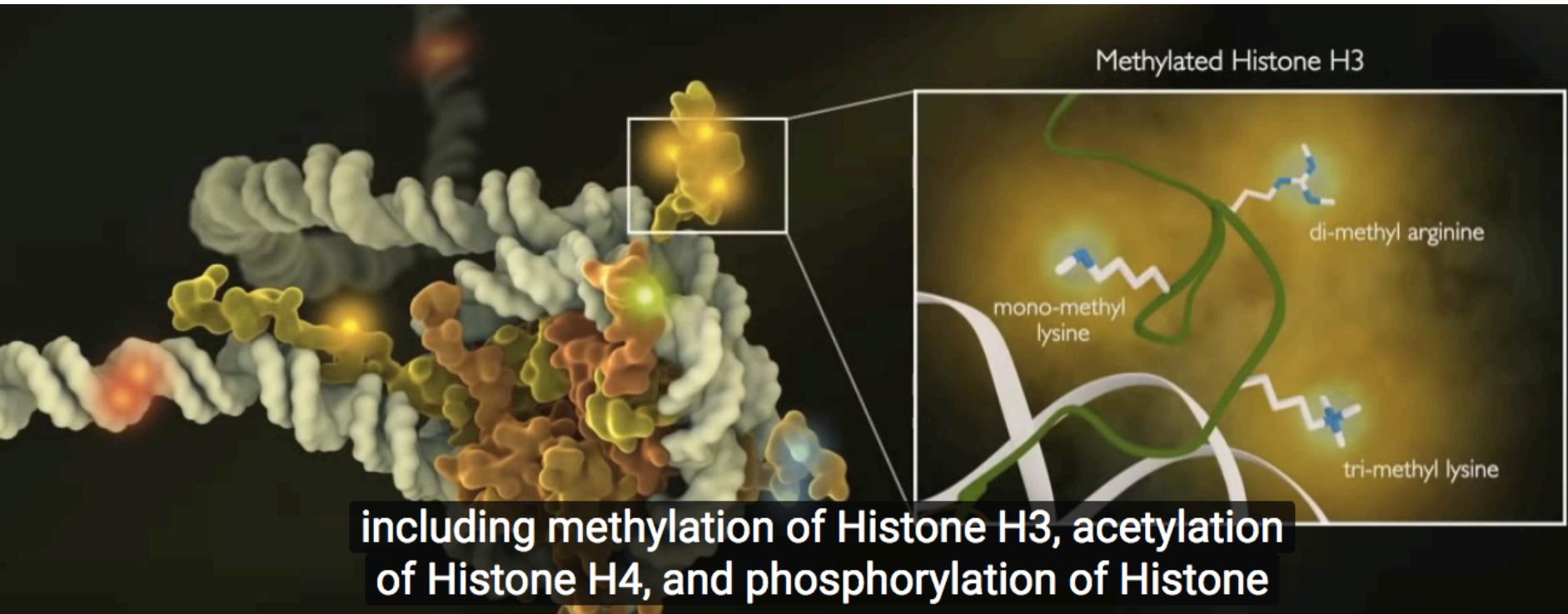
Second Layer of Genome Annotation



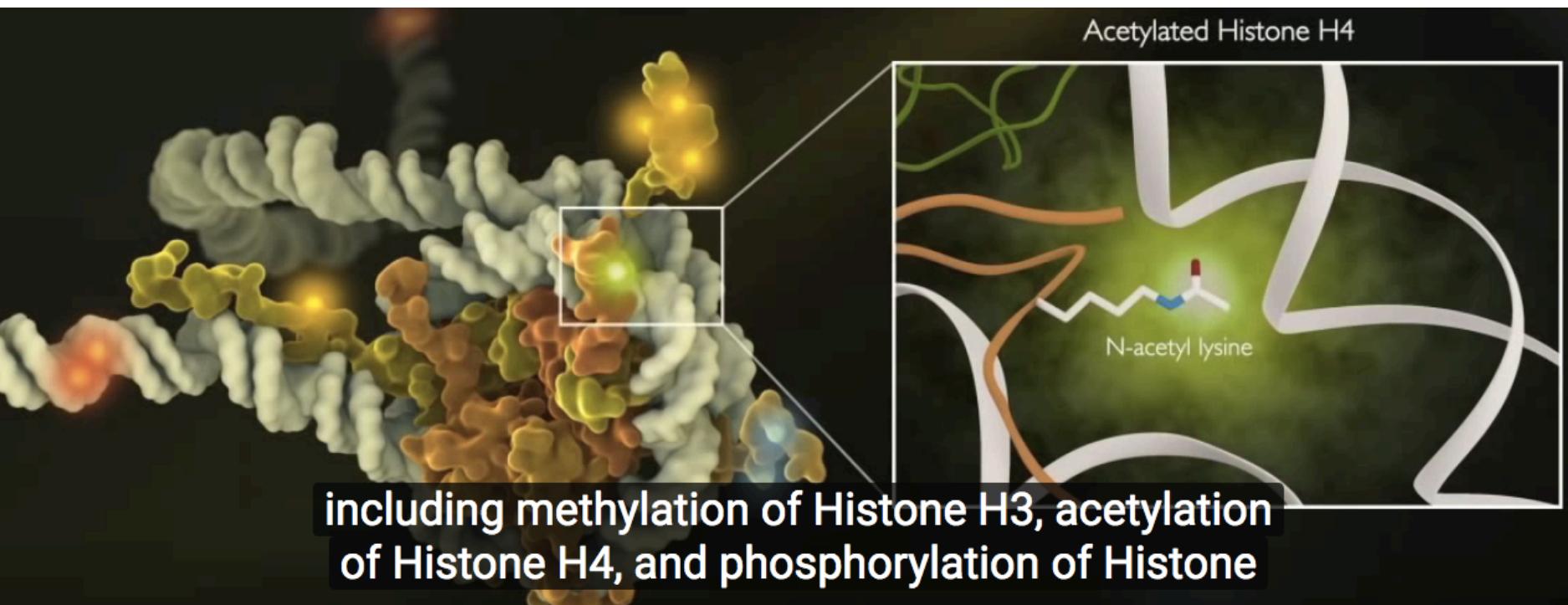
Second Layer of Genome Annotation



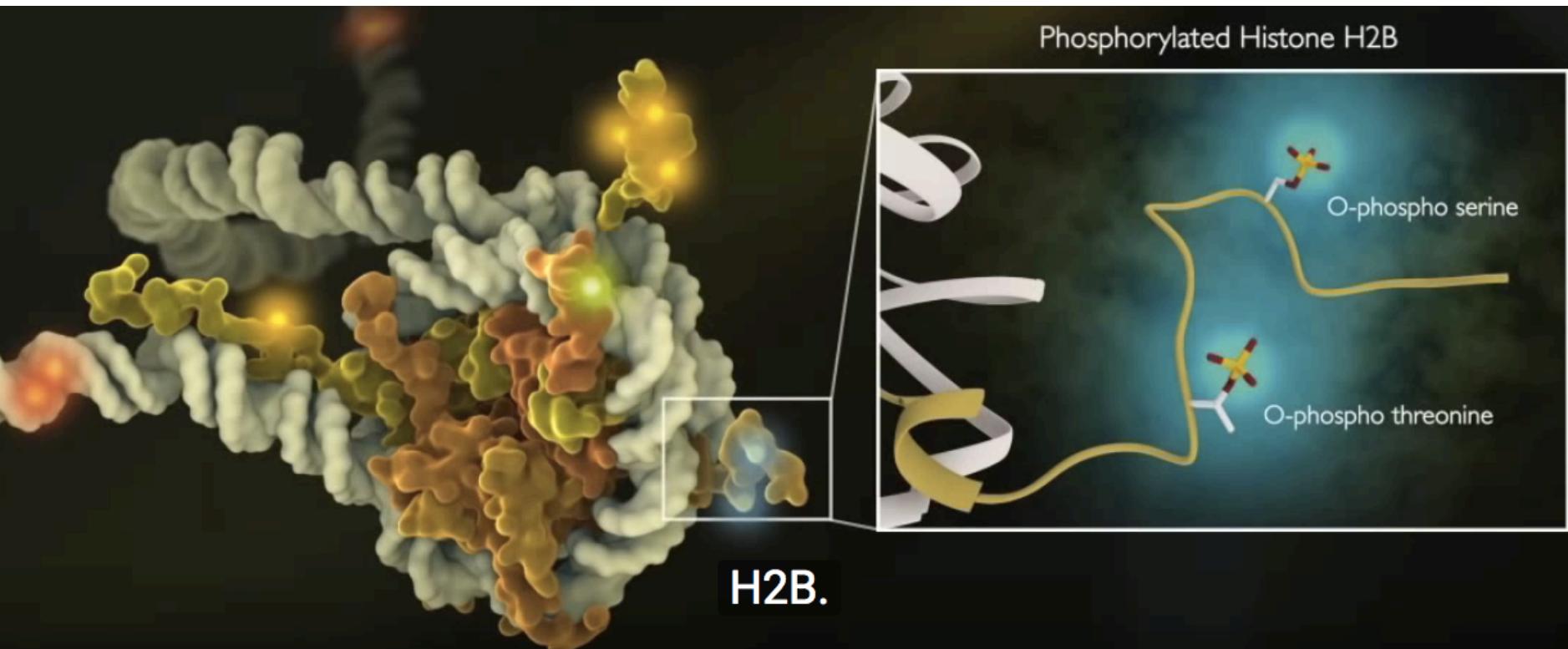
Second Layer of Genome Annotation



Second Layer of Genome Annotation



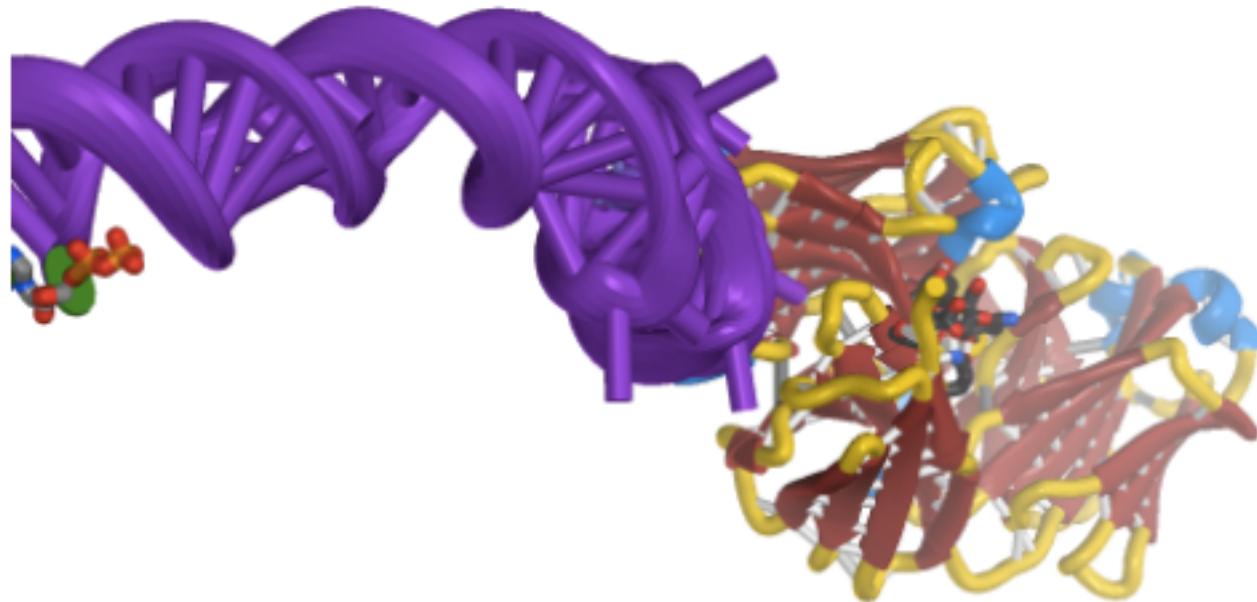
