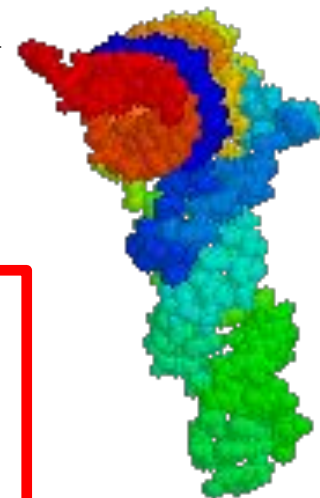


HTGAA 10-Feb-2026,
Reading & Writing Life

3-3:20, 4:40-5 pm
George Church



@geochurch



DOE
1987

2001



Blavatnik Institute
Regenesis Institute
SIAT

BROAD
INSTITUTE

WYSS INSTITUTE



Lipper 1997
Foundation

Full pfCOI list:
arep.med.harvard.edu/t

Core concepts

Why read, write & measure polymers?

Natural & designed properties

Chirality, reactivity, etc.

Codes:

Basepair code: A=T, C=G etc.

Ribosomal translation: 4 RNA $\square$ 20+AA

Interactions: Protein $\square$ DNA

How to Read almost anything?

De novo vs ID, in vitro vs in situ imaging

Size of side chains, Fluorescence

Single molecules: Why?

How to Write almost anything?

Sequential, Exponential, Editing

Post-polymerization modifications

How to Grow/Read Almost Any Polymer ...

Easy for cells

Polynucleotides

Polypeptides

Polyterpenes

Polyketides

Polyhydrocarbon

Polyynes

Polysaccharides

Poly amino acids

Polypyrroles

Polyesters

Uses

DNA, RNA

Collagen, Hb, silk

Rubber

Antibiotics

Wax, fats

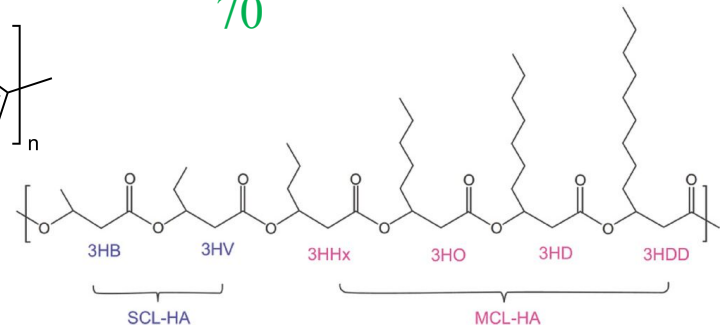
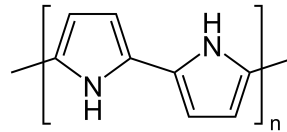
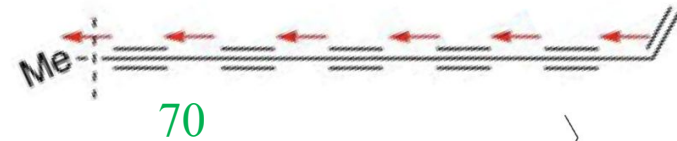
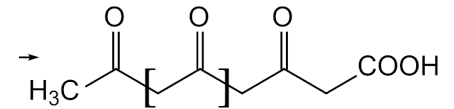
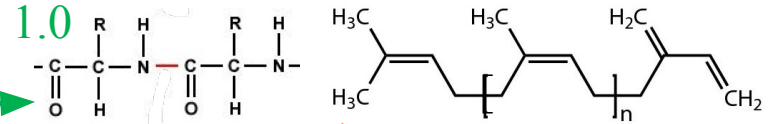
Pigment, antibiotic

Cellulose

Lignin

Heme

PHA



Hard (so far)

Peptide nucleic acid

Graphenes(nanotubes)

Diamonoids(nanothreads)

Uses

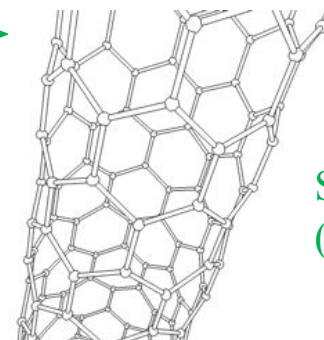
Reverse translate

Strong

Strong



60



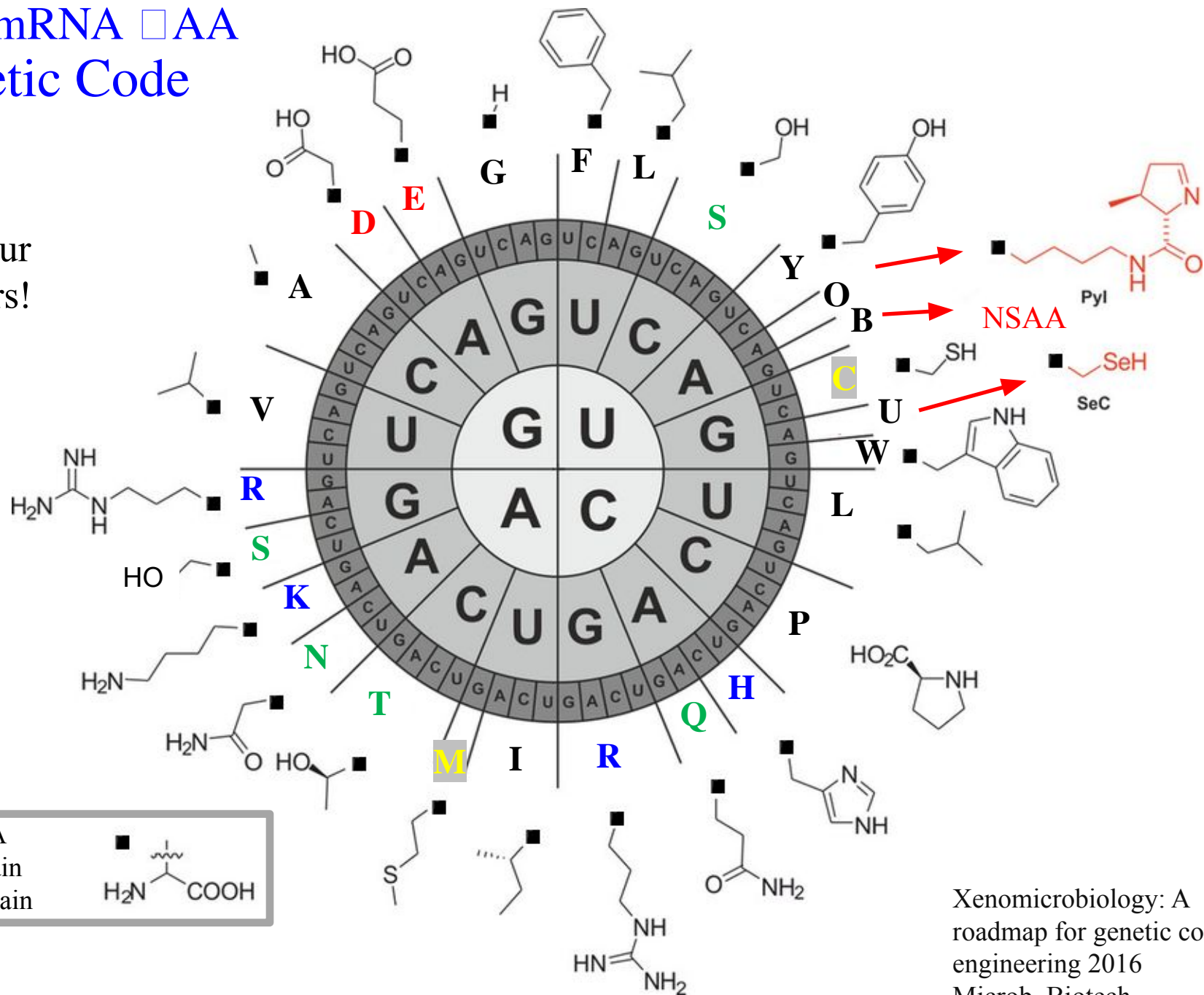
50

Specific strength
(kN·m/kg)

DNA □ mRNA □ AA

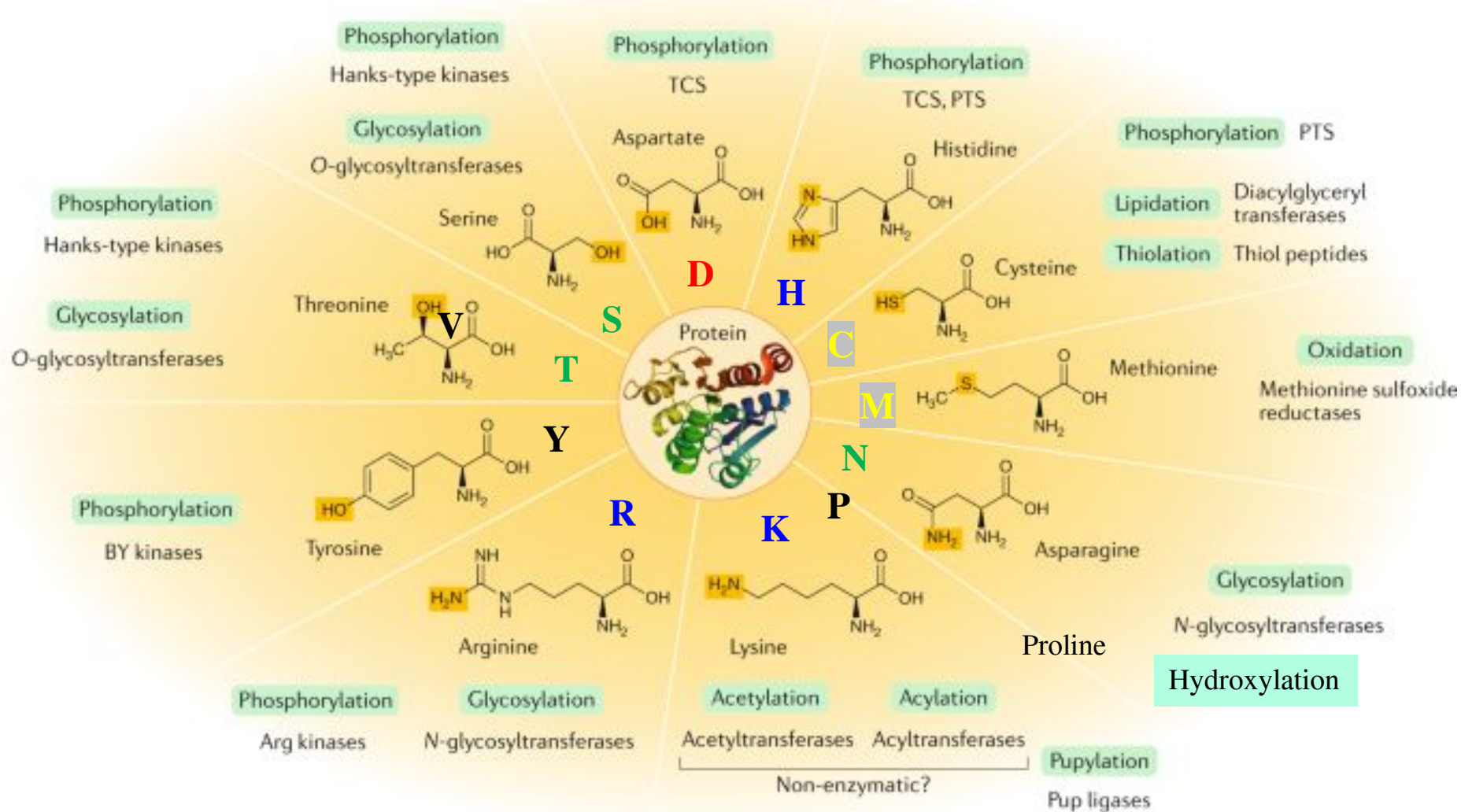
Genetic Code

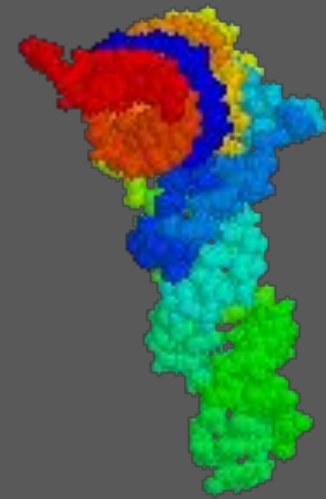
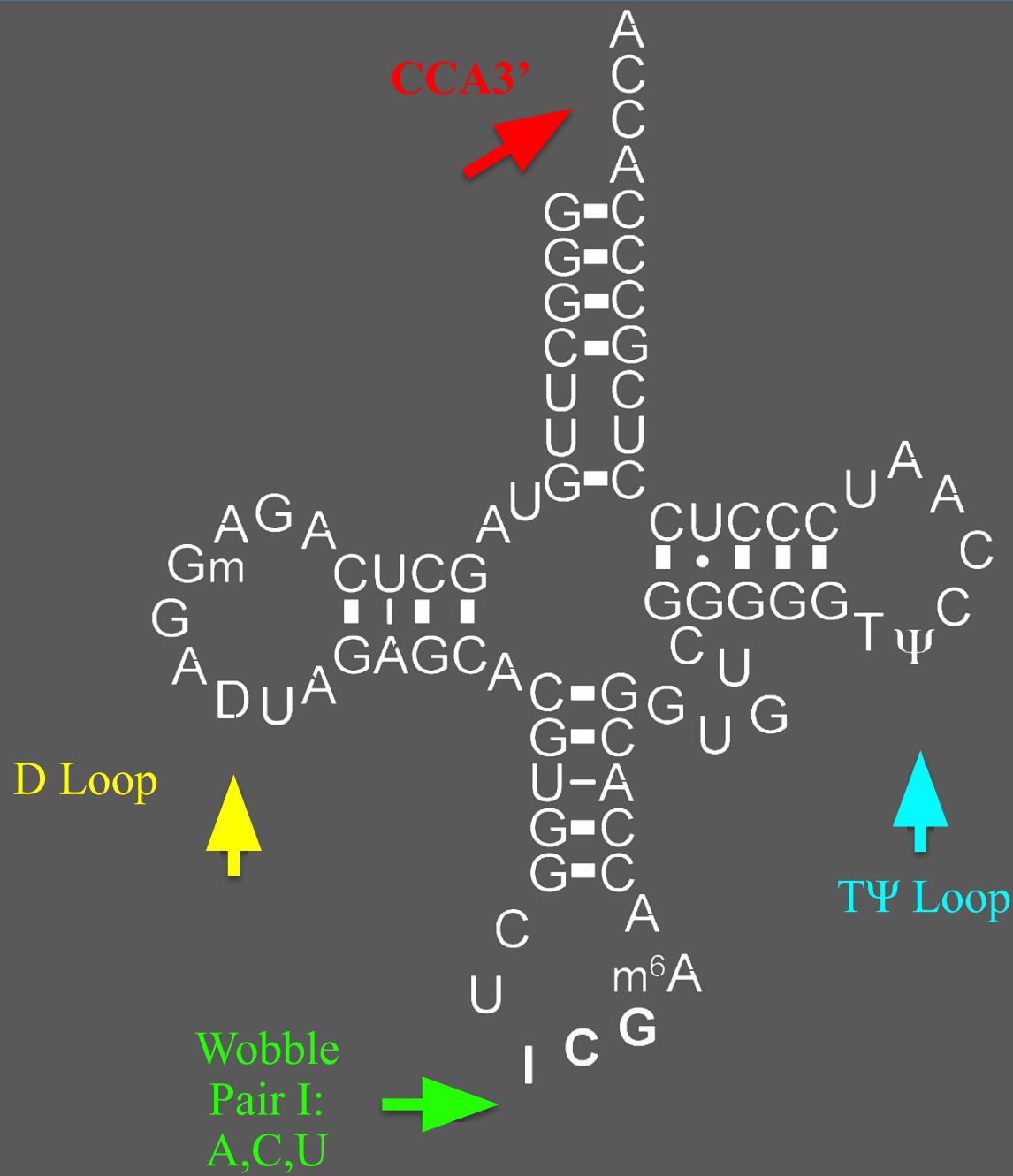
Know your monomers!



Xenobiology: A
roadmap for genetic code
engineering 2016
Microb. Biotech

Protein Posttranslational modification (PTM)

