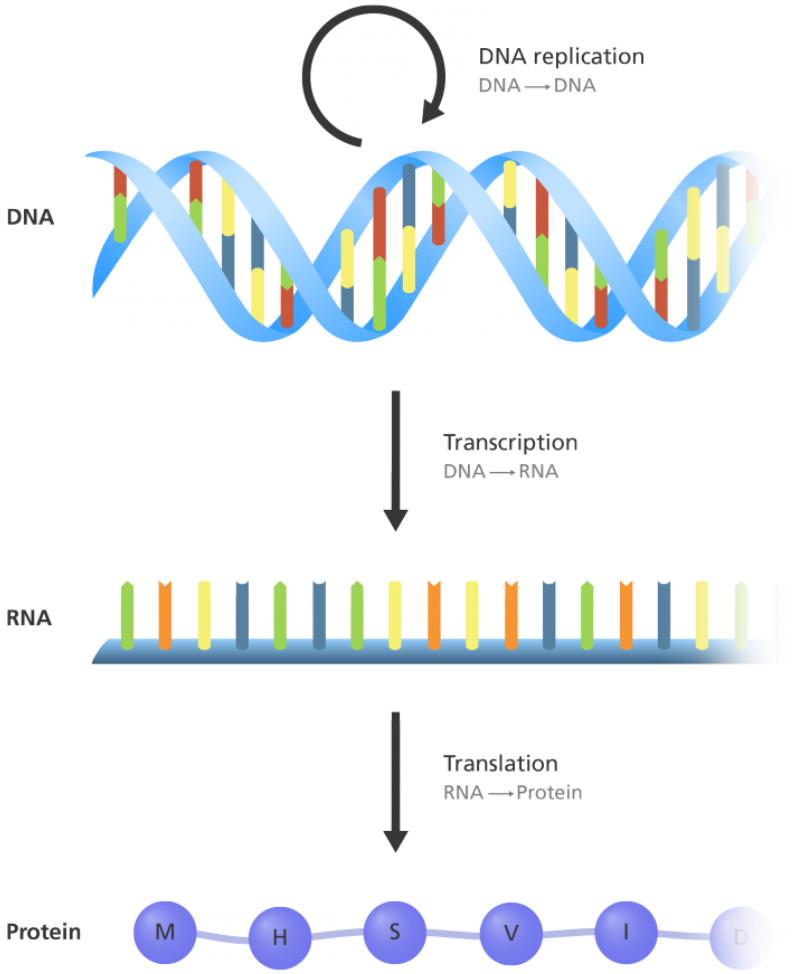
Measuring proteins with DNA: Twisting the central dogma for better understanding of biology

Nebojsa Janjic

May 13, 2021

somalogic™

The central dogma of biology



Genome (where proteins are encoded)

Transcriptome (copies of genes that encode proteins)

Proteome (proteins in action)

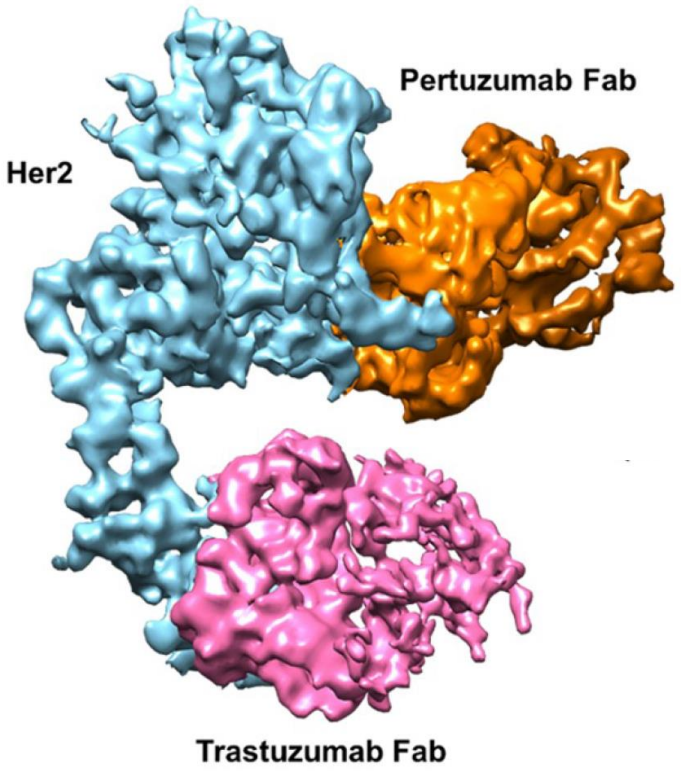
Proteins do things

- Protein functions:
  - Structural components (collagen)
  - Transport (hemoglobin)
  - Storage (ferritin)
  - Communication
    - Hormones (insulin)
    - Growth factors (epidermal growth factor, EGF)
    - Cell surface receptors (EGF receptor)
    - Signal transduction (kinases)
  - Immune response (antibodies)
  - Catalysis (thrombin)
  - Gene regulation (HIF-1α)
  - Many other things

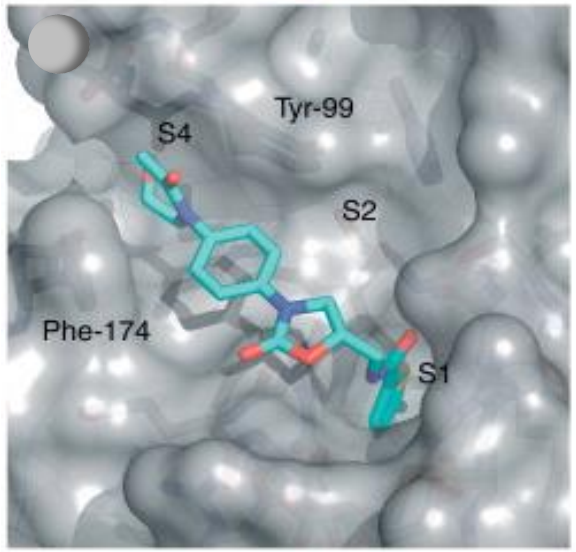
More than 95% of drug targets are proteins

- Protein imbalances cause disease and can be corrected with drugs (pharmaceuticals)
- Most drugs work by directly binding to proteins thereby interfering with their function
- It is through protein measurements that we gain insight into disease and health

Herceptin/Perjeta : Her2



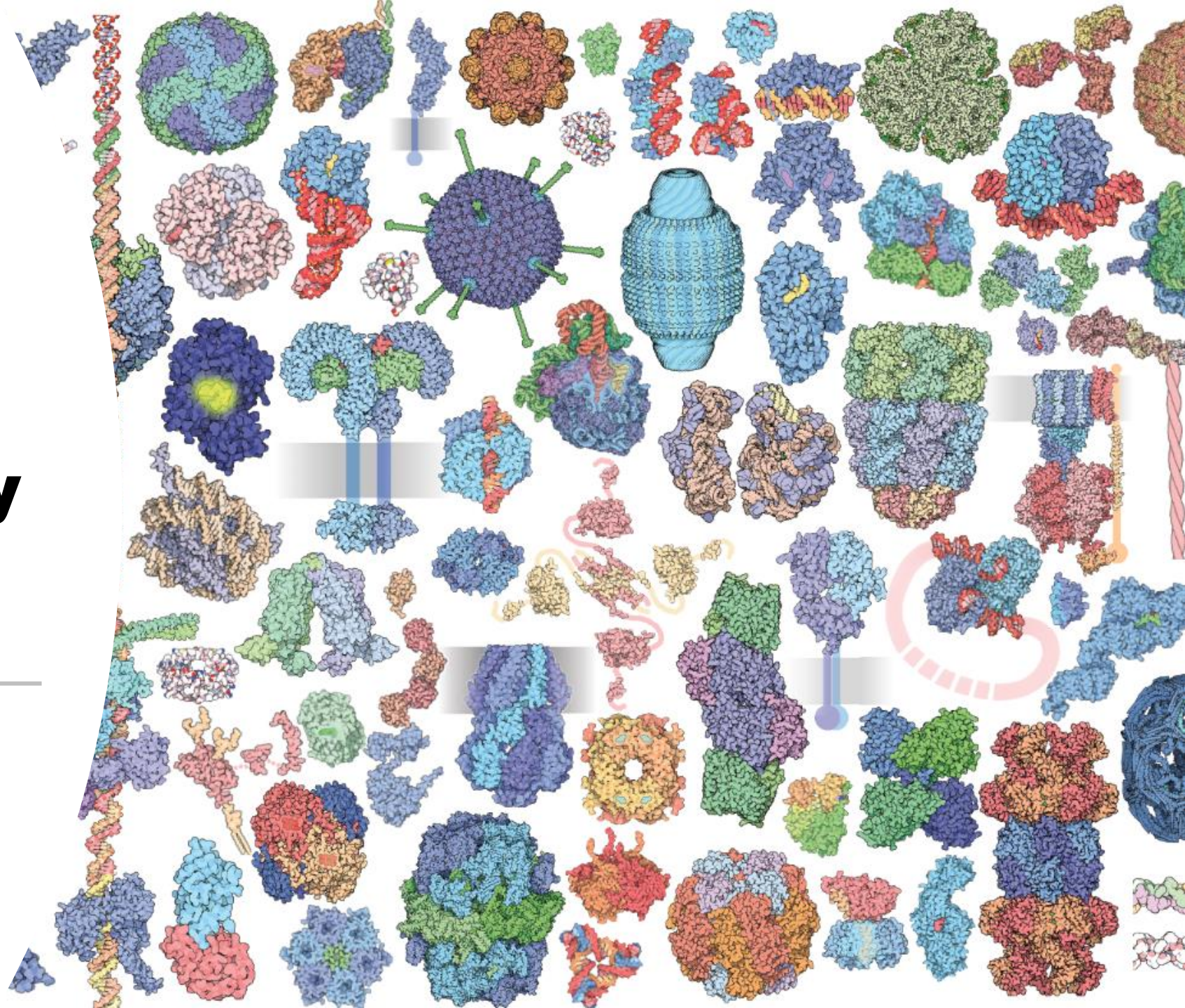
Apixaban (Eliquis) : Factor Xa



Santos *et al*. (2017) *Nat. Rev. Drug Disc*. **16** 19

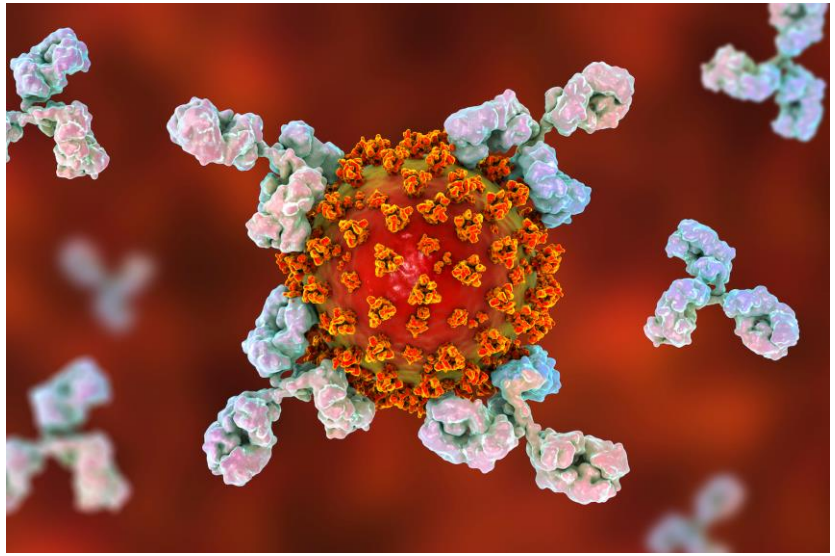
To measure proteins, we need a collection of complementary shapes

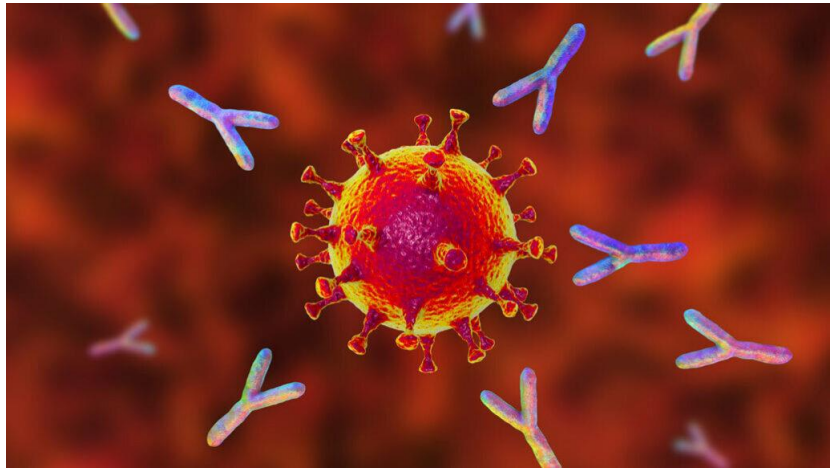
There are 20,000 genes that encode proteins

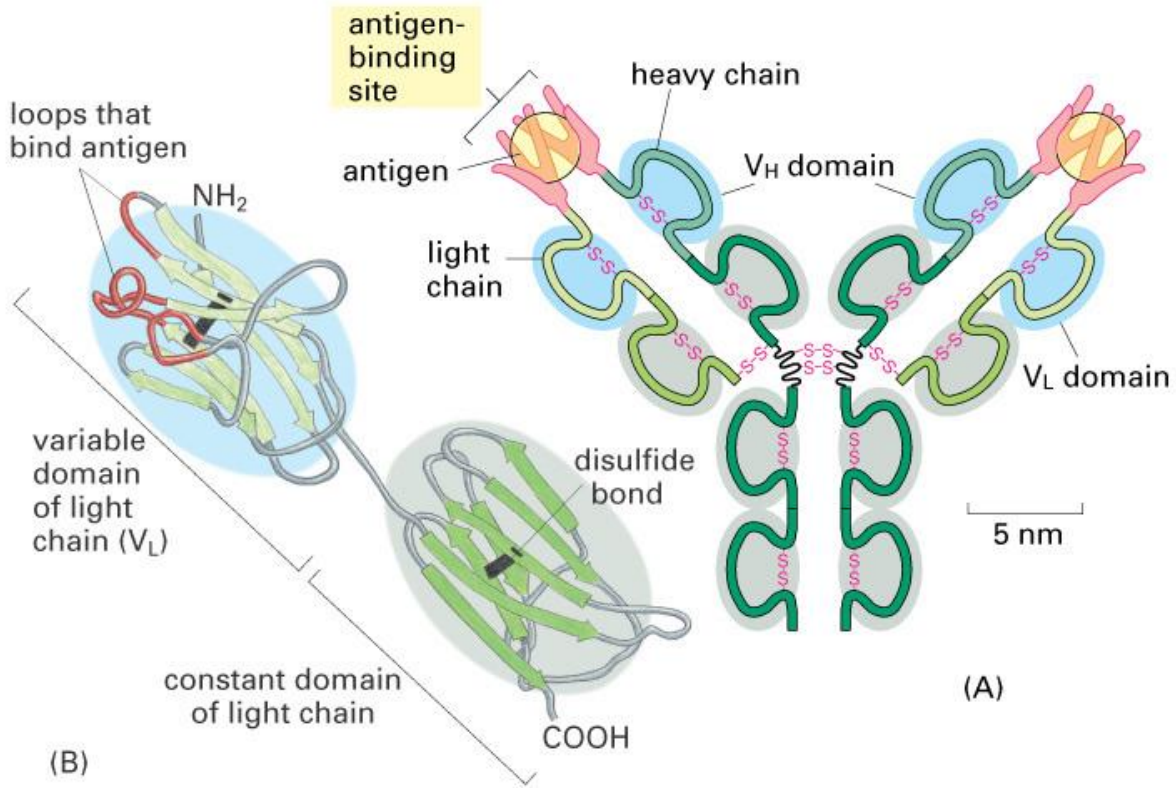


somalogic

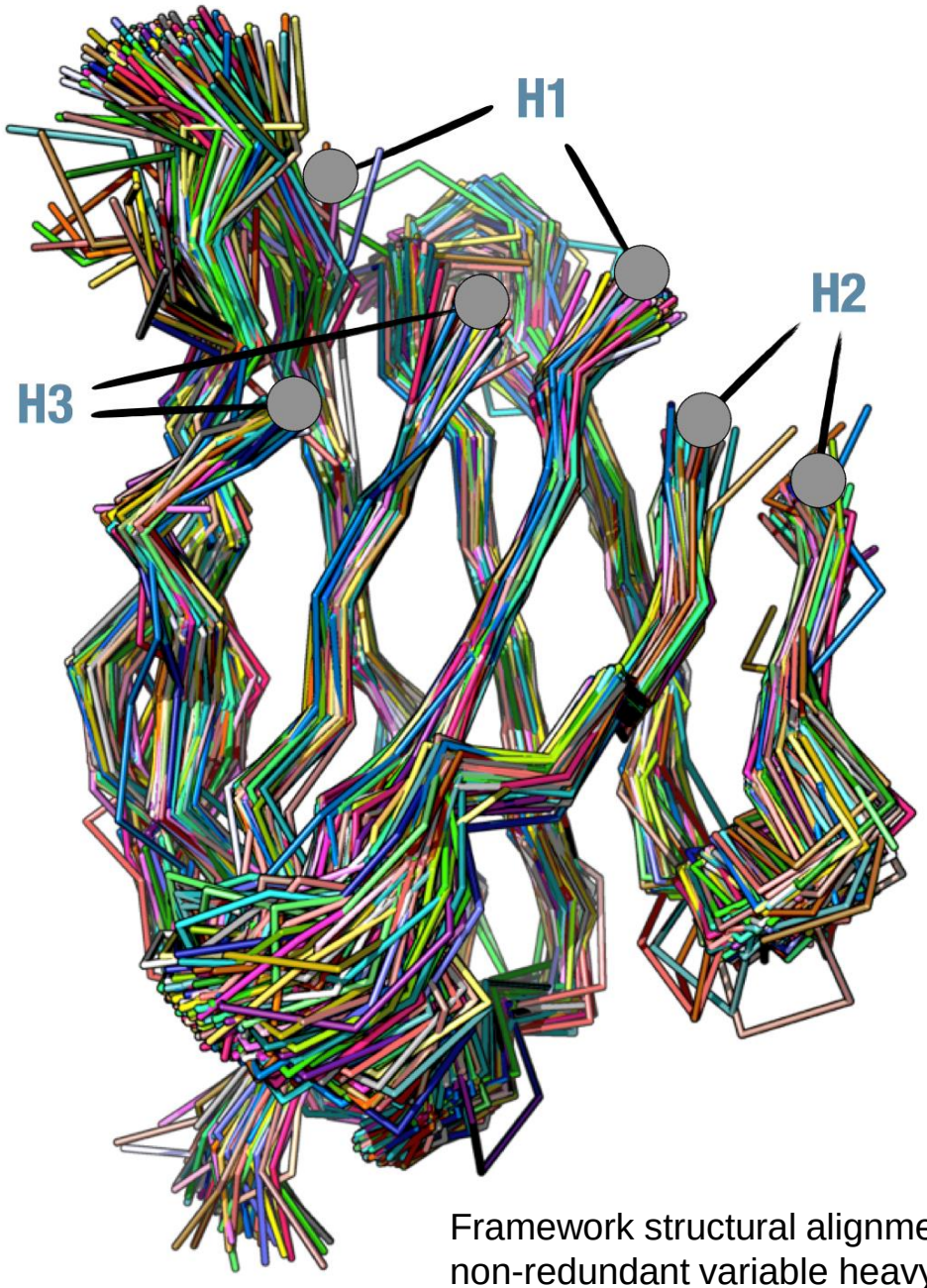
Antibodies: nature's professional ligands