Second Layer of Genome Annotation



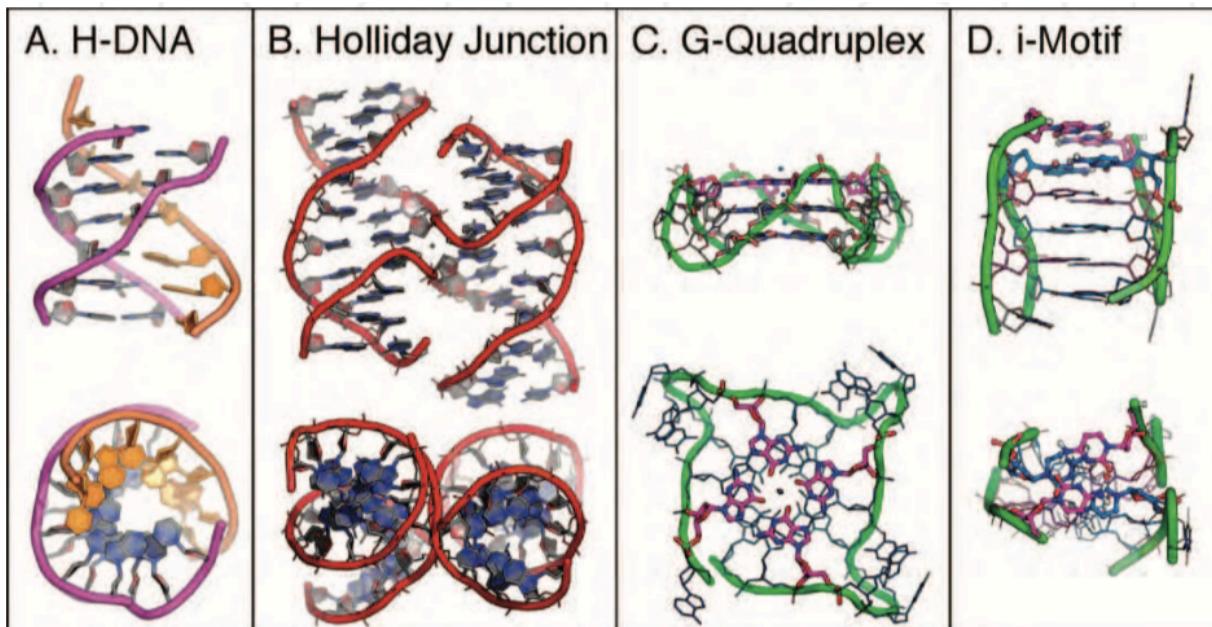
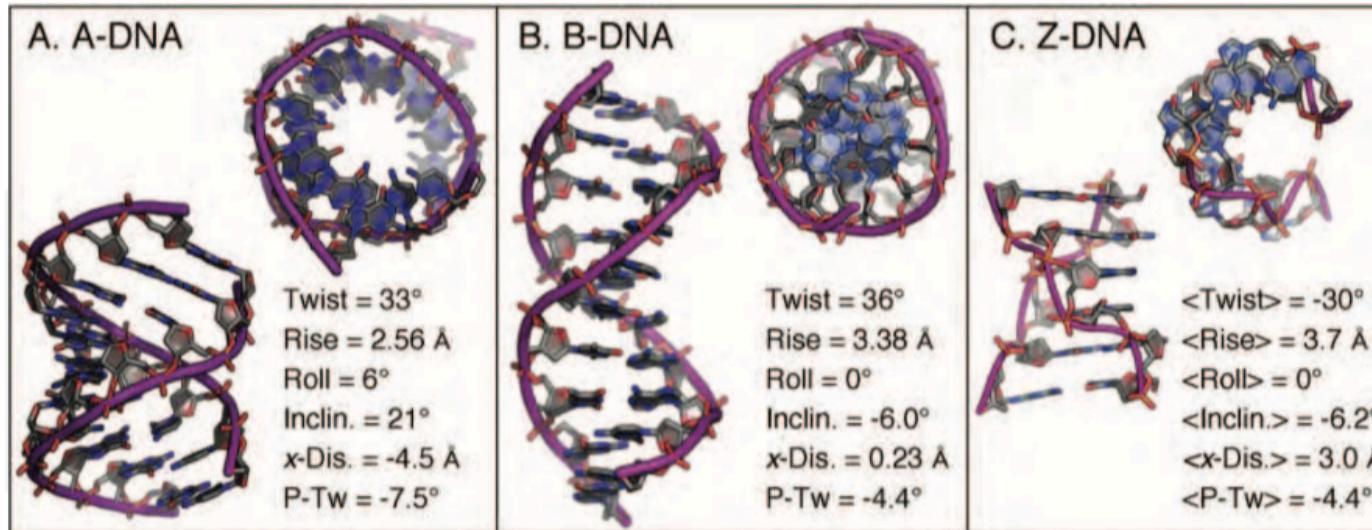
Layers of genomic encoding

- DNA structure
 - quadruplexes, triplexes, Z-DNA, stem-loops, cruciforms,

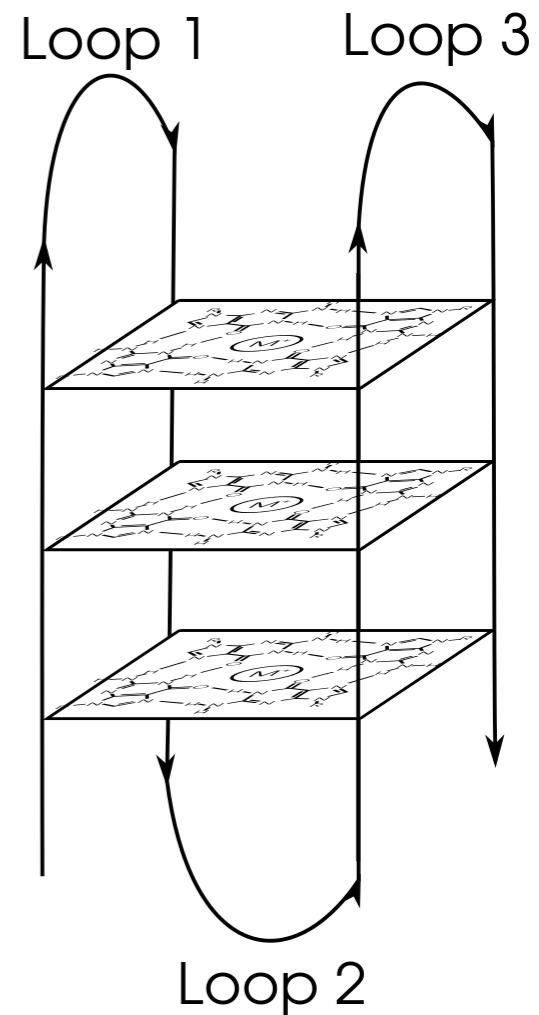
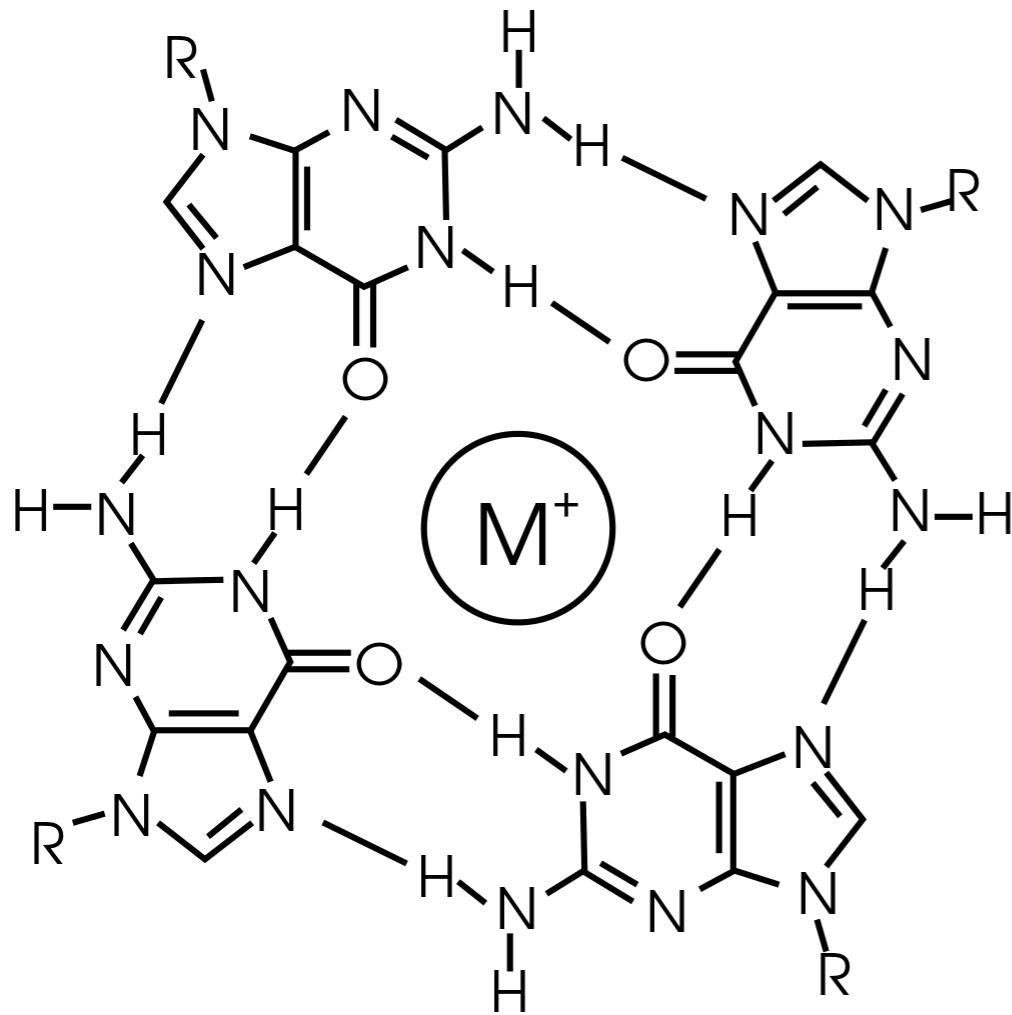
ChIP-seq, G4-seq, ssDNA-
Seq, KAS-seq