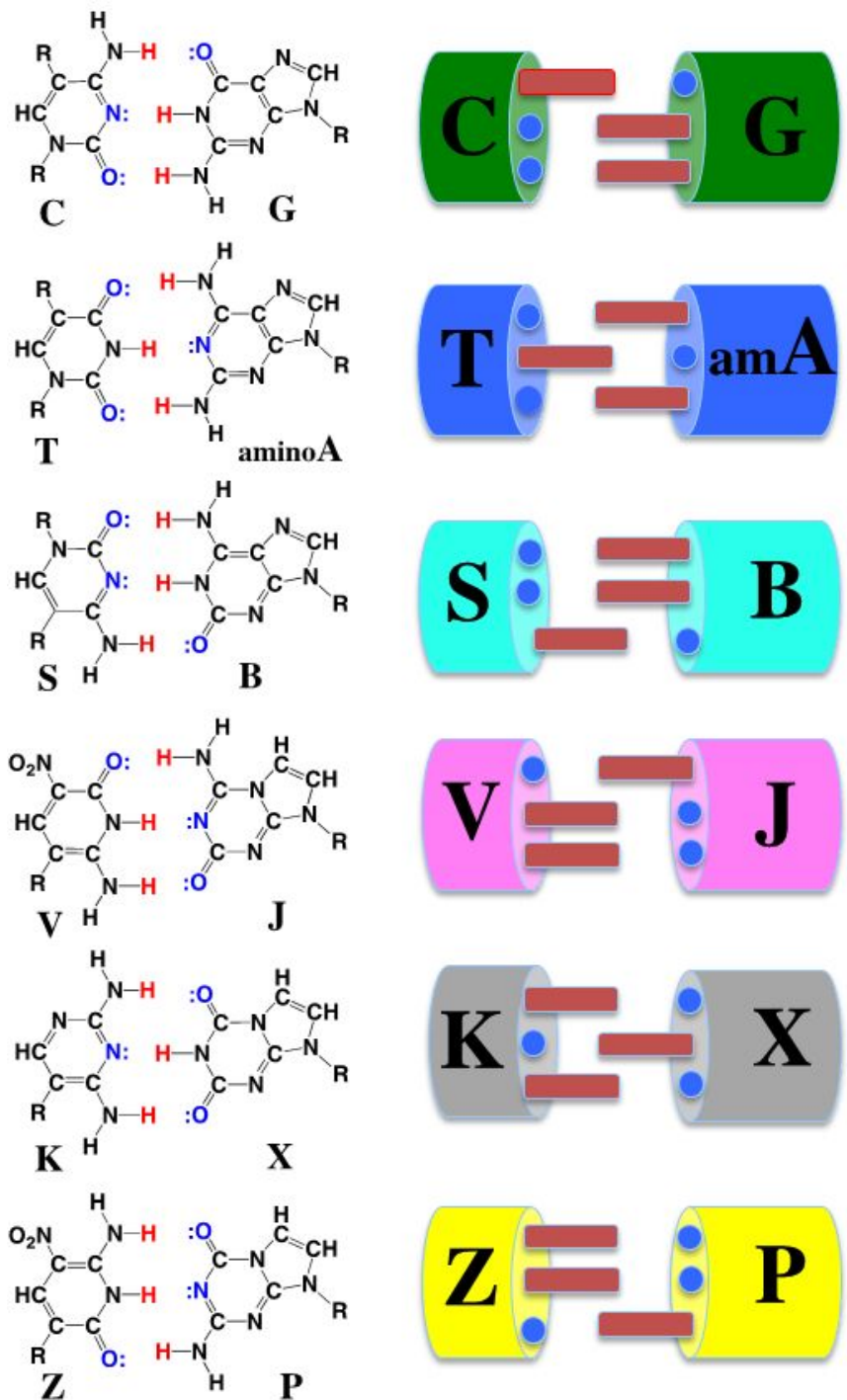




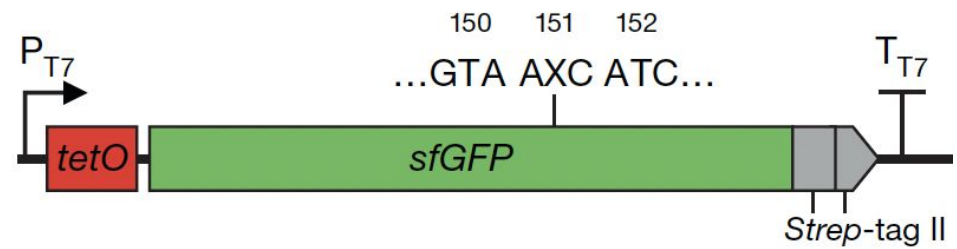
AARS Code
AA□tRNA
Synthase

Watson-Crickoid from 2 to 6 basepairs ...emphasis on H-bonds

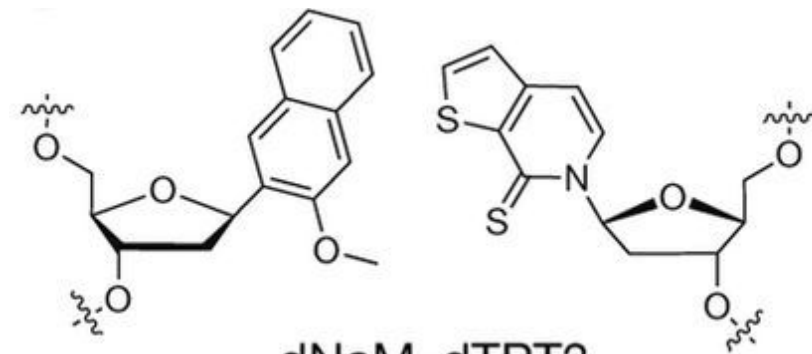
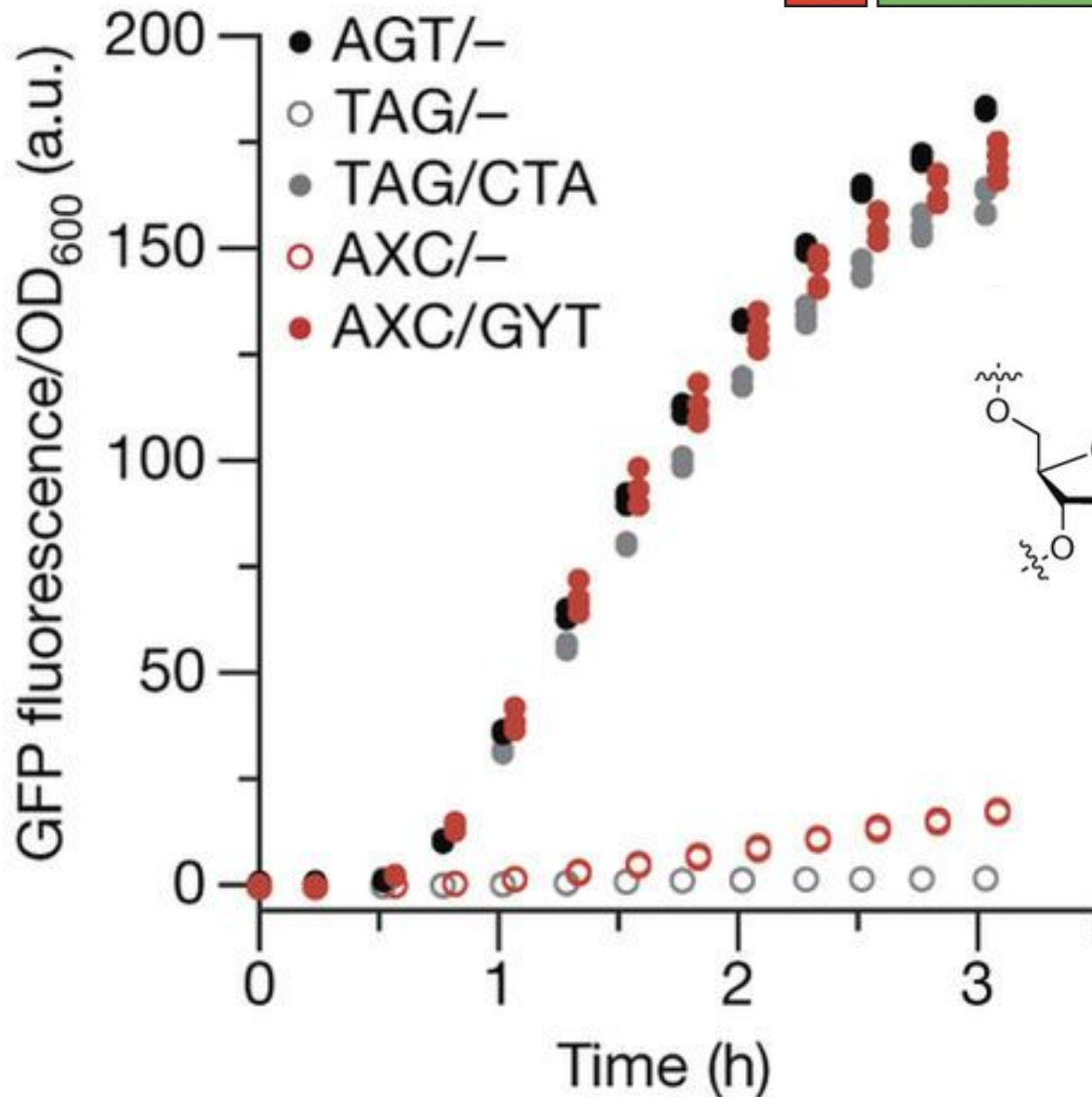
Artificially Expanded Genetic
Information Systems for New
Aptamer Technologies
Biondi & Benner (2018)
Biomedicines.



Or no H-bonds!



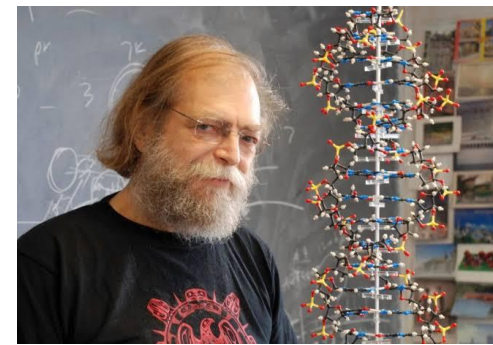
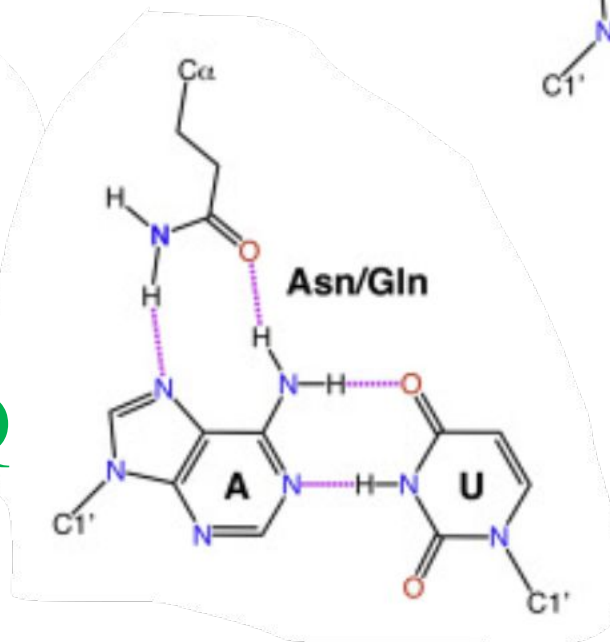
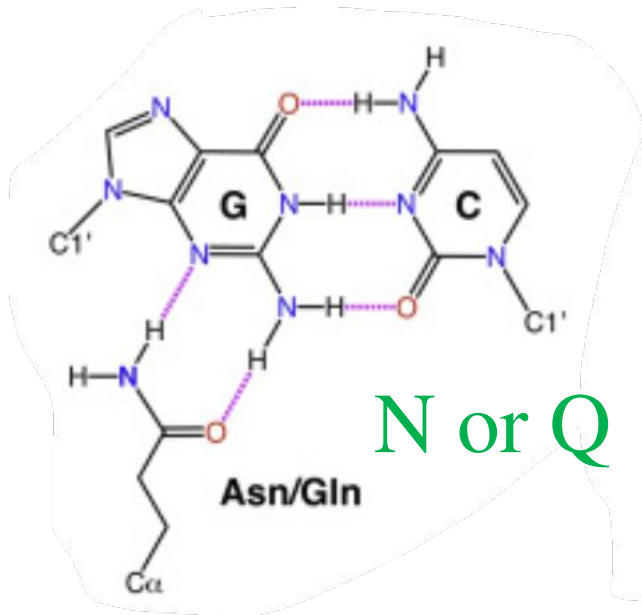
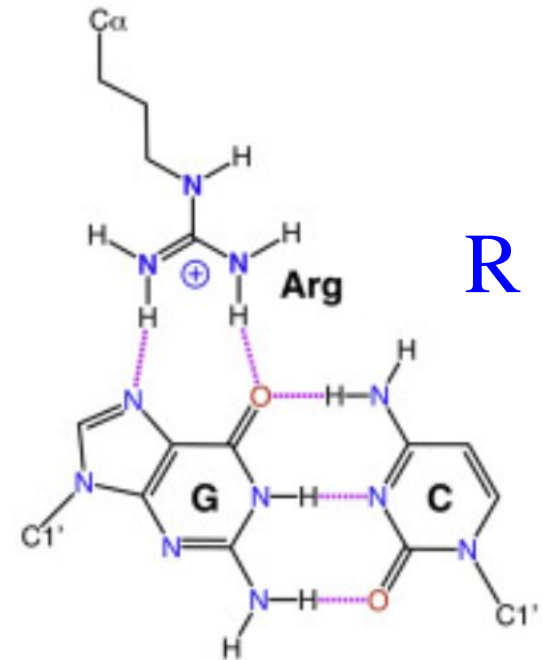
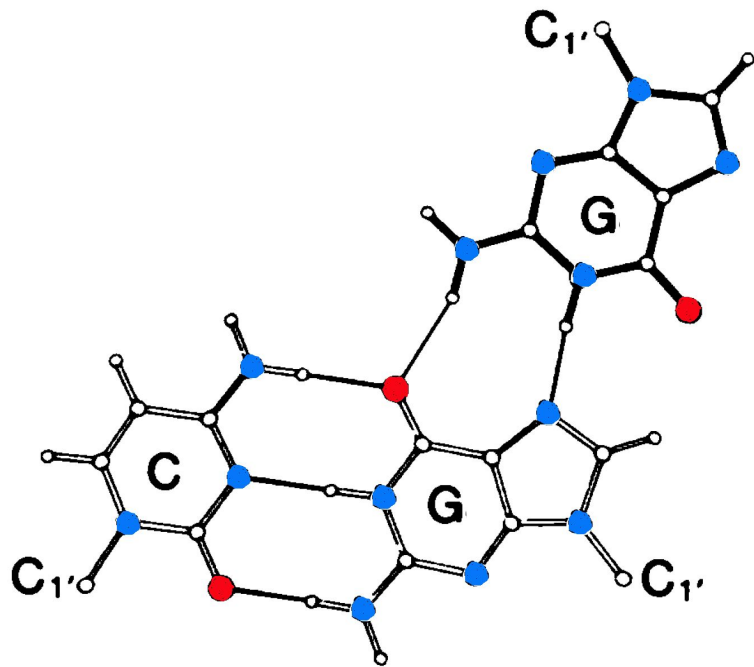
DNA □ mRNA □ AA

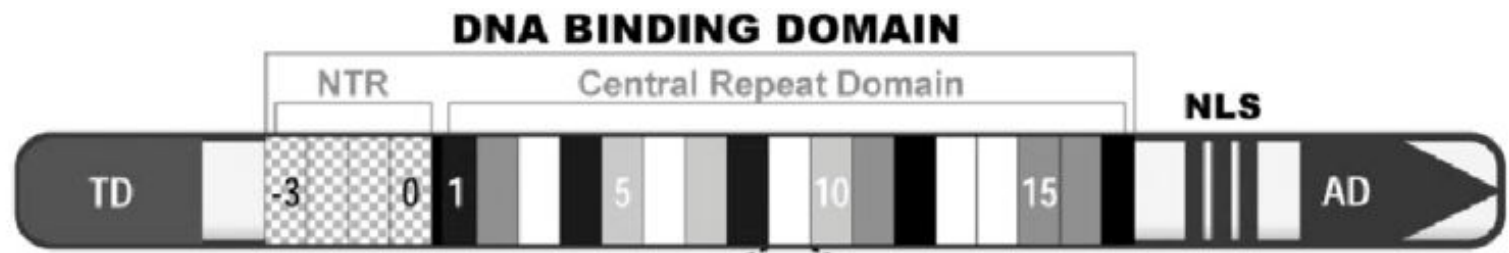


X - Y

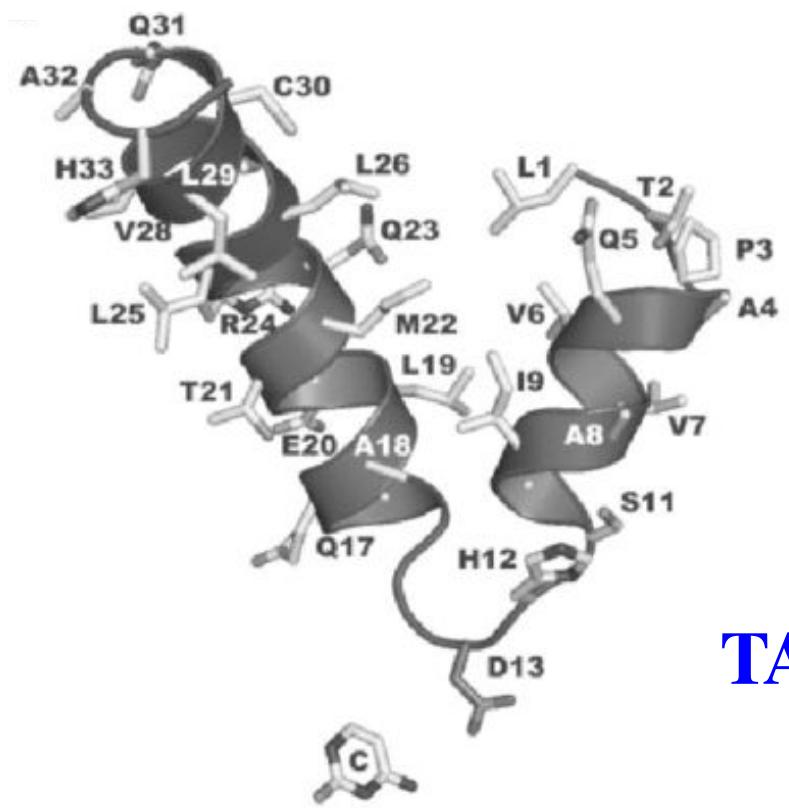
Zhang ...Romesberg
2017 Nature.

Ned Seeman Predictions PNAS 1976 based on non-WC bp





12/13
 1 LTPEQVVAIAS NG GGKQALETVQRLLPVLCQAHG 34
T



C

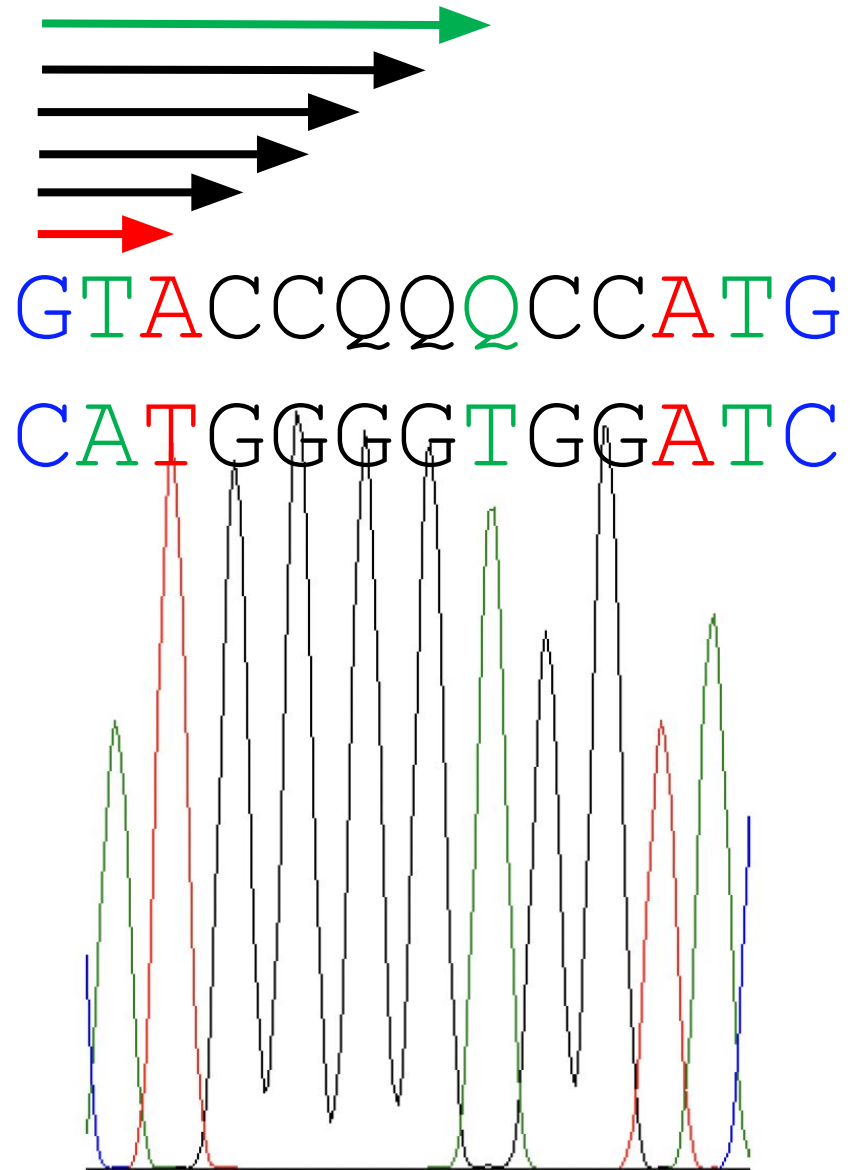
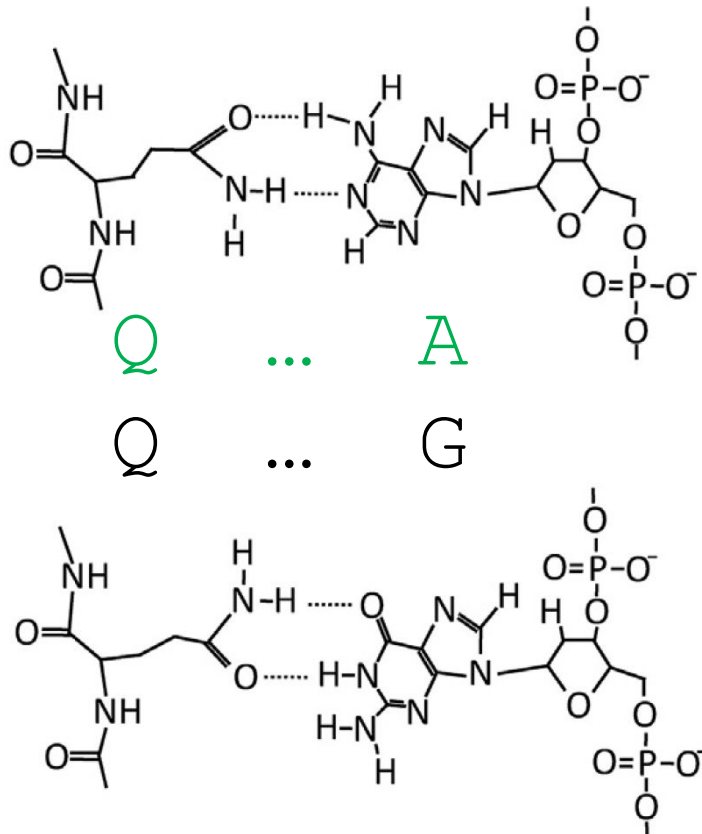
standard RVDs

HD	NH	NK	NG	NI	NN	NS
C	G	G	T ^MC	A	G A	G C A

TALE Code 2 AA □ 1 DNA bp

PCR seq PNA (Sanger ABI)

Kuwayama BBRC 2021



DNA, RNA reading strategies

Seq by Synthesis

Illumina+

MGI+

Element

UltimaGenomics

Singular

454-hv

Ion Torrent-H⁺

Sanger-ABI !