Main-chain conformations in CDRs generally adopt canonical structures

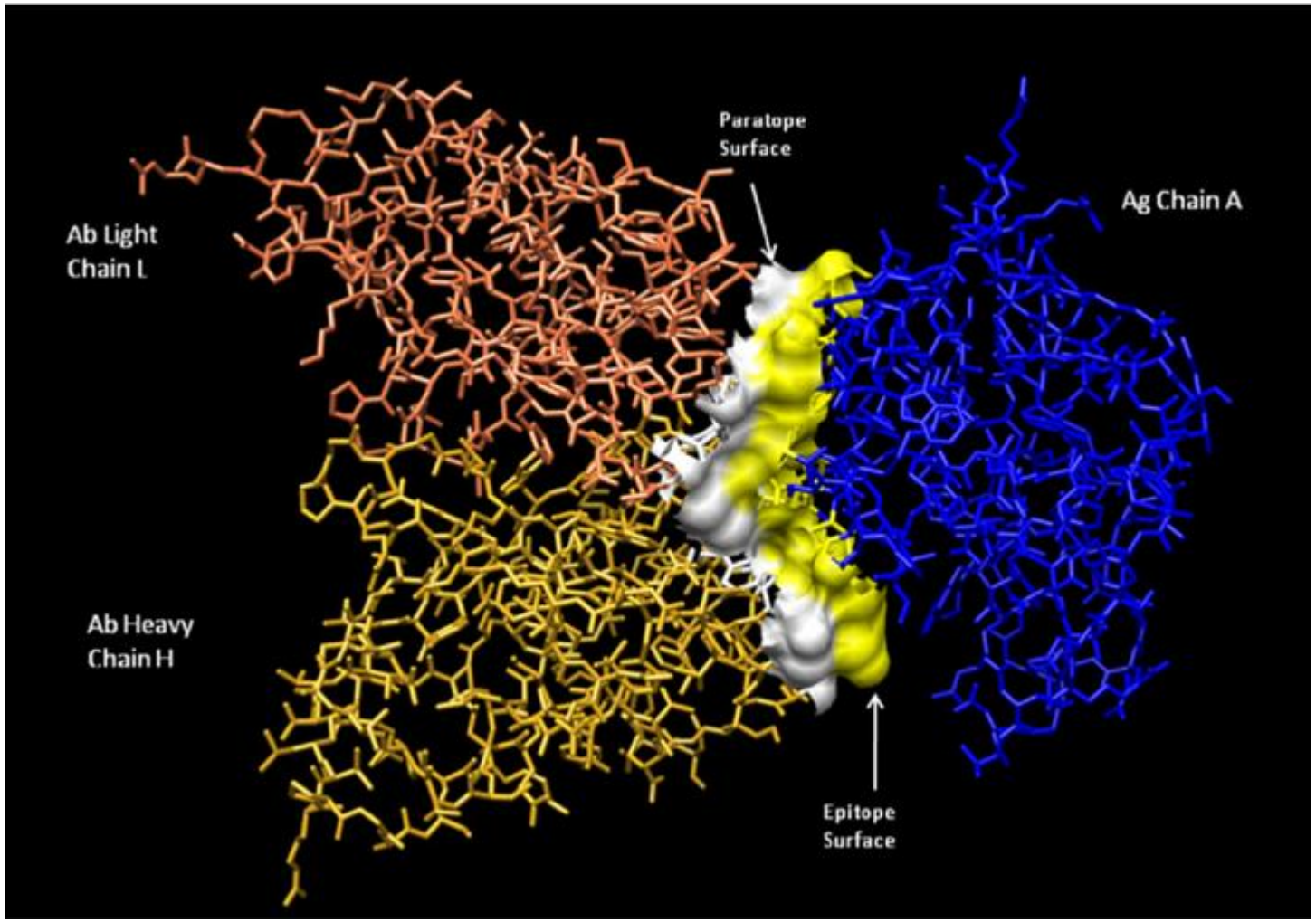


Framework structural alignment of 605 non-redundant variable heavy loops

- So where is shape diversity?
- All is OK, because
  - Shape of CDR scaffolds is augmented with side chains
- Side chains needed:
  - Large
  - Greasy (hydrophobic)
  - Rigid (limited number of rotatable bonds)
- For example: tyrosine, tryptophan, phenylalanine

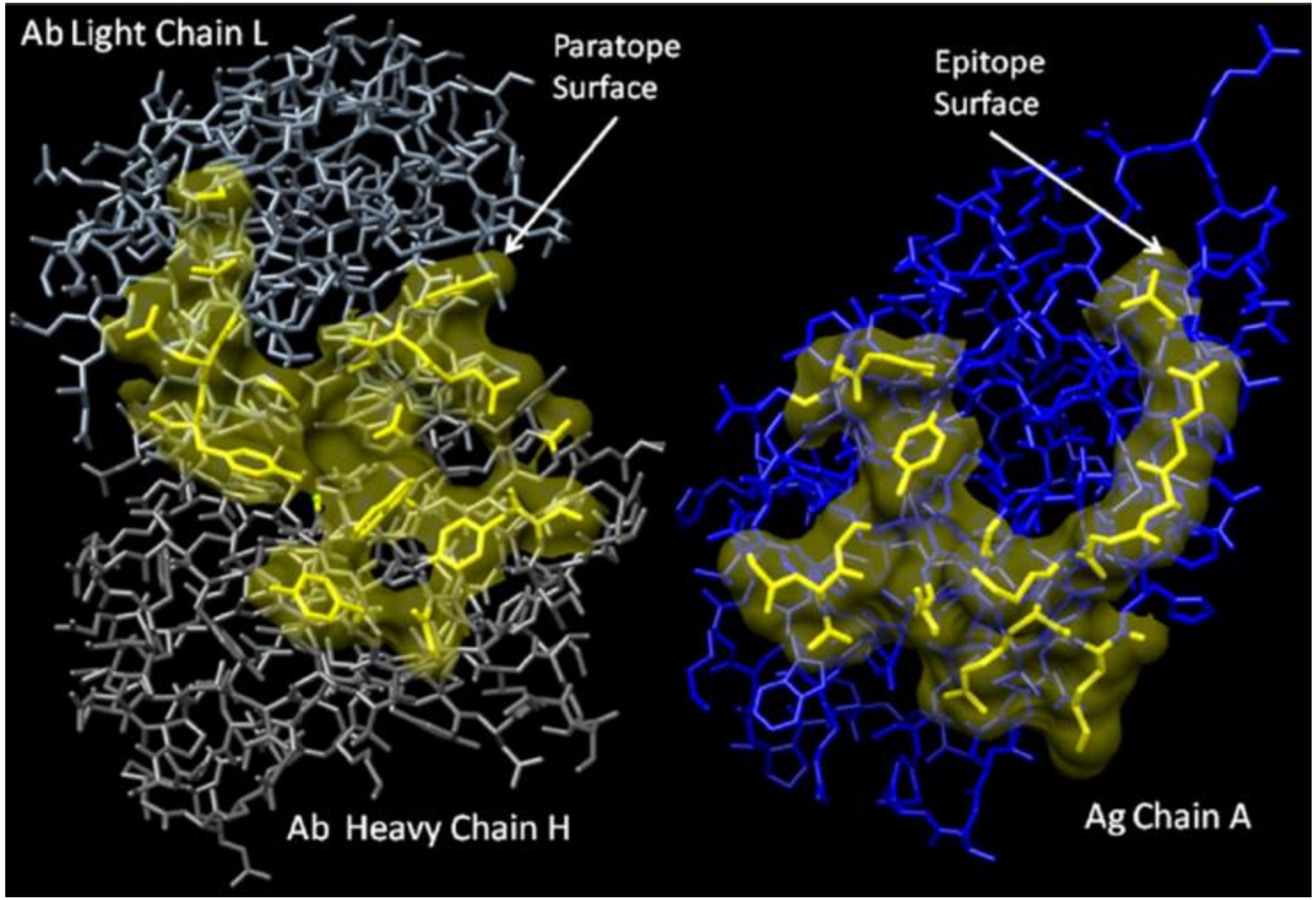
Mian, S. *et al.* (1991) *J. Mol. Biol.* 217, 133

Lessons from antibodies: antibody paratopes engage protein epitopes



Ramaraj, T. *et al.* (2012) *Biochem. Biophys. Acta* **1824**, 520

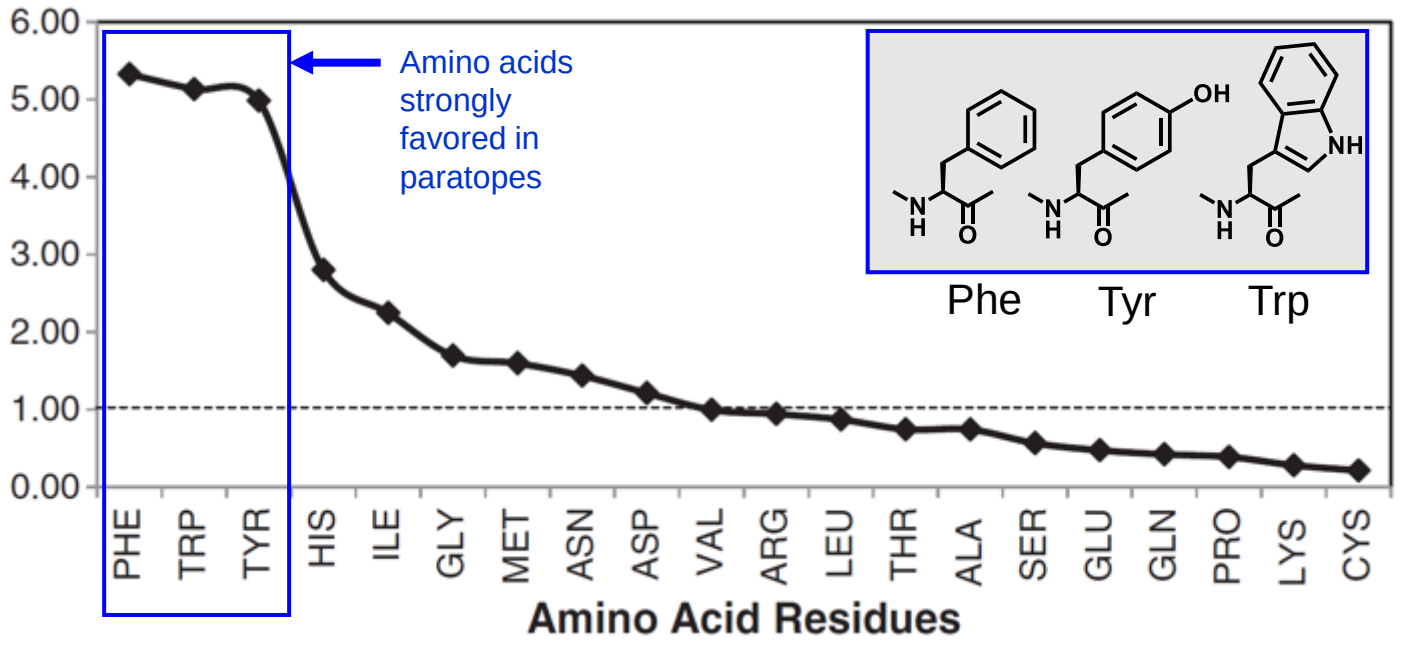
Lessons from antibodies: antibody paratopes engage protein epitopes



Ramaraj, T. *et al.* (2012) *Biochem. Biophys. Acta* **1824**, 520

Side chains favored in antibody paratopes

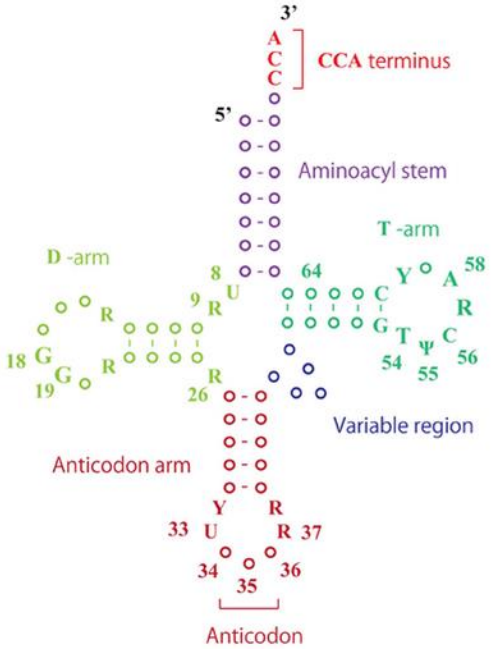
Frequency of Amino Acids in Paratopes (Relative to Rest of Ab Surface)

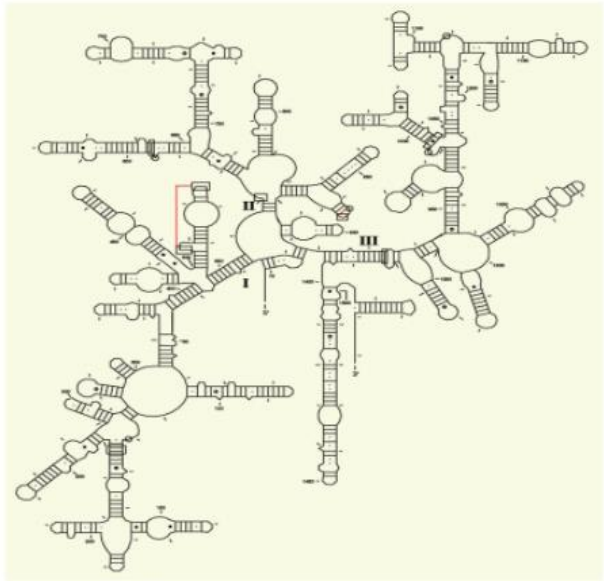


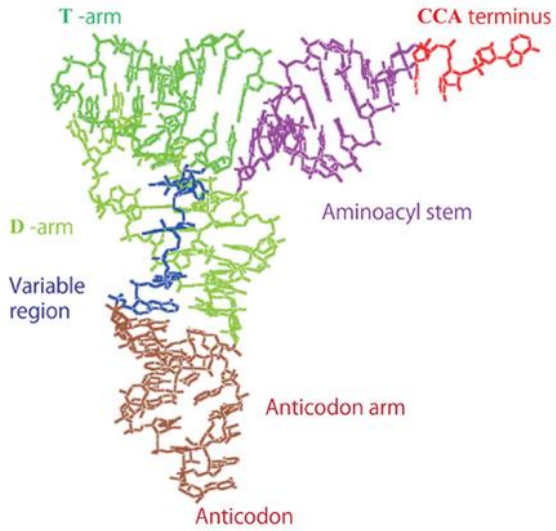
Ramaraj, T. *et al.* (2012) *Biochem. Biophys. Acta* 1824, 520

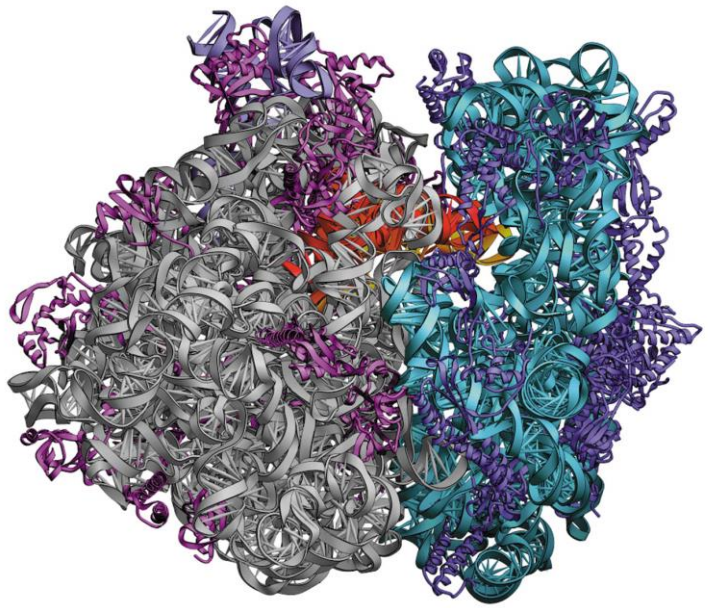
Can single-stranded nucleic acids fold into unique shapes?

Absolutely!

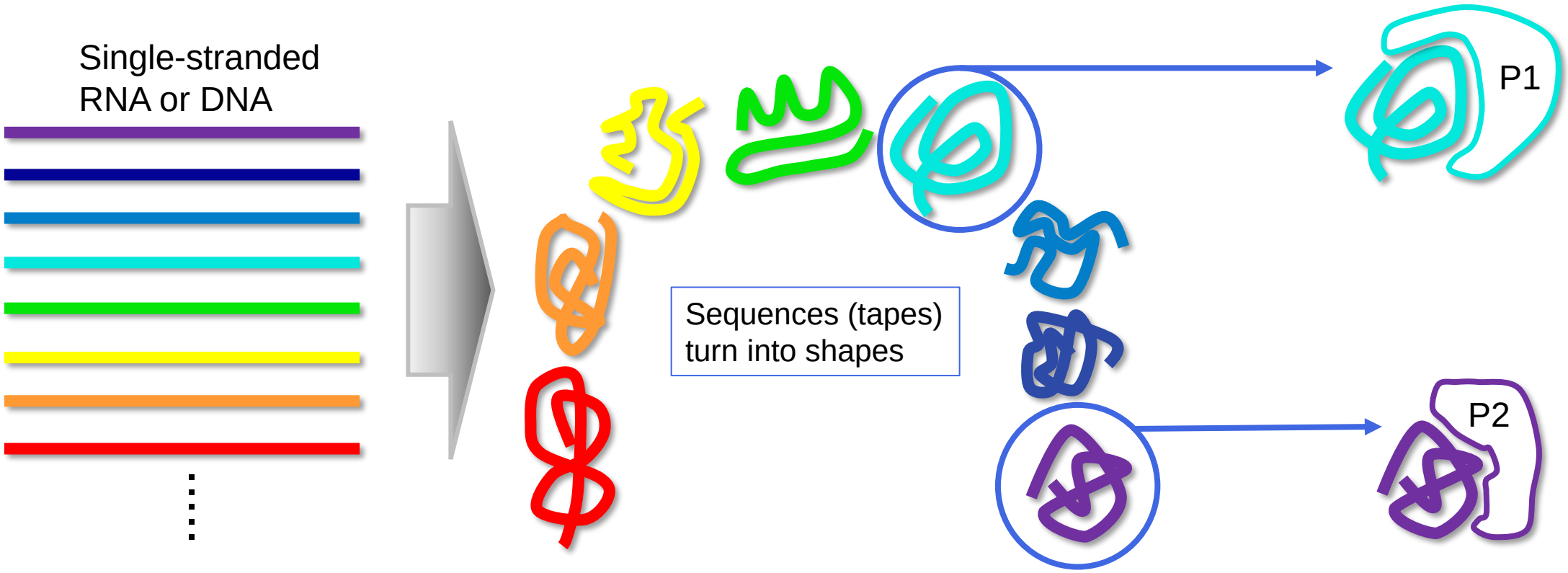








Aptamers: nucleic acid ligands



Make many sequences (10^{15})

Sequences (tapes) turn into shapes

Understanding of folding rules not required

Select shapes (sequences) that (happen to) bind to desired targets

Tuerk and Gold (1990) *Science* **249**, 505
Ellington and Szostak (1990) *Nature* **346**, 818

SELEX and aptamers

Research Articles

Systematic Evolution of Ligands by Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase

SELEX

CRAIG TUERK AND LARRY GOLD

Science 249, 505 (1990)

ARTICLES

In vitro selection of RNA molecules that bind specific ligands

Aptamers

Andrew D. Ellington & Jack W. Szostak*

Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

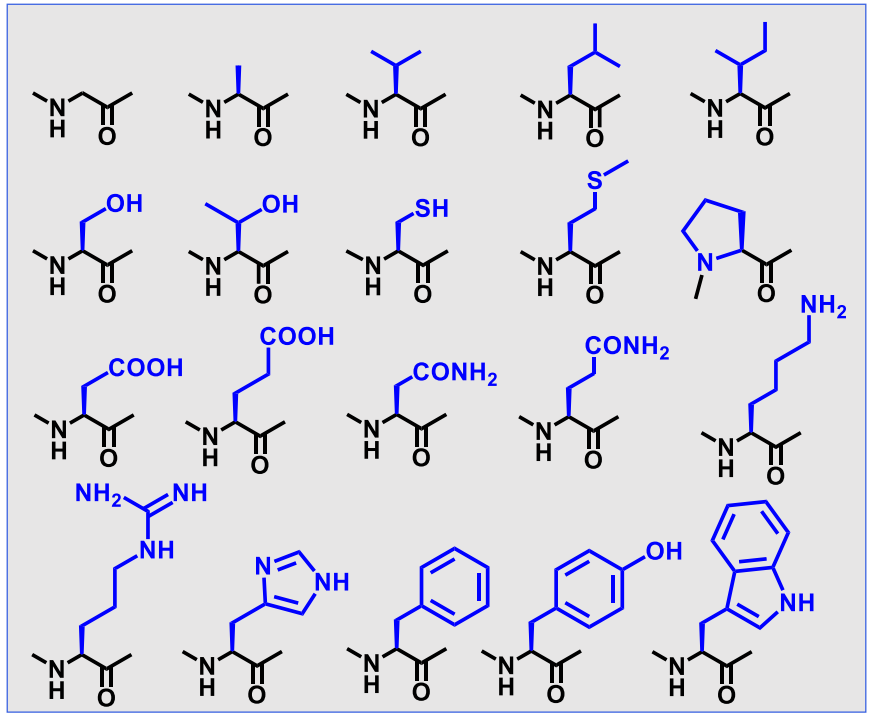
Nature 346, 818 (1990)

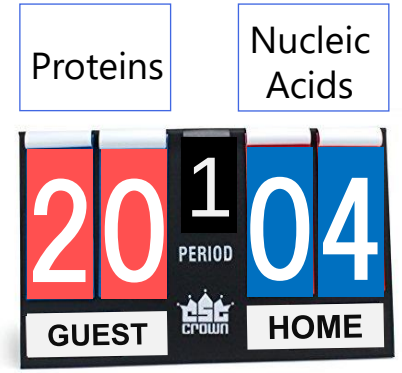
Can aptamers ever be as good as antibodies?

Can nucleic-acid ligands ever be as good as protein ligands?