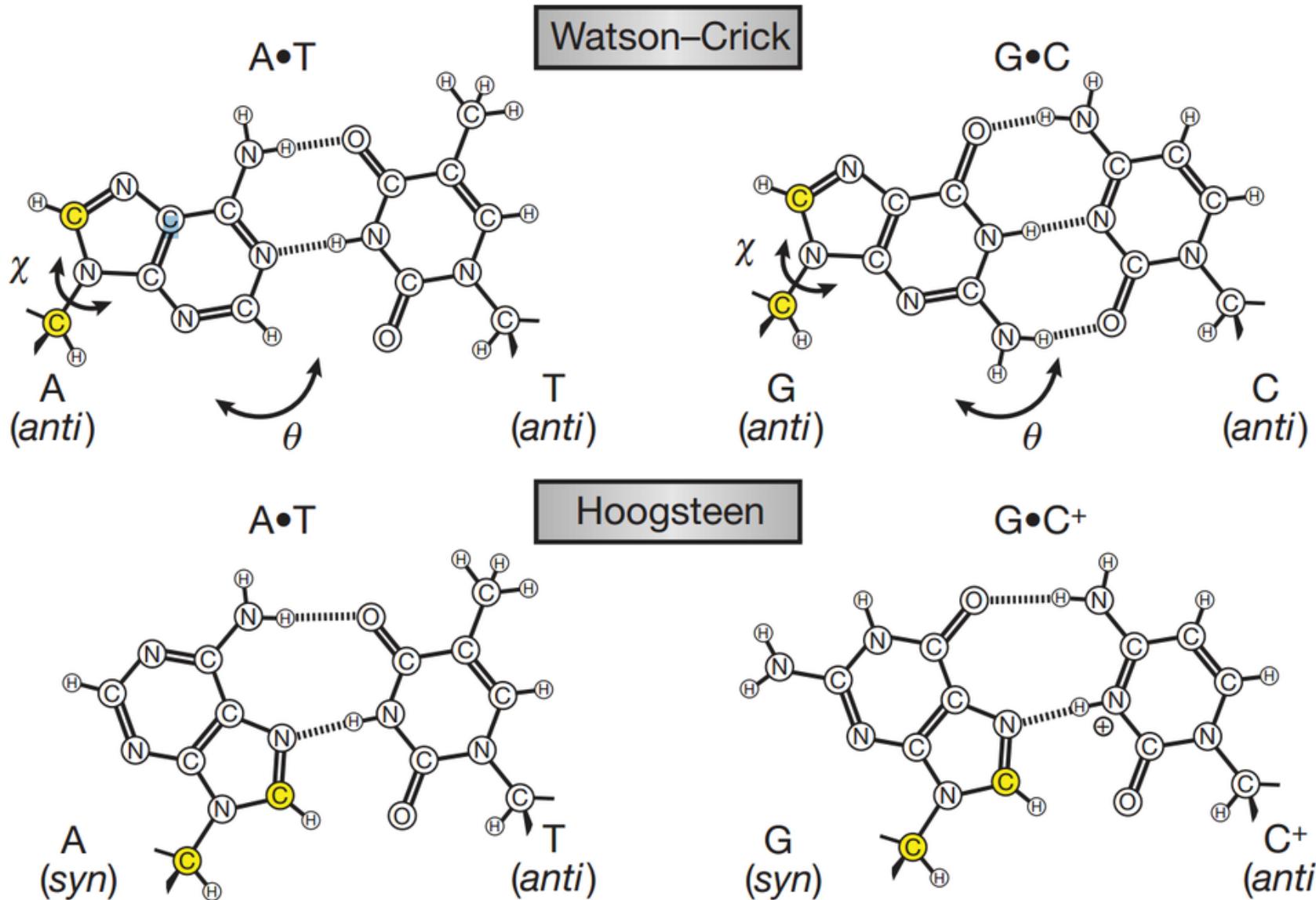


Third Layer – DNA structures

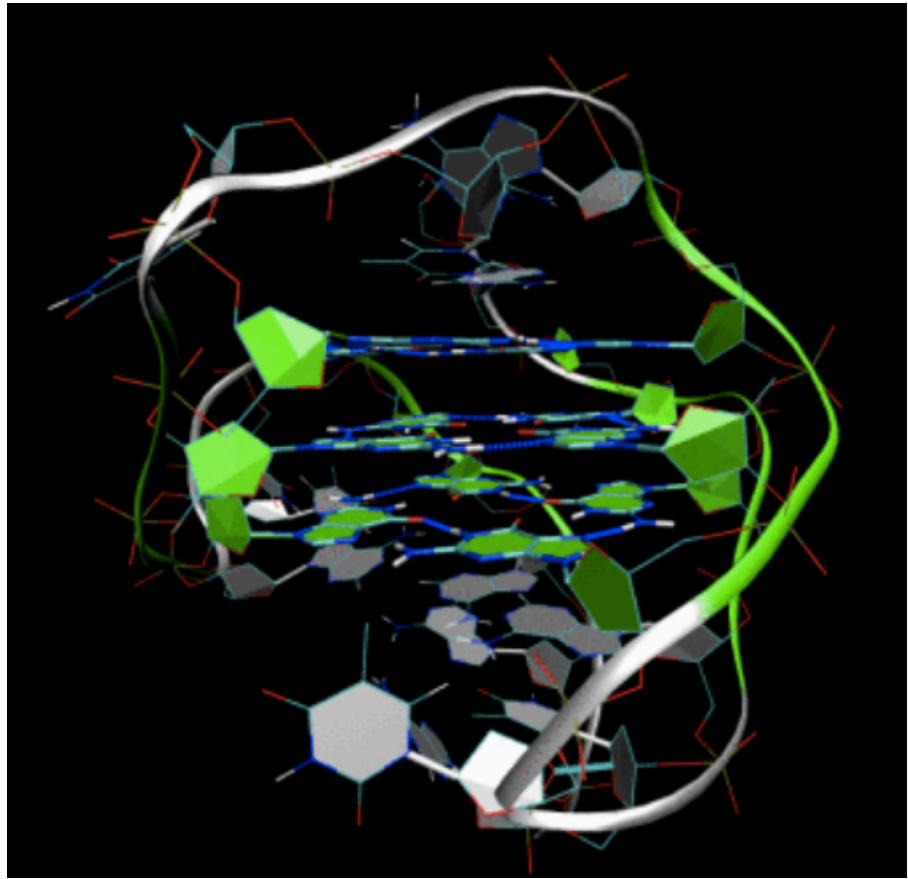
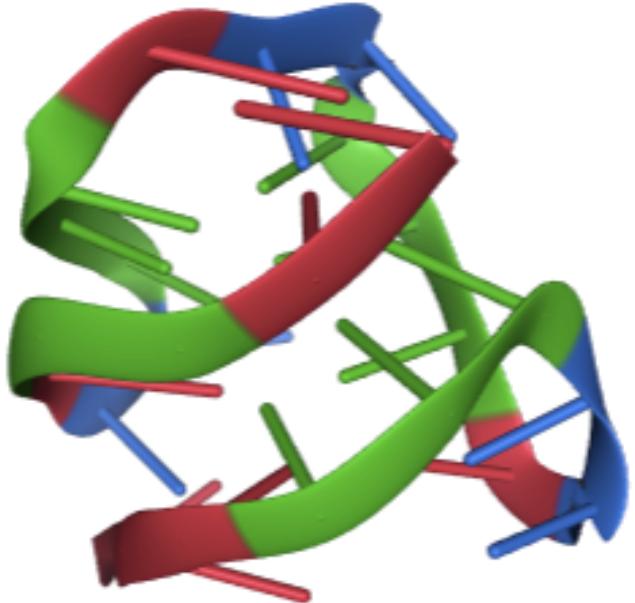


Квадруплексы

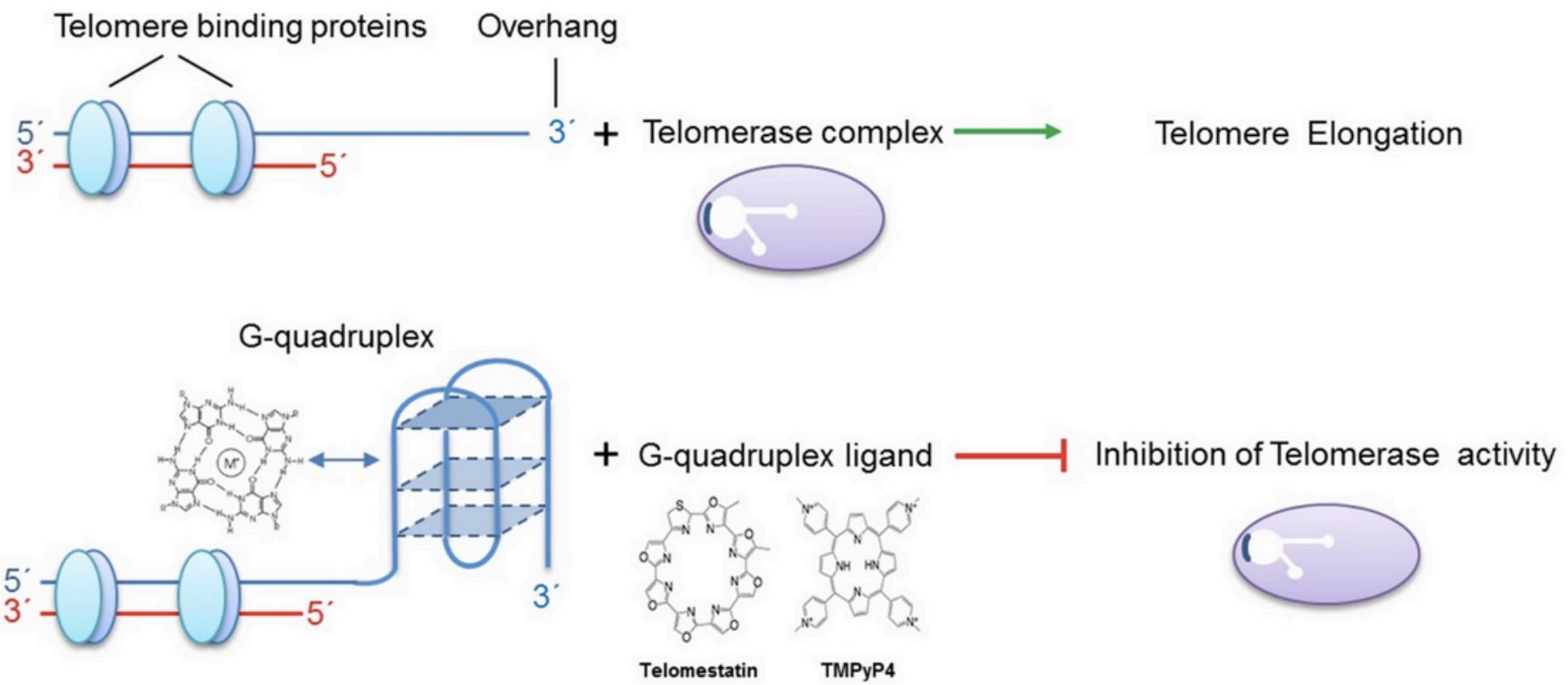


a

Chemical structures for Watson-Crick and Hoogsteen A•T and G•C⁺ base pairs. The Hoogsteen geometry can be achieved by purine rotation around the glycosidic bond (χ) and base-flipping (θ), affecting simultaneously C8 and C1' (yellow).