Single Molecule

PacBio

ONT nanopore

Roche nanopore

Roswell transistor

Fluorescent 3D

Bruker Vutura

10X Xenium

Protein reading

Single Molecule

Aminopeptidase

Fluor-labeled ☐

Mass Spectrometry

LC-MS-MS CAD

Edman Degradation+

+ = Reversible

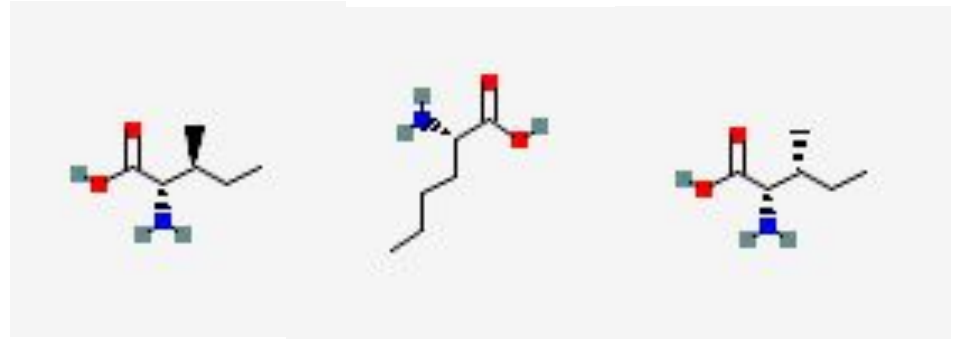
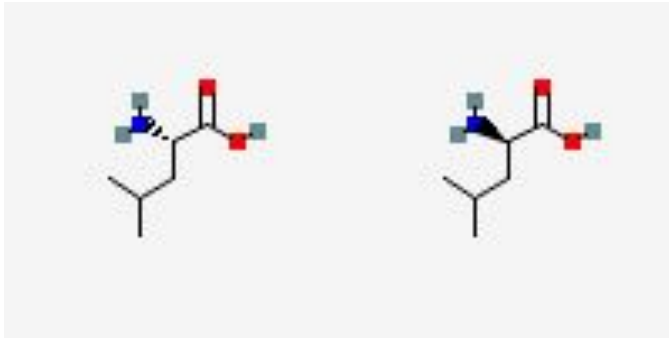
Terminators

! = Irreversible

Mass-spectrometry

L-Leu, D-Leu,

Ile, NorLeu, AlloLeu

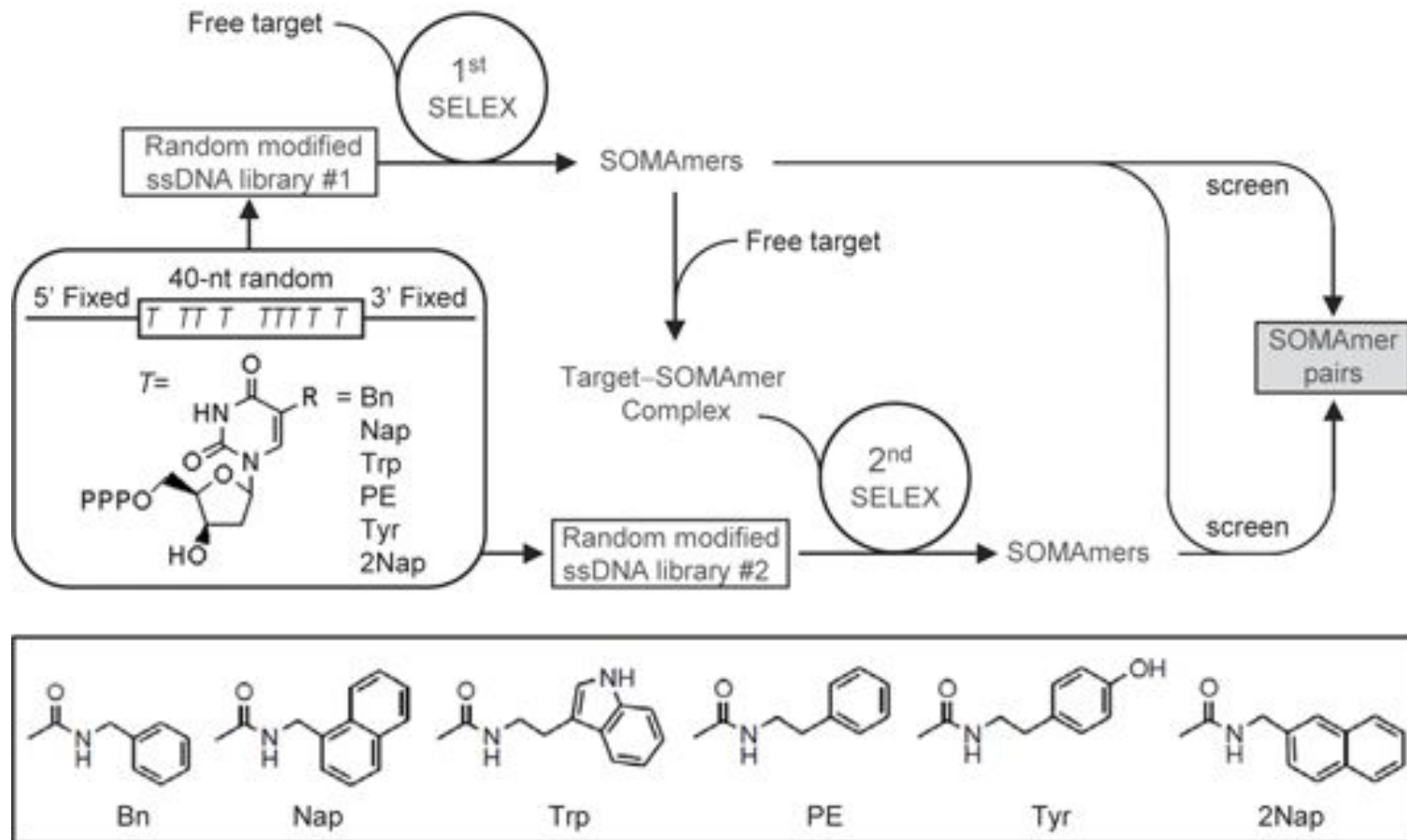


Same Average mass 131.173 Da

Same Monoisotopic mass 131.095 Da

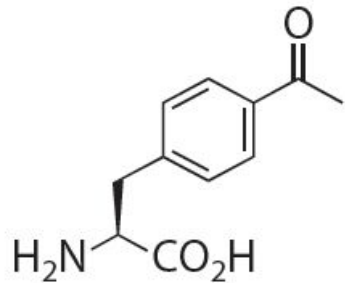
Fixes: Chiral chromatography

Collisionally activated dissociation (CAD)

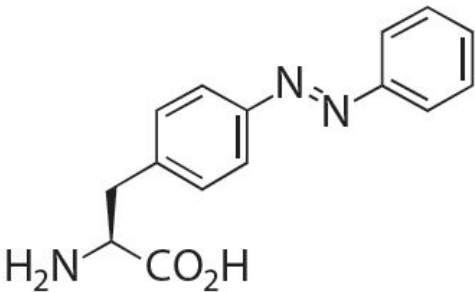


‘Aptamers’: Improving low chemical diversity of DNA Ochsner ...Janjic. 2014 Biotechniques.

Non-standard Amino Acids



AA	Side Chain	#atoms(non-H)	Key AA Features
Ketone		9	Orthogonal chem



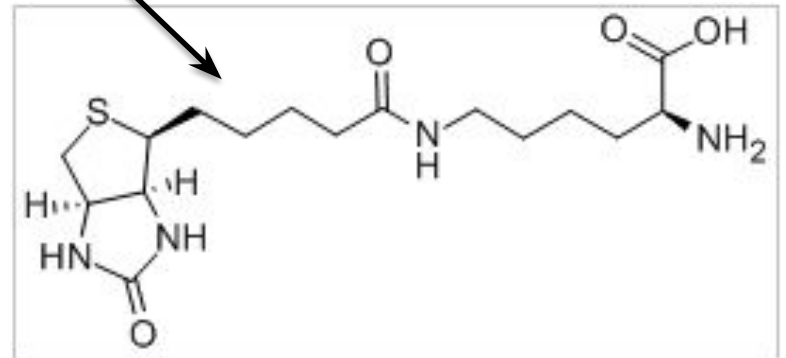
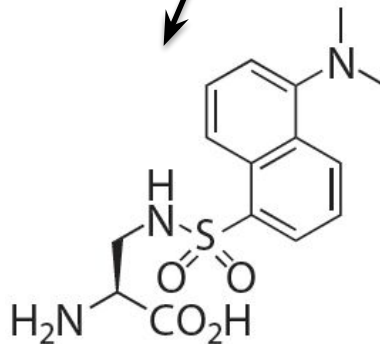
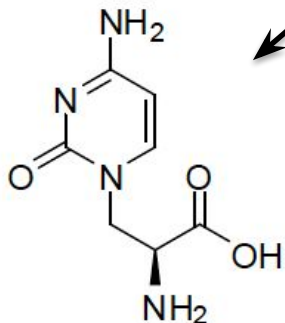
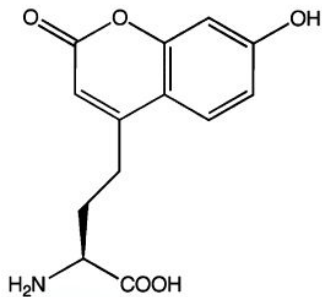
Azobenzyl	14	Photoisomers
-----------	----	--------------

Coumarinyl	13	Redox
------------	----	-------

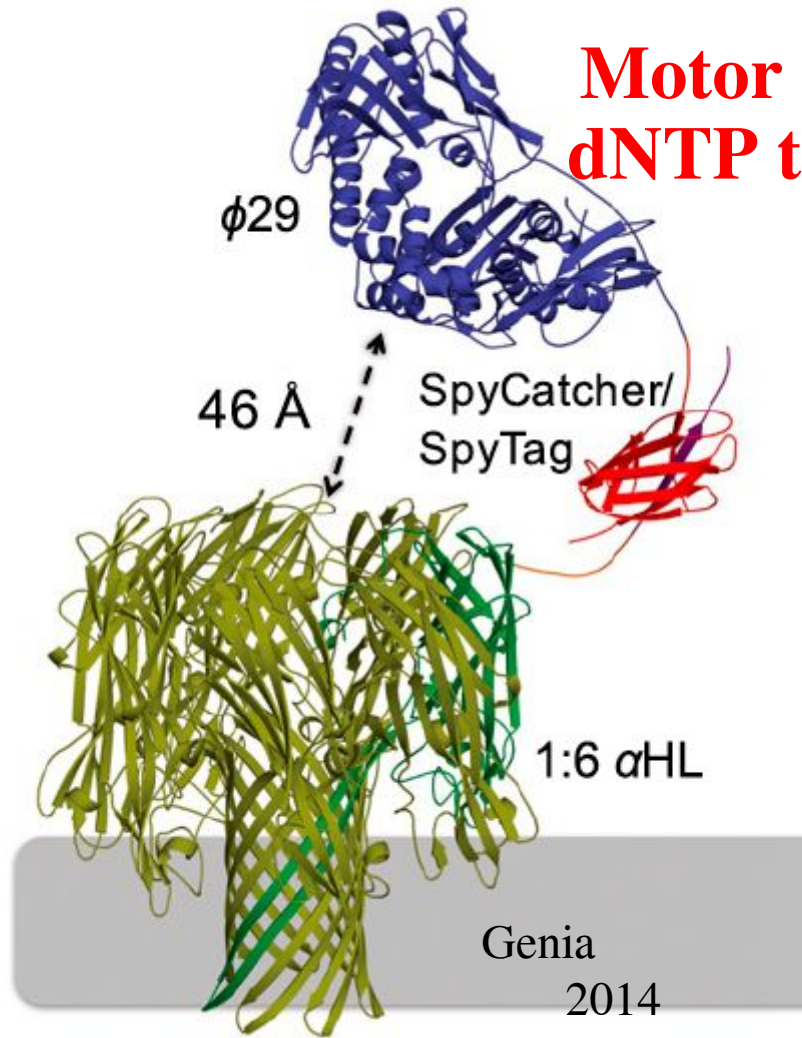
Cytosinyl	8	Nucleopeptide
-----------	---	---------------

Dansyl	17	Fluorescent
--------	----	-------------

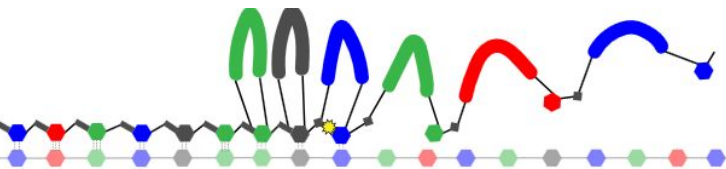
Biotinyl-Lys	19	Binds Avidin
--------------	----	--------------



Motor (pol) + pore: dNTP tags v. ssDNA

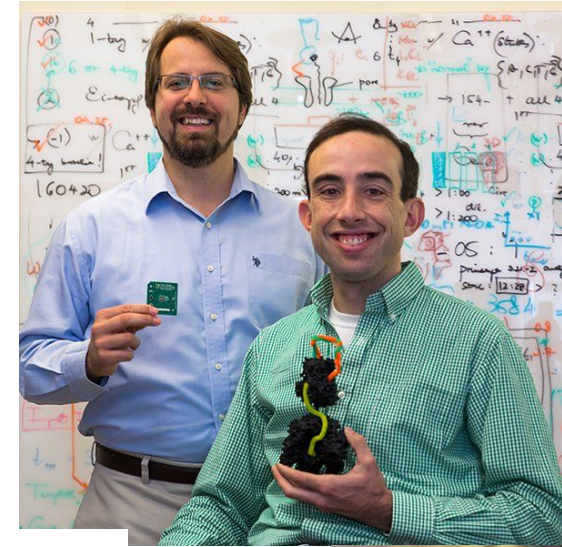


2020
Stratos

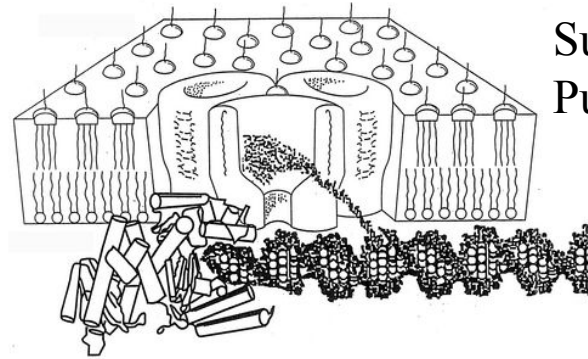


2016 PNAS
2020 ACS Nano

Mirko Palla
Ben Stranges



Sukanya
Punthambaker



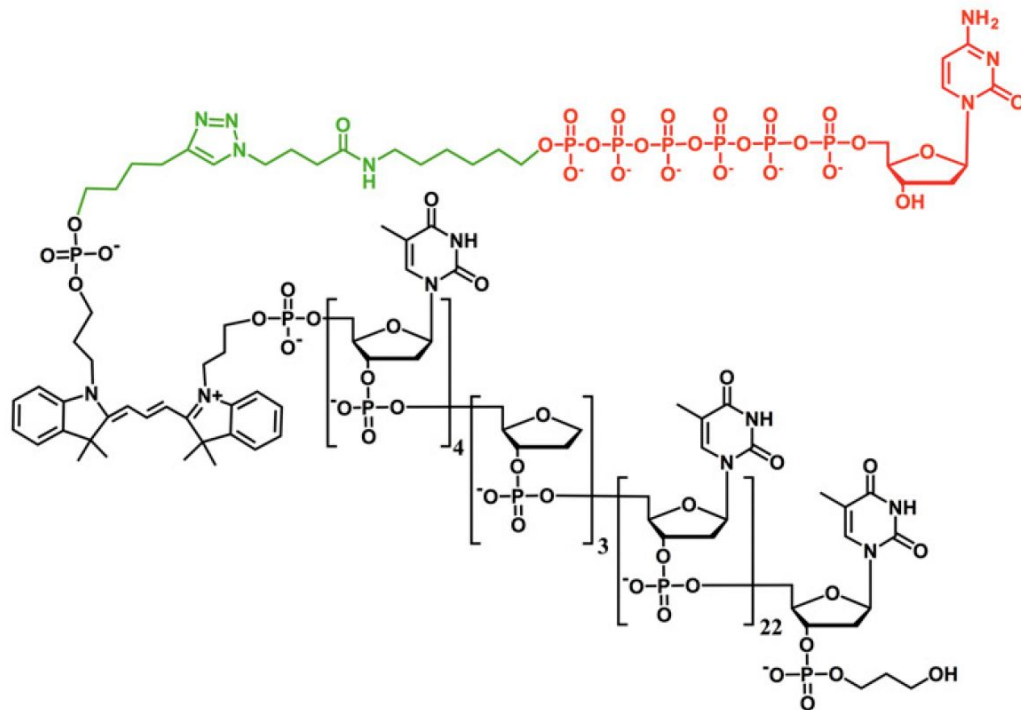
1987-98: US5795782 Church, Deamer,
Branton, Baldarelli, Kasianowicz

Motor (helicase)
ss-DNA through pore

Oxford
Nanopore

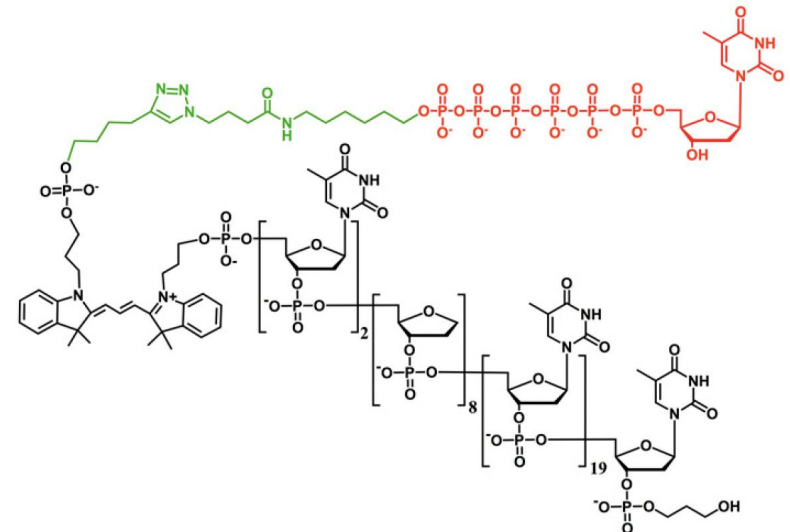
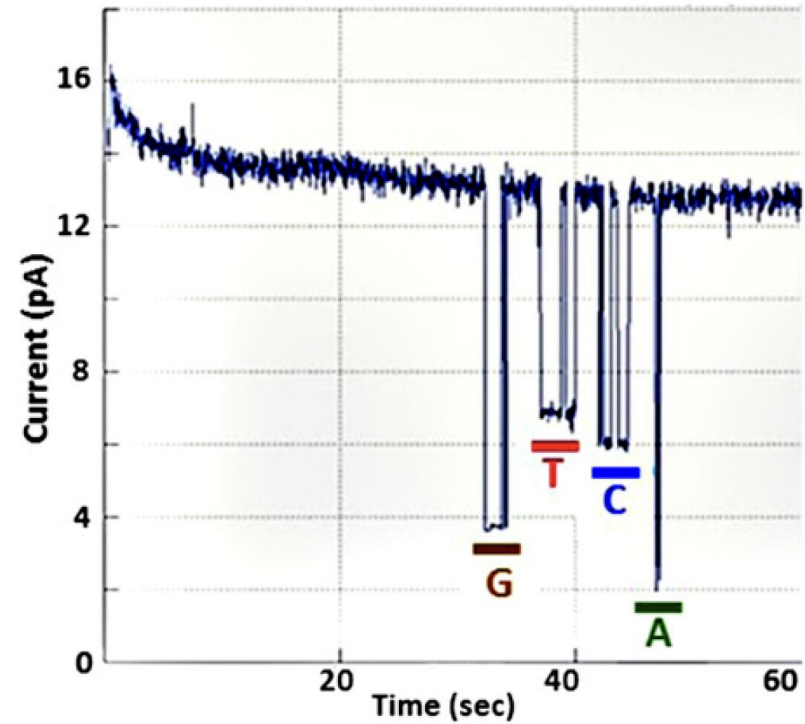
Motor (pol) + pore
dNTP tags
1 molecule, 1 AMU