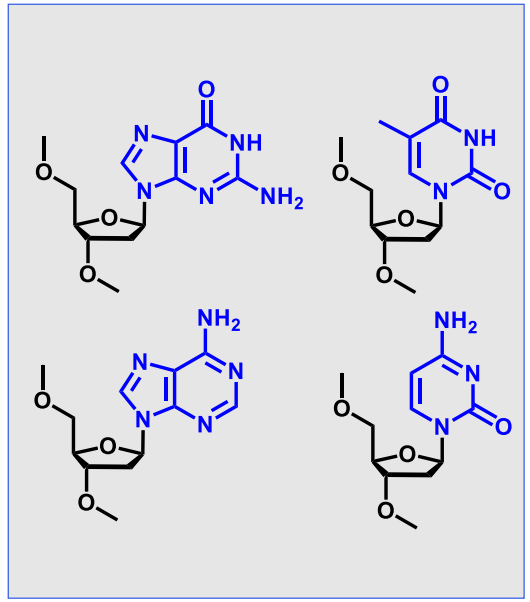
- Proteins: 20 amino acids
- Nucleic acids: 4 bases
  - The bases are kind of similar, in terms of chemical diversity

Amino Acid Side Chains





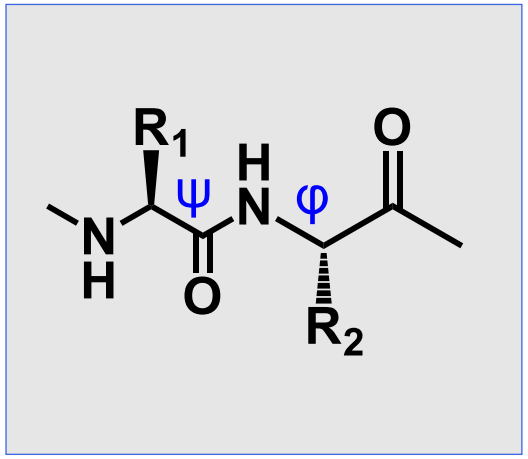
Nucleic Acid Bases

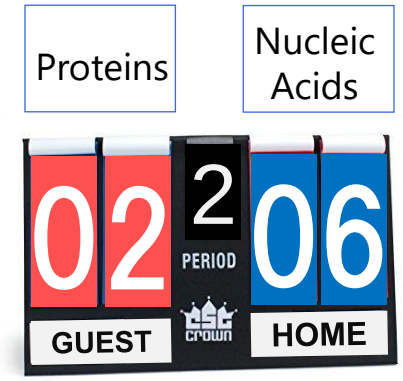


Nucleic acid backbone has more rotational degrees of freedom

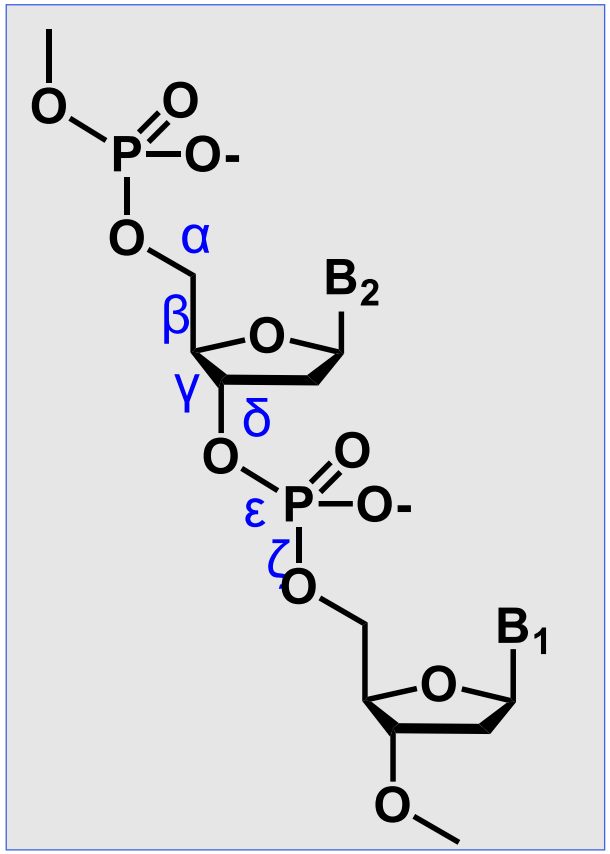
- Phosphodiester backbone has higher conformational flexibility per monomer building block than the peptide backbone (6 *vs.* 2 rotatable bonds)
- More backbone shape diversity per monomer
- Unlike antibodies, backbone conformations of aptamers do not have to assume canonical structures

Peptide backbone





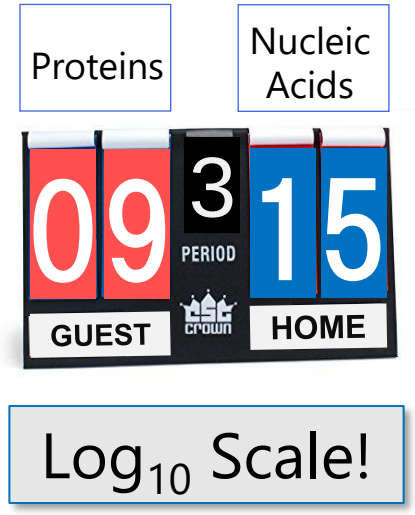
Nucleic acid backbone



Much larger nucleic acid libraries can be screened

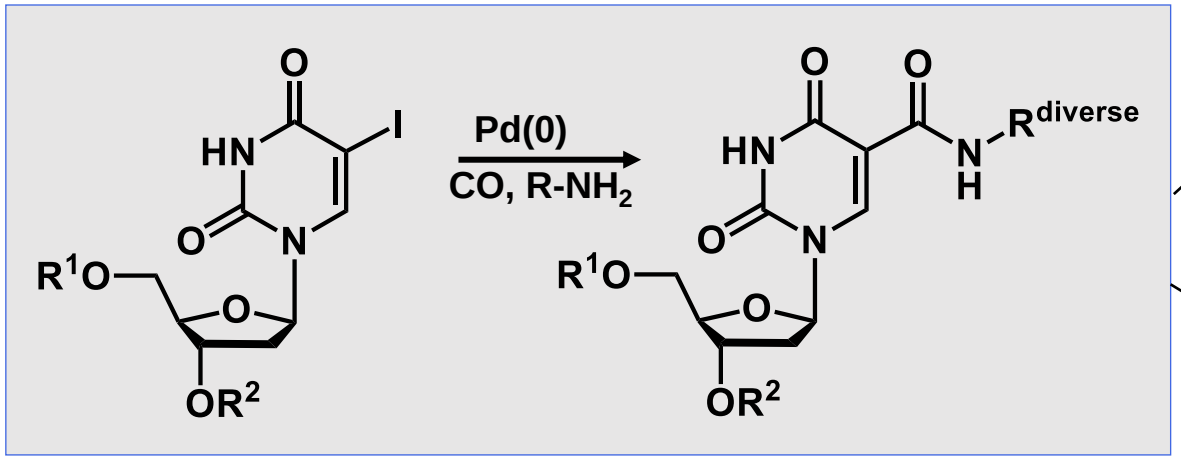
- Unique antibodies in the germline repertoire: $10^8 – 10^9$

- Typical random nucleic acid libraries for SELEX: $10^{13} – 10^{15}$



Introducing additional diversity to aptamers

- Additional diversity can be readily introduced at the 5-position of pyrimidines
  - Many natural variants (m^5U, cm^5U, mcm^5U, ncm^5U, hm^5C, etc.)
  - Away from H-binding face, compatible with enzymatic steps of SELEX
  - Tolerates a variety of functional groups

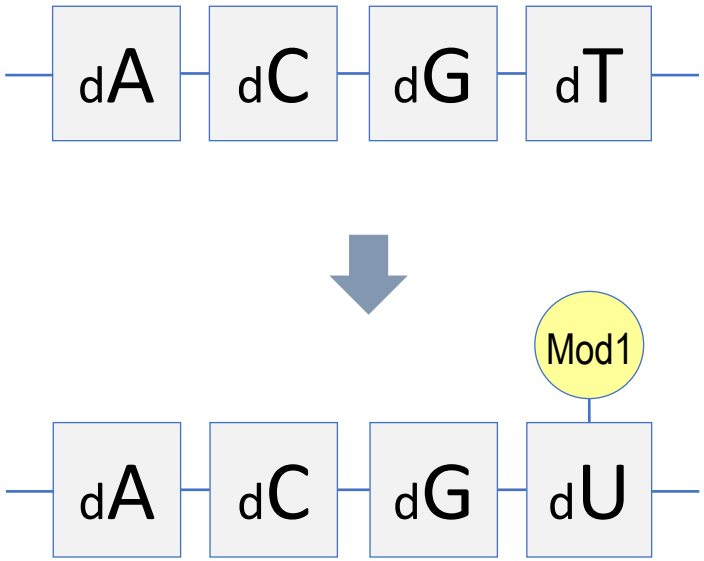


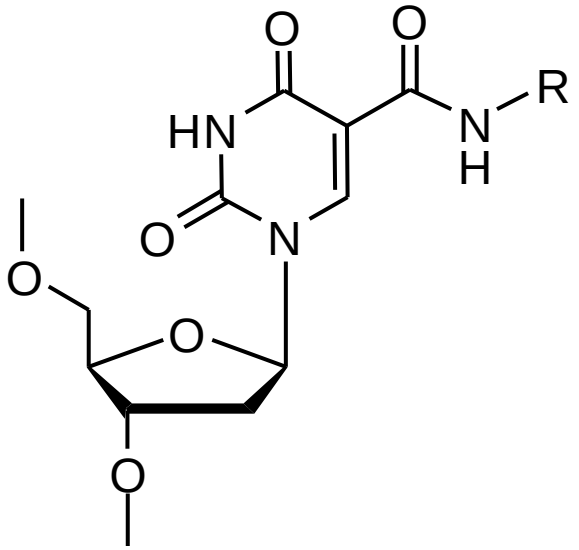
Triphosphates (SELEX)

Phosphoramidites (synthesis)

Dewey *et al*. (1996) *Nucleotides & Nucleosides* **15**, 1611
Vaught *et al.* (2010) *J. Am. Chem. Soc.* **132**, 4141

Modified DNA for SELEX





Wide range of side chains possible

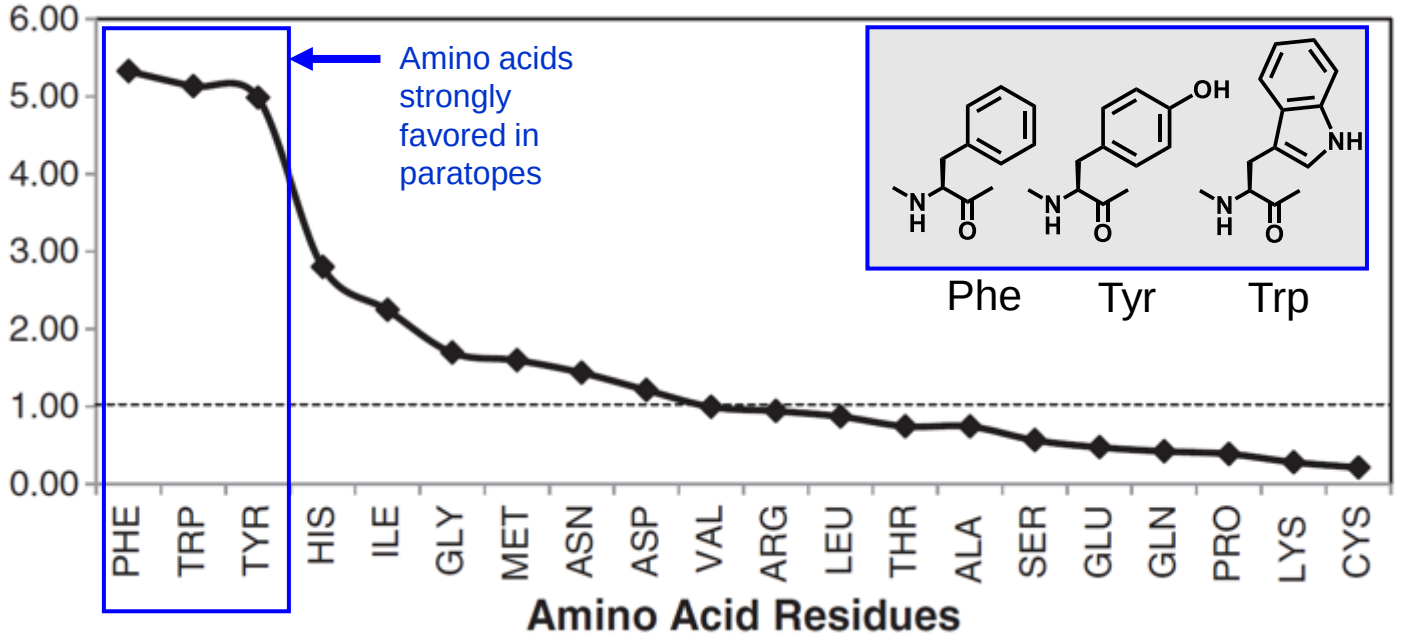
5-position-modified dU

- One base modified
- DNA library uniformly substituted with the same modification
- What are the best modifications for SELEX?

Vaught *et al.* (2010) *J. Am. Chem. Soc.* **132**, 4141
Gold *et al*. (2010) *PloS ONE* **5**, e15004

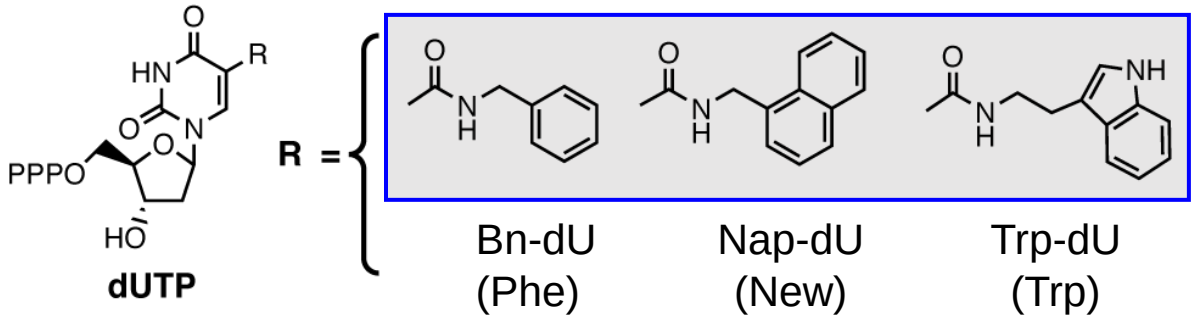
Side chains favored in antibody paratopes
Also make the biggest impact in SELEX

Frequency of Amino Acids in Paratopes (Relative to Rest of Ab Surface)