3D Structure of the intramolecular human telomeric G-quadruplex in potassium solution. The backbone is represented by a tube. The center of this structure contains three layers of G-tetrads. The hydrogen bonds in these layers are represented by blue dashed lines. (PDB: 2HY9)



The Nobel Prize in Physiology or Medicine 2009



© The Nobel Foundation. Photo: U. Montan

**Elizabeth H.
Blackburn**
Prize share: 1/3



© The Nobel Foundation. Photo: U. Montan

Carol W. Greider
Prize share: 1/3



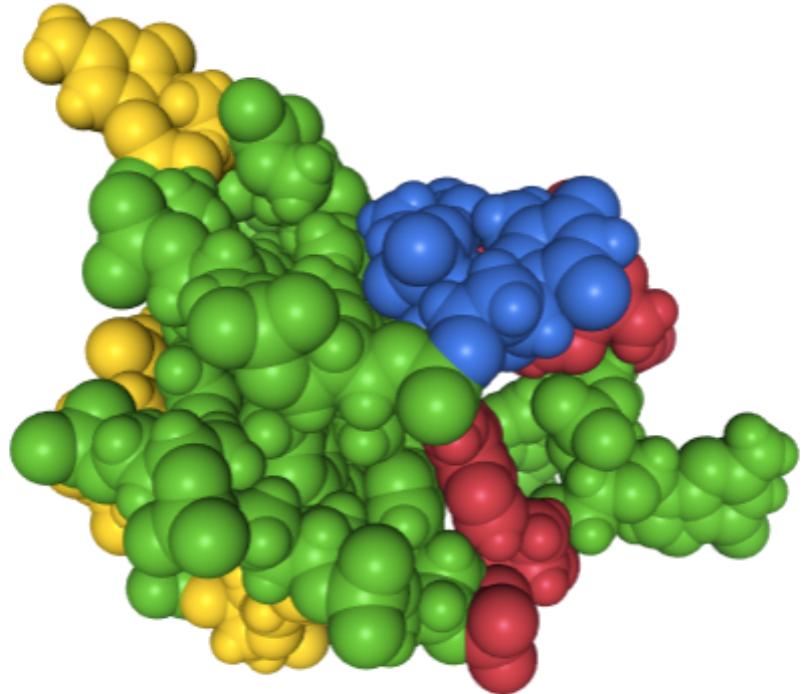
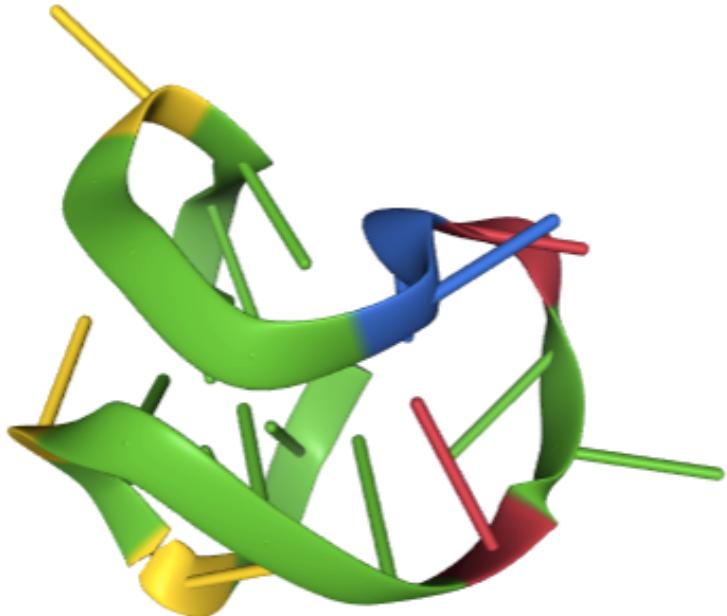
© The Nobel Foundation. Photo: U. Montan

Jack W. Szostak
Prize share: 1/3

The Nobel Prize in Physiology or Medicine 2009 was awarded jointly to Elizabeth H. Blackburn, Carol W. Greider and Jack W. Szostak "for the discovery of how chromosomes are protected by telomeres and the enzyme telomerase."

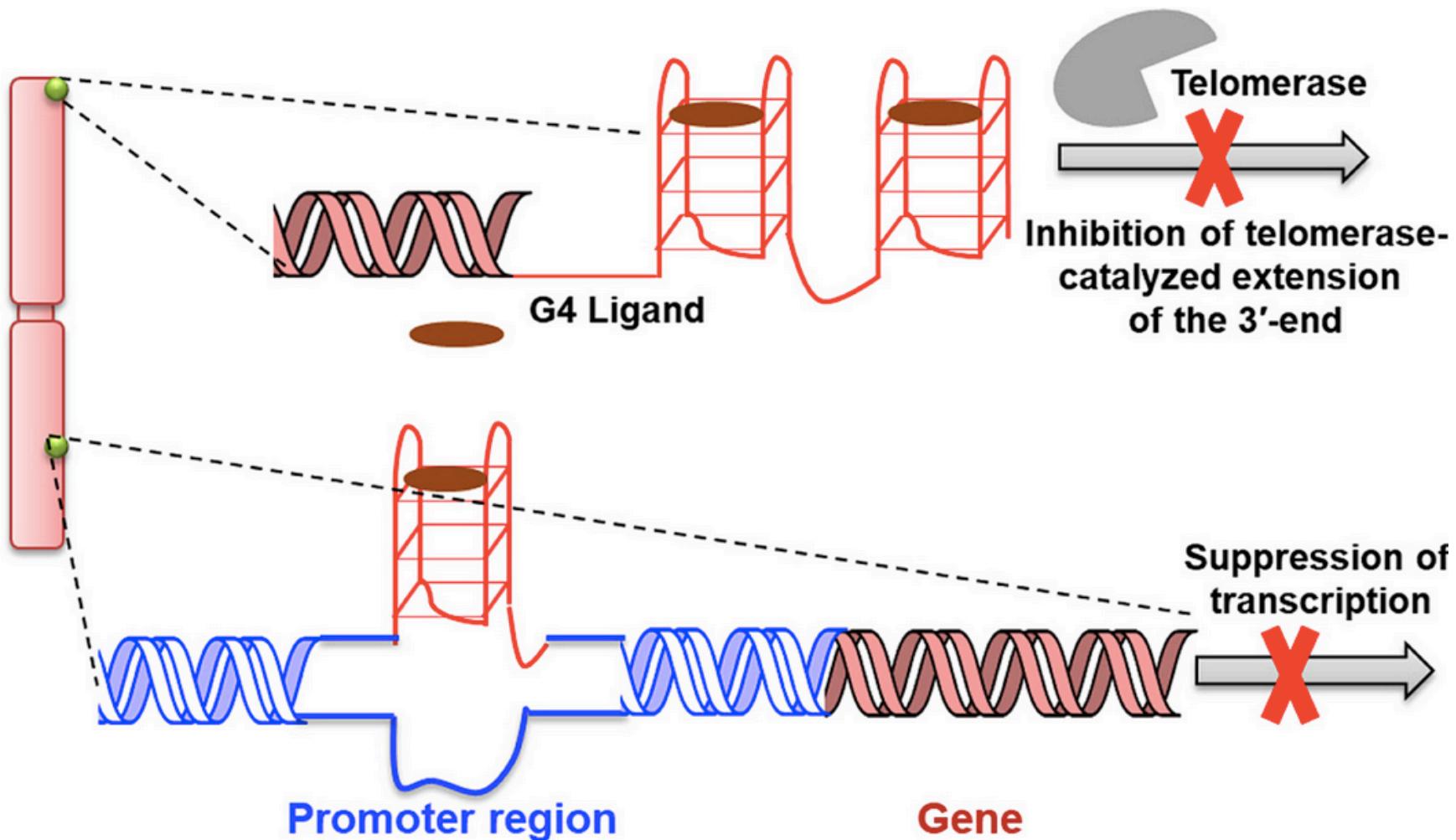
2F8U

G-quadruplex structure formed in human Bcl-2 promoter, hybrid form



The BCL2 associated agonist of cell death[5] (BAD) protein is a pro-apoptotic member of the Bcl-2 gene family which is involved in initiating apoptosis.

Targeting G-quadruplex DNA structures in the telomere and oncogene promoter regions by benzimidazole–carbazole ligands

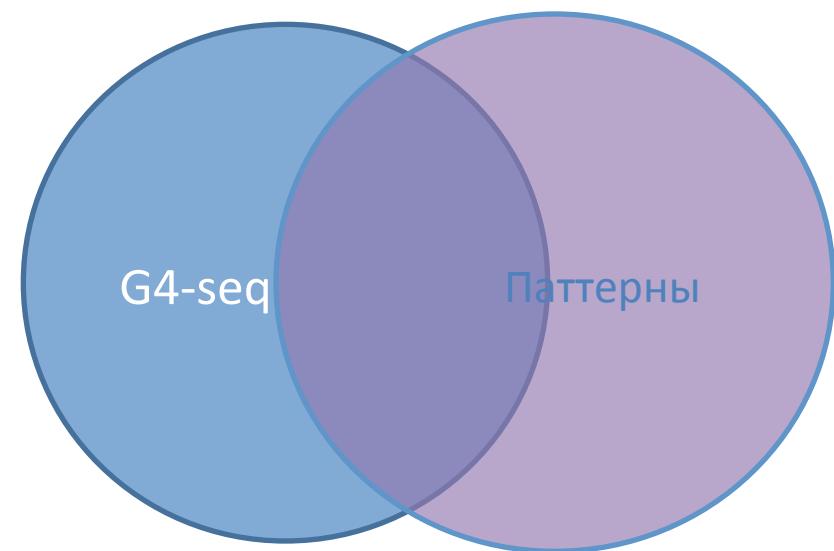


Pattern Search Methods

- $G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}$ - классический паттерн
- $G_{3+}N_{1-12}G_{3+}N_{1-12}G_{3+}N_{1-12}G_{3+}$ - расширенный паттерн

	TPR	FPR
Классический паттерн	36.86%	46.37%
Расширенный паттерн	65.56%	50.96%

Тестирование на G4 Seq



G4-seq

High-throughput sequencing of DNA G-quadruplex structures in the human genome

Vicki S Chambers^{1,5}, Giovanni Marsico^{2,5}, Jonathan M Boutell³, Marco Di Antonio^{1,2}, Geoffrey P Smith³ & Shankar Balasubramanian^{1,2,4}

We identified 716,310 distinct G4 structures, 451,646 of which were not predicted by computational methods.

Read 1: reference (Na^+)