Stranges ... Church GM (2016) PNAS.
Fuller ... Ju J (2016) PNAS



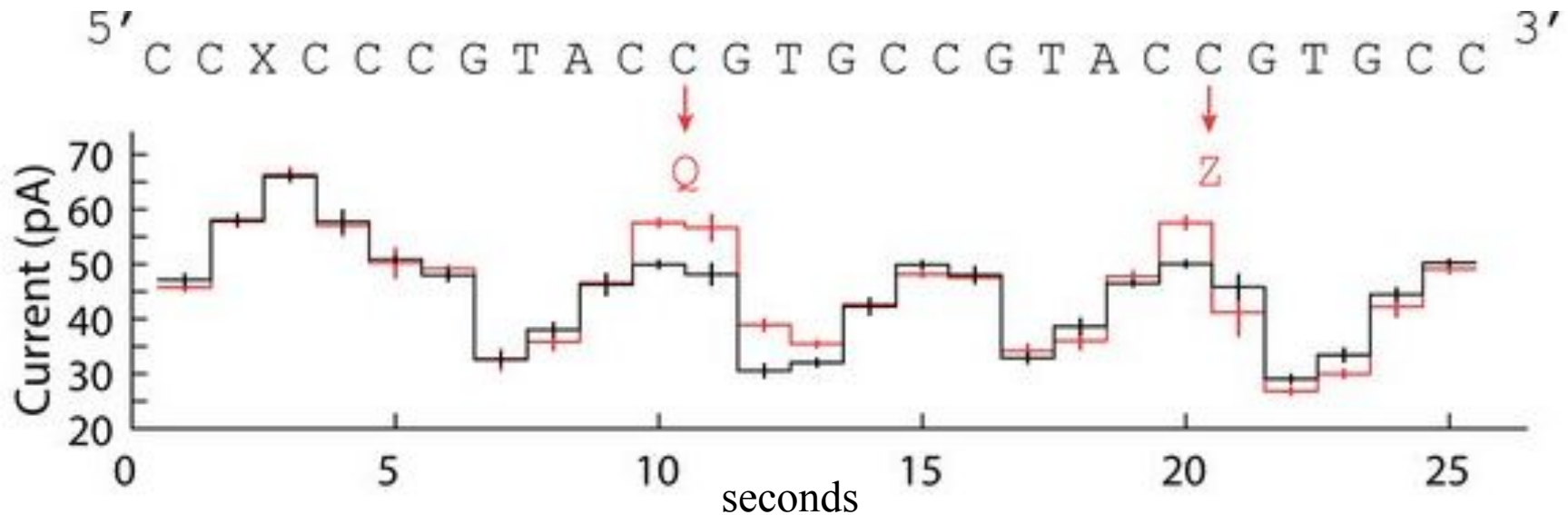
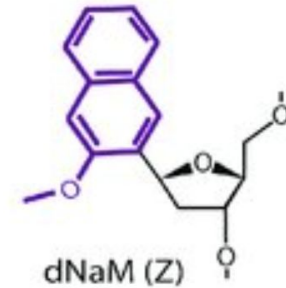
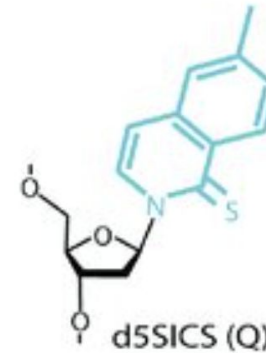
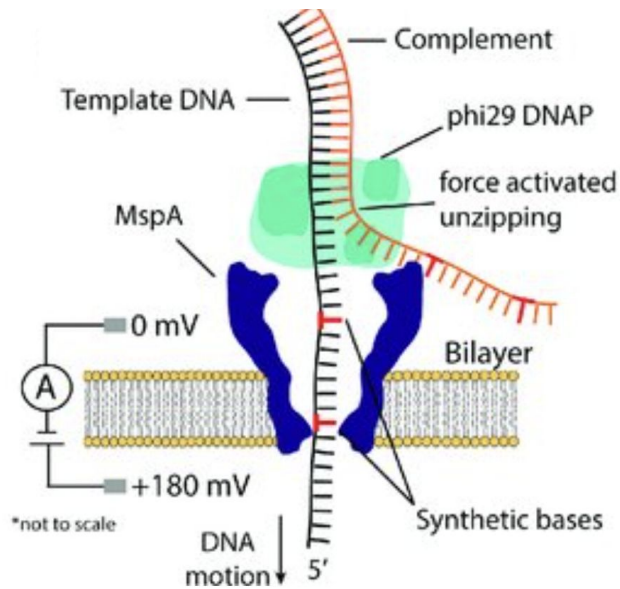
Extended Primer: 5'—GTCA—3'

Template: 3'—CAGT—5'



Nanopore: Modified & Non-standard bases

Craig ... Gundlach 2015 Direct detection of unnatural DNA nucleotides dNaM and d5SICS using the MspA nanopore. PlosOne



Miniaturize: Nanopores in space

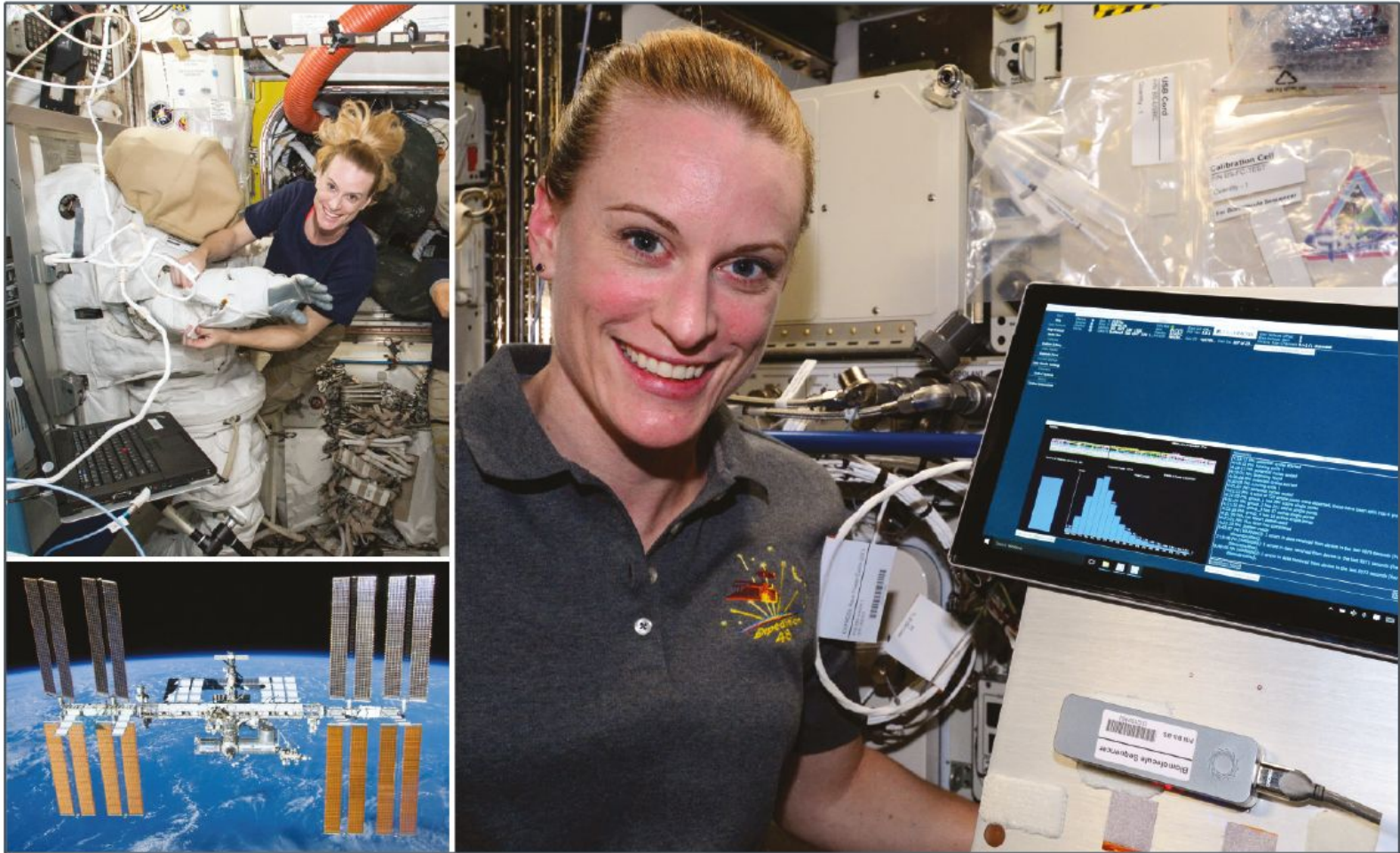


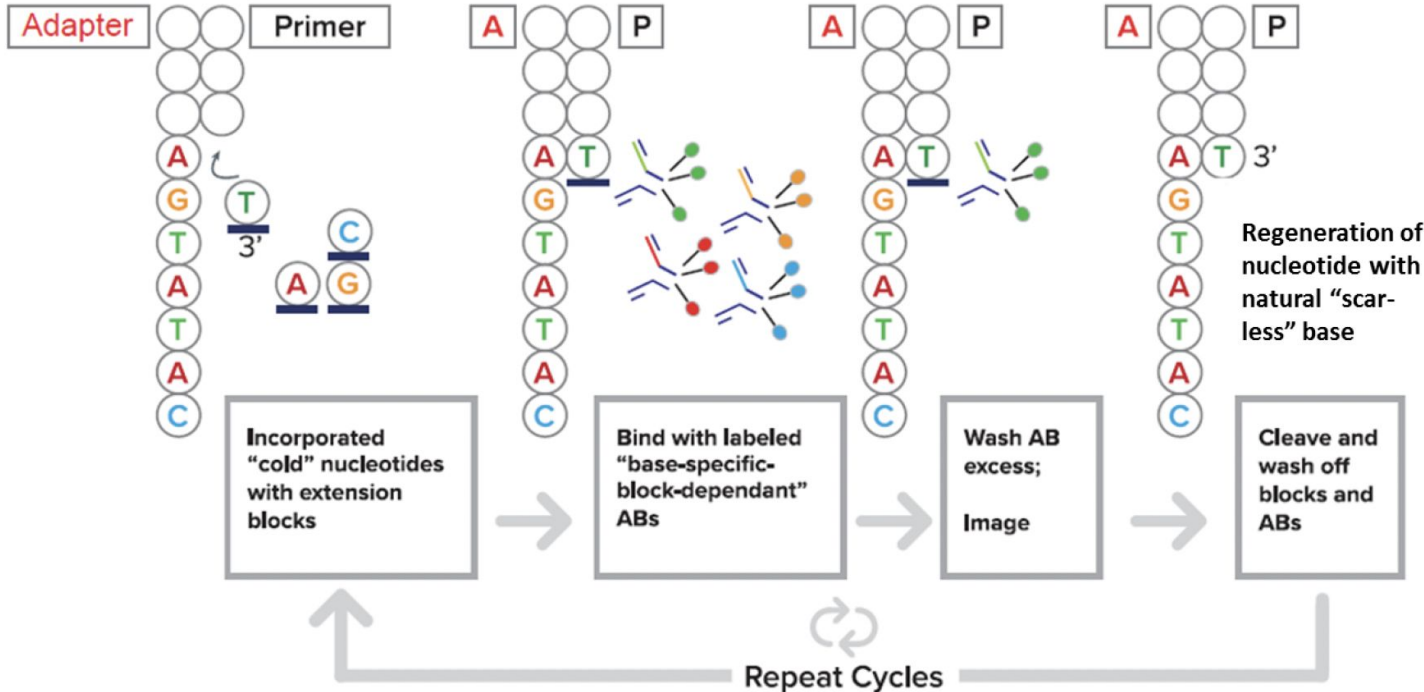
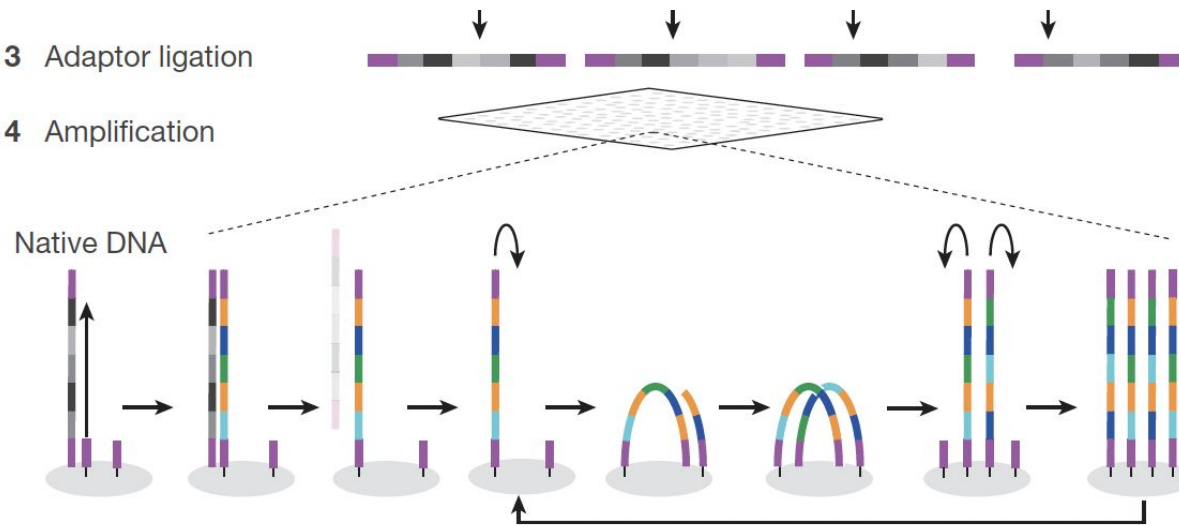
Fig. 1 Astronaut Kate Rubins on the ISS

Read DNA via Fluorescent Microscopy

3 Adaptor ligation

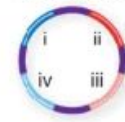
4 Amplification

Native DNA



a Library generation

Library molecule

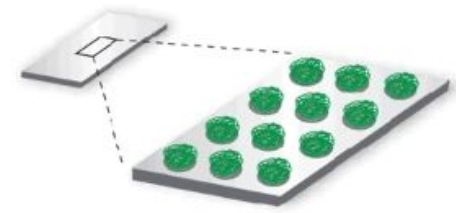


Rolling circle amplification



DNA nanoballs

b Arraying



Patterned array in low-volume flowcell

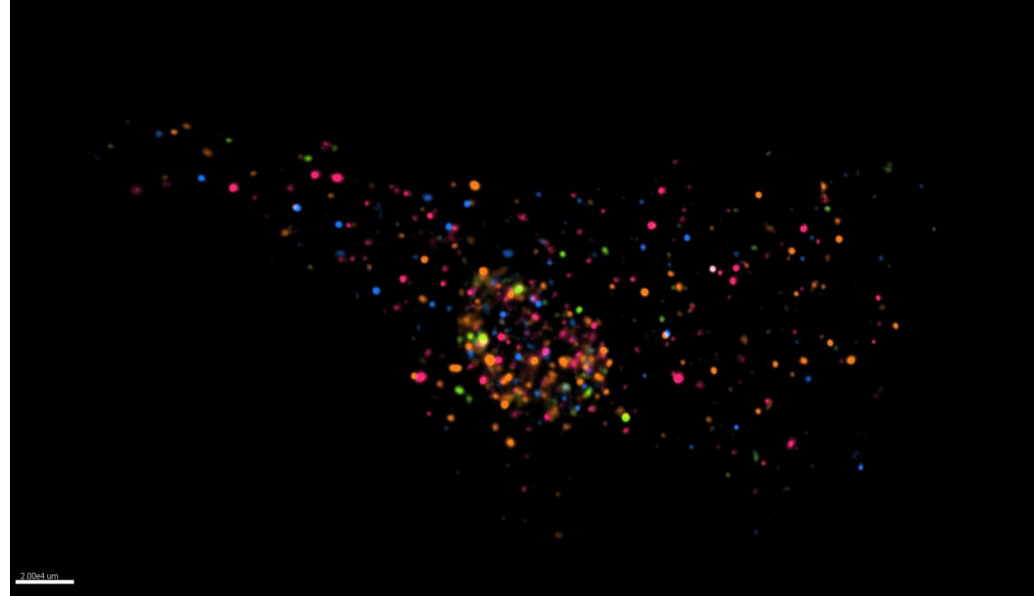
Drmanac et al.
CoolMPS:
evaluation of
antibody labeling
based massively
parallel
non-coding RNA
sequencing NAR
2020

Spatial Omics Microscopy

DNA, RNA, protein
in situ 3D (hyb, lig, pol)
2 nm to 40 μm thick
0.2 - 400 nm precision

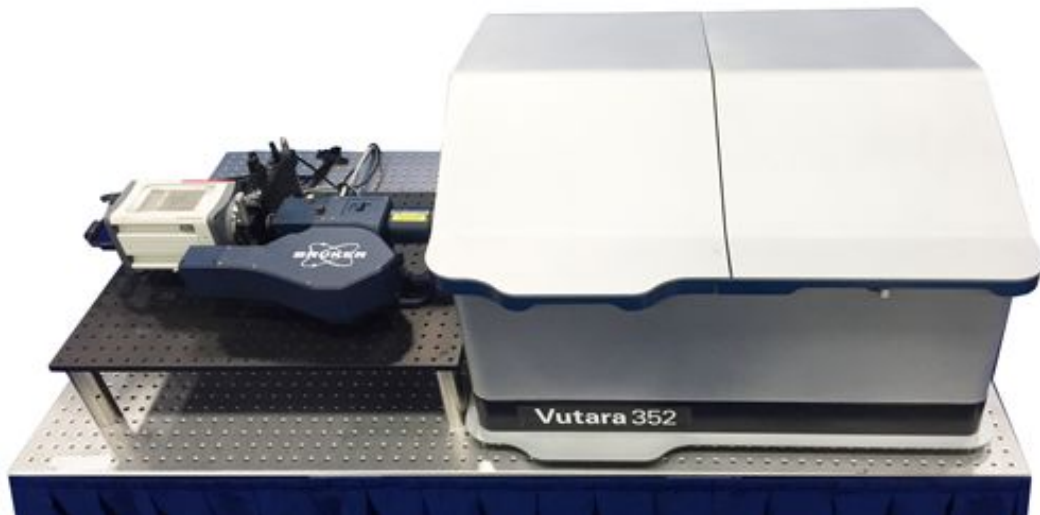


Lee, Daugharthy ...
Church
Science 2014,
Nature Protocols. 2015



10X Genomics (Church lab)

Bruker Vutara (Wu lab)