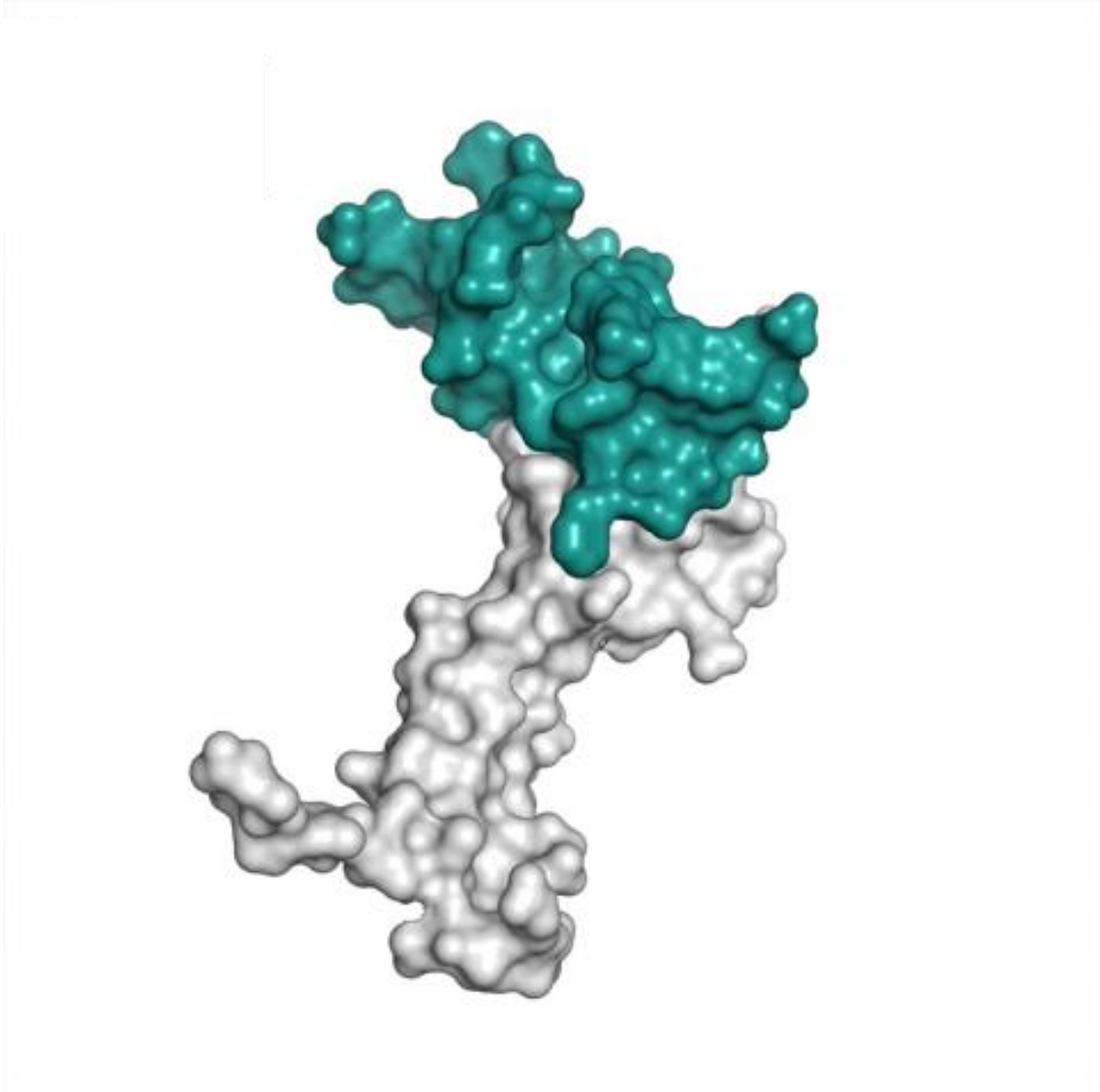


Ramaraj, T. *et al.* (2012) *Biochem. Biophys. Acta* 1824, 520

Types of base modifications with best performance in SELEX

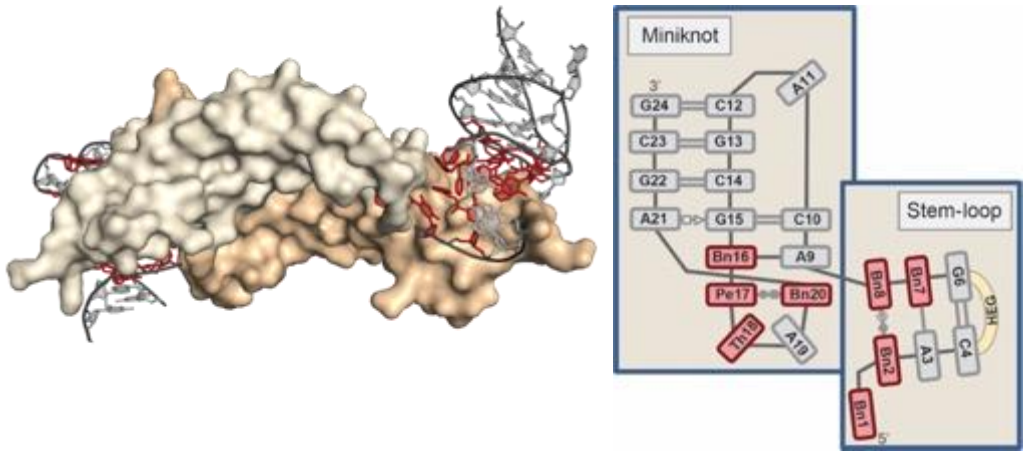


Gold *et al*. (2010) *PloS ONE* **5**, e15004

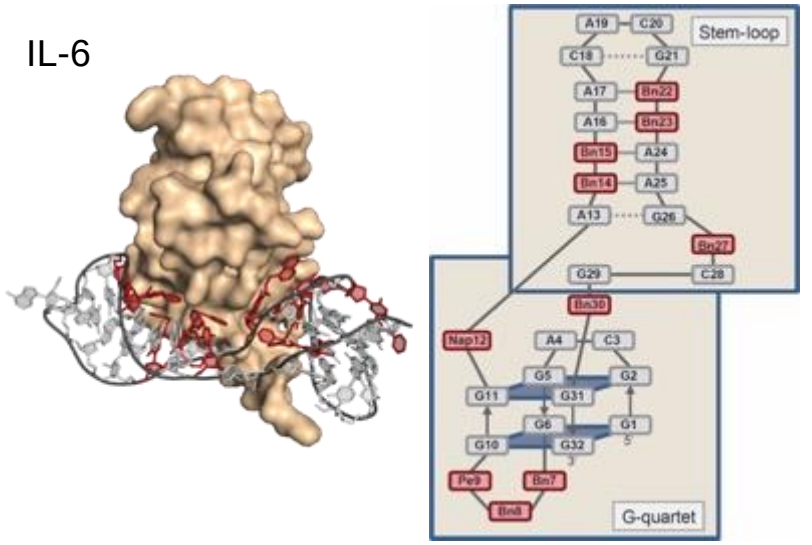


Side chains play a role in folding and binding

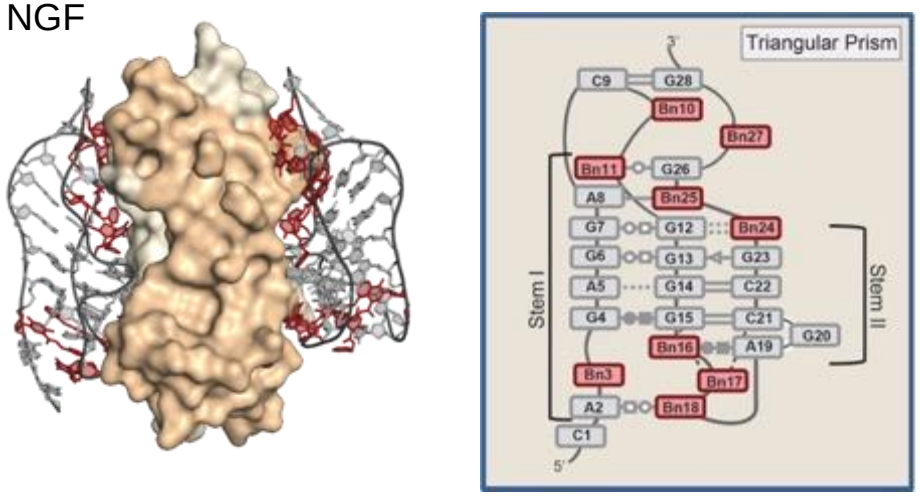
PDGF-BB



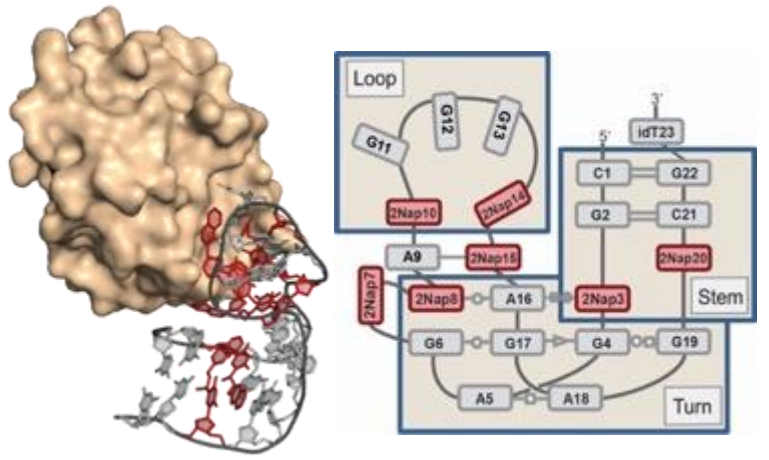
IL-6



NGF



IL-1α



somalogic

Davies *et al.* (2012) *PNAS* **109**, 19971

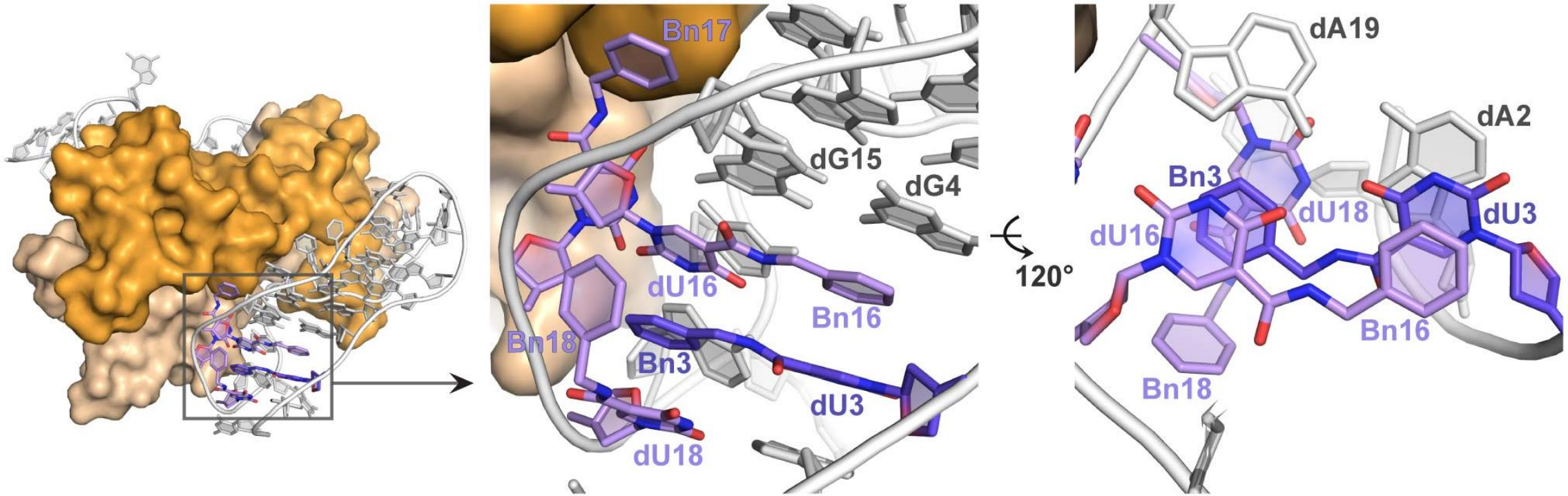
Gelinas *et al.* (2014) *J. Biol. Chem.* **289**, 8720

Jarvis *et al.* (2015) *Structure* 23, 1293

Ren *et al.* (2017) *Nat Commun* **8**, 810

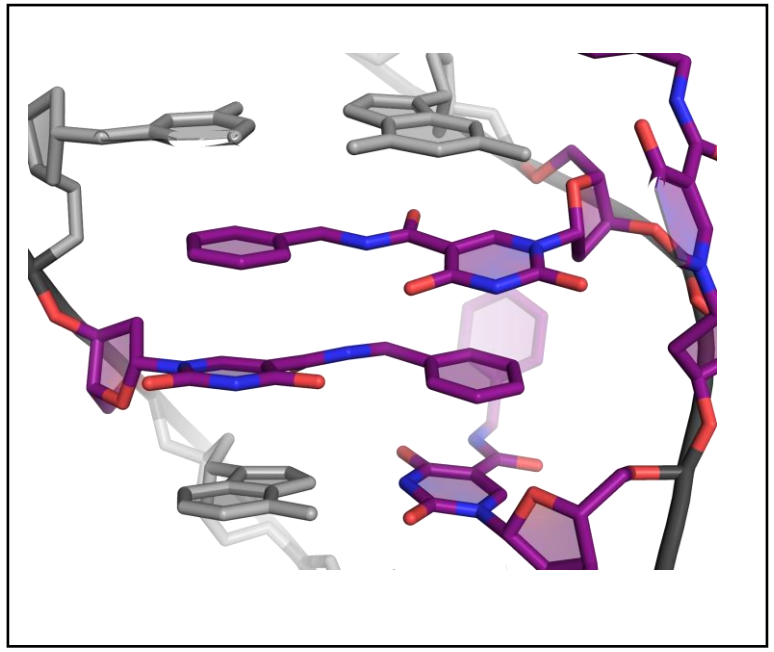
Side chains stabilize folded structures

Zipper Motif (NGF SOMAmer Reagent)

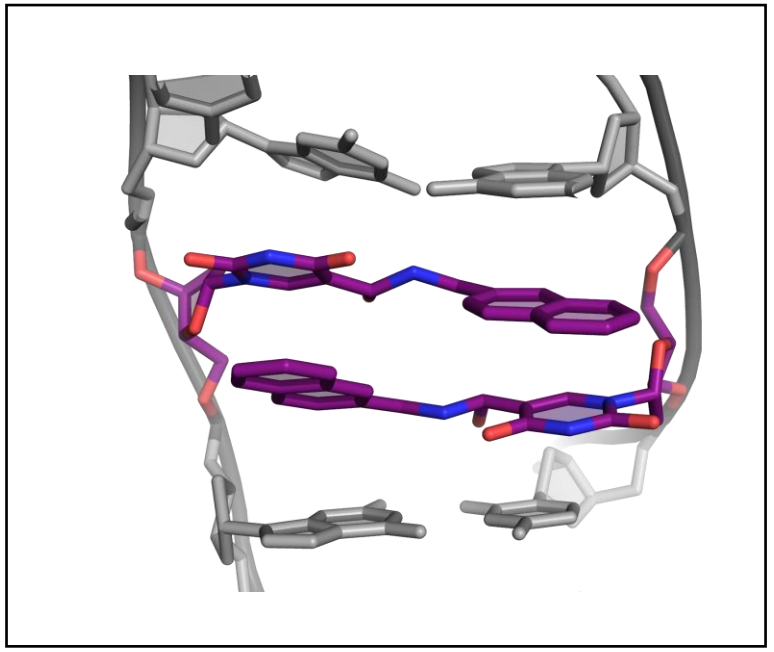


Some novel motifs are recurring

Zipper Motifs



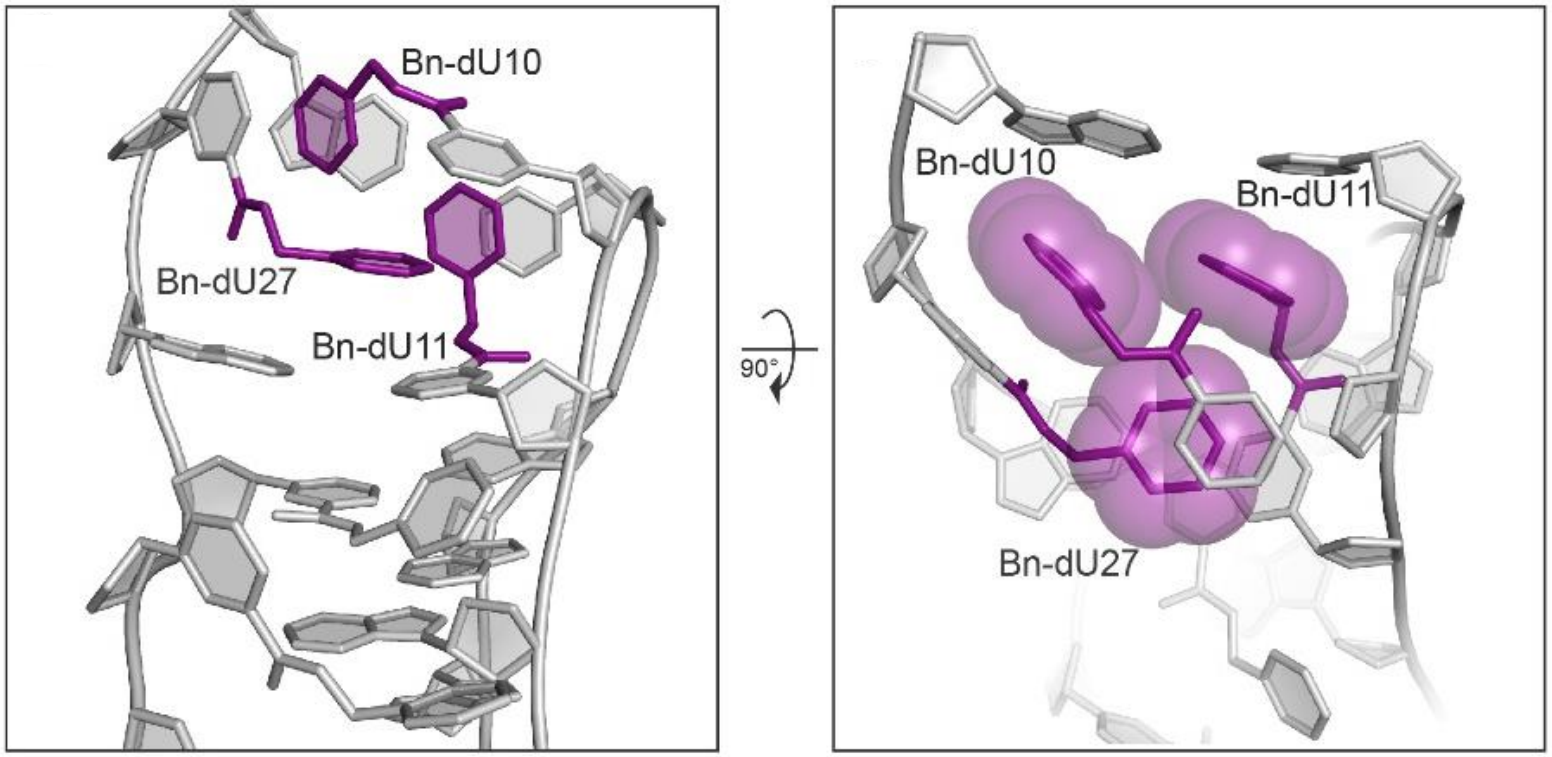
Benzyl zipper
(NGF SOMAmer)



Naphthyl zipper
(IL-1α SOMAmer)

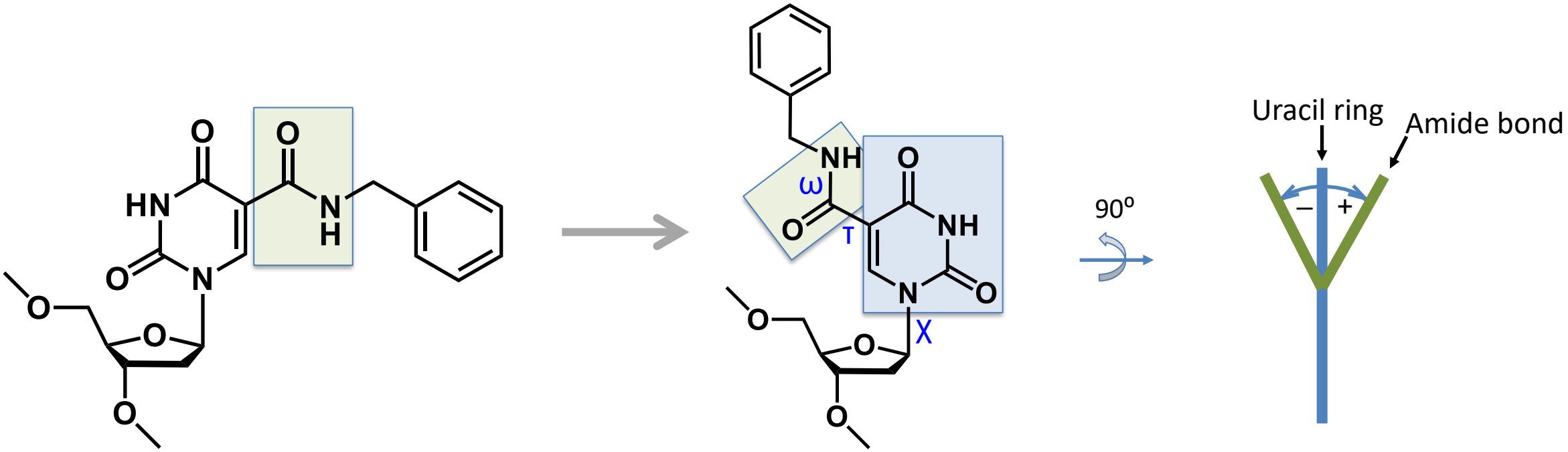
Side chains stabilize folded structures

Hydrophobic cluster motifs



Amide linker is not just a passive spacer

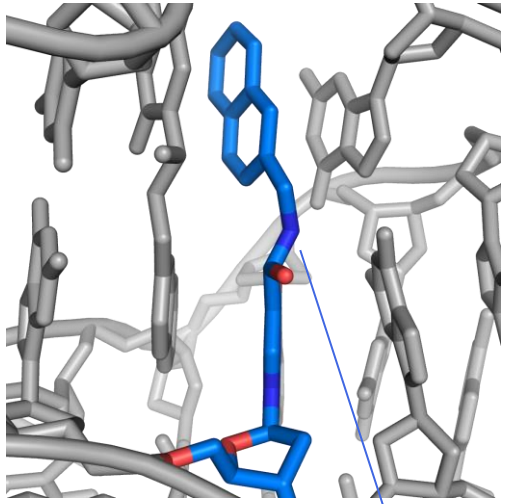
- Imposes conformational rigidity at the point of attachment to the uracil ring
  - Anecdotal evidence suggests this may be important for success rate of SELEX
- Participates in making internal motifs and protein contacts

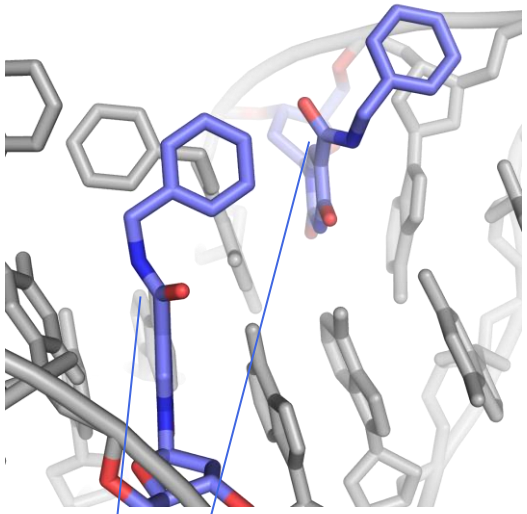


χ = glycosidic bond (88% *anti*)
τ = carbonyl orientation (100% *anti*)
ω = amide bond (94% *trans*)