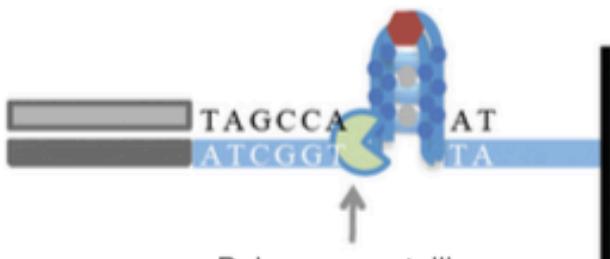


Melt and remove primer

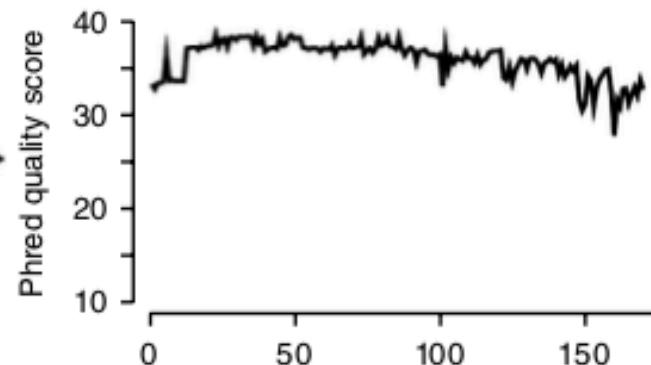
Reprime template

Stabilizing agent

Read 2: G4-stabilization conditions



Polymerase stalling
at G4 start site

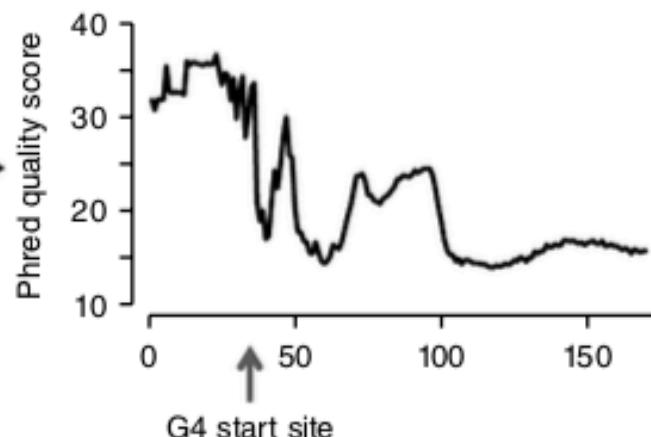


Mismatch analysis

Read 1 - TAGCCACCCCTCCCACCCCTCCCAT

Read 2 - TAGCCA**TCC**ATCTGCCT**TAT**TTG

↑ G4 start site ■ Base mismatch

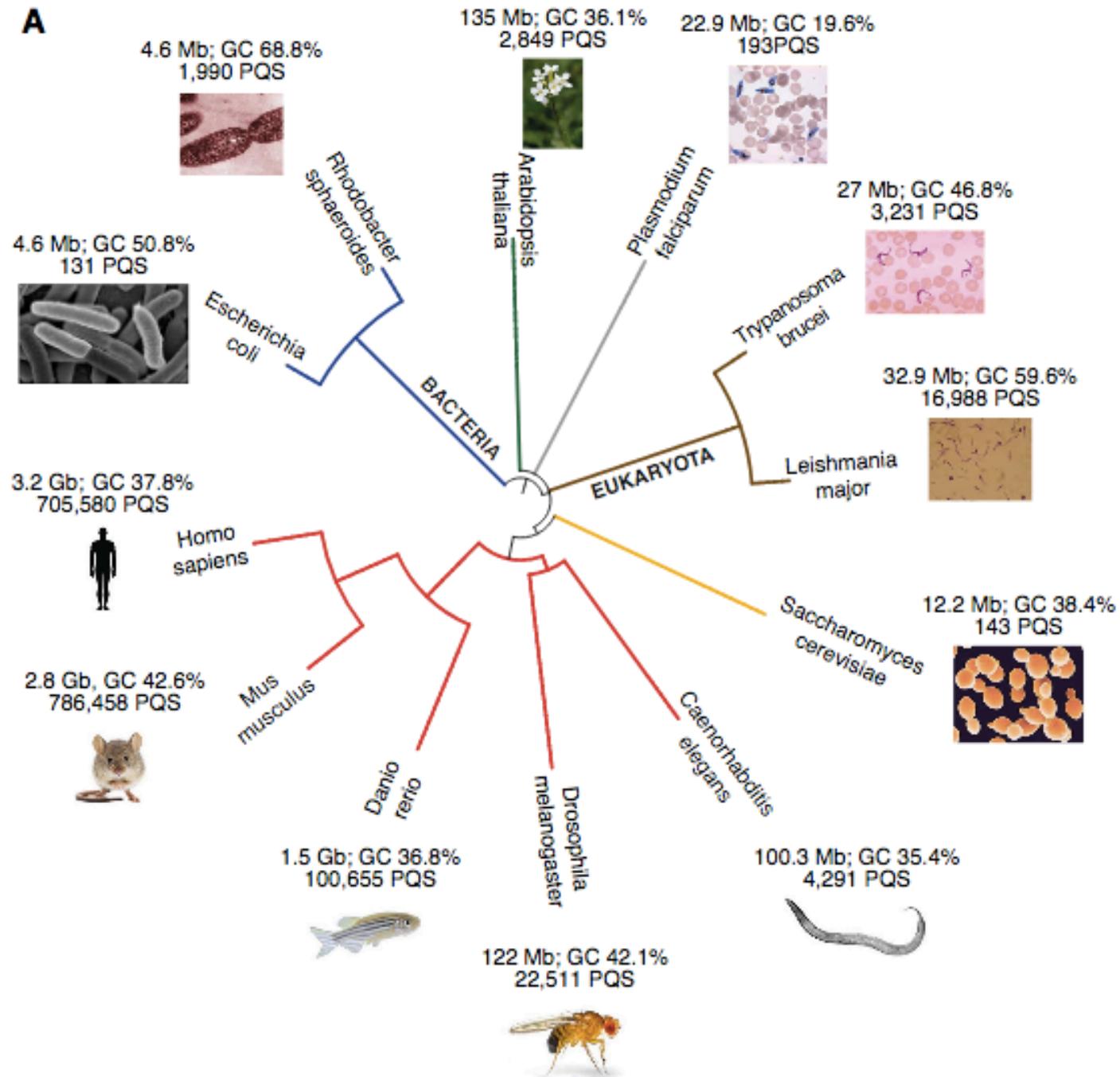


G4-stabilizing ligand pyridostatin (PDS)

Whole genome experimental maps of DNA G-quadruplexes in multiple species

Giovanni Marsico¹, Vicki S. Chambers^{2,3}, Aleksandr B. Sahakyan^{1,2}, Patrick McCauley³, Jonathan M. Boutell³, Marco Di Antonio² and Shankar Balasubramanian^{1,2,4,*}

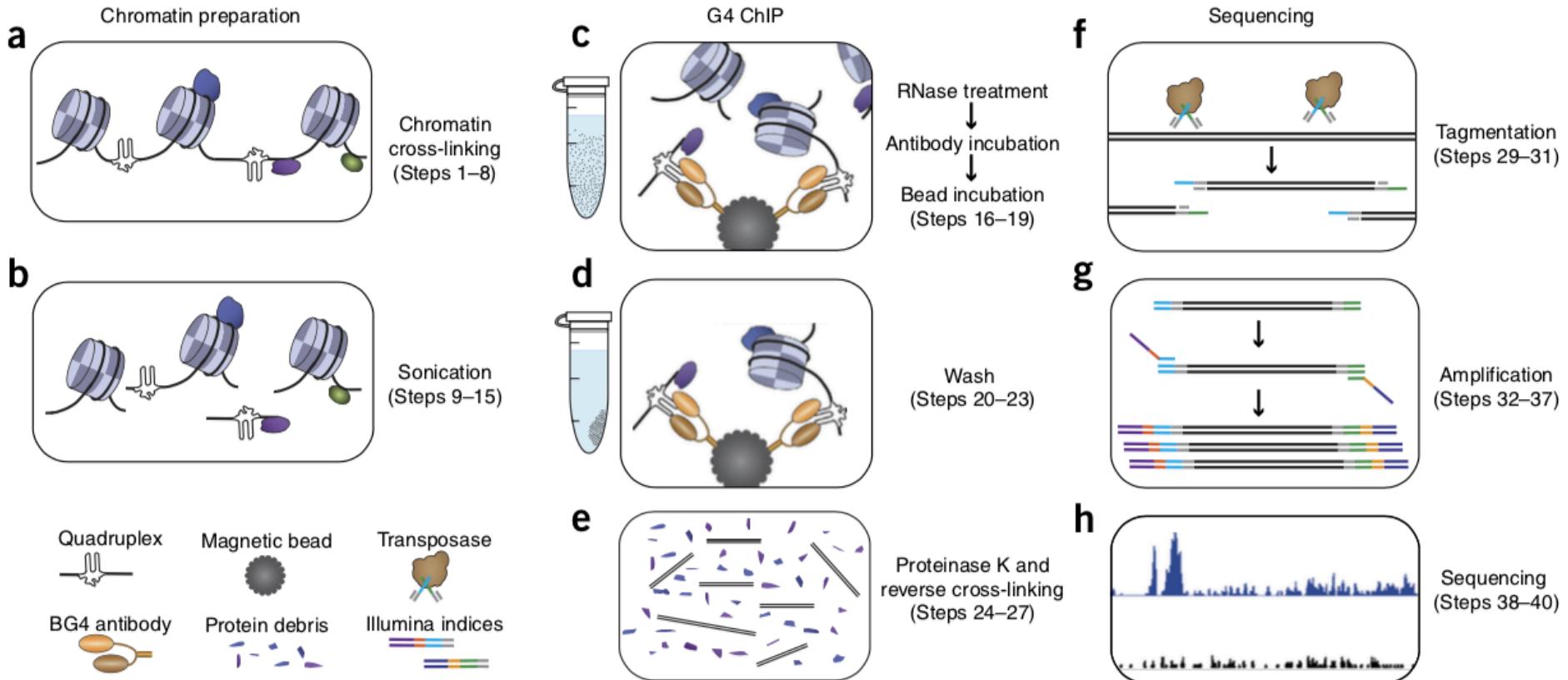
Genome	# all OQs
<i>human</i>	434 272
<i>mouse</i>	797 789
<i>zebrafish</i>	141 637
<i>Drosophila</i>	19 399
<i>C. elegans</i>	4144
<i>Saccharomyces</i>	103
<i>Leishmania</i>	17 343
<i>Trypanosoma</i>	3236
<i>Plasmodium</i>	173
<i>Arabidopsis</i>	2407
<i>E. coli</i>	47
<i>Rhodobacter</i>	291

A

G4-Chip

Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing

Robert Hänsel-Hertsch^{1,4}, Jochen Spiegel^{1,2,4}, Giovanni Marsico¹, David Tannahill¹ & Shankar Balasubramanian^{1–3} 



G4-specific antibody BG4 and immobilized on magnetic beads

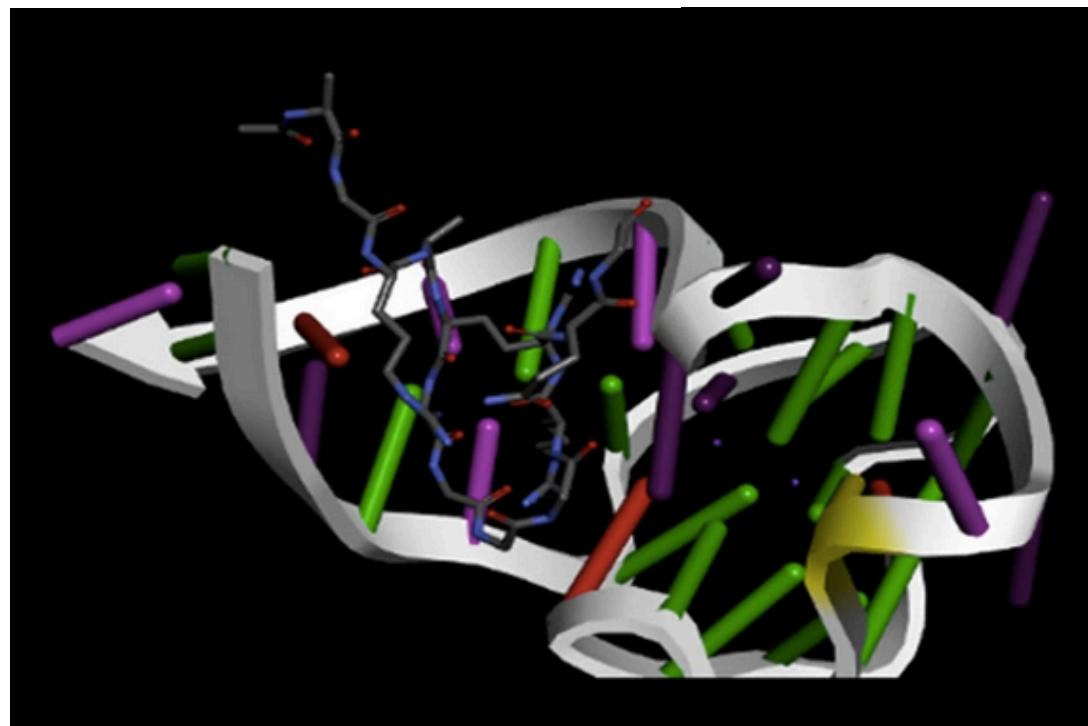
G4IPDB: A database for G-quadruplex structure forming nucleic acid interacting proteins