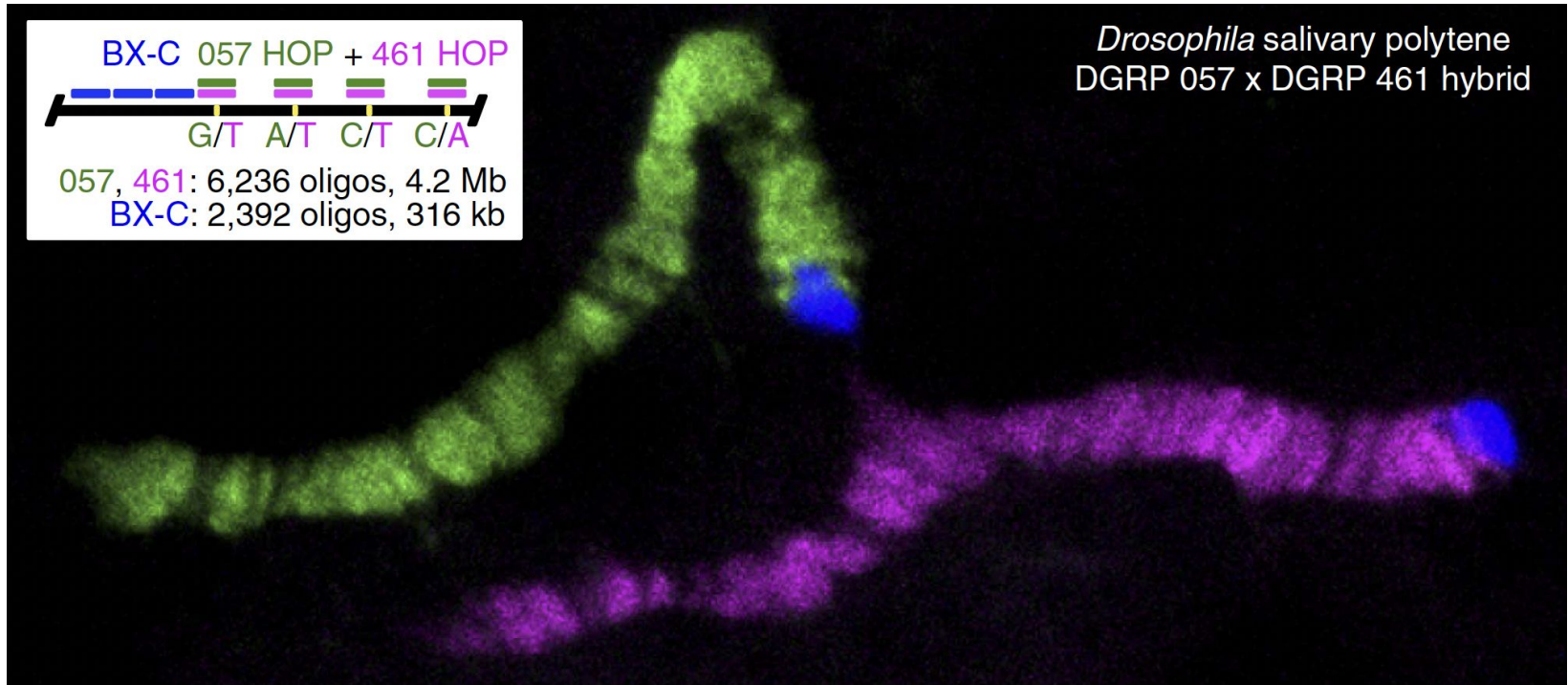


10X Xenium 2022

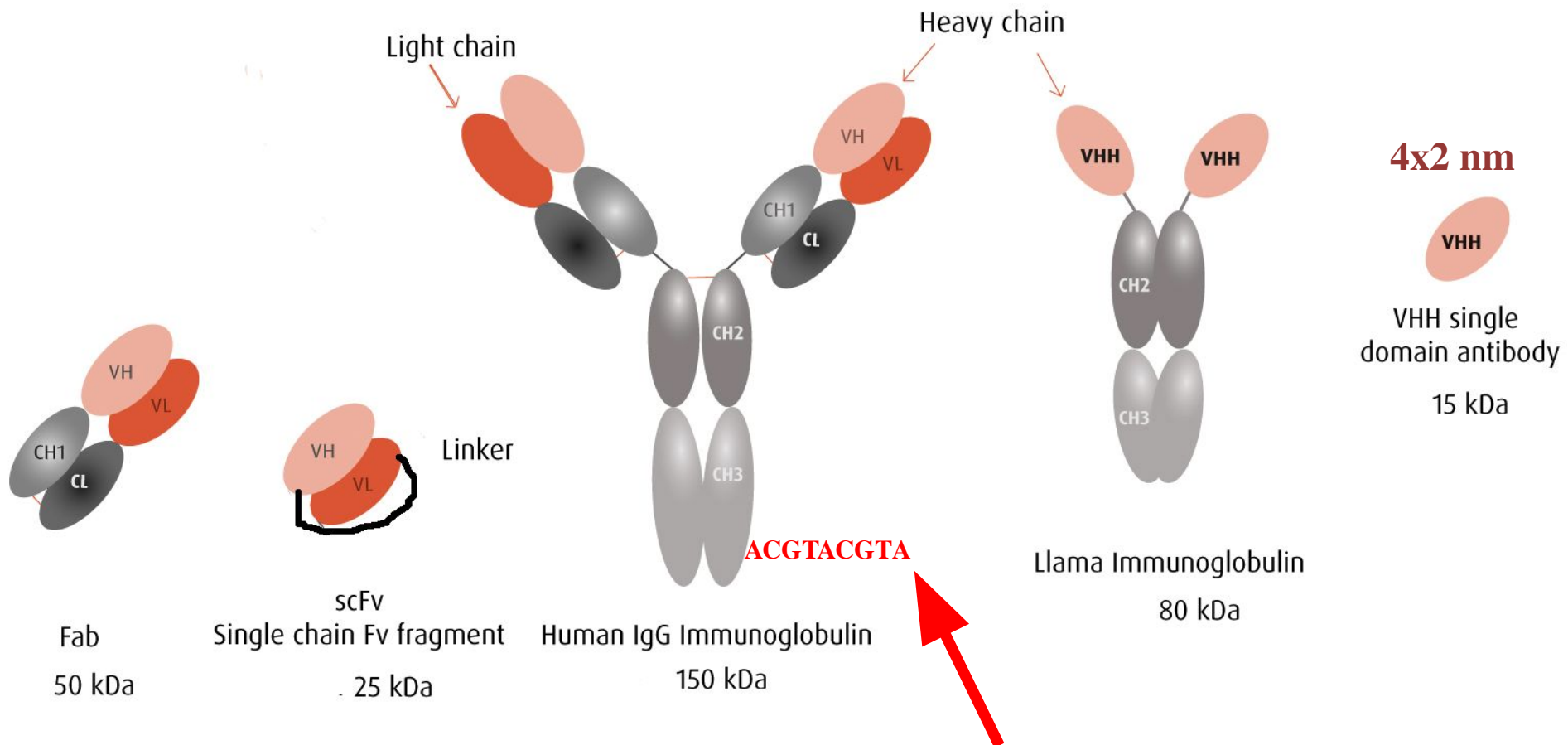


Single-molecule super-resolution imaging of chromosomes & in situ haplotype visualization using Oligopaint FISH probes

Beliveau ...Wu (2015) Nature Comm.

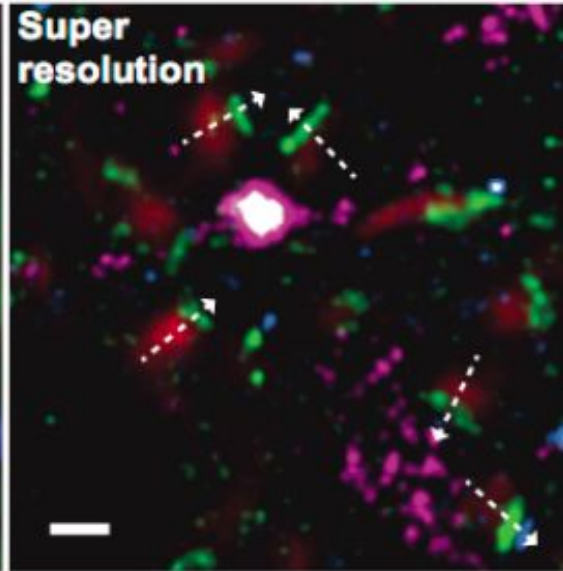
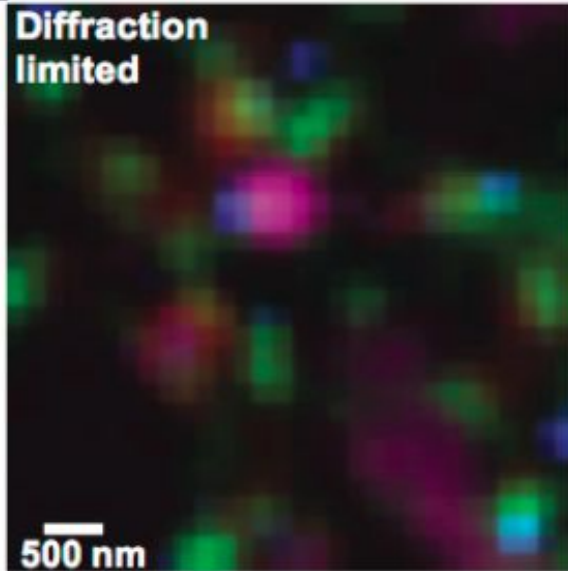
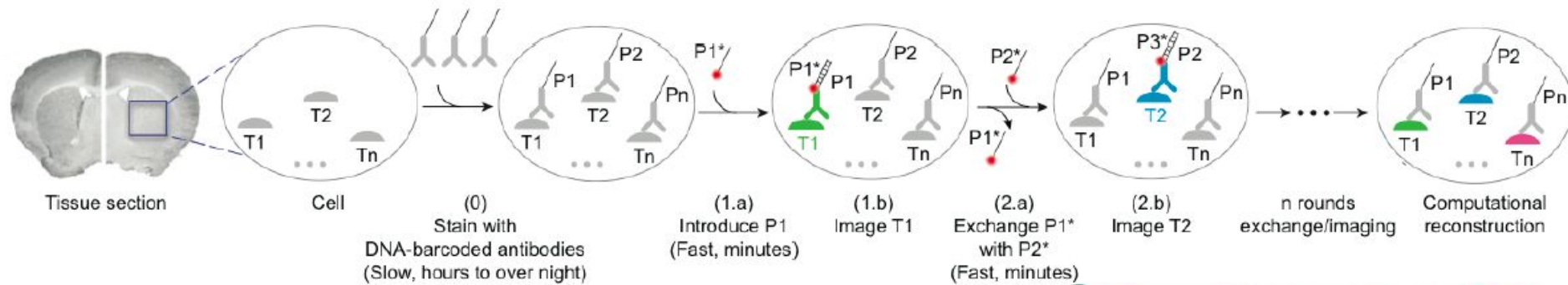


Oligonucleotide-labeled **Protein-Protein** Recognition

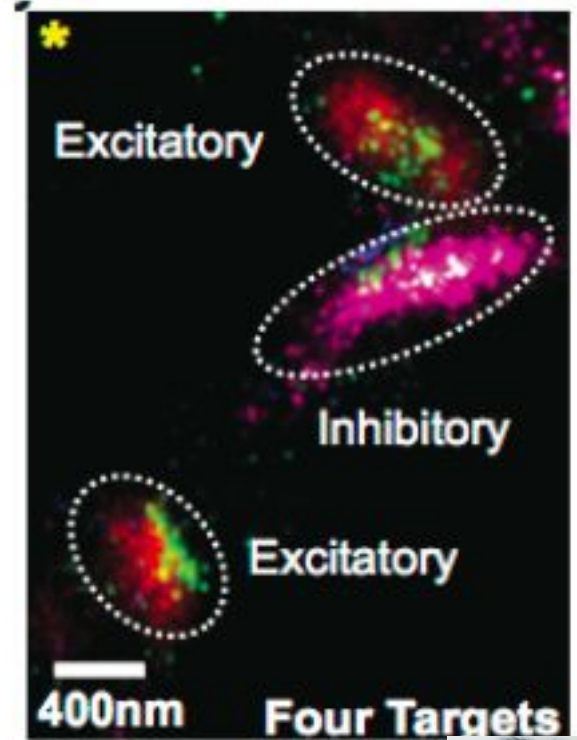


DNA Exchange Imaging (DEI): Super-resolution
In progress: Multiplex ML-Syn of nanobodies.

Proteins: *in situ* DNA-Exchange multiplexing 20



VGAT
 Gephyrin
 SynapsinI
 Bassoon



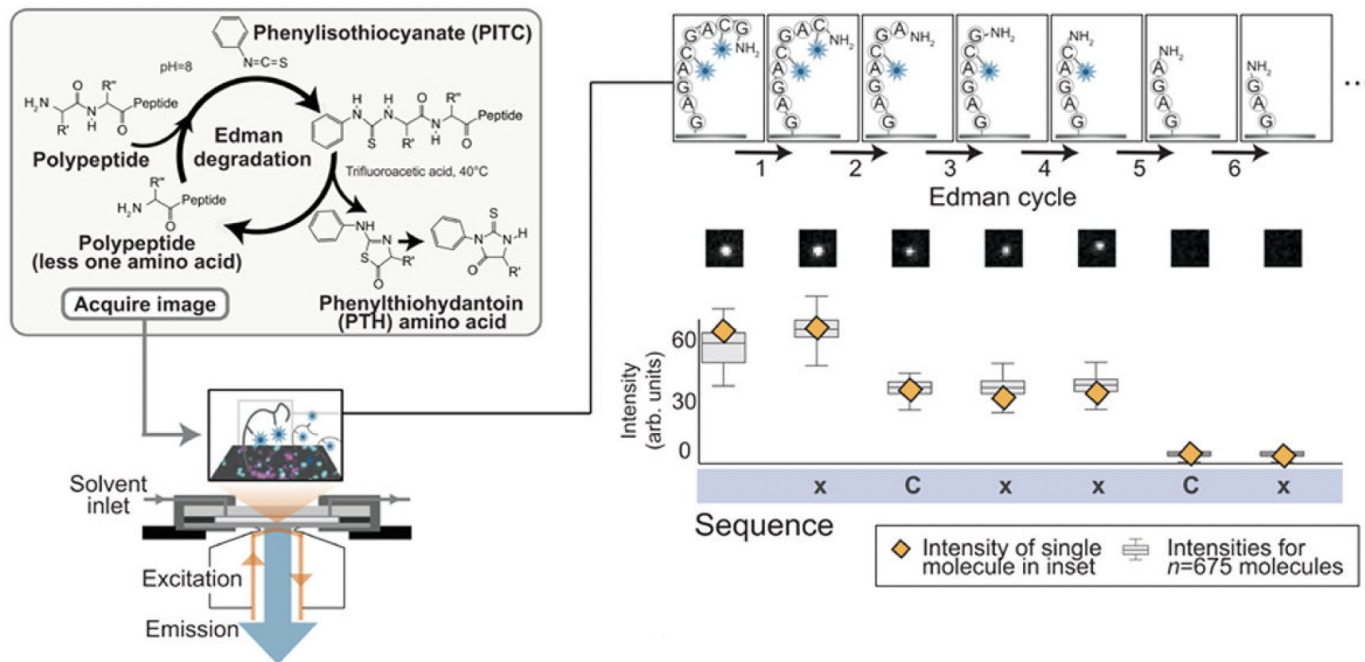
Wang, Woehrstein, Donoghue, Dai, Avendaño, Schackmann, Zoeller, Wang, Tillberg, Park, Lapan, Boyden, Brugge, Kaeser, Church, Agasti, Jungmann, Yin (2017) Nano Lett.

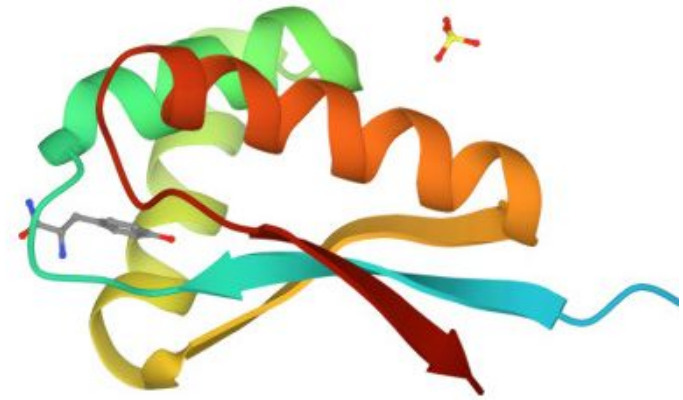
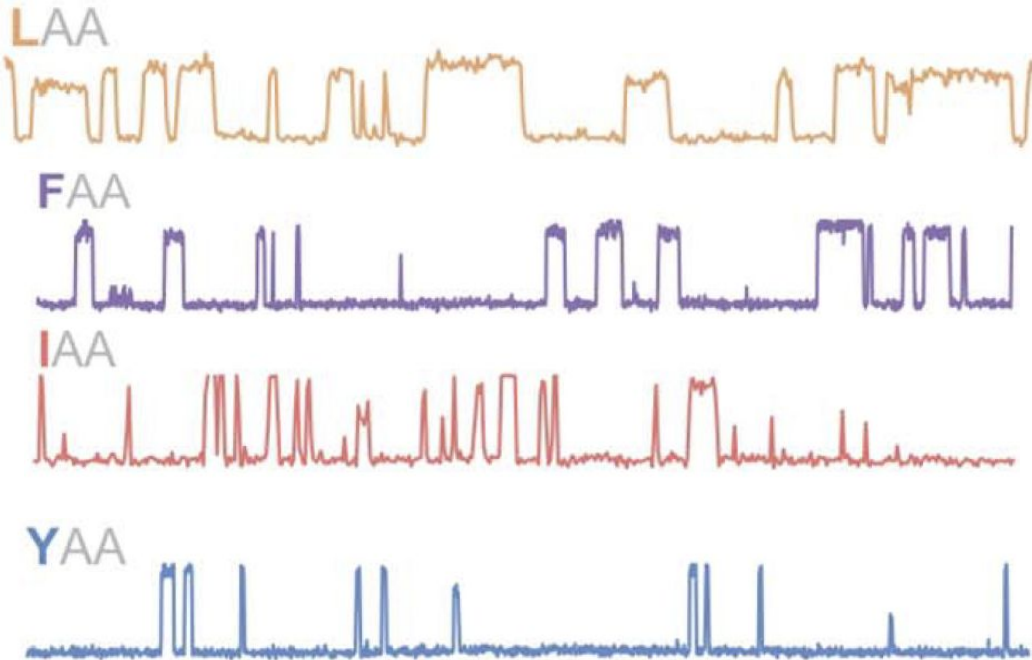
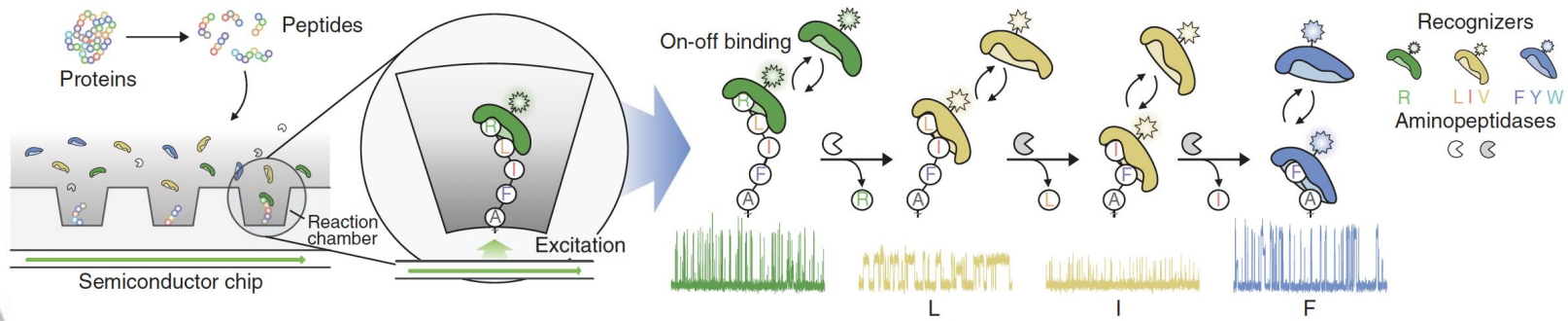