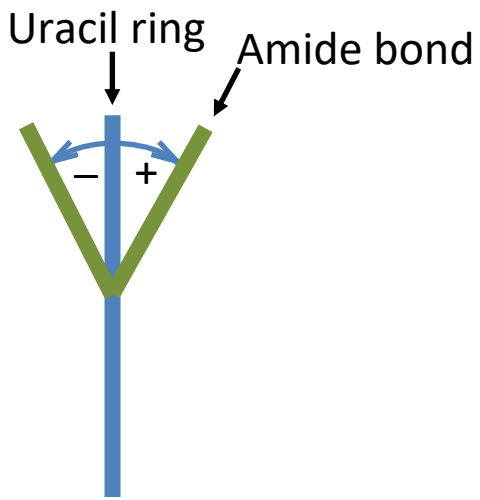
Amide linker imposes rigidity

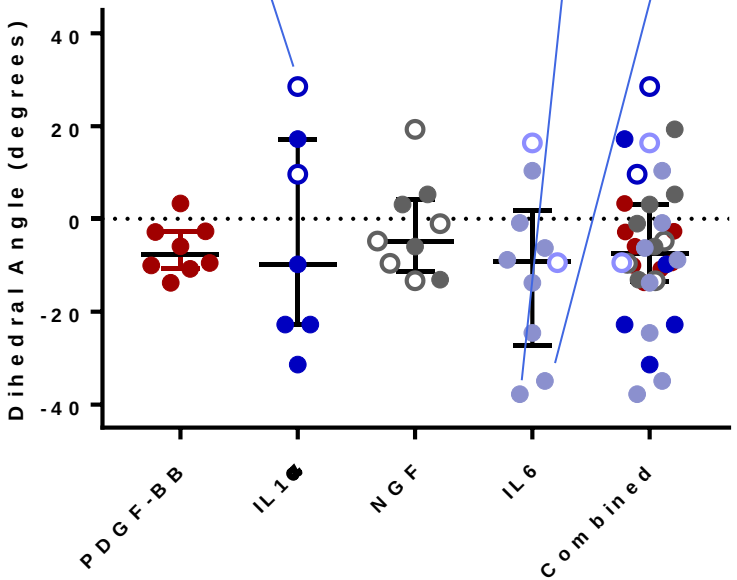
IL-1α:
Internal motif





IL-6:
Protein contacts



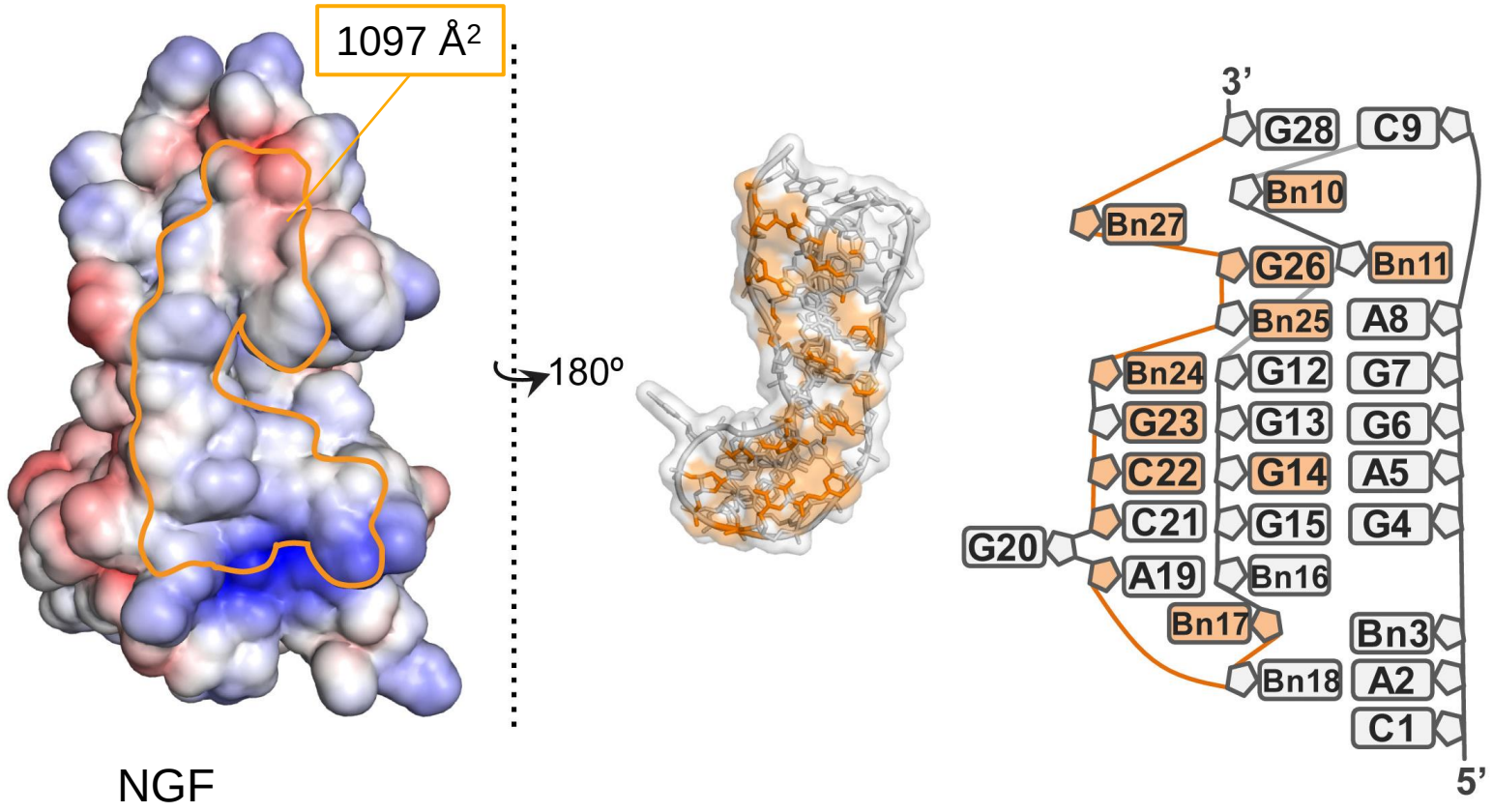


Closed symbols:
protein contacts

Open symbols:
internal motifs or
solvent-exposed

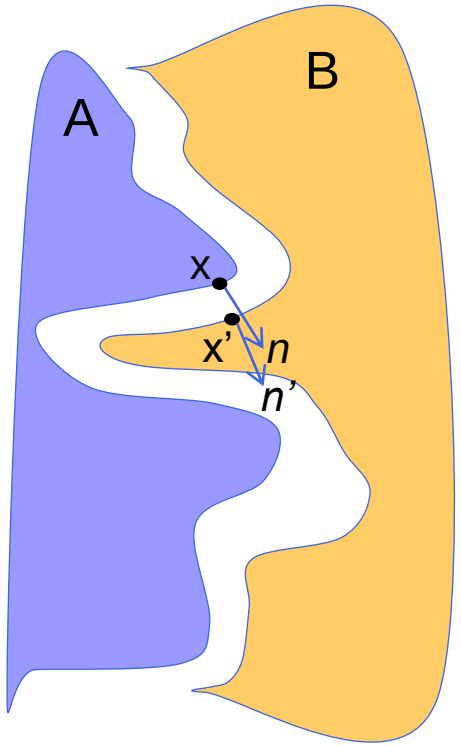
Epitopes for aptamers: size and shape

- Interaction surfaces for aptamers and antibodies have similar size
  - Antibodies: 560 – 1300 Å^2
  - Aptamers: 350 – 1280 Å^2



Gelinas et al (2016) *Curr. Opin. Struct. Biol.* **36**, 122

Aptamers and antibodies engage proteins with exquisite shape complementarity



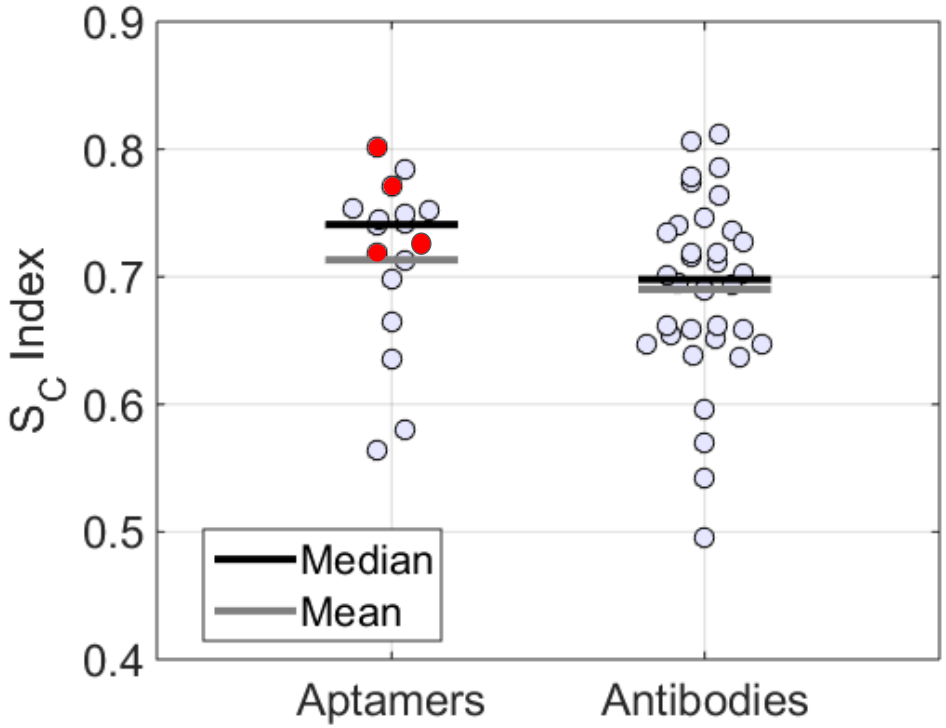
Shape complementarity index (S_c)

$S^{A\text{-}B}(x_A) = n_A \cdot n'_A \exp[-w(|x_A - x'_A|)^2]$

$S_c = (S^{A\text{-}B} + S^{B\text{-}A})/2$

Oligomeric protein interfaces: S_c=0.70-0.76

SOMAmers: red symbols



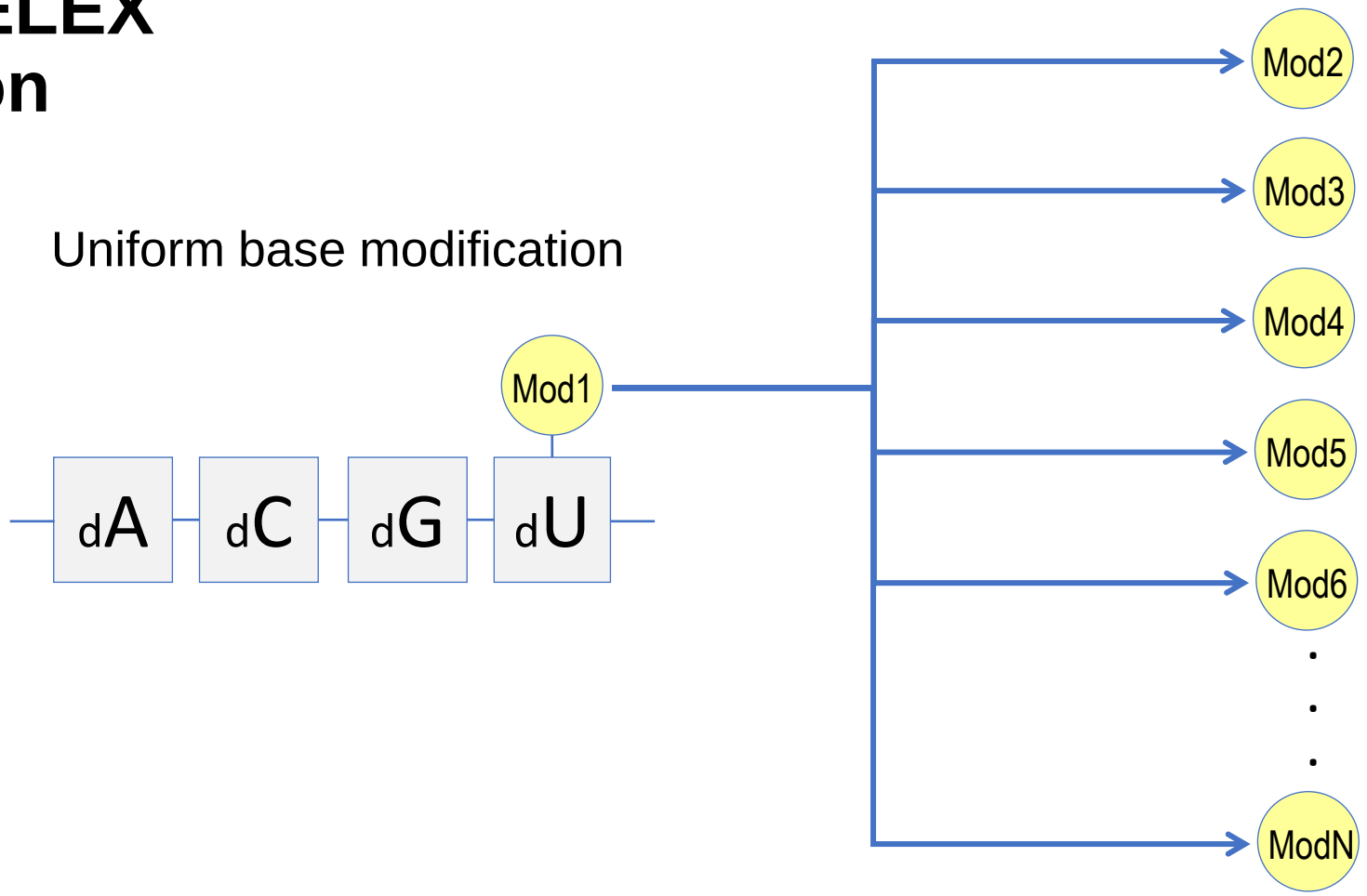
Lawrence & Colman (1993) *J. Mol. Biol.* **234**, 946

Gelinas et al (2016) *Curr. Opin. Struct. Biol.* **36**, 122

Side chains are convenient handles for post-SELEX optimization

Akin to SAR in small molecules,
or affinity maturation in antibodies

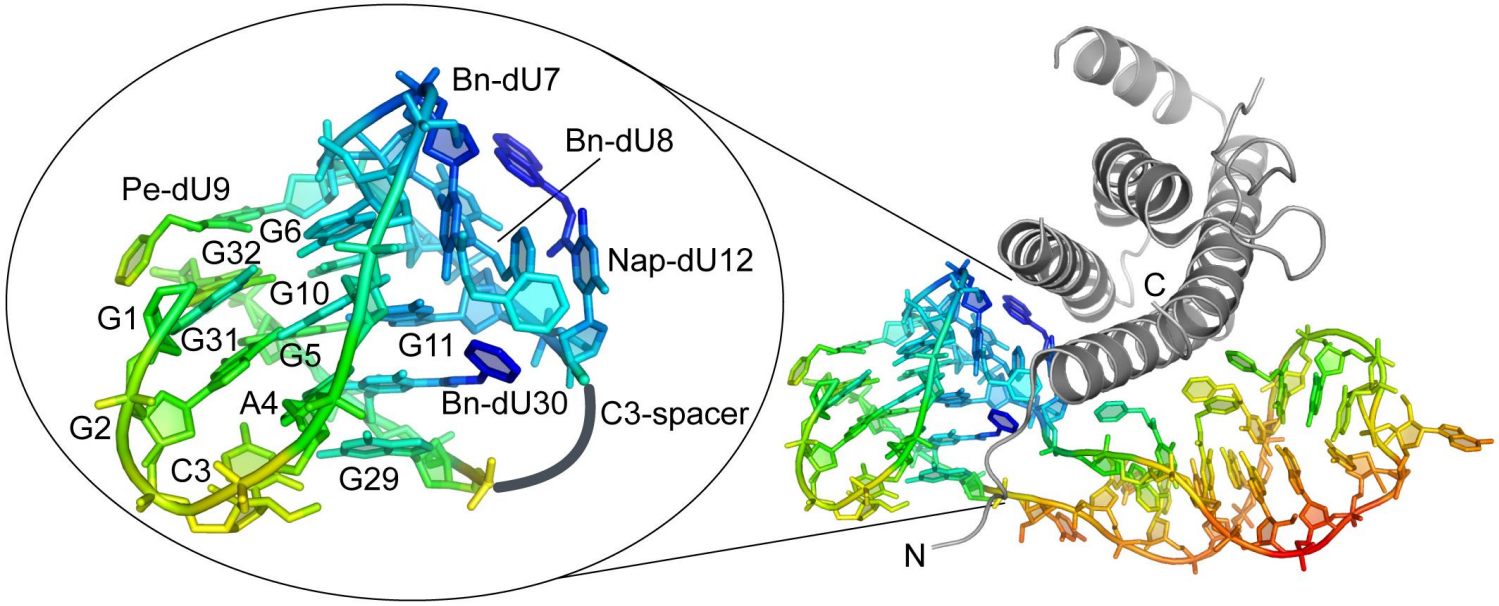
Uniform base modification



Gelinas *et al.* (2014) *J. Biol. Chem.* **289**, 8720

IL-6 SOMAmer: disconnecting the G-quartet domain

B-factors: 31-95 Å^2 (blue-red)



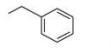
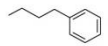
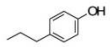
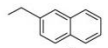
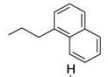
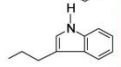
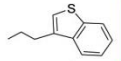
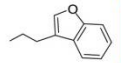
K_d = 270 nM

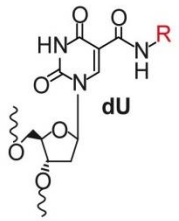
K_d = 0.2 nM

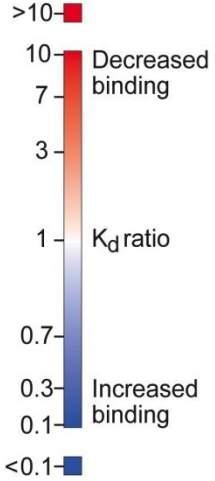
SeqID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
SL1025	G	G	mC	A	G	mG	Bn	Bn	Pe	G	G	Nap	A	Bn	Bn	mA	A	C	mA	mC	G	Bn	Bn	A	A	G	Bn	mC	G	Bn	G	G
SL1028	G	G	mC	A	G	mG	Bn	Bn	Pe	G	G	Nap	C3 Spacer																G	Bn	G	G

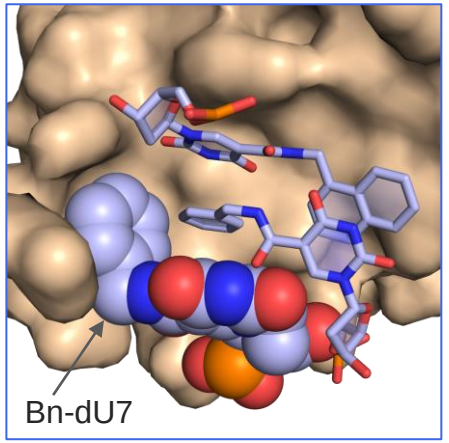
Gelinas *et al.* (2014) *J. Biol. Chem.* **289**, 8720

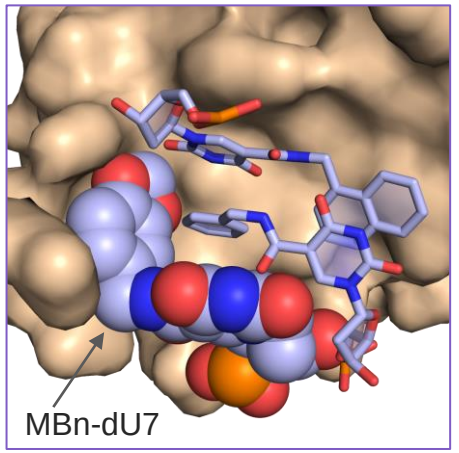
Post-SELEX optimization of G-quartet domain

5-dU Modification	R group	Bn-dU7	Bn-dU8	Pe-dU9	Nap-dU12	Bn-dU30
Bn-dU		1	1	N.D.	1	1
iBu-dU		>100	>100	1.4	>100	0.15
FBn-dU		4.1	1.3	1.4	>100	1.2
Pe-dU		0.82	>100	1	>100	2.4
Pp-dU		1.7	>100	1.1	>100	1.3
Tyr-dU		3.8	>100	1.5	>100	>100
MBn-dU		0.03	0.70	1.3	1.1	2.9
Nap-dU		0.15	>100	1.3	1	3.4
2Nap-dU		0.56	>100	0.64	1.2	0.72
Ne-dU		0.44	>100	0.79	2.4	1.3
Trp-dU		0.34	>100	0.81	>100	2.3
Bt-dU		0.23	1.1	>100	1	>100
MOE-dU		>100	>100	1.4	>100	>100
Bf-dU		0.04	>100	0.97	>100	1.8
RTHF-dU		>100	>100	1.3	>100	2.4
STHF-dU		>100	>100	2.5	>100	3.8





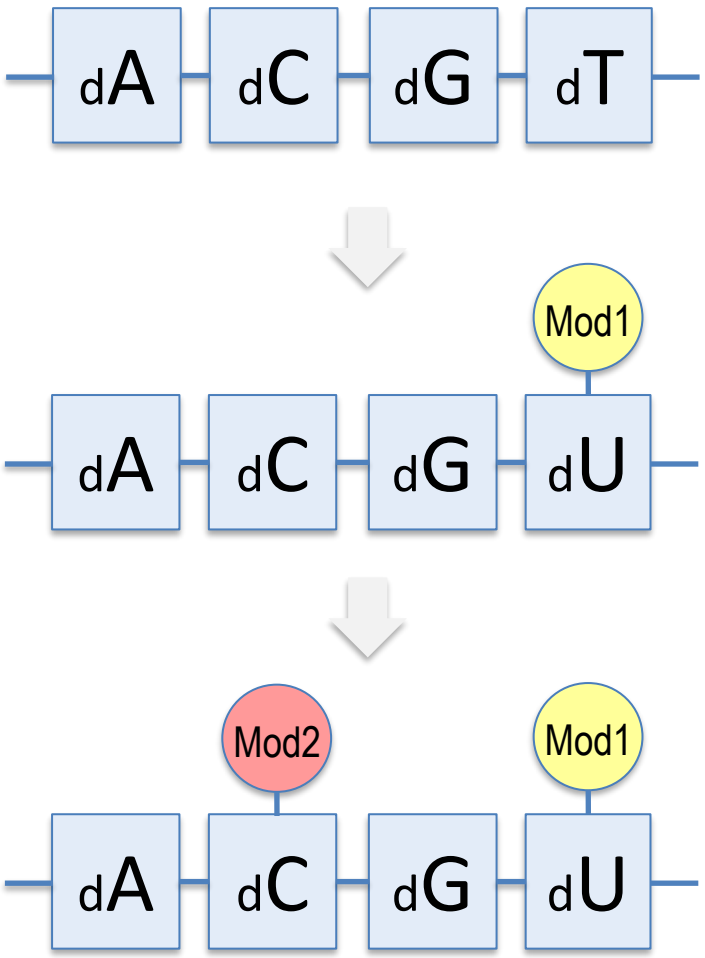




K_d = 270 nM K_d = 7.4 nM

Gelinas *et al.* (2014) *J. Biol. Chem.* **289**, 8720

Introducing a second point of diversity



Unmodified DNA Library

One base uniformly modified

Two bases modified

doi:10.1016/j.jmb.2005.03.041 J. Mol. Biol. (2005) 348, 1153–1162

JMB

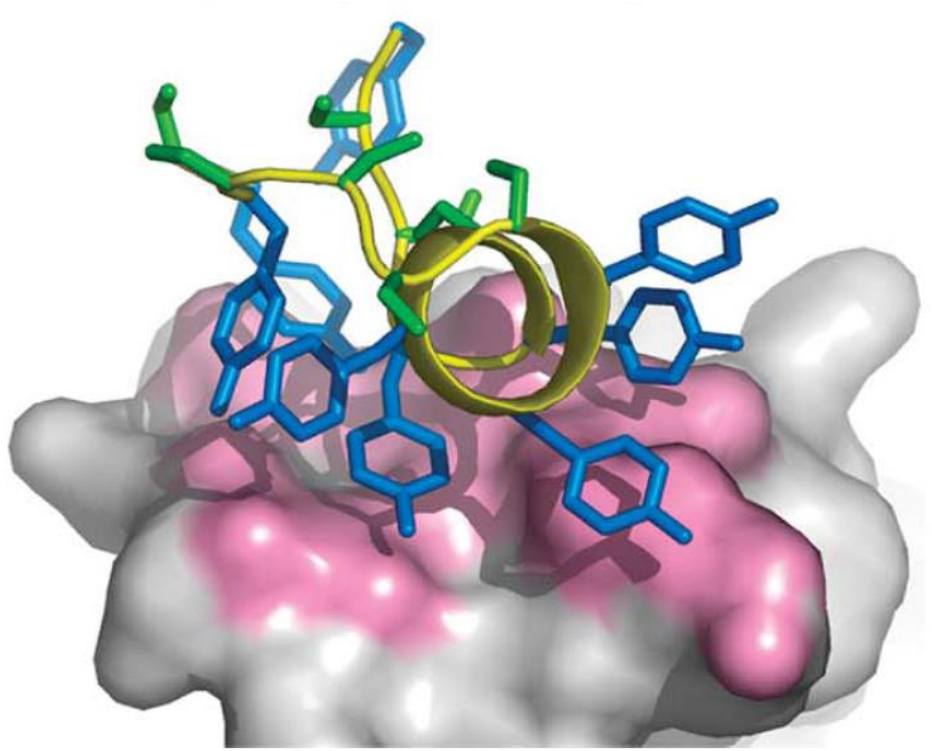
Available online at www.sciencedirect.com
SCIENCE DIRECT®