Subodh Kumar Mishra, Arpita Tawani, Amit Mishra & Amit Kumar

Scientific Reports **6**, Article number: 38144 (2016) | [Cite this article](#)

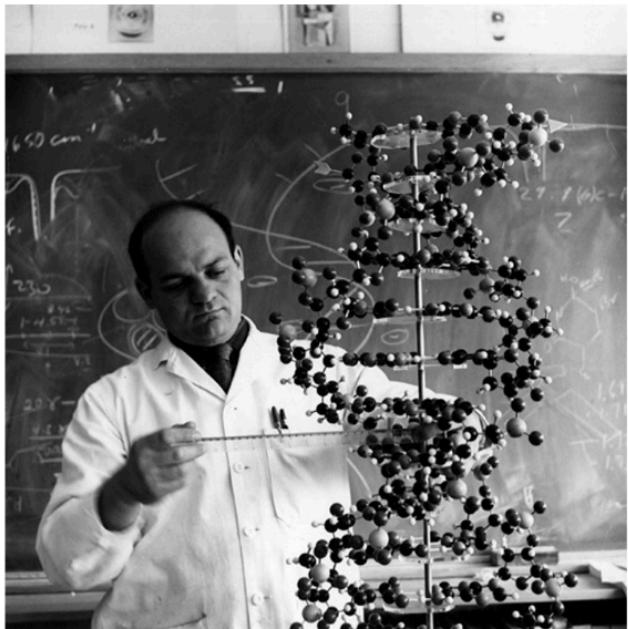
1769 Accesses | **30** Citations | **13** Altmetric | [Metrics](#)

This database
contains more
than 200 entries



Z-DNA

Александр Рич (MIT) –
открыл Z-ДНК (1979)



20 October 2005 | www.nature.com/nature | 510 THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

nature

THE B TO Z
OF DNA

Left-handed meets
right-handed DNA

SOLAR SYSTEM SURFACES
Europa's craters
lose their impact

THE HUMAN CONDITION
Are we still evolving?

MICROARRAY TECHNOLOGY
Biochips with
everything

NATUREJOBS
Go to work in genes

A detailed 3D molecular model of Z-DNA, showing its characteristic left-handed helical twist. The model consists of orange spheres representing phosphate groups and grey spheres representing deoxyribose sugars. Some yellow and green spheres are interspersed along the backbone, likely representing specific nucleotides or binding proteins.

A Z-DNA binding domain present in the human editing enzyme, double-stranded RNA adenosine deaminase

(RNA editing/negative supercoiling/B-Z junction/transcription)

ALAN HERBERT*, JENS ALFKEN*, YANG-GYUN KIM*, I. SAIRA MIAN†, KAZUKO NISHIKURA‡,
AND ALEXANDER RICH*

*Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139; †Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720; and ‡The Wistar Institute, 3601 Spruce Street, Philadelphia, PA 19104

- Z-DNA binding proteins
 - ADAR1
 - ZBP1
 - E3L
 - PKZ



Computer Methods to detect Z-DNA

The EMBO Journal vol.5 no.10 pp.2737–2744, 1986

A computer aided thermodynamic approach for predicting the formation of Z-DNA in naturally occurring sequences

Pui S.Ho, Michael J.Ellison, Gary J.Quigley and
Alexander Rich

Z-Hunt

Department of Biology, Massachusetts Institute of Technology, Cambridge,
MA 02139, USA

Table I. Experimentally determined energetic parameters for dinucleotides associated with the B to Z transition

Nearest neighbor interactions		ΔG (kcal/mol) per dinucleotide	Reference
5' anti-syn 3'	5' syn-anti 3'		
5'-C G-3'	5'-G C-3'		
⋮ ⋮	⋮ ⋮		
3'-G C-5'	3'-C G-5'	0.66	Peck and Wang, 1983
5'-C A-3'	5'-A C-3'		
⋮ ⋮	⋮ ⋮		
3'-G T-5'	3'-T G-5'	1.34	Vologodski and Frank-Kamenetskii, 1984
5'-T A-3'	5'-A T-3'		
⋮ ⋮	⋮ ⋮		
3'-A T-5'	3'-T A-5'	2.4	Ellison <i>et al.</i> , 1986a
5'-C C-3'	5'-C C-3'		
⋮ ⋮	⋮ ⋮		
3'-G G-5'	3'-G G-5'	2.4	Ellison <i>et al.</i> , 1985
5'-C T-3'	5'-T C-3'		
⋮ ⋮	⋮ ⋮		
3'-G A-5'	3'-A G-5'	2.5	Ellison <i>et al.</i> , 1985
B-Z		5.0/junction	Peck and Wang, 1983
Z-Z junctions		4.0/junction	Ellison <i>et al.</i> , 1986b

Computer Methods to detect Z-DNA

Z-Hunt II

Mapping Z-DNA in the Human Genome

COMPUTER-AIDED MAPPING REVEALS A NONRANDOM DISTRIBUTION
OF POTENTIAL Z-DNA-FORMING SEQUENCES IN HUMAN GENES*

Gary P. Schroth, Ping-Jung Chou‡, and P. Shing Ho§

From the Department of Biochemistry and Biophysics, Oregon State University, Corvallis, Oregon 97331

THE JOURNAL OF BIOLOGICAL CHEMISTRY

© 1992 by The American Society for Biochemistry and Molecular Biology, Inc.

Chip-Seq for Z-DNA

Human genomic Z-DNA segments probed by the Z α domain of ADAR1

Heng Li¹, Jie Xiao², Jinming Li², Le Lu³, Shu Feng¹ and Peter Dröge^{1,*}

¹Division of Genomics and Genetics, ²Structural and Computational Biology, School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551 and ³Bioinformatics Division, TNLIST and Department of Automation, Tsinghua University, Beijing 100084, PR China

Nucleic Acids Research, 2009, Vol. 37, No. 8 2737–2746

2009

doi:10.1093/nar/gkp124

Z-DNA binding domain, Za, of the RNA editing enzyme ADAR1

we identified 186 genomic Z-DNA hotspots.

46 hotspots were located in centromeres of 13 human chromosomes.

Chip-Seq for Z-DNA

Z-DNA-forming sites identified by ChIP-Seq are associated with actively transcribed regions in the human genome

**So-I. Shin^{1,†}, Seokjin Ham^{1,†}, Jihwan Park^{2,†}, Seong Hye Seo¹,
Chae Hyun Lim², Hyeongrin Jeon¹, Joungyun Huh¹, and
Tae-Young Roh^{1,2,*}**

¹Department of Life Sciences, and ²Division of Integrative Biosciences & Biotechnology, Pohang University of Science and Technology (POSTECH), Pohang, Republic of Korea

DNA Research, 2016

Zaa with two Z-DNA-binding domains

391 ZFSs - A large portion of ZFSs was enriched in the promoter regions.

Genes containing ZFSs were occupied by RNA polymerase II at the promoters and showed high levels of expression.

ZFSs were significantly related to active histone marks such as H3K4me3 and H3K9ac.

RNA-DNA Triplex Formation by Long Noncoding RNAs

Yue Li,¹ Junetha Syed,¹ and Hiroshi Sugiyama^{1,2,*}

¹Department of Chemistry, Graduate School of Science, Kyoto University, Kitashirakawa-oiwakecho, Sakyo-ku, Kyoto 606-8502, Japan

²Institute for Integrated Cell-Material Sciences (WPI-iCeMS), Kyoto University, Yoshida-ushinomiyacho, Sakyo-ku, Kyoto 606-8501, Japan

*Correspondence: hs@kuchem.kyoto-u.ac.jp

<http://dx.doi.org/10.1016/j.chembiol.2016.09.011>

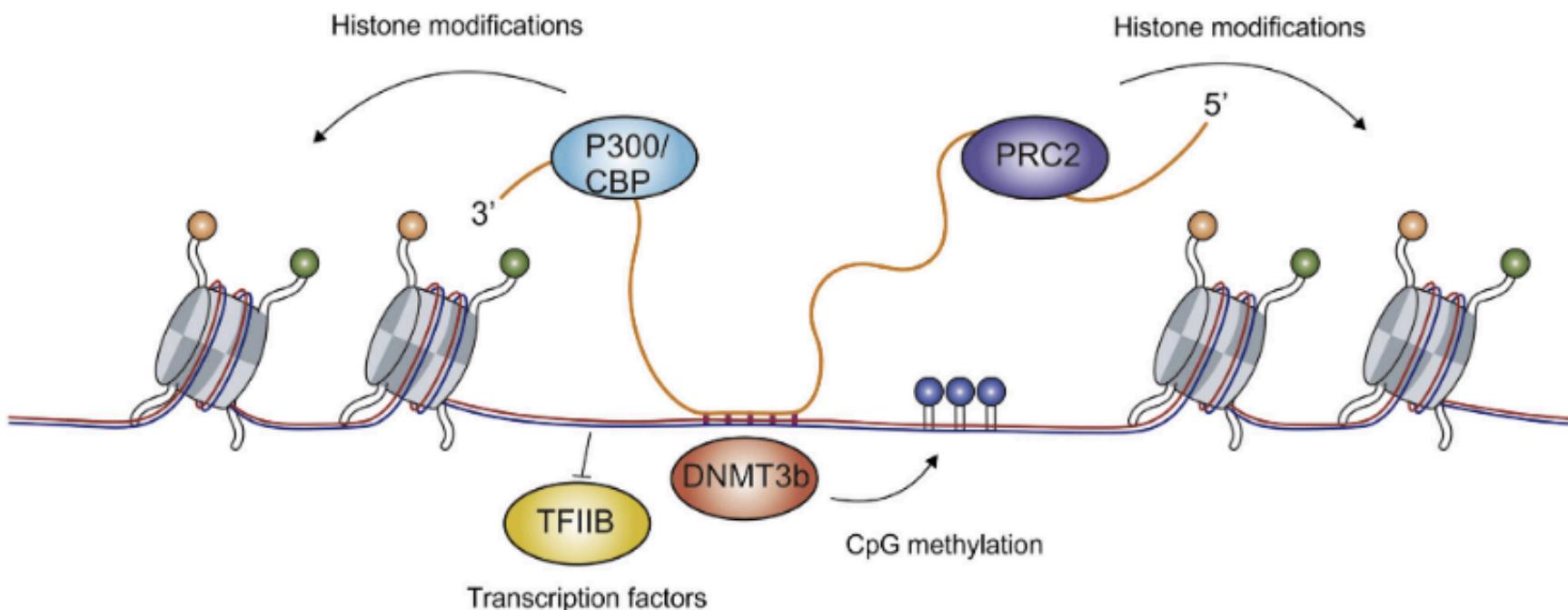
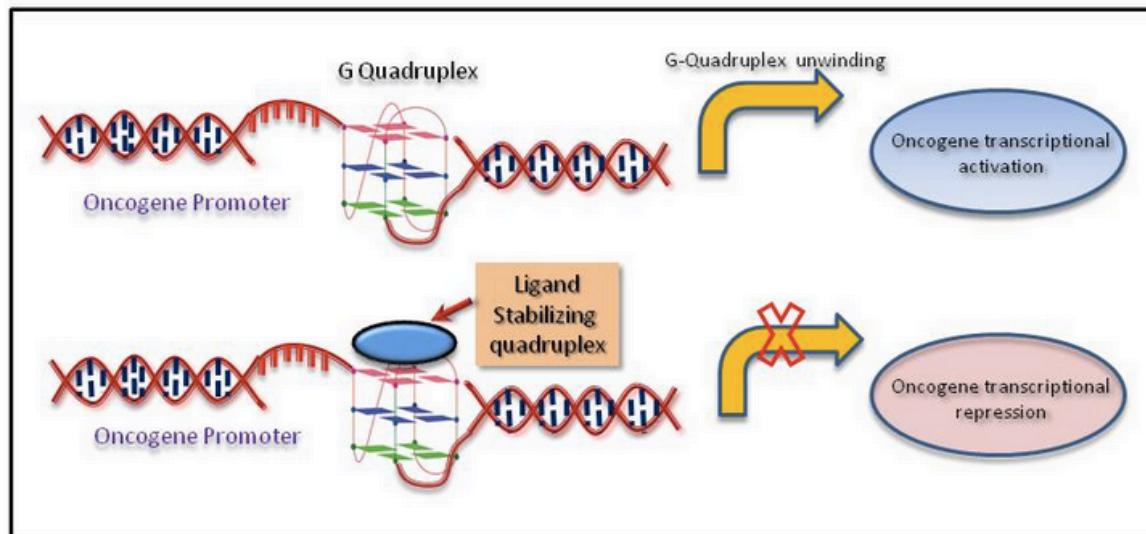


Figure 2. Examples of lncRNAs that Are Reported to Form RNA-DNA Triplexes

lncRNAs can form a triplex structure with a DNA sequence in either neighboring or distant genes and regulate their expressions. lncRNAs also interact with various regulatory proteins such as transcription factors and chromatin modifiers.