Protein Sequencing, One Molecule at a Time

Floyd & Marcotte (2022) Ann. Reviews Biophys

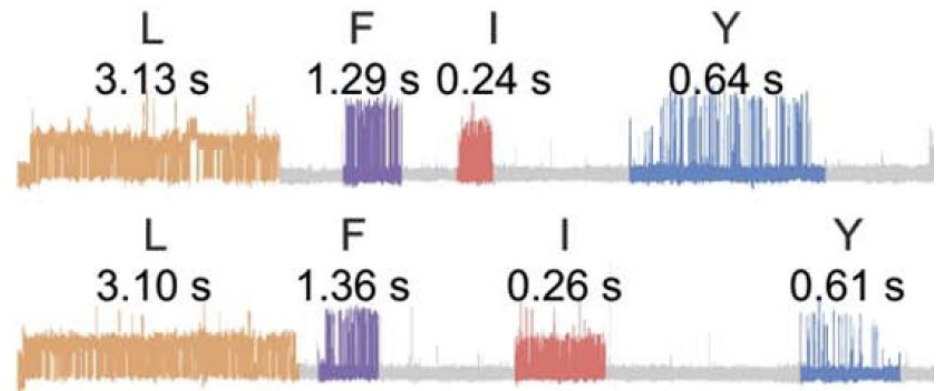
Sequencing Category	DNA Sequencing Method	Protein Sequencing Method
Sequencing-by-synthesis	Polymerase-based sequencing (Sanger, Illumina, Pacific Biosciences, etc)	None, as no practical chemical or enzymatic copying mechanisms are known for proteins
Sequencing-by-degradation	Enzymatic digest sequencing (26, 53)	Fluorosequencing NAAB binding Nanopore mass spectrometry
Sequencing-by-transit	Nanopore sequencing	Nanopore sequencing Electron tunneling ClpXP based fingerprinting
Sequencing-by-affinity	DNA microarrays Fluorescence <i>in situ</i> hybridization	DNA nanotechnology (DNA PAINT, FRET X, DNA nanoscope)





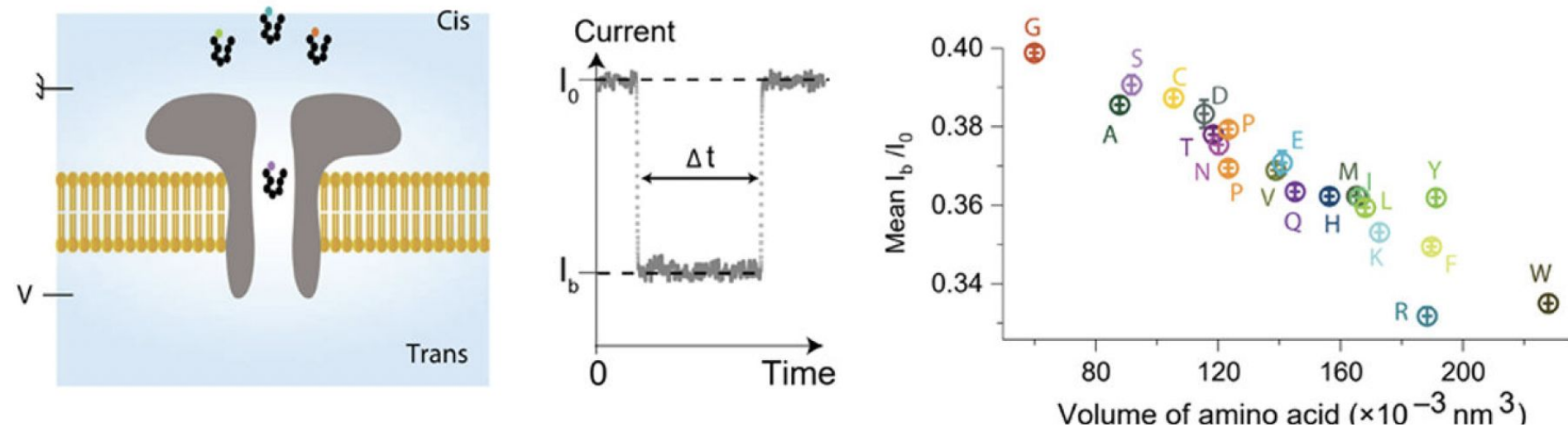
N-term-Tyr – ClpS2

LAQFASIAAYASDDD



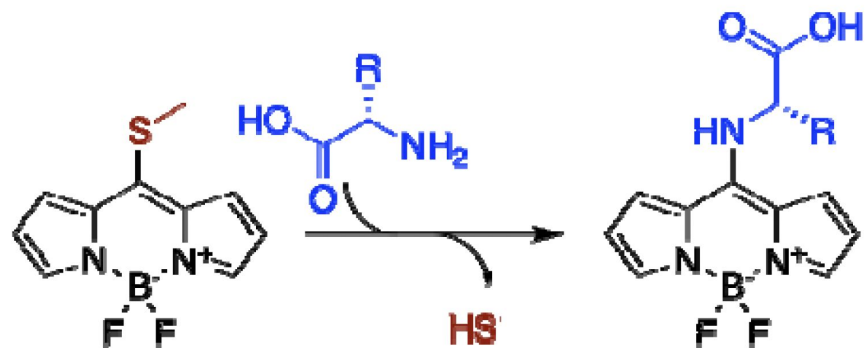
Reed ... Rothberg Science
Oct-2022 (Quantum-Si Inc.)

Sequencing by Pore Transit



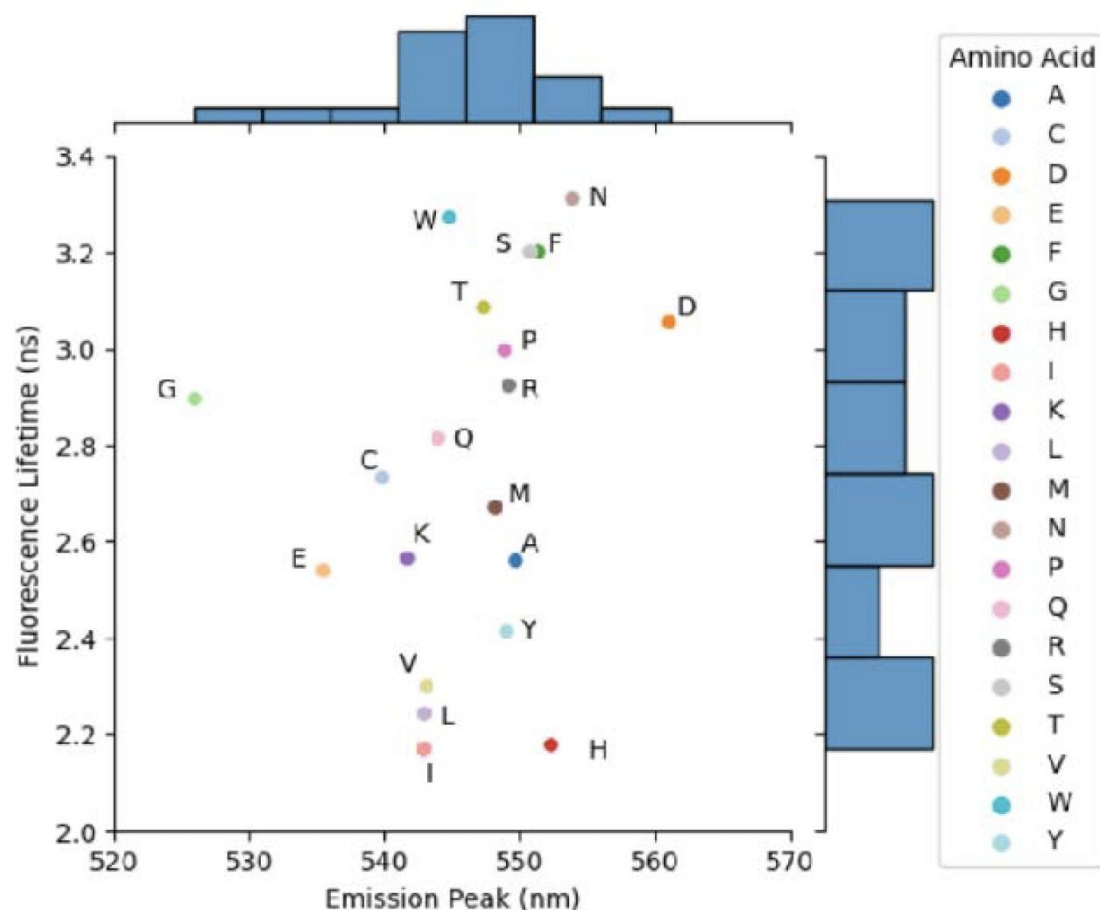
Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore (2019) Ouldali et al.
[nature.com/articles/s41587-019-0345-2](https://www.nature.com/articles/s41587-019-0345-2)

41j.com/blog/2021/07/dreampore-nanopore-protein-sequencing



boron-dipyrro...

Development of BODIPY-based fluorescent probes for highly selective amino acid identification



Hendrik et al. [BU & Prisma Tx]

biorxiv.org/content/10.1101/2024.04.02.587799v1.full.pdf



Core concepts

Why read, write & measure polymers?

Natural & designed properties

Chirality, reactivity, etc.

Codes:

Basepair code: A=T, C=G etc.

Ribosomal translation: 4 RNA $\square$ 20+AA

Interactions: Protein $\square$ DNA

How to Read almost anything?

De novo vs ID, in vitro vs in situ imaging

Size of side chains, Fluorescence

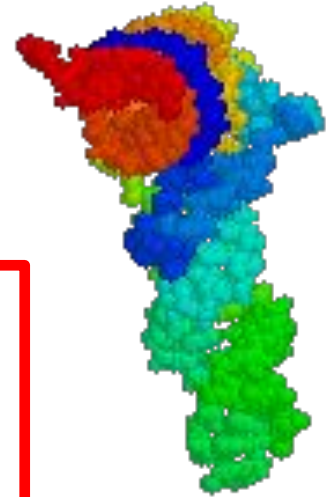
Single molecules: Why?

How to Write almost anything?

Sequential, Exponential, Editing

Post-polymerization modifications

HTGAA 10-Feb-2026, 3-3:20, 4:40-5 pm
Reading & Writing Life George Church



@geochurch



v.ht/UX2i6



DOE
1987

2001

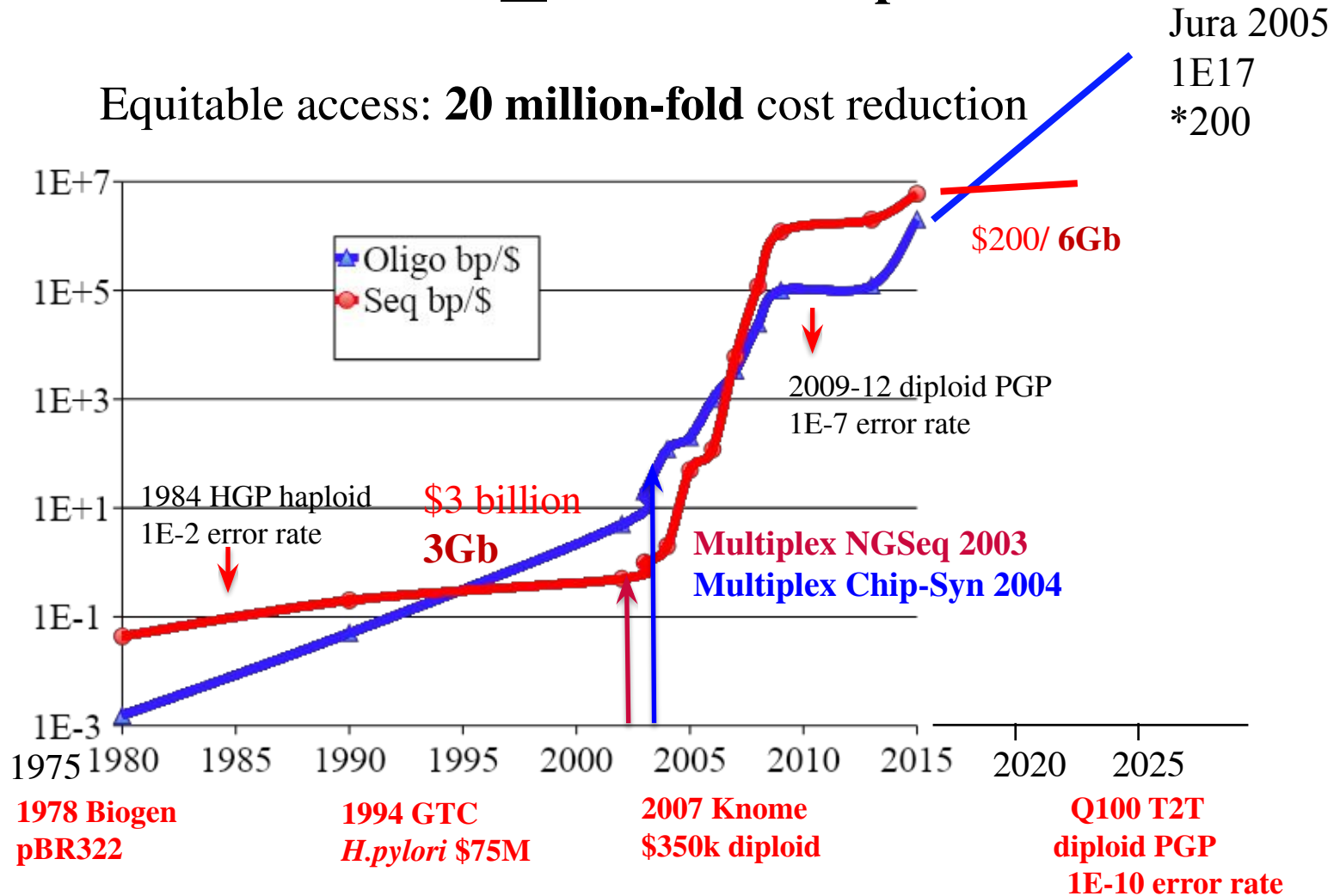


Lipper 1997
Foundation



Full pfCOI list:
arep.med.harvard.edu/t

DNA/RNA **Read** & **Write** tech exponentials



Next 10 years



CRISPR-based treatments
in later stages of clinical trials



FDA approval of
sickle cell therapy



FDA approval of
additional CRISPR cell therapies



Increased nutritional
value of more foods



Improved
in vivo delivery



Multigenic traits in more
plants and animals



Expansion of CRISPR-
modified crops



Disease resistance and
improved crop yields

90% of GeneTx are non-CRISPR

\$2.2 M Casgevy; \$3.1M Lyfgenia (non-Cas)

☐ Organ transplantation

Genetic counseling \$300

Ag

☐ Aging reversal

Delivery

☐ Vaccines (non-Cas)

☐ Caspid protein design (non-Cas)

Ag

☐ DNA data storage

☐ Endangered species

☐ Resistance to all viruses

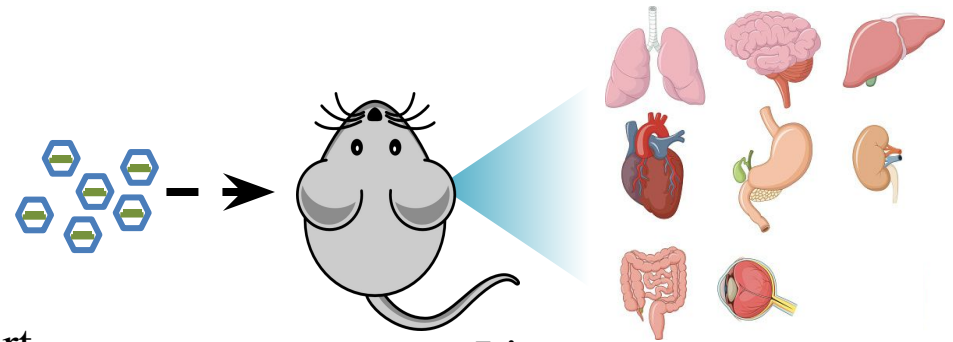
Ag

Ag

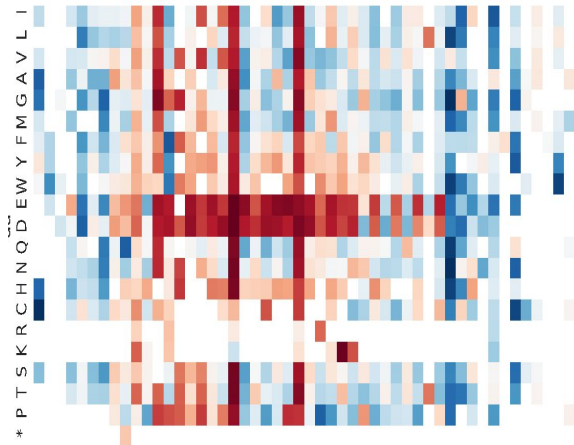
CRISPR Wang & Doudna 2023 Science



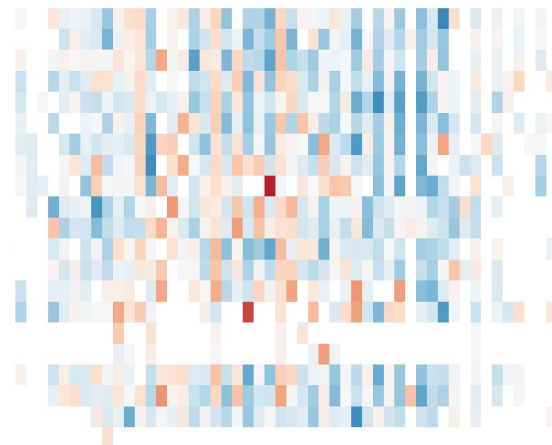
Mapping the space of tissue tropism modifications