ELSEVIER

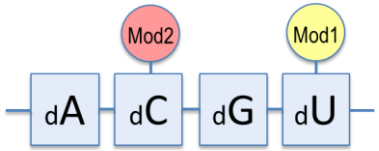
Molecular Recognition by a Binary Code

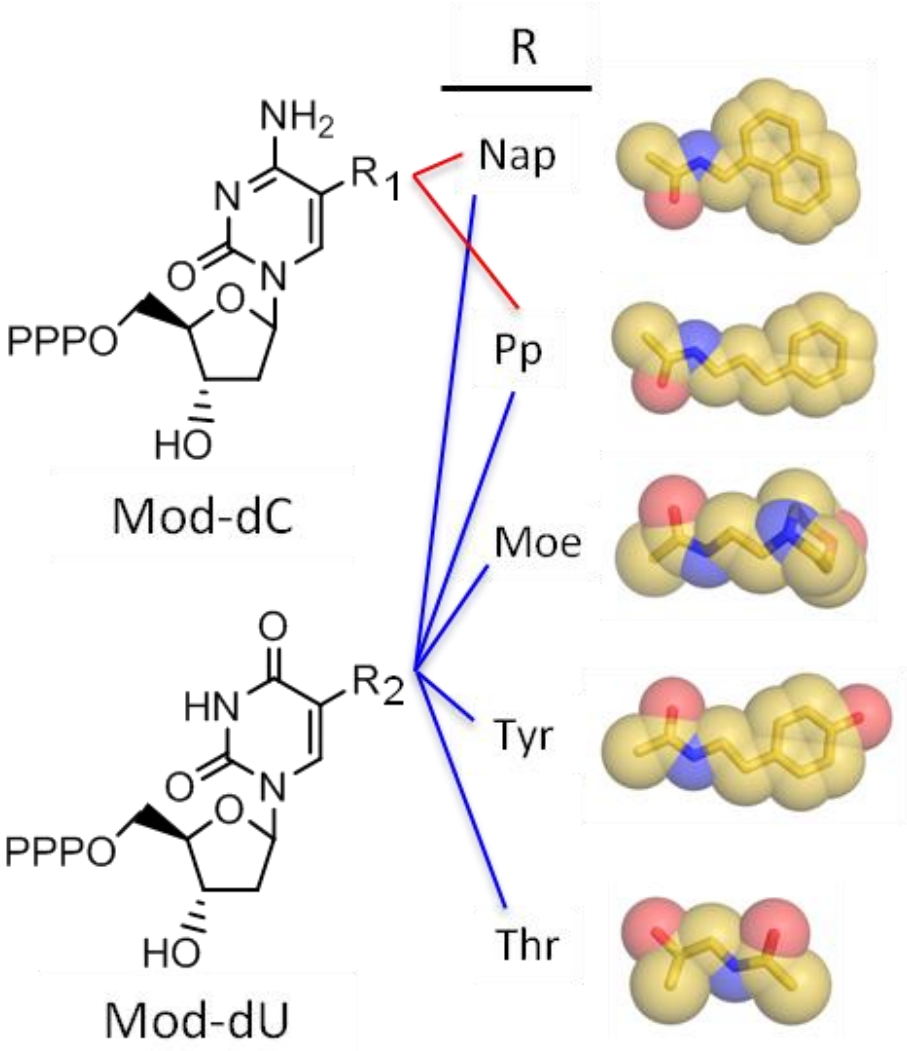
Frederic A. Fellouse, Bing Li, Deanne M. Compaan, Andrew A. Peden Sarah G. Hymowitz and Sachdev S. Sidhu*

"Functional antibodies were obtained from a library of antigen-binding sites…restricted to tyrosine and serine…which may be particularly well suited to work together in antigen recognition."

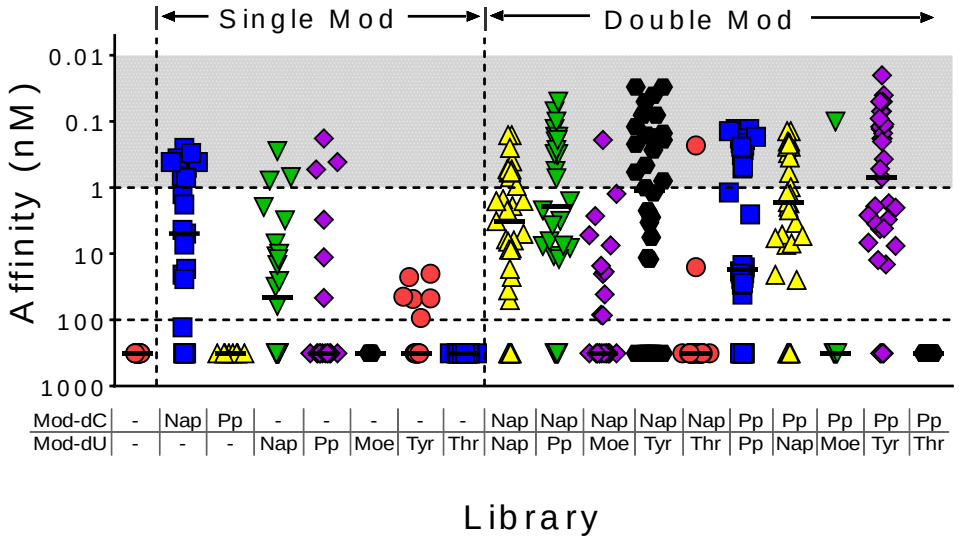


Libraries with two modifications

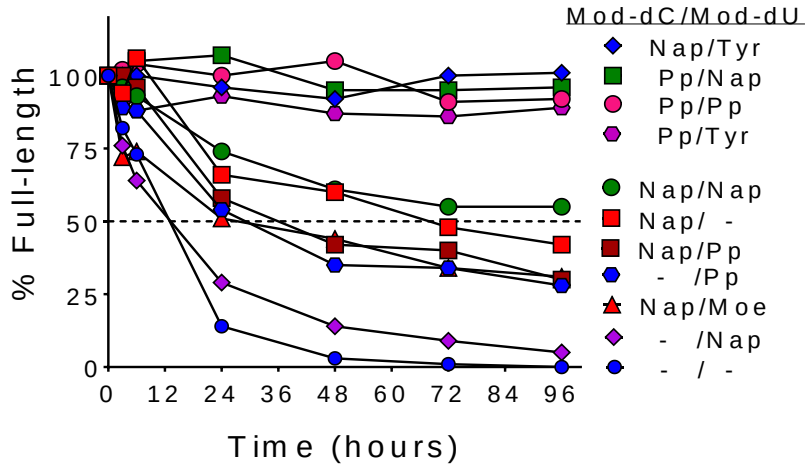




Affinity for target (PCSK9)

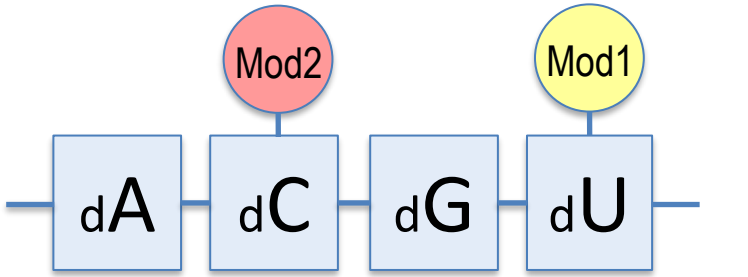


Stability in human serum



Gawande *et al.* (2017) *Proc. Natl. Acad. Sci.* **114**, 2898

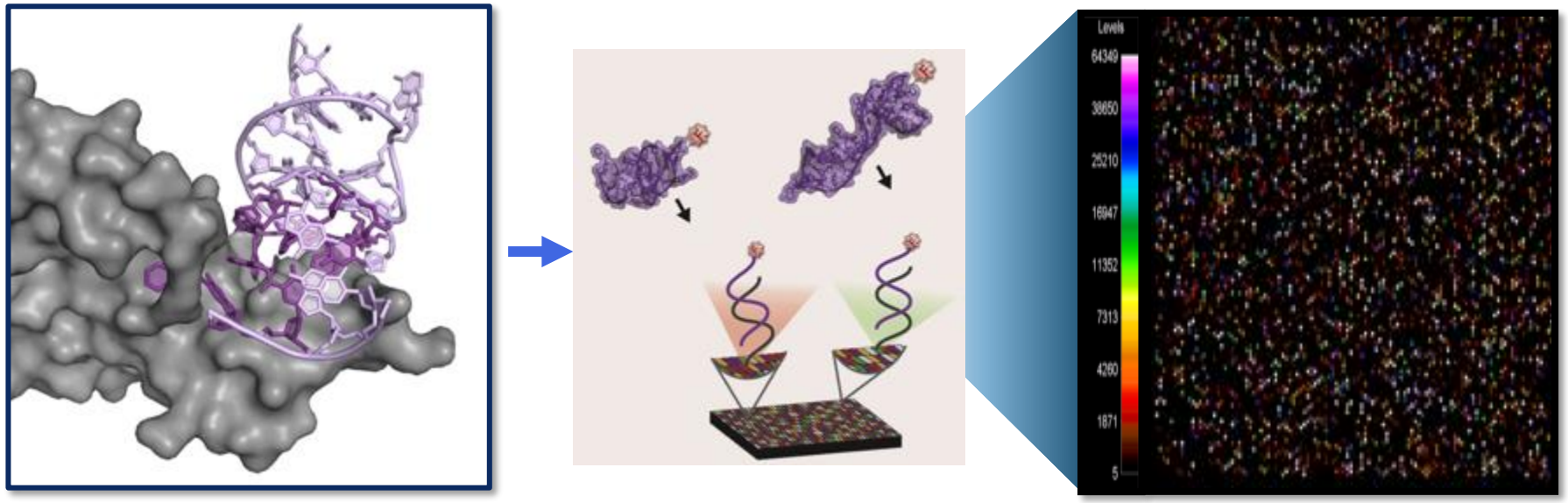
Advantages of two modifications



- Higher affinity
- More complete epitope coverage (easier to identify sandwich pairs)
- Enhanced nuclease resistance
- Binding properties encoded in shorter sequences

So, what do we do with all this stuff?

SomaScan assay: unbiased interrogation of the proteome



Shapes to tapes
Proteins to DNA

- Measures 7,000 proteins simultaneously (current version)
- Proteins measured via nucleic acid signal
- Dynamic range (>10 logs) and sensitivity (buffer LLOD 125 fM)
- Low sample requirement (55 µL plasma or serum)
- Allows systems-level understanding of biology

Gold *et al*. (2010) *PloS ONE* **5**, e15004
Rohloff *et al.* (2014) *Mol. Ther. Nucl. Acids* **3**, e201

Unrecognized medical conditions overrepresented among scientists

***Ahypothesemia*.** Characterized by the absence of a hypothesis. Some scientists have hypothesized that this is a problem. See also *hypothesosis*.

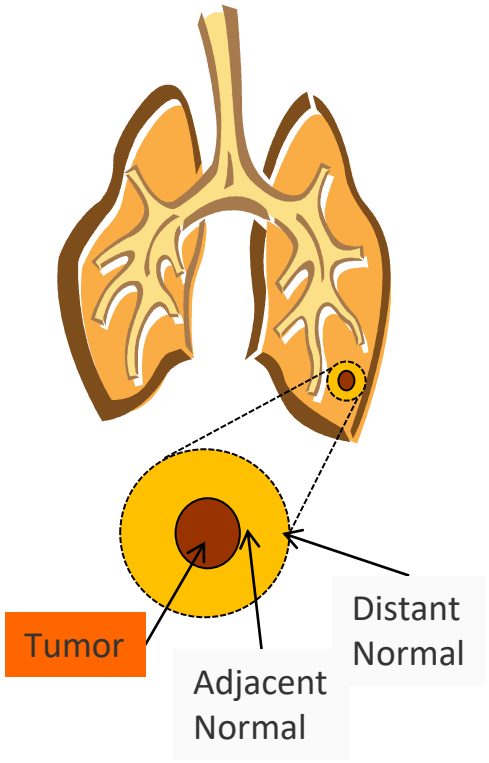
***Hypothesosis*.** Characterized by an inability to recognize that not all research requires a hypothesis.

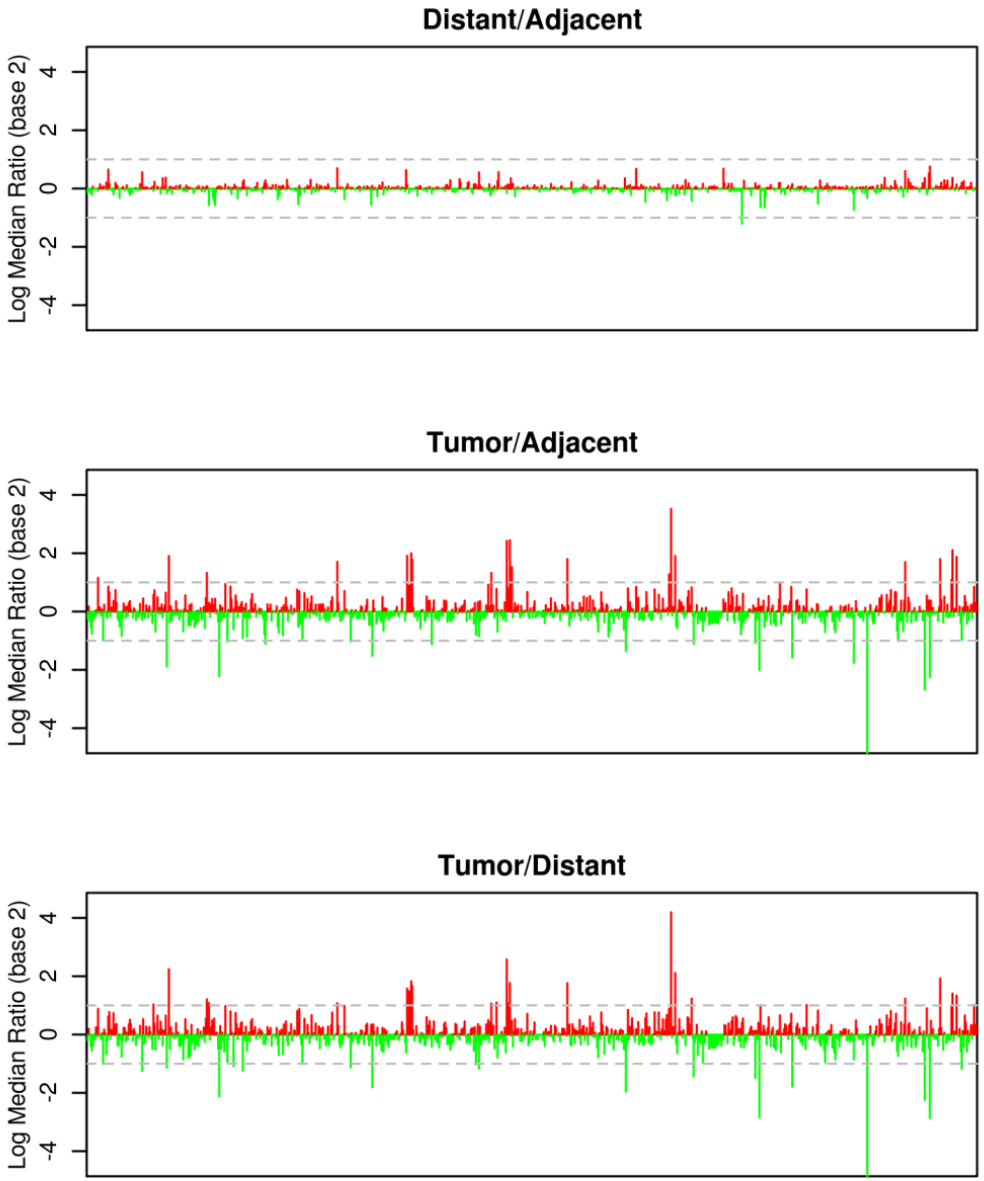
From “Diseased science” by Casadevall and Fang (2014)

Casadevall and Fang (2014) *Microbe* **9**, 390

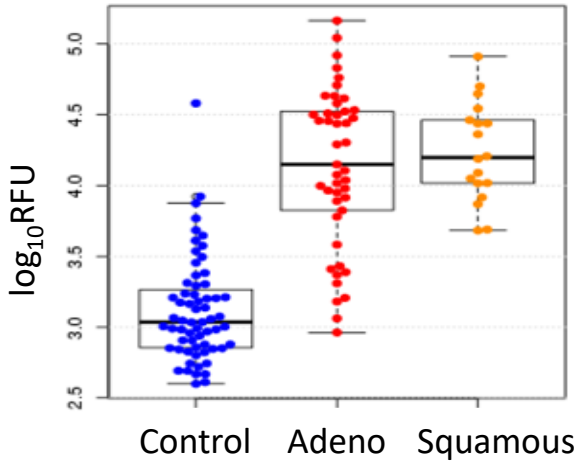

SYMPATHY CARDS FOR SCIENTISTS
CONDOLENCES ON YOUR FAILED EXPERIMENT
SORRY TO HEAR ABOUT YOUR REJECTED PAPER
WITH DEEPEST SYMPATHY ON YOUR DISPROVED THEORY
TOM GAULD