Functional role of DNA-structures

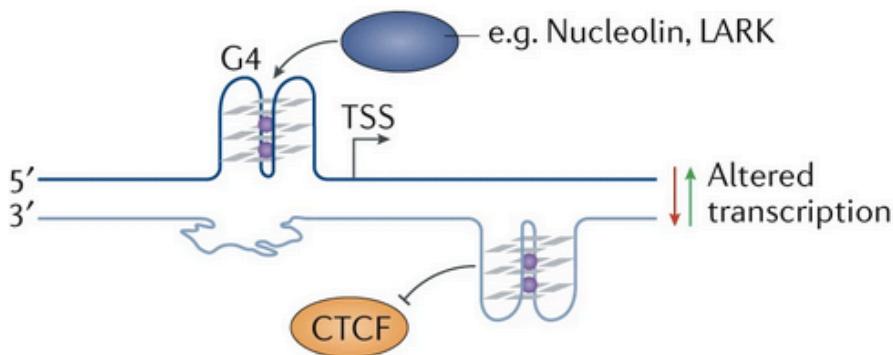
- Dynamical switches
 - Located near promoter – can up or down regulate
 - Examples:



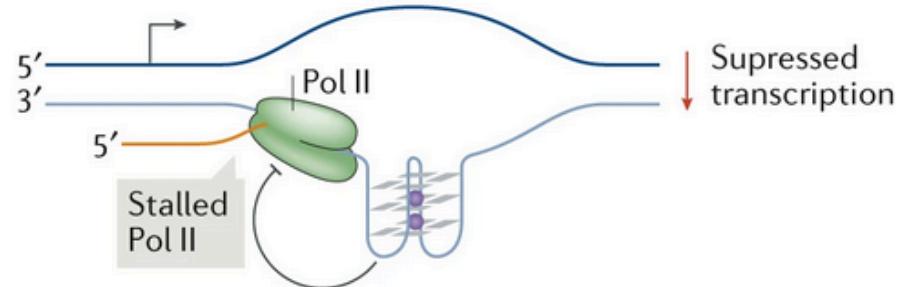
G-Quadruplex targeting ligands: A hope and a new horizon in Cancer Therapeutics

Models of G-quadruplex involvement in transcription

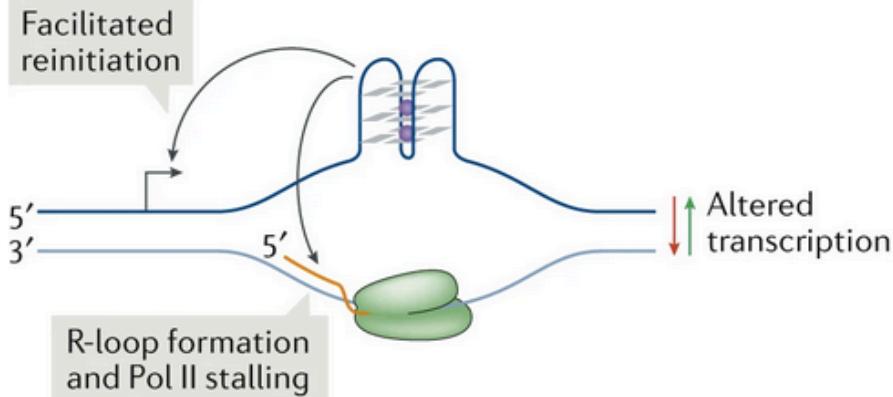
a Binding or displacement of transcription factors



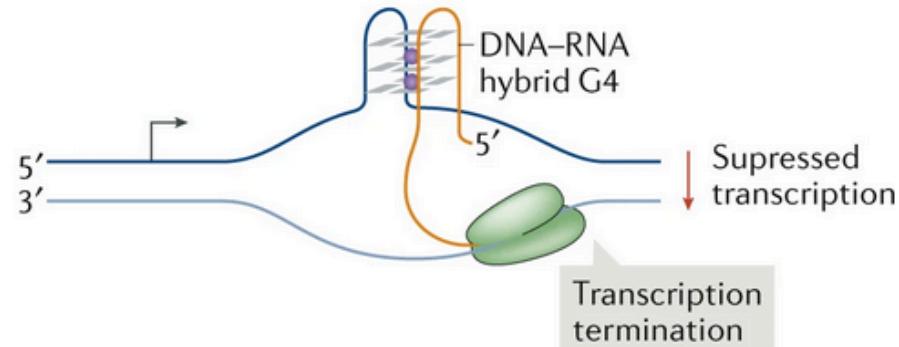
b Gene-body, template-strand DNA G4



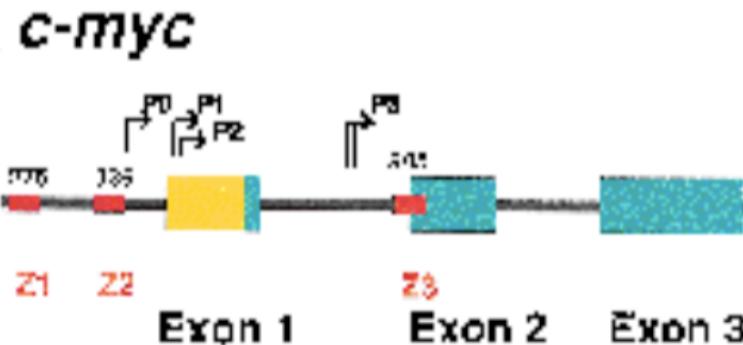
c Gene-body, non-template-strand DNA G4



d Gene-body, non-template-strand hybrid G4



Functional role of Z-DNA



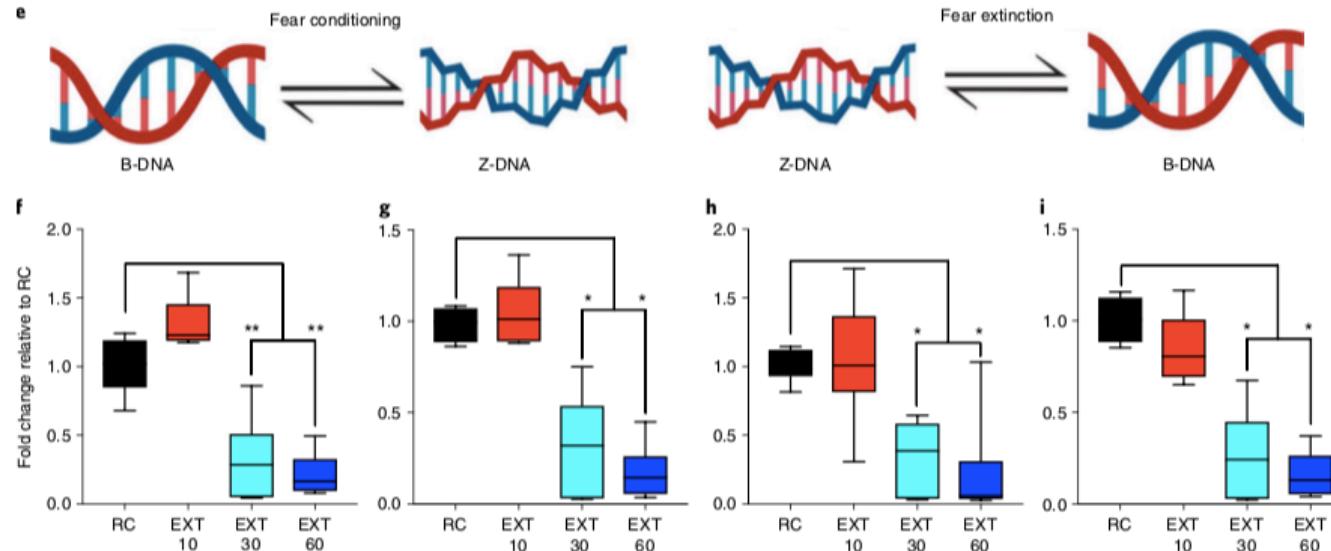
Z-DNA-forming segments, shown in *red*, can be detected in the genes encoding corticotropin-releasing hormone (*crh*) and the *c-myc* 67-kDa protein product. Z-DNA forms only when these genes are transcriptionally active. The promoters for each gene are labeled with an *arrow* and numbered according to their location within the gene. Translated regions of exons are shown in *green* and untranslated parts in *yellow*. Introns are shown by a *heavy black line*.

A. Herbert and A. Rich. The Biology of Left-handed Z-DNA*

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 271, No. 20, Issue of May 17, pp. 11595–11598, 1996
© 1996 by The American Society for Biochemistry and Molecular Biology, Inc.
Printed in U.S.A.