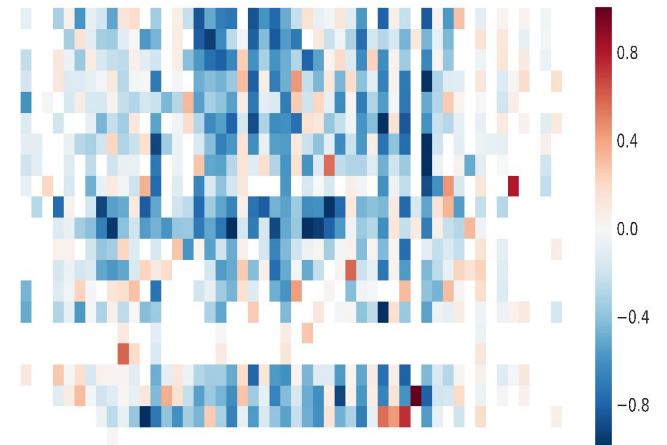
Blood



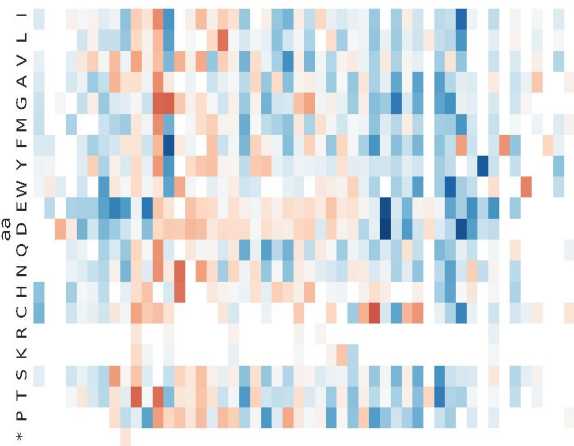
Heart



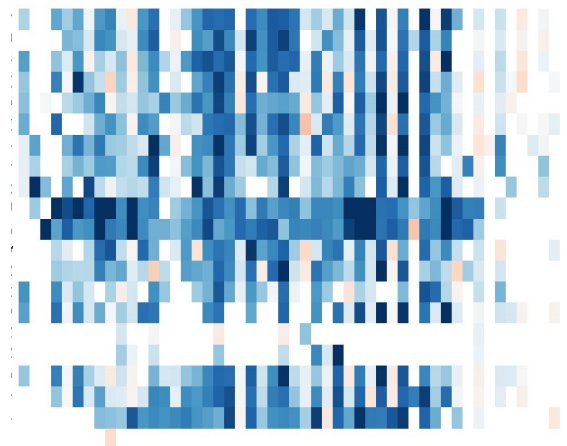
Liver



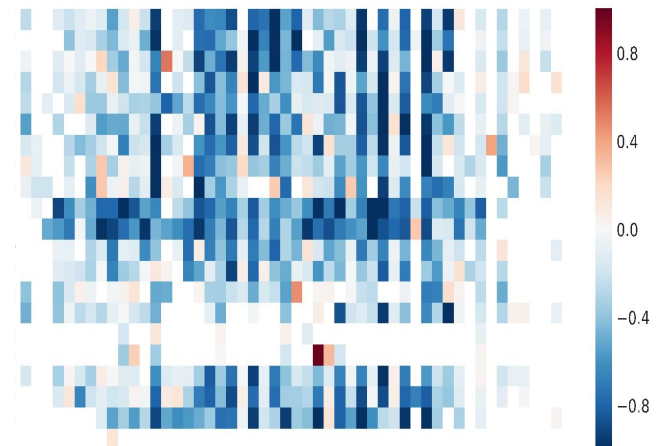
Kidney



Spleen



Lung



570

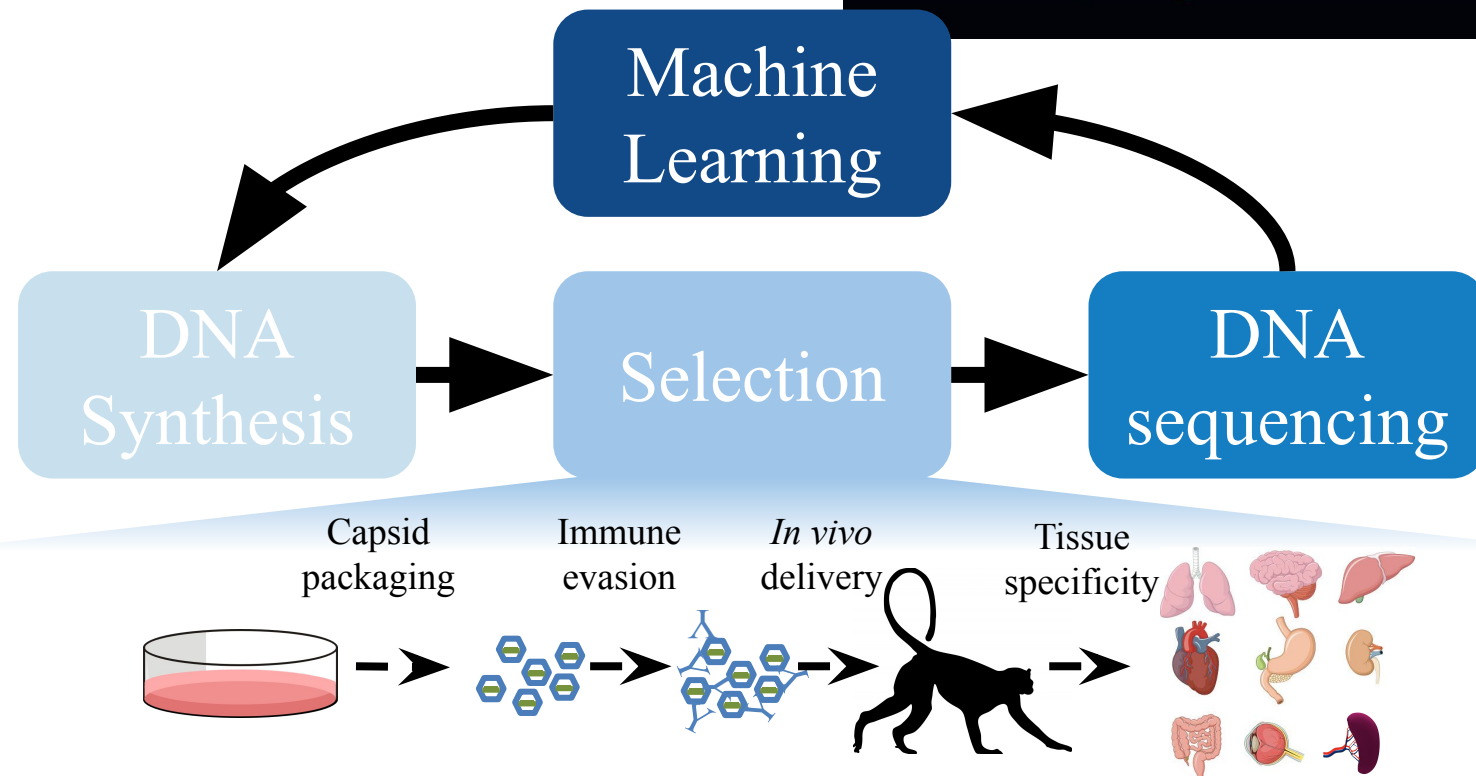
Position

605

Immediate primate testing
intravenous delivery
ahq: 280x AAV9 pan-brain
50x liver detargeting

yp2: 94% of neurons in
humanized TfR1 mice

28/28 AA libraries

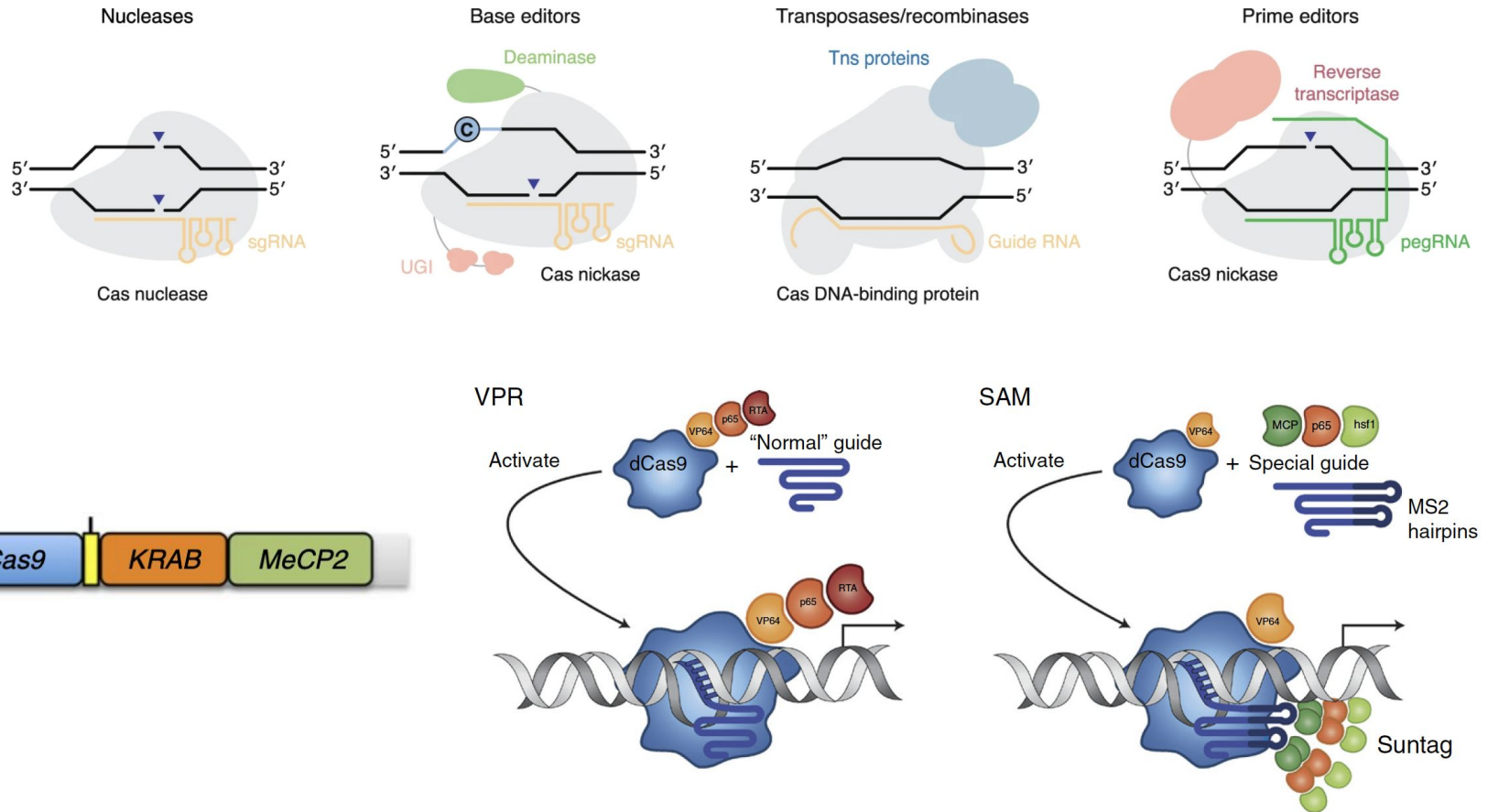


Pierce Ogden
Eric Kelsic
Science 2019

dynotx.com/gene-delivery



CRISPR as RNP scaffold

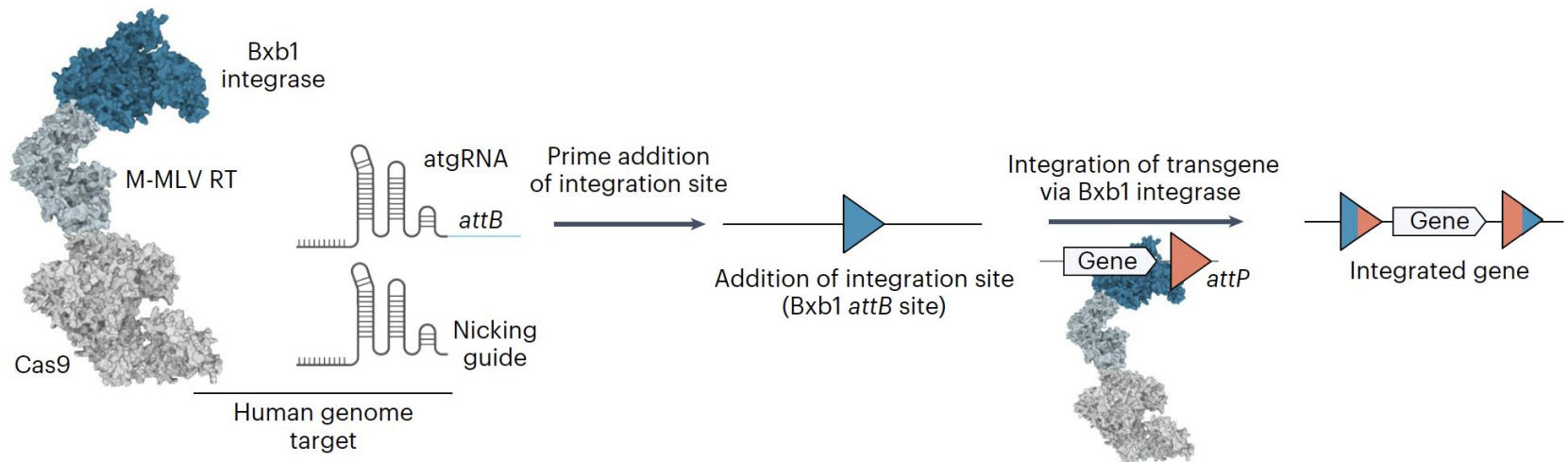


Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors Anzalone et al. 2020 Nature Biotech
 An enhanced CRISPR repressor ... Yeo et al. 2018 Nature Meth.
 Comparison of Cas9 activators ... Chavez et al. 2016 Nature Meth.

Drag-and-drop genome insertion of large sequences without double-strand DNA cleavage using CRISPR-directed integrases

Yarnall et al. 2023 Nature Biotech

“**P**rogrammable **A**ddition via **S**ite-specific **T**argeting **E**lements (PASTE), which uses a CRISPR–Cas9 nickase fused to both a reverse transcriptase and serine integrase for targeted ... integration ... as large as ~36 kilobases at multiple genomic loci ...we discovered 25,614 serine integrases and cognate attachment sites”



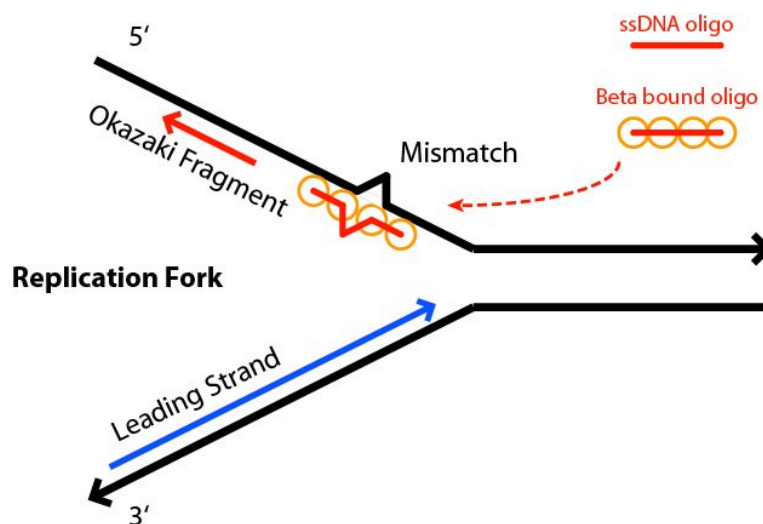
Technologies : Multiplex, Precise, Large Edits via Mega-synthesis, Recombinases & Integrases

Gabe Filsinger &
Tim Wannier

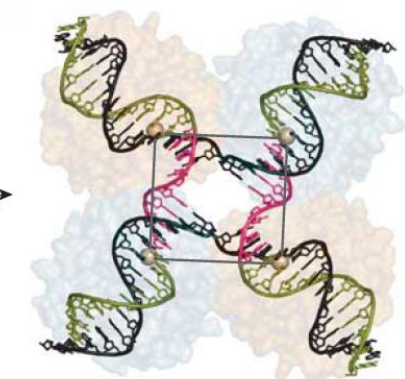


Single-Strand Annealing Proteins (SSAP) &
Single-Strand Binding proteins (SSB) interactions

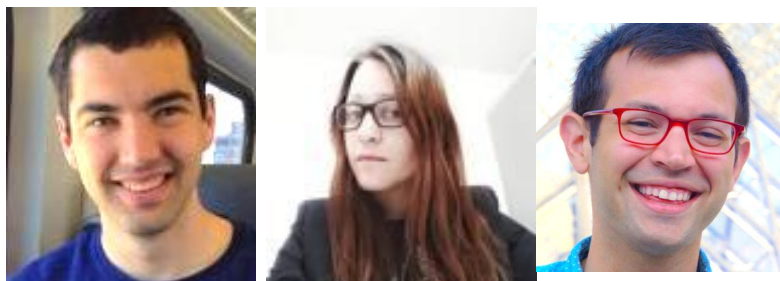
MAGE



Ser,Tyr Integrases: ML-Syn



Kale Kundert, Tina Lebar, Ben Weinberg.



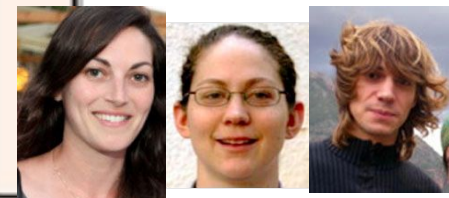
Genomically Recoded Organisms (GRO)

2 +1 codons
(Fredens..Chin 2019)

6 +1 codons

4 Mbp synthesized
3kb □ 50 kb

Ostrov,
Norville,
Guell, ..
Church 2016



		2nd base					
		U	C	A	G		
5' base	U	F 22445 30204	82 11672 11396	Y 16567 21888	C 8687 6948	3' base	C
		L 18824 18488	S 9620 12064	RF1 RF2 2765 0 RF1 321	* 4 RF2 1232 W 20688		A
							G
	C	L 21 15077 29 14953 5260 71864	P 3 7401 9476 11425 31603	H 13140 17534 20842 39234	R 29866 28424 4744 3 7273		C
							U
							A
	A	I 34177 41335 1 5797	T 133 31766 12065 9492 19572	N 29304 23948 45743 13937	S 21748 11843 R 4 2737 1 1491		C
		M 37698		K			U
							A
	G	V 59 20757 24817 14745 35620	A 65 34747 20713 27460 45816	D 25982 43739 53822 24211	G 40263 33623 10675 14975		C
							U
							A
							G

Multiplexed Genome Engineering (MAGE)

4 Million bp Recoding

321 □ 18,214 □ 62,214 codons 1□3□7

1,2. Genetic & metabolic isolation

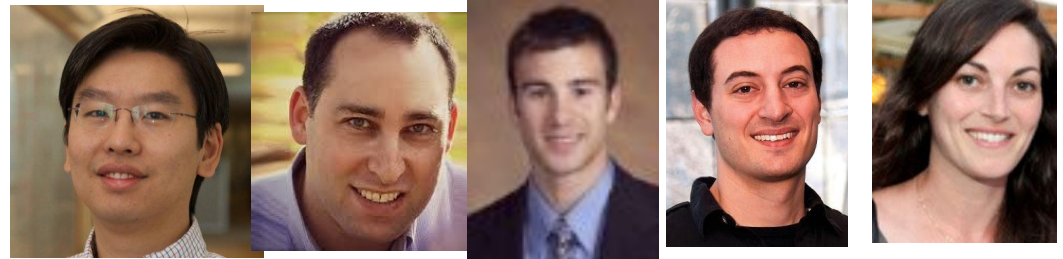


3. Non-standard amino acids (NSAA)

4. Multi-Virus resistance

COSB 2020 Ostrov, et al
Science 2016 Ostrov, et al
Nature 2015 Mandell et al
Science 2013 Lajoie et al
Science 2011 Isaacs et al
Nature 2009 Wang et al

Immune tolerance
Stability
Fluorogenic



SynGenome Technologies:

3 forms of “biocontainment”

Cautions: #1: Need extensive sampling of wild viruses

#2: Reduced code could work in wild

#3 multivirus resistance advantage in the wild, should be contained.

Infection, Conjugation, Transformation

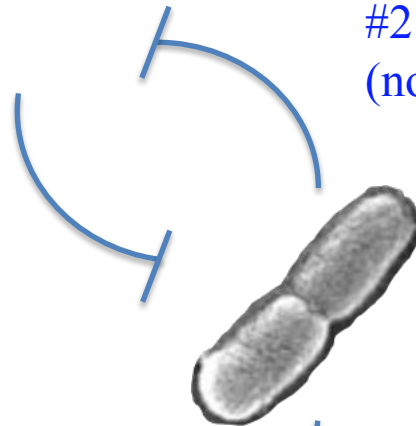
Environmental **DNA**

RNA

#2 BipA or swap in all proteins
(not just mobile element proteins)

#1 see next slides

S⇌L 2 codon swap



#3 BipA or swap in a few
essential proteins

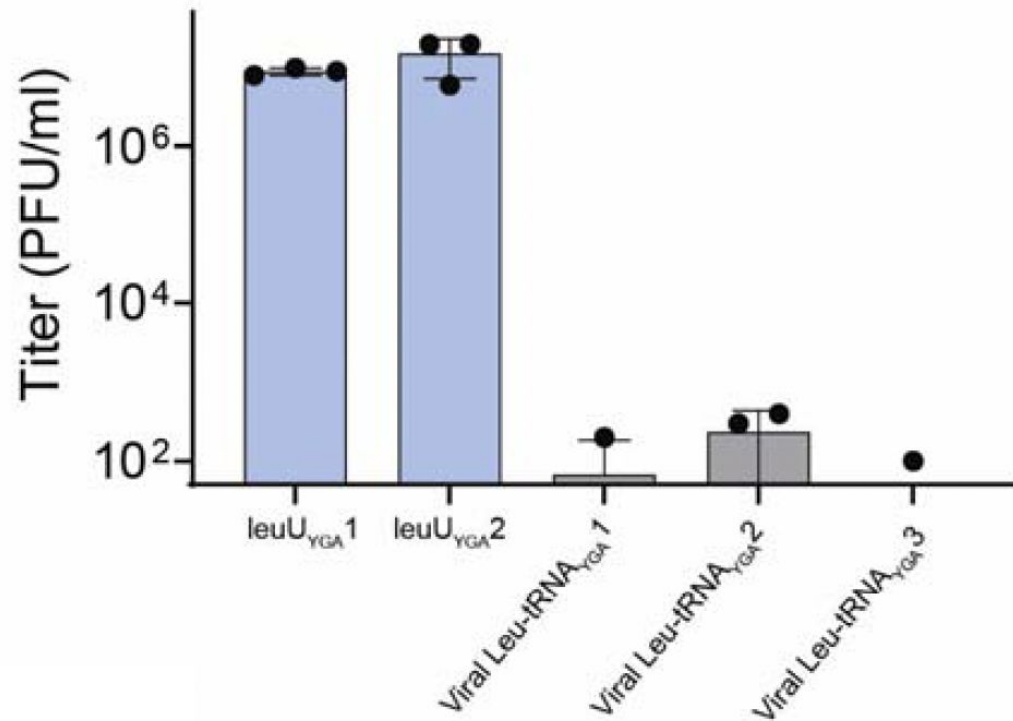
Metabolism
(Replication)

12 viruses replicate in Syn61 Δ 3

but not in 2 codon swap

TCA TCG:

Ser $\square$ Leu



3a, 3b, 3c, 3d

7a, b

8a

9a

13a, b, c, d

Wastewater (primary effluent)
Wastewater (primary effluent)
River water
Soil from horse and alpaca enclosure
Water from porta-potty overflow
Irrigation water near wild rat nest
Soil from pasture-raised chicken enclosure
Soil from egg-laying chicken shed
Fresh pig feces
Mixed fresh horse and cow feces
Farm compost pile
Farm compost pile
Water from pig and goose water trough

Akos Nyerges, Svenja Vinke, Regan Flynn...Church, Jul-2022 Biorxiv.
Swapped genetic code blocks viral infections & gene transfer.