Proteome analysis in non-small cell lung cancer

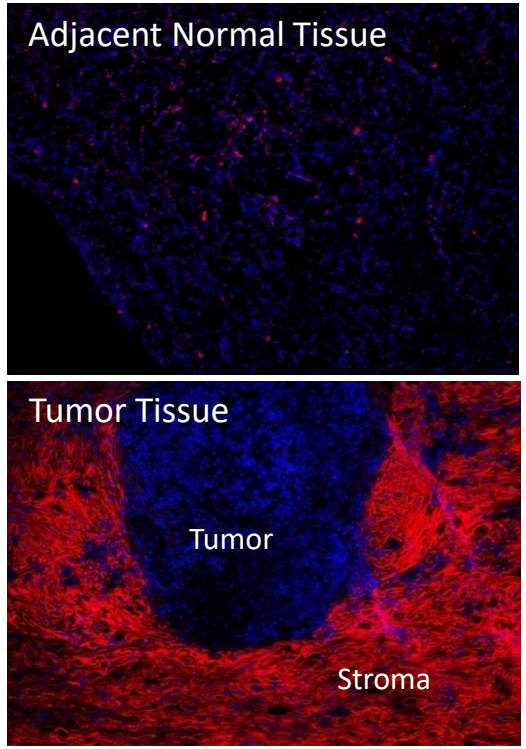




Thrombospondin 2, SomaScan

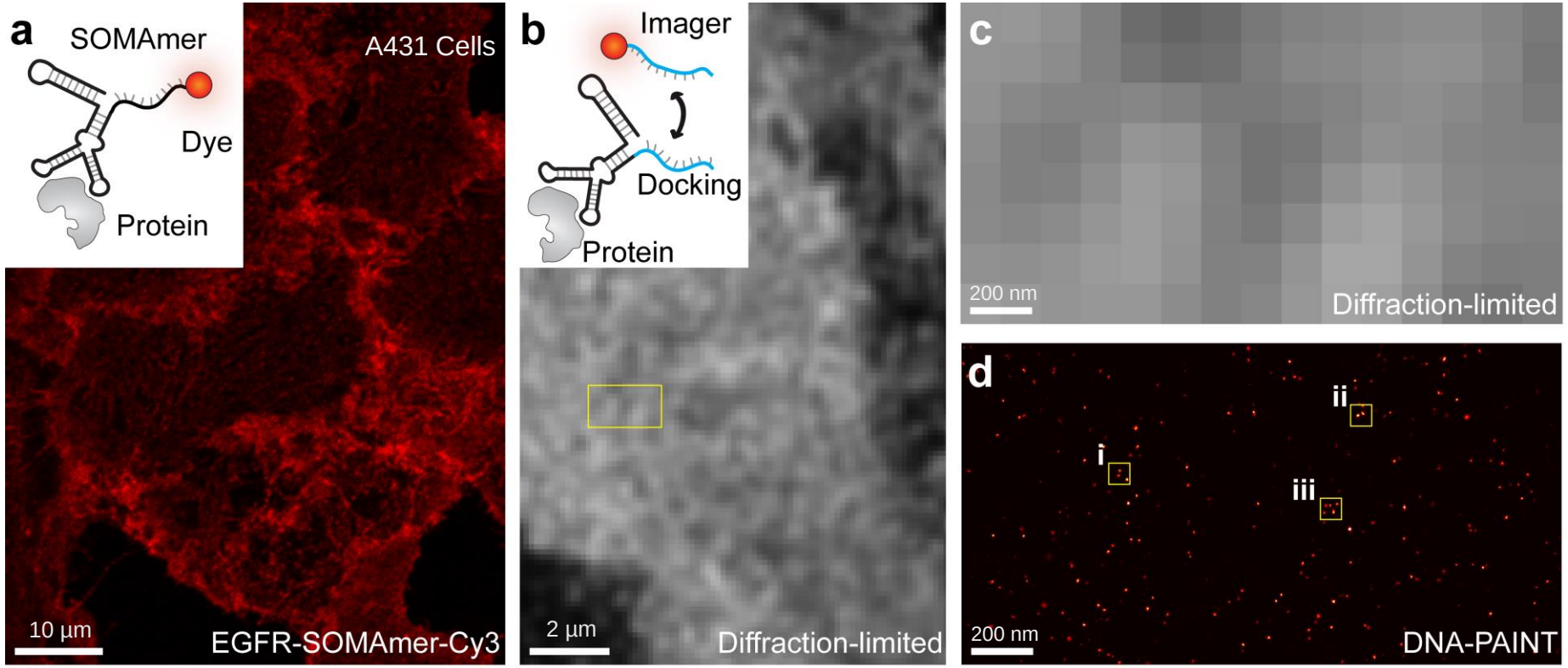


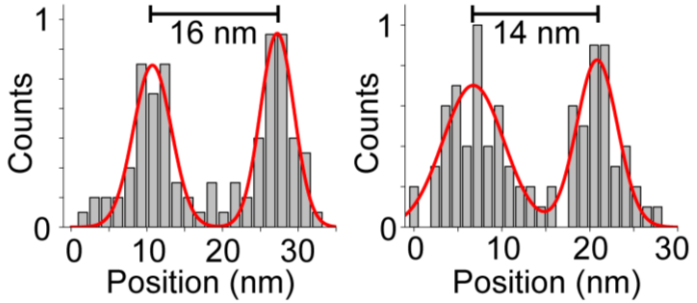
Thrombospondin 2, histochemistry



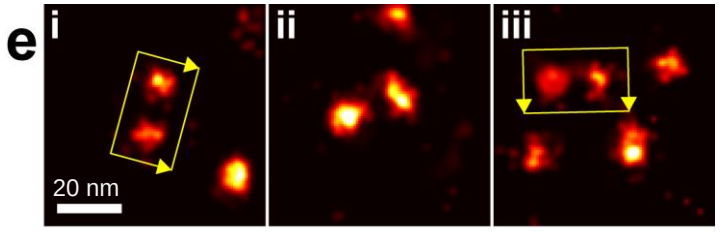
Mehan *et al*. (2012) PLoS ONE **7**, e35157

Super-resolution (sub-10 nm) imaging





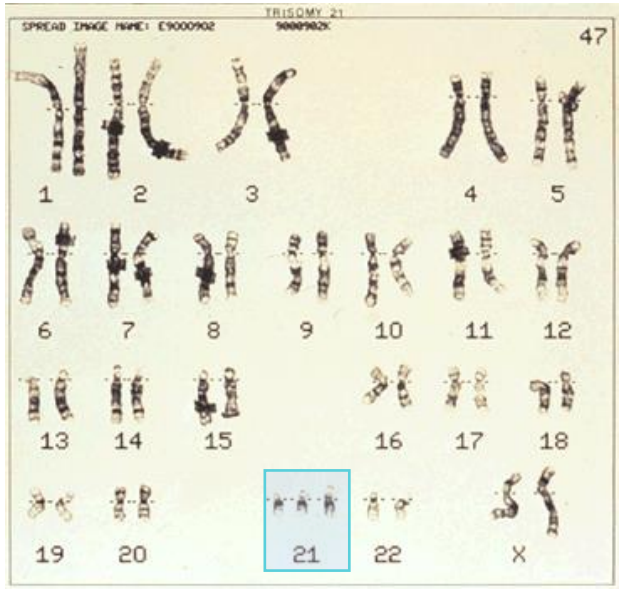
Single-molecule localization

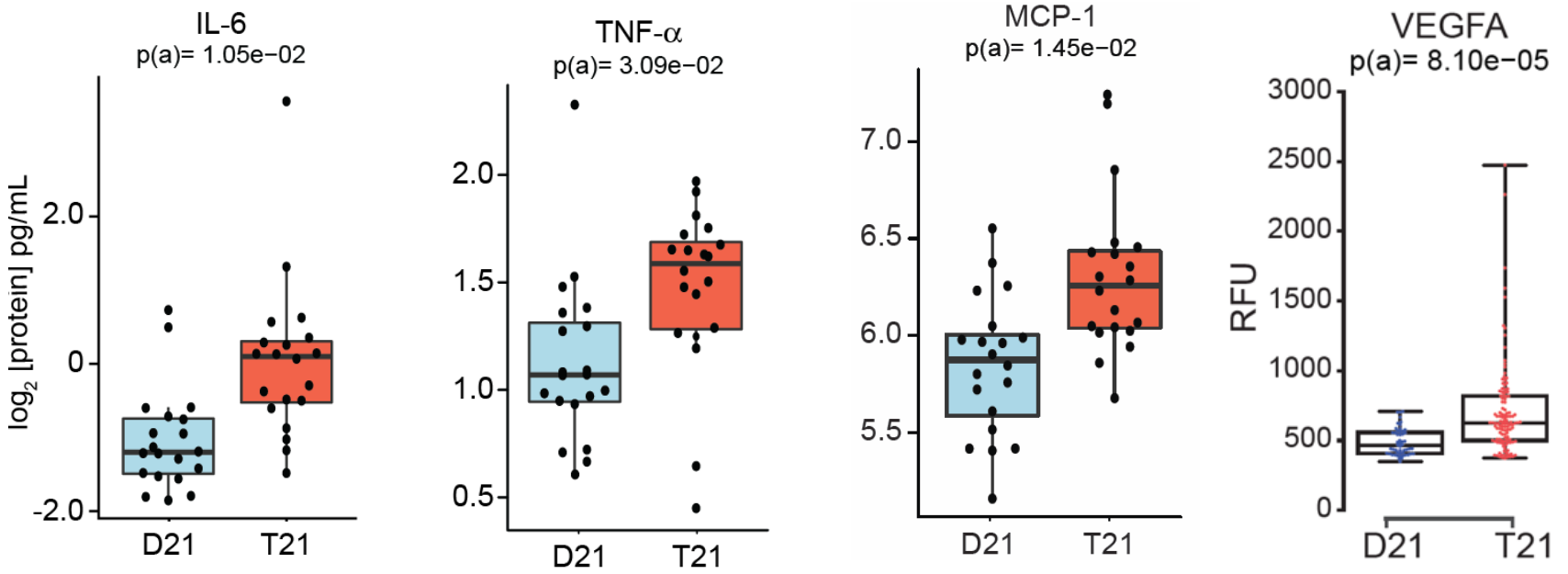


Strauss *et al*. (2018) *Nat Methods* **15**, 685

Down syndrome is associated with chronic inflammation

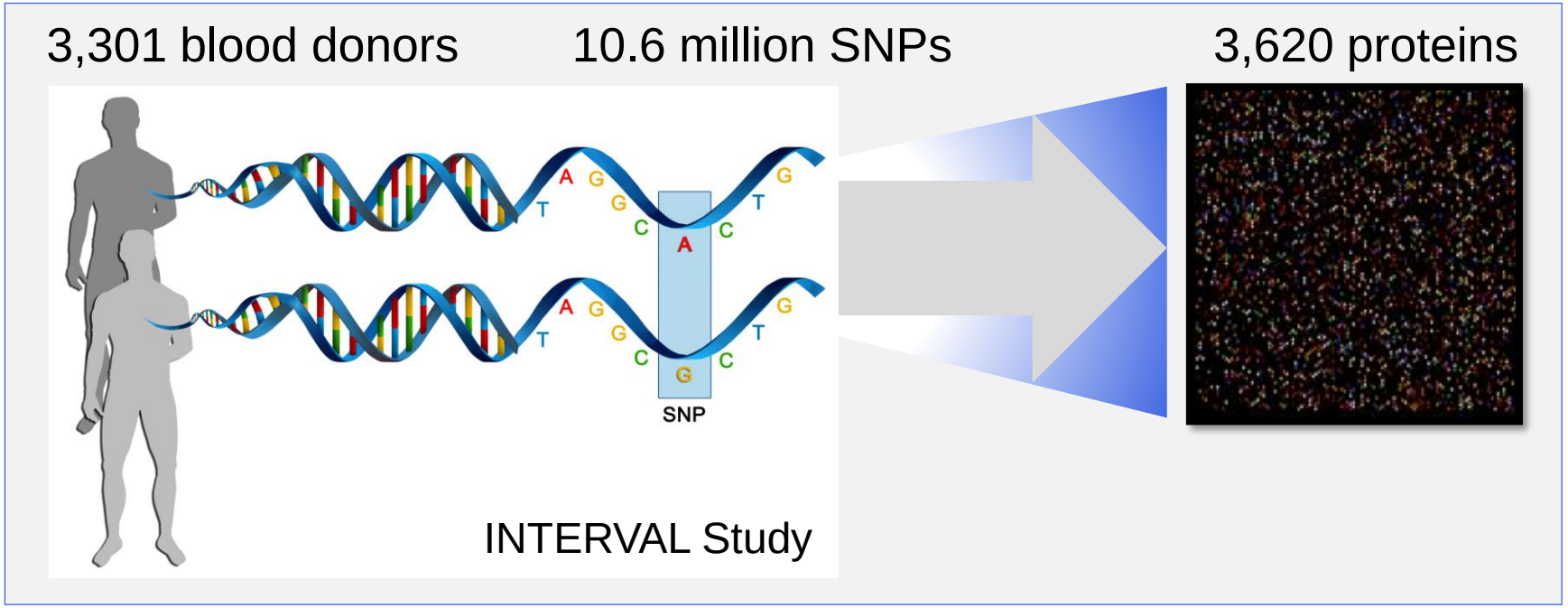
- Trisomy at chromosome 21 (T21)
  - 60% of the 300 proteins different between T21 and D21 were downregulated
  - Of 50 proteins measured on chromosome 21, only 9 are changed (all upregulated)
  - Chronic pro-inflammatory state, disregulated angiogenesis, complement consumption





Sullivan *et al*. (2018) *Sci. Reports* **7**, 14818

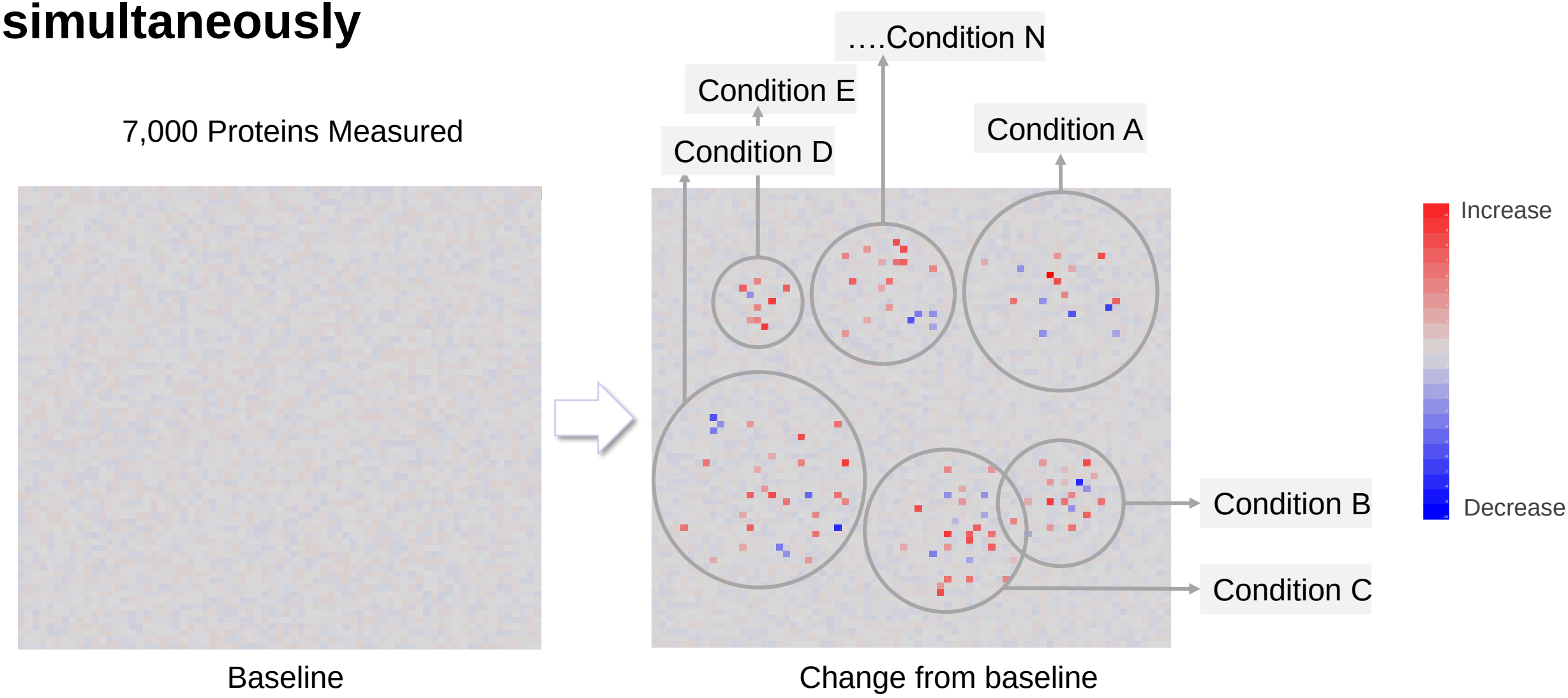
Linking natural variation in genomes with protein expression



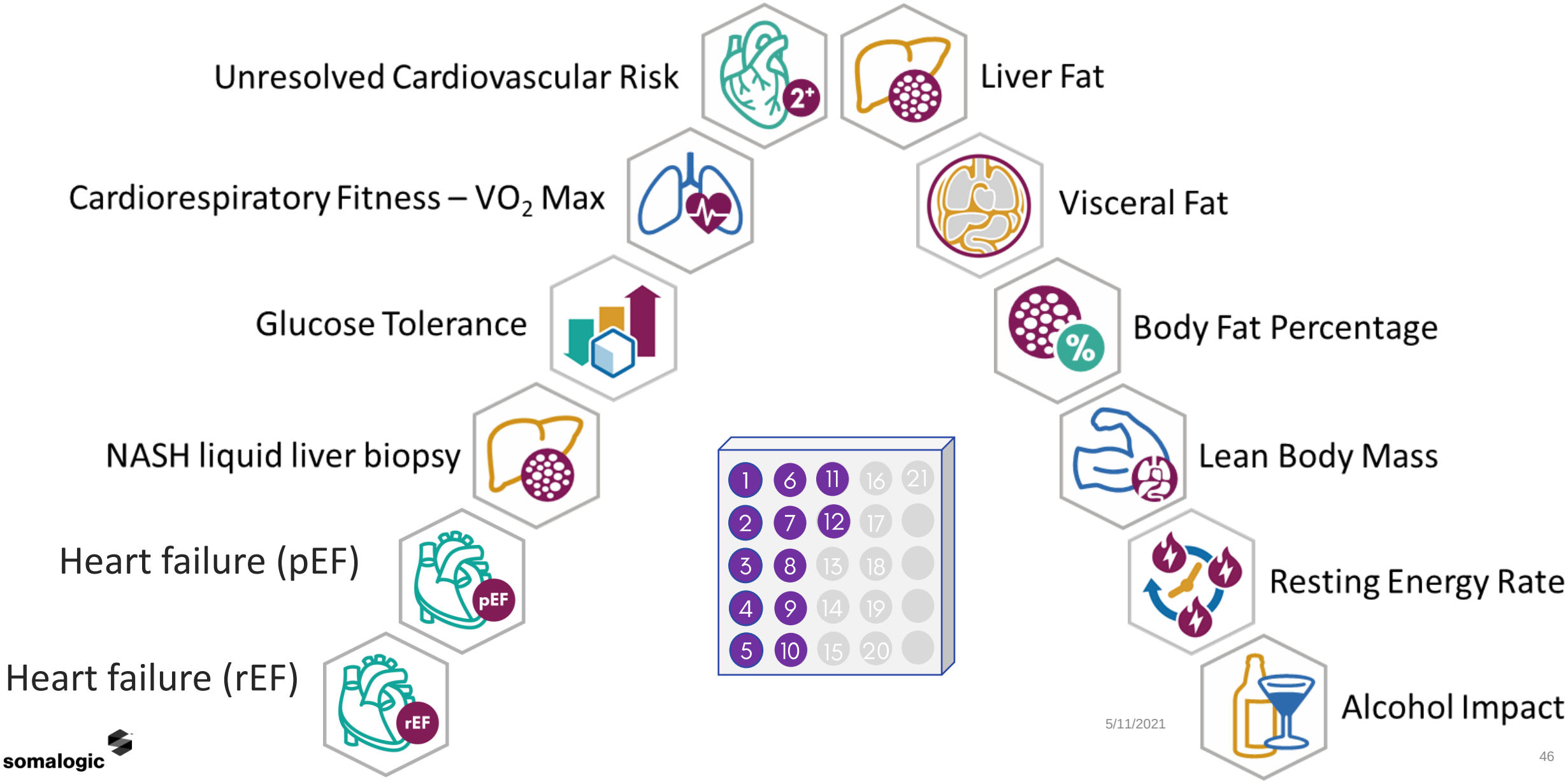
- 1,927 associations between proteins and SNPs
- Both *cis* and *trans* associations
  - 1,104 *trans* associations, role of protein networks
  - 89% of associations previously unreported