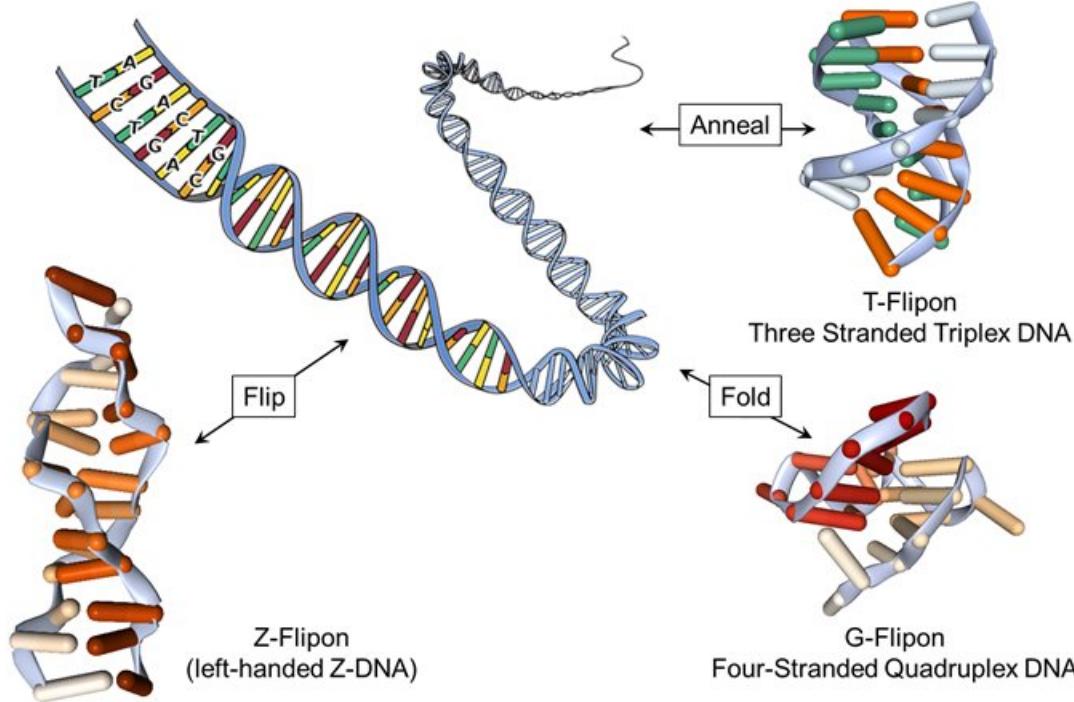
Dynamic regulation of Z-DNA in the mouse prefrontal cortex by the RNA-editing enzyme Adar1 is required for fear extinction

Paul R. Marshall¹, Qiongyi Zhao^{ID 1}, Xiang Li¹, Wei Wei¹, Ambika Periyakaruppiah¹, Esmi L. Zajaczkowski¹, Laura J. Leighton¹, Sachithrani U. Madugalle¹, Dean Basic¹, Ziqi Wang¹, Jiayu Yin¹, Wei-Siang Liau¹, Ankita Gupte², Carl R. Walkley² and Timothy W. Bredy^{ID 1}✉



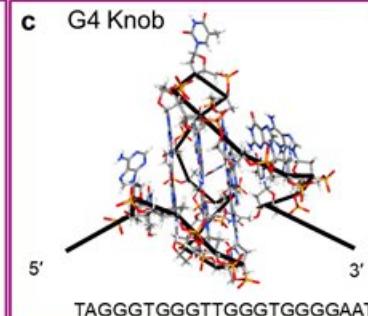
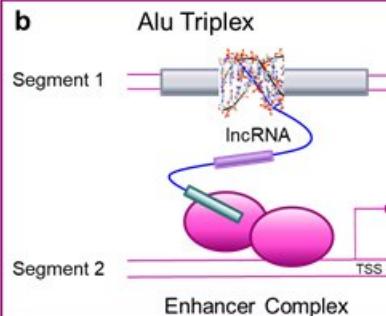
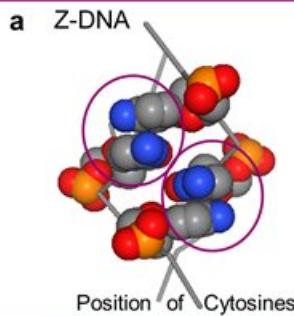
Flipon concepts

Flipons are Genetic On-Off Switches



THE DIGITAL GENOME

Flipons act as digital switches to turn one message on or another message off. They are scattered throughout the genome and associated with ALU Repeat elements. They switch between the regular B-conformation and an alternate DNA structure. There are different classes of flipons with different numbers of DNA strands (2 for Z-DNA, 3 for a triplex and 4 for a quadruplex) and ALU sequence requirements. Flipons alter how a chromosome is read out. The message produced depends on the flipon conformation.

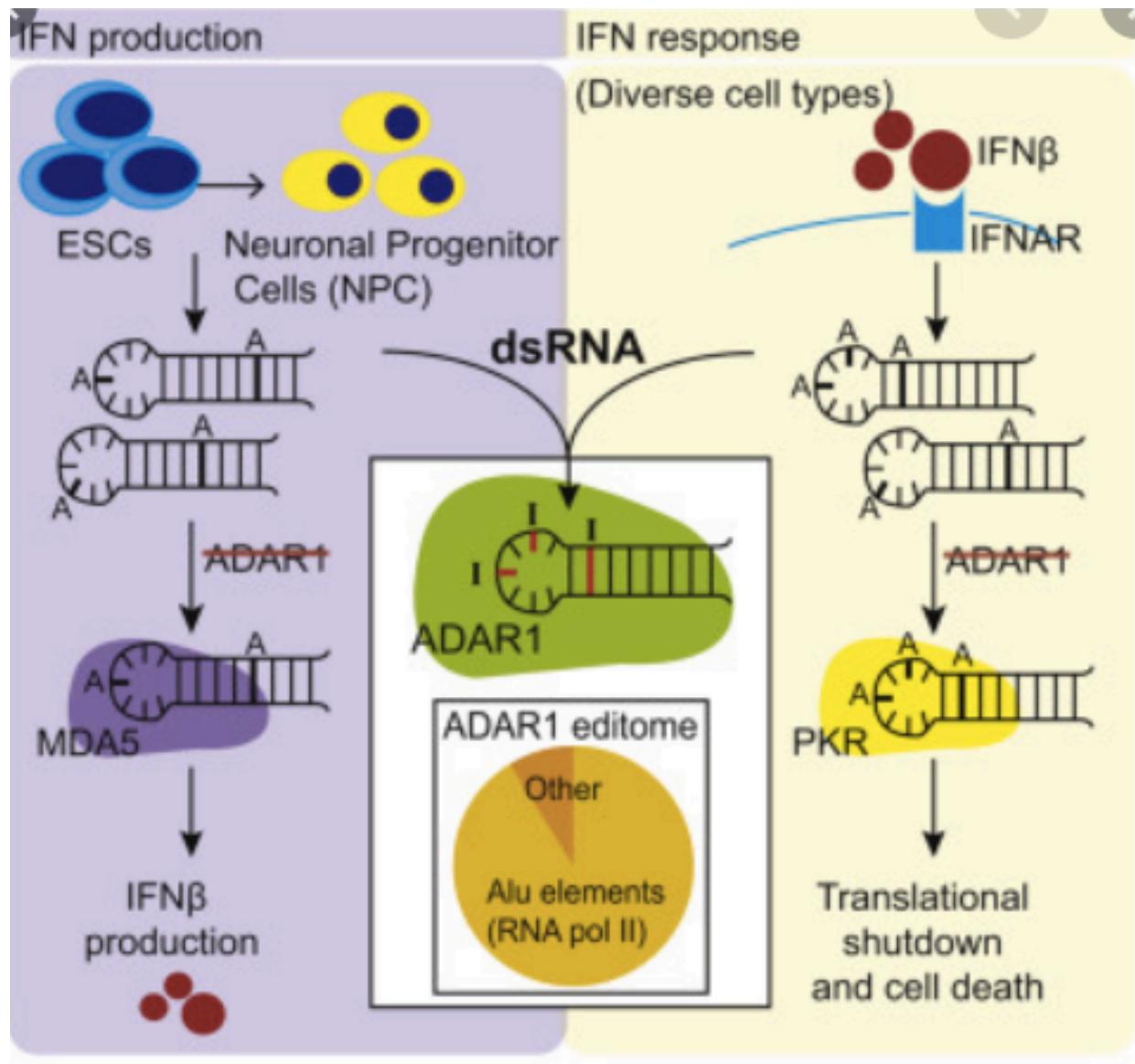


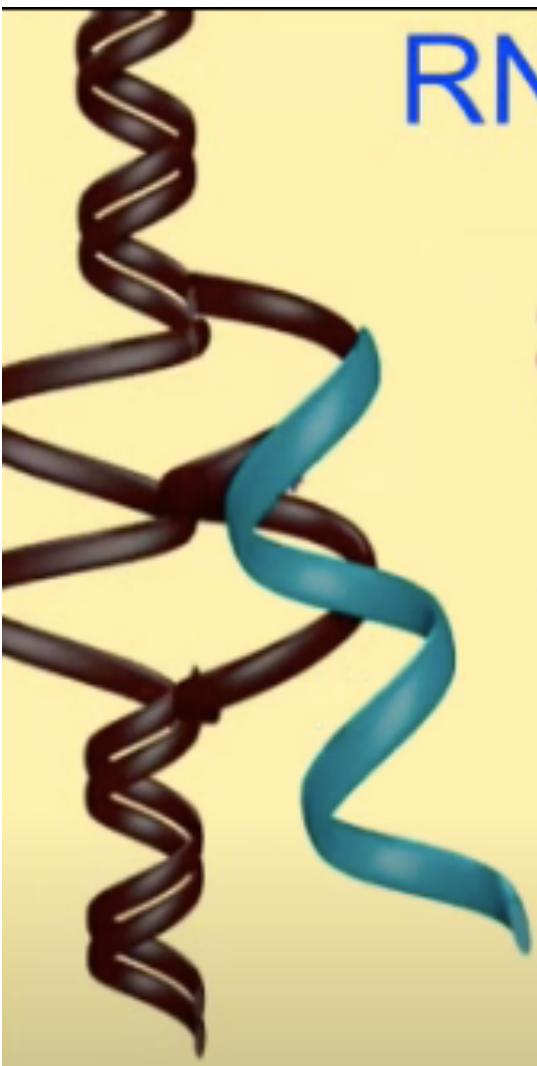
Flipons are programmable by a mix of epigenetic modifications

Additional more sophisticated functions

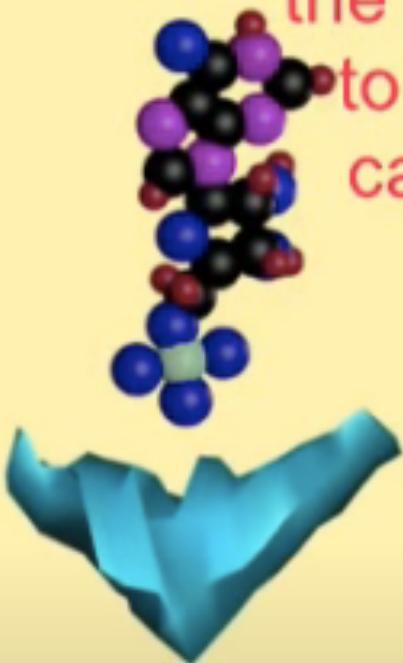
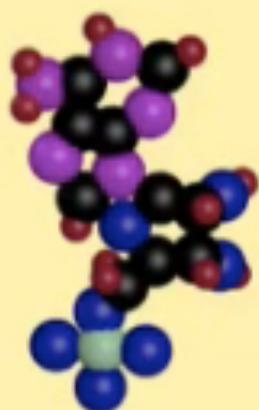
- Transposon recognition
- Z-DNA as a sensor of antiviral response
- Quadruplexes in telomers

Innate Immune Response





RNA EDITING



the adenine coded by
the DNA is changed
to inosine (which
can function like
guanine)

A, B and Z: The Structure, Function and Genetics of Z-DNA and Z-RNA

May 19th, 2021 [ONLINE]

Starting time: 5:00 am PDT / 6:00 am MDT / 7:00 am CDT / 8:00 am EDT / 1:00 pm BST / 2:00 pm CEST / 3:00 pm MSK / 9:00 pm JST

[Free Registration](#)



In celebration of 50 years of discoveries on Z-DNA and Z-RNA, we invite researchers to the inaugural ABZ meeting ("A, B and Z: The Structure, Function and Genetics of Z-DNA and Z-RNA") to be held on May 19th, 2021. This year's meeting will be virtual.

We will feature speakers from many different fields to increase awareness of the many innovative approaches available to further advance the field. Our aim to build community and to foster multidisciplinary collaborations.