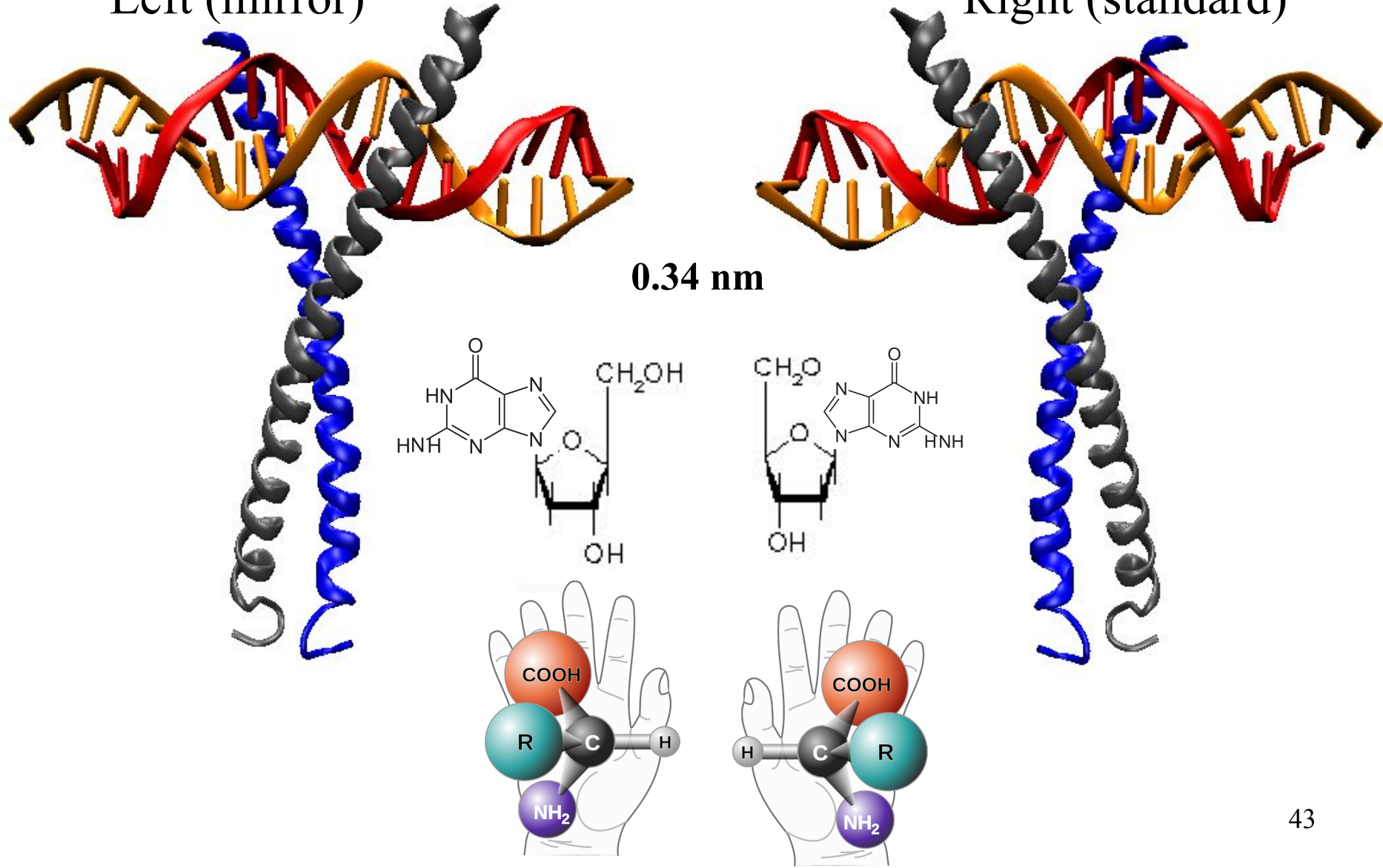
Also: Zürcher et al. Oct-2022 Science. Refactored genetic codes enable bidirectional genetic isolation

Mirror world : ~~blow your~~ mirror

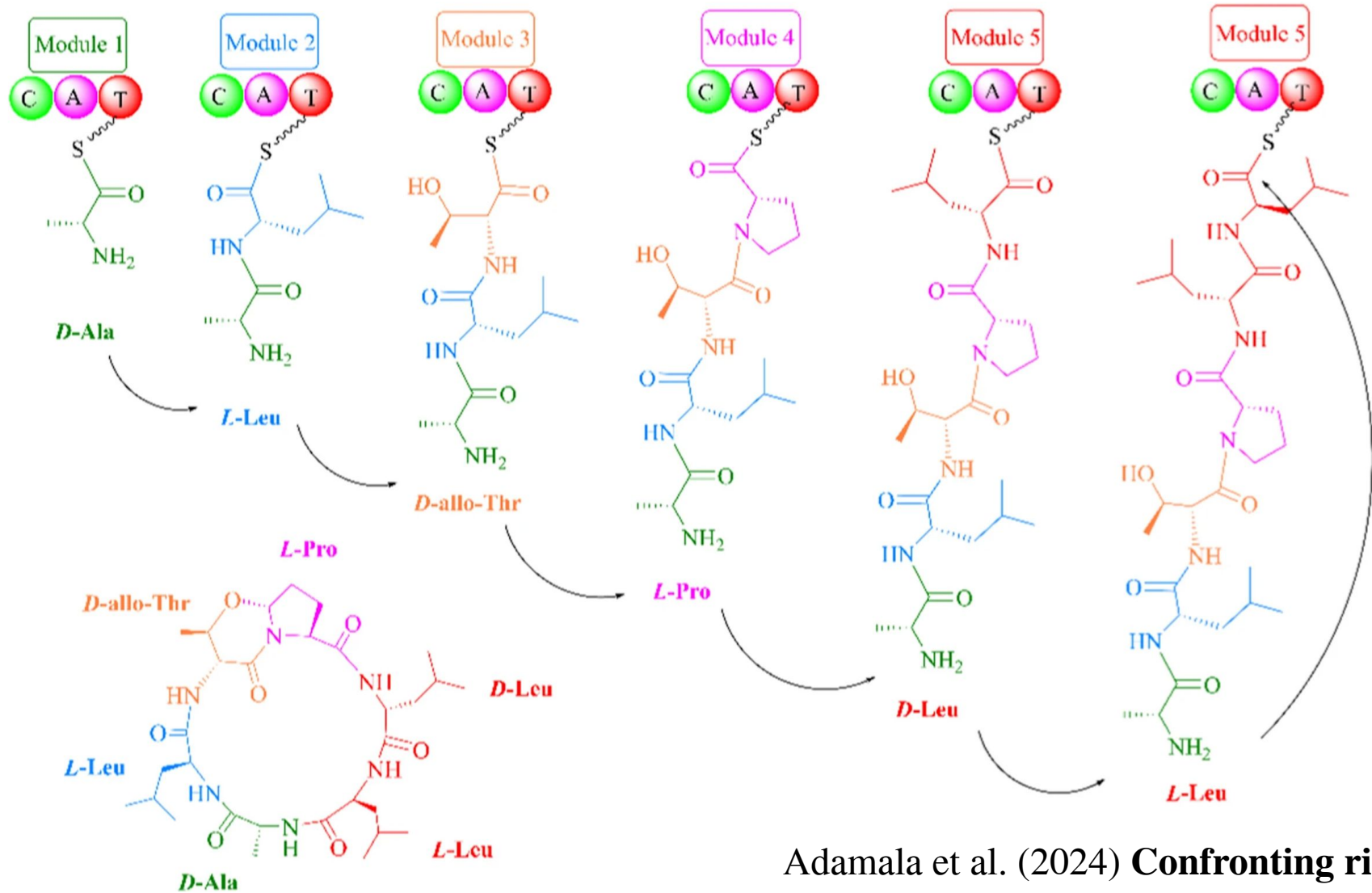
B-DNA & α -helical protein chirality

Left (mirror)

Right (standard)



Non-ribosomal peptide synthesis (NRPS) & D-AA



Adamala et al. (2024) **Confronting risks of mirror life.** Science

Baker D, Church G (2024) **Protein design meets biosecurity.** Science

Nonribosomal Peptide Synthesis.

Sussmuth & Mainz (2017). Huang et al. (2024)

Storing temporal (4D) data in DNA

Narrative, image/video...lineage, physiology

in vitro:

2002-10 Church & **Shendure**. USPTO

2012 Church, Gao, Kosuri. Science (syn DNA, book jpg)

in vivo:

2016 Kalhor, Mali, Church. Nature Meth. (hgRNA)

2016 Perli, Cui, Lu Science (stgRNA)

2016 McKenna...**Shendure** Science (Cas9)

2016 **Shipman**...Church Science (Cas1/2)

2017 Frieda...Cai L Nature (Cas9)

2018 Raj...Schier Nature Biotech (Cas9)

2019 Chan...Weissman Nature (Cas9)

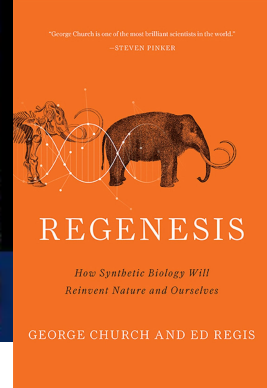
2018 Alemany...van Oudenaarden Nature (Cas9)

2018 Spanjaard...Junker. Nature Biotech (Cas9)

2020 Smith...Church. Nucl. Acids Res. (ABE)

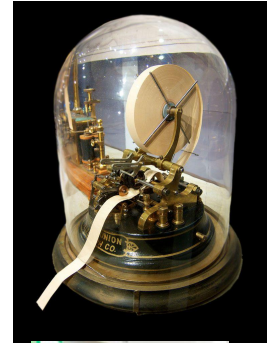
2022 Bhattarai-Kline...**Shipman** Nature (retron)

2022 Choi...**Shendure** Nature (pegRNA)



Tickertape

**Molecular
Recording**



Jay Shendure
UW Seattle



Seth Shipman
UCSF



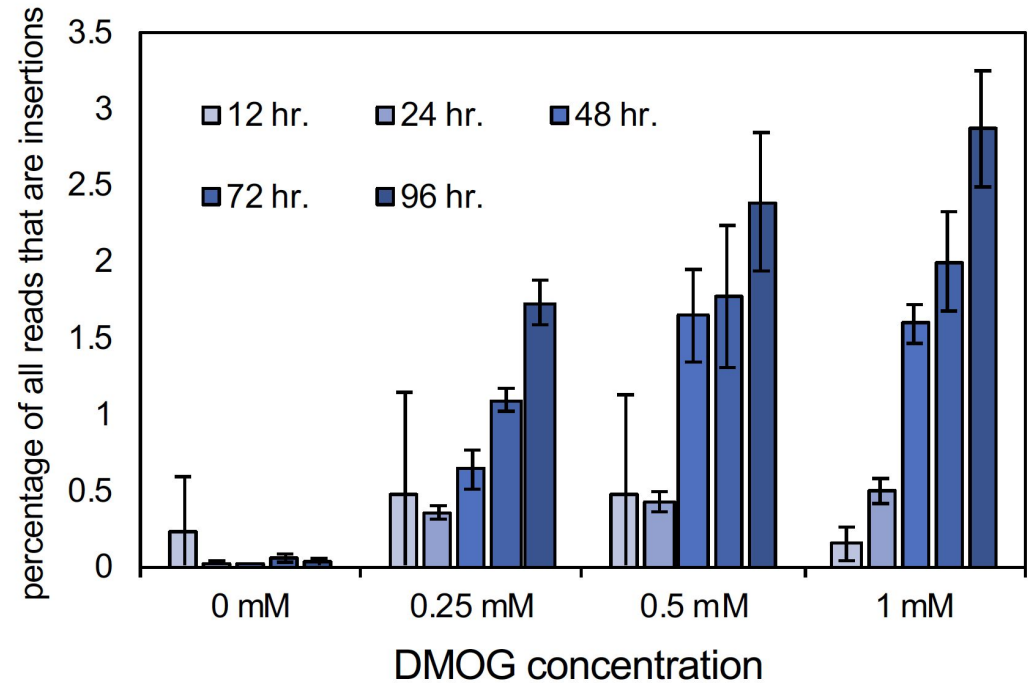
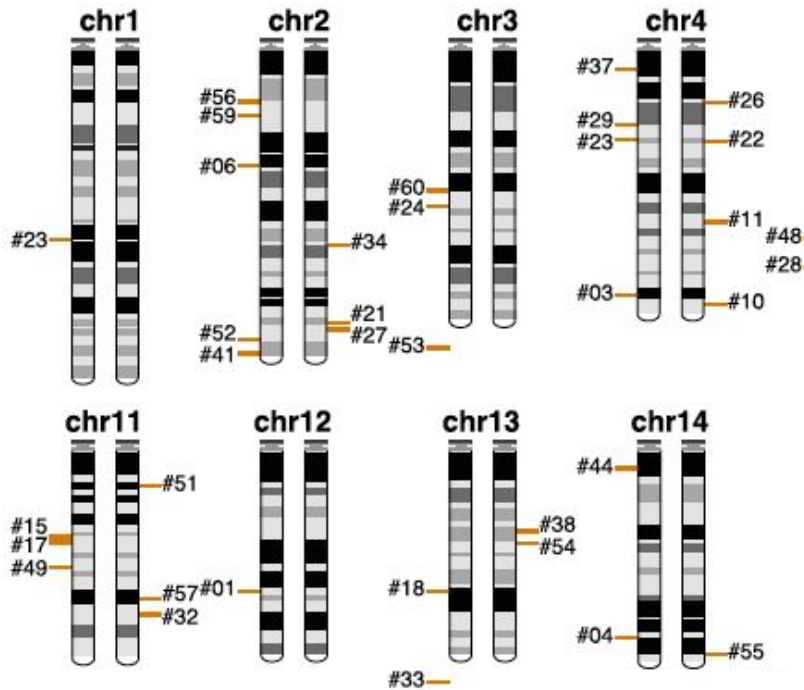
[biorxiv.org/content/10.1101/2024.01.22.576745v1.full](https://www.biorxiv.org/content/10.1101/2024.01.22.576745v1.full)

[biorxiv.org/content/10.1101/2024.01.22.576756v1](https://www.biorxiv.org/content/10.1101/2024.01.22.576756v1)

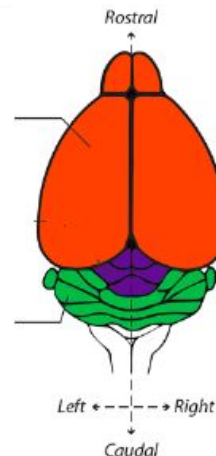
Time-series data in DNA of 3 billion cells = 0.2TeraB $\square$ 1ExaB (1E18)

Loveless et al. Nature Chem. Bio. 2021 TdT in cell culture 14.6 bits per site

Kalhor, et al. Science 2018 in mice. 24k edits iPSCs Smith et al NAR 2019



Fore/hindbrain vs Left/right

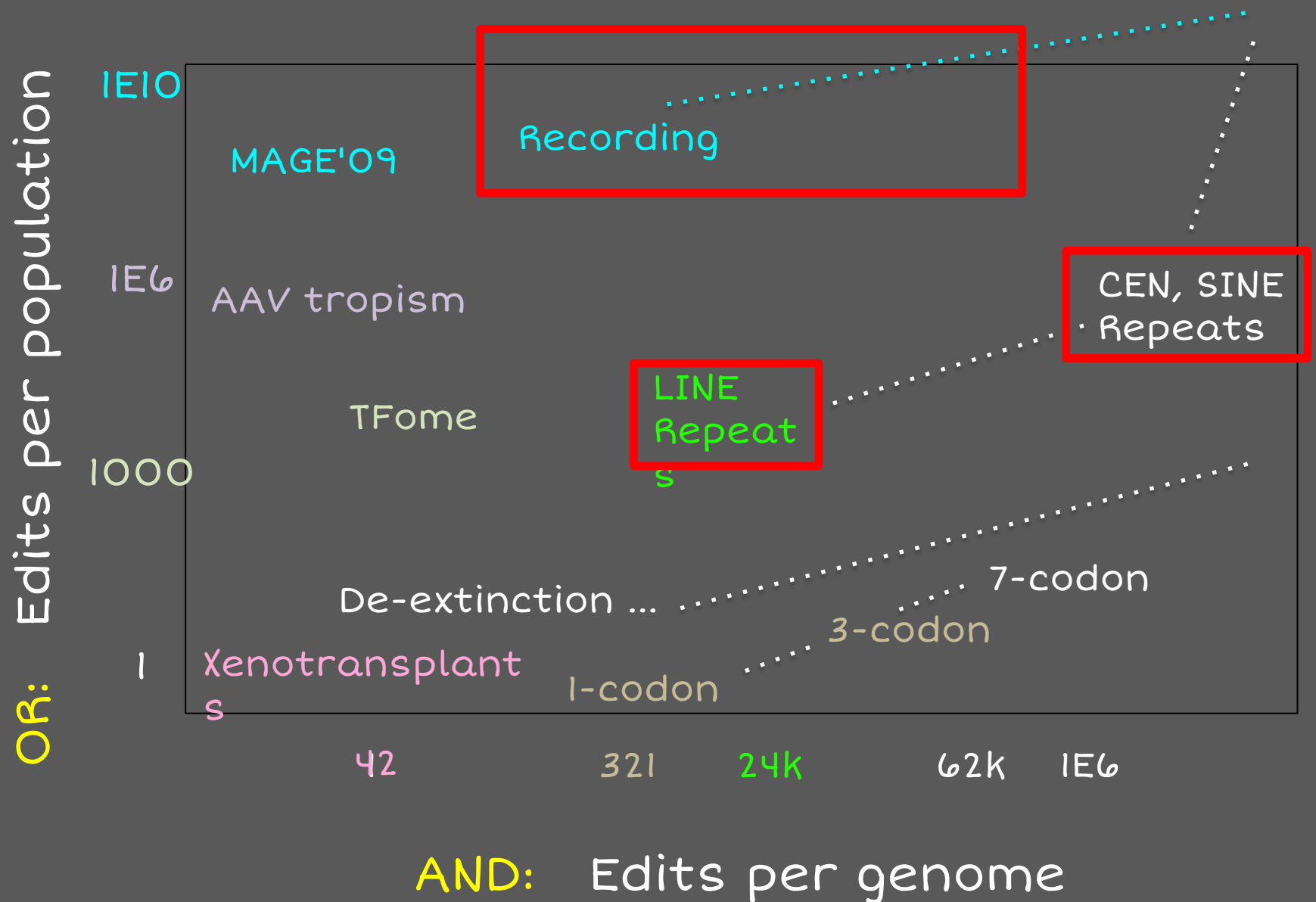


Base edit windows 1 to 12 bp
A $\square$ G, A $\square$ C+T, C $\square$ T, C $\square$ G

$$3e9 * 14 \text{ bits} * 40/8 = \mathbf{0.2 \text{ TeraByte}}$$

$$3e9 * 2e9 \text{ bits} / 8 = \mathbf{1 \text{ ExaByte}}$$

Multiplex (epi) Genome Engineering



Core concepts

Why read, write & measure polymers?

Natural & designed properties

Chirality, reactivity, etc.

Codes:

Basepair code: A=T, C=G etc.

Ribosomal translation: 4 RNA $\square$ 20+AA

Interactions: Protein $\square$ DNA

How to Read almost anything?

De novo vs ID

Size of side chains, Fluorescence

Single molecules: Why?

How to Write almost anything?

Sequential, Exponential, Editing

Post-polymerization modifications