Sun *et al*. (2018) *Nature* **558**, 73

Many conditions (SomaSignal™ Tests) can be analyzed simultaneously

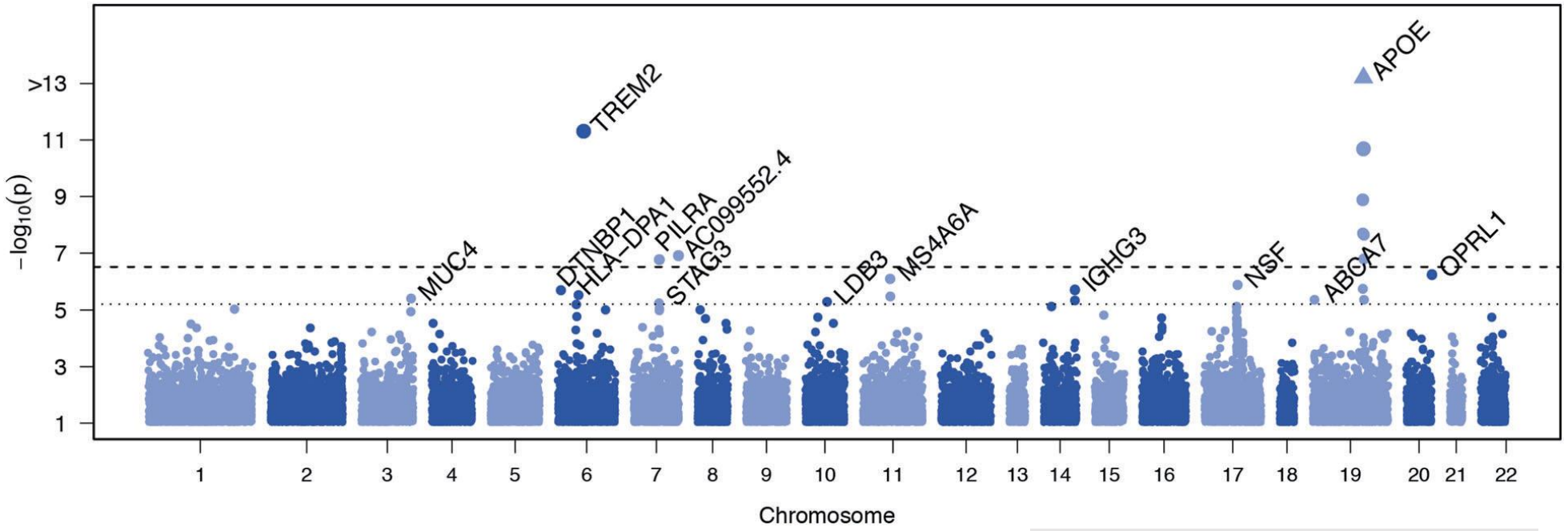


A single sample, tests for many things



Many diseases have polygenic inheritance

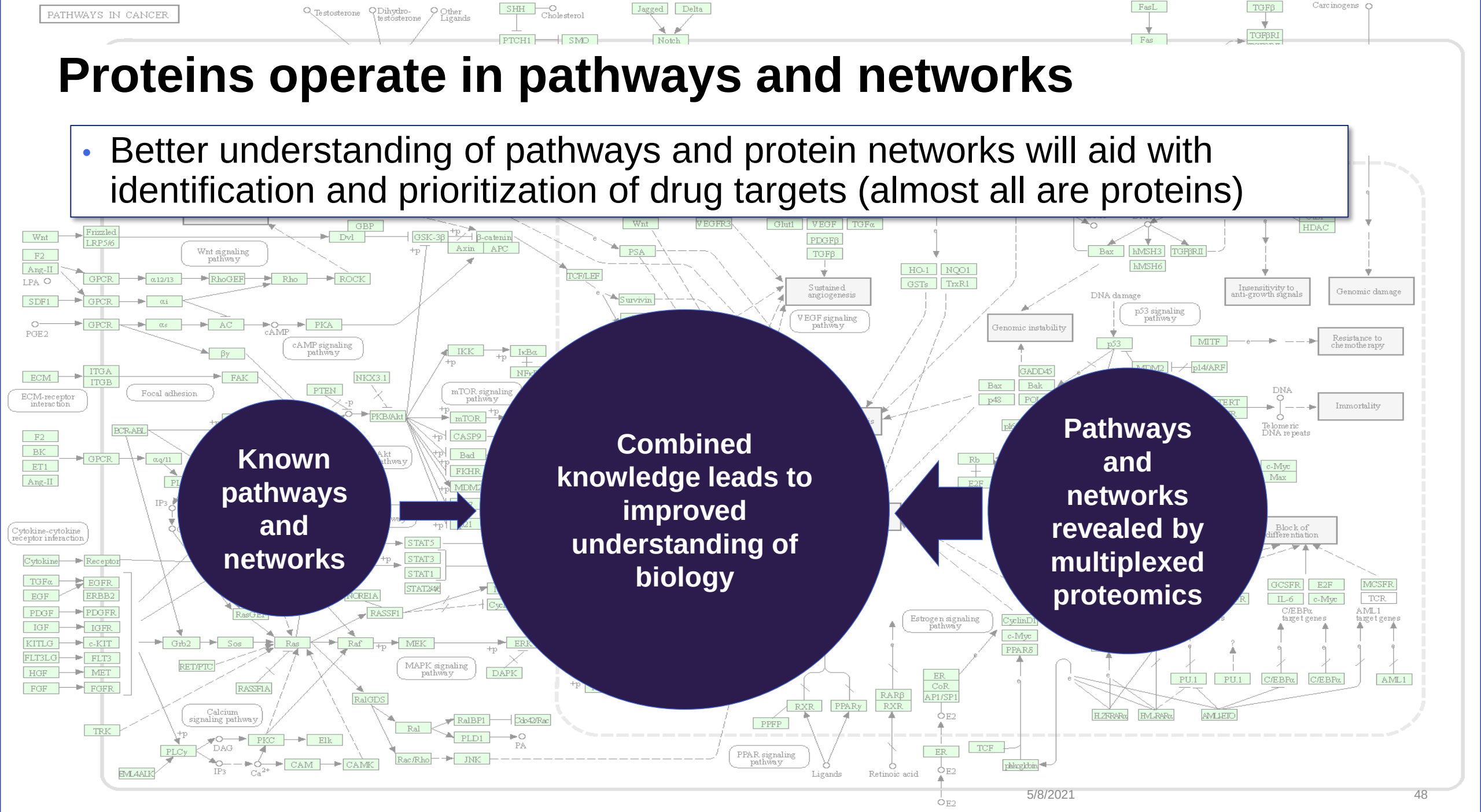
- Example: Alzheimer's Disease is highly polygenic (>30 genes)
  - But highly heritable (h^2 = 0.58 – 0.79)
- Other examples: coronary artery disease, type 2 diabetes, inflammatory bowel disease, autism
- Pinpointing the right targets for therapeutic intervention can be a hard problem
- Concurrent measurements of multiple proteins for fuller understanding of polygenic conditions



Bis *et al*. (2018) *Mol. Psychol.* **25**, 1859

Proteins operate in pathways and networks

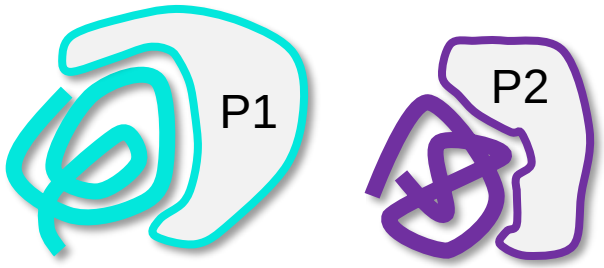
- Better understanding of pathways and protein networks will aid with identification and prioritization of drug targets (almost all are proteins)



A common theme: make and test all, select the best

- Select from all possible variations, rather than (only) relying on predictions from limited knowledge

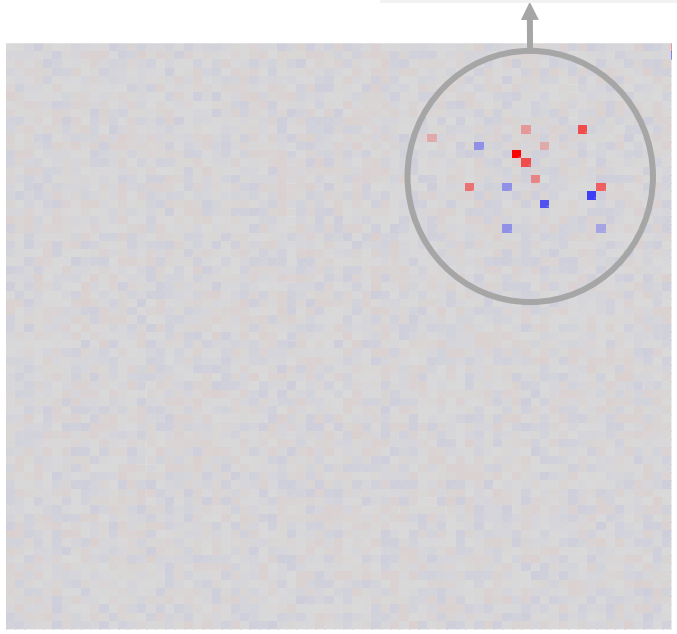
SELEX and SOMAmer reagents



Make all sequences, select those that bind to desired targets

SomaScan assay

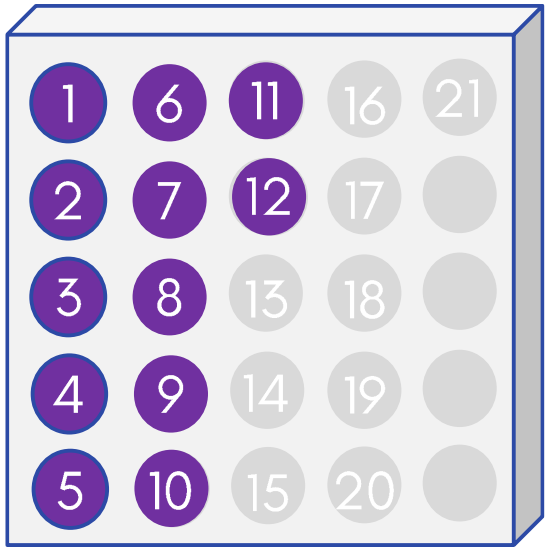
SomaSignal test



Change from baseline

Measure all proteins, select best markers

SomaSignal tests to best health management



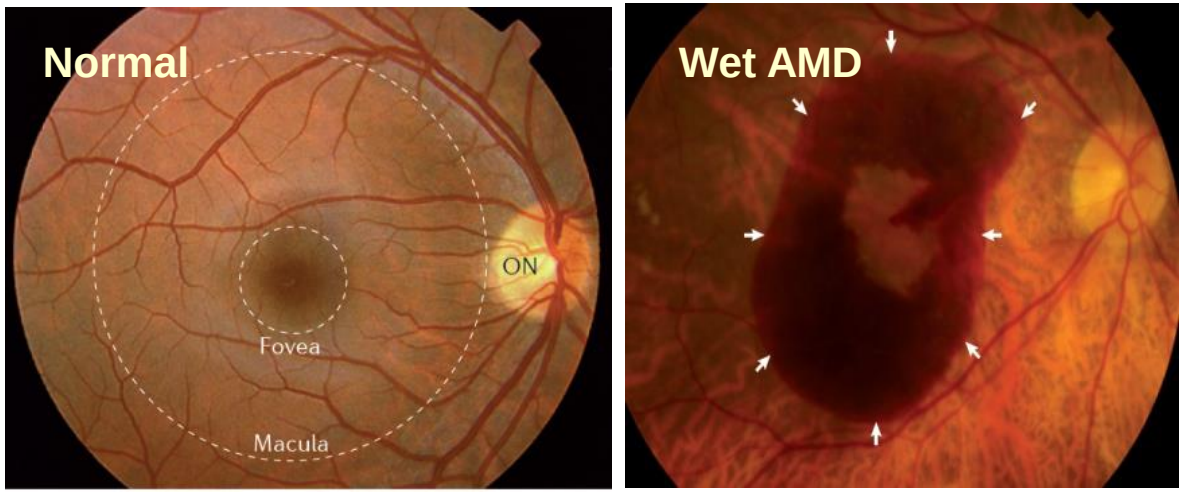
Make many tests, select those with biggest impact on health

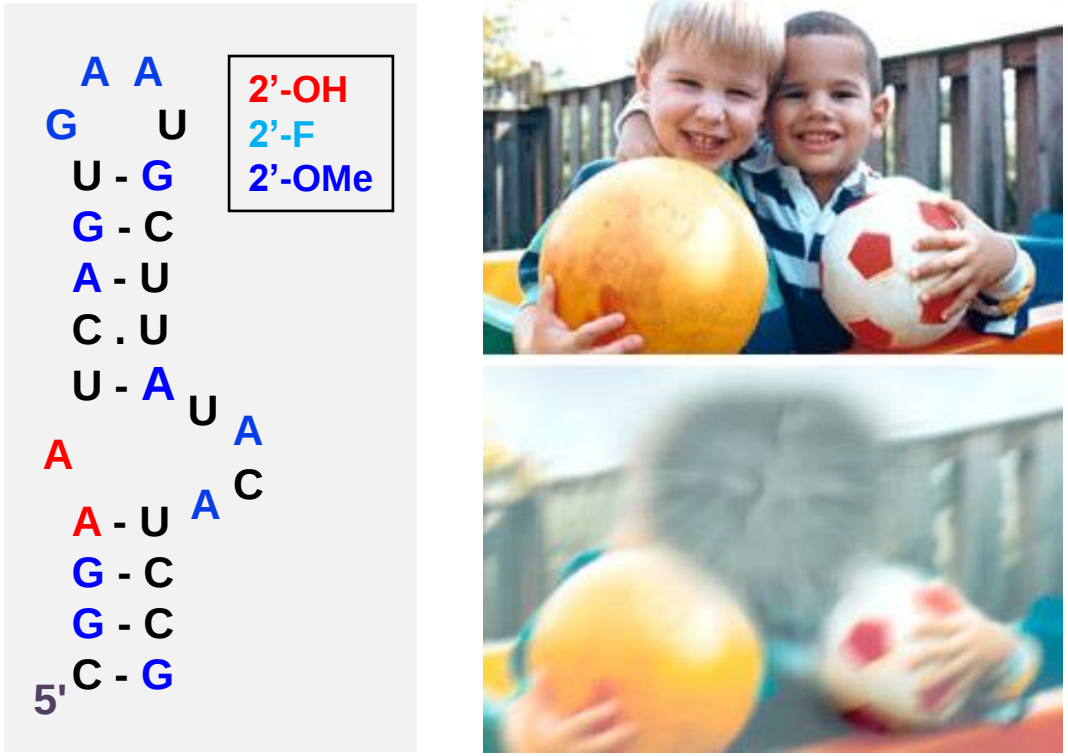
Aptamers can be drugs

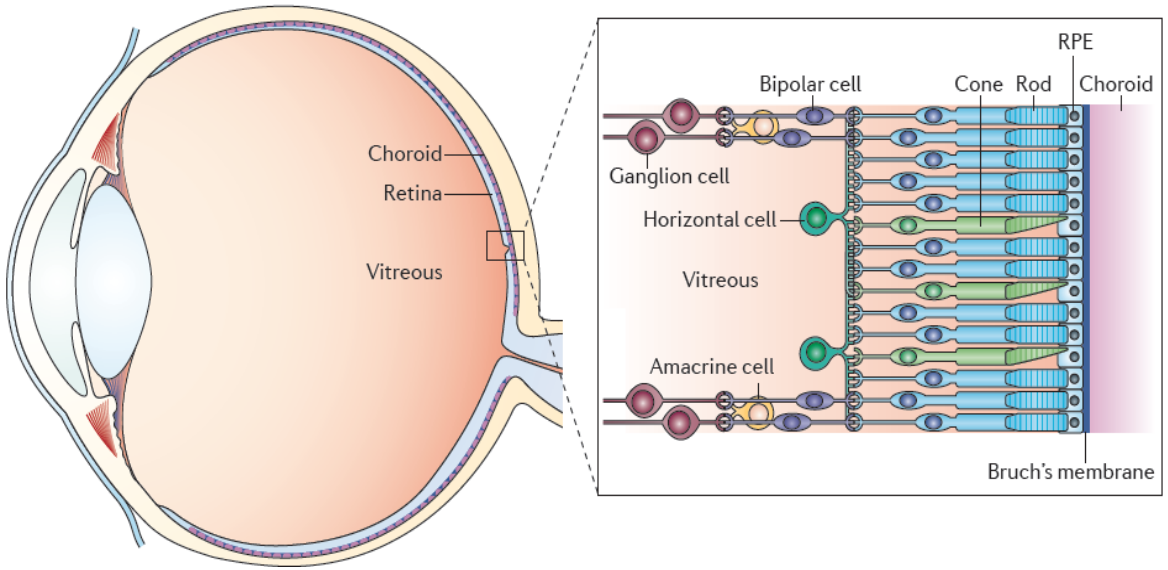
MACUGEN®

(pegaptanib sodium injection)

MACUGEN is indicated for the treatment of neovascular (wet) age-related macular degeneration.



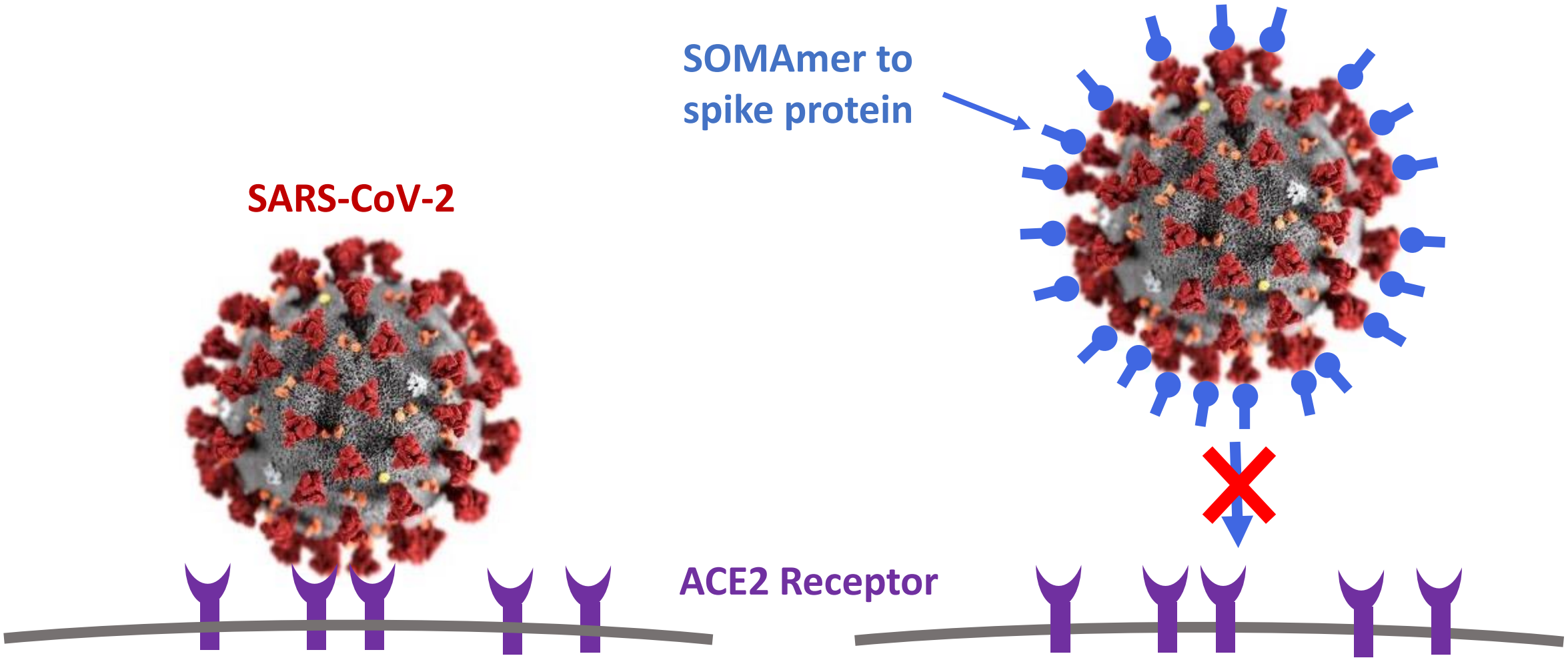




somalogic

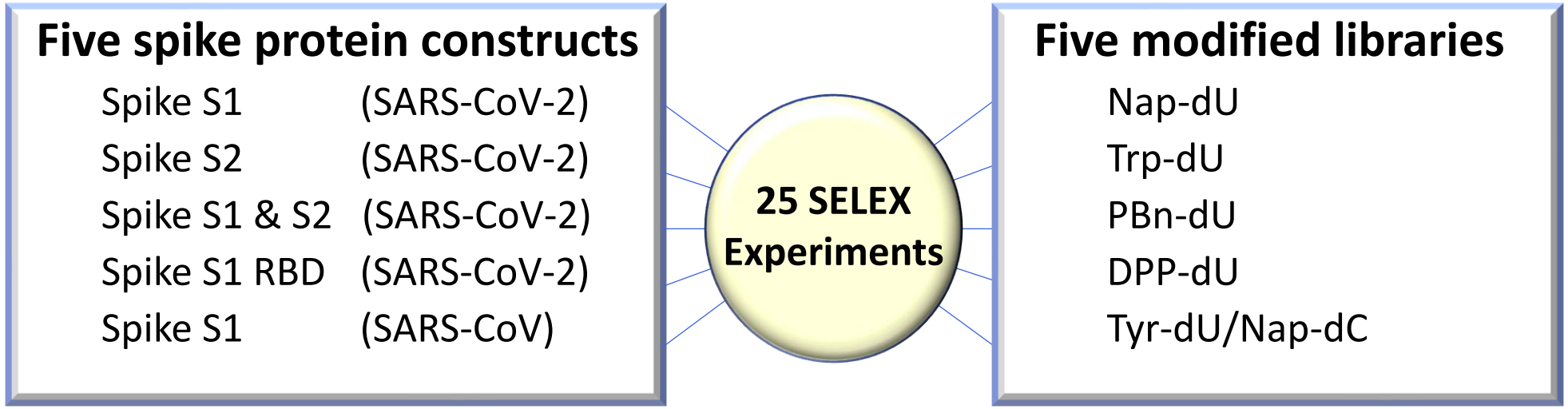
Ruckman *et al*. (1998) *J. Biol. Chem*. **273**, 20556
Drolet *et al*. (2016) *Nucl. Acid. Therap*. **26**, 127

Prevent virus from binding to ACE2 receptor



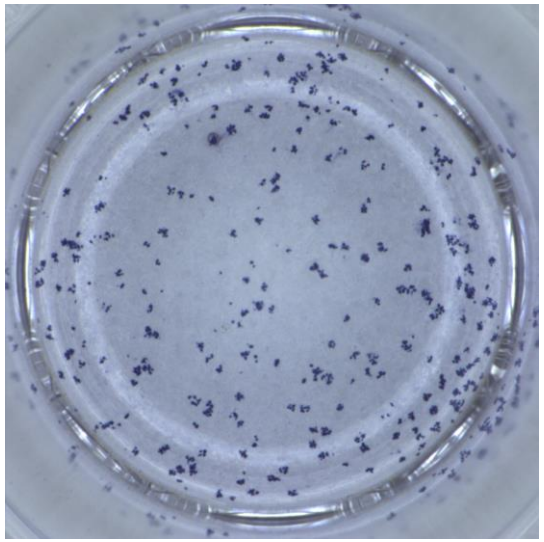
We have identified more than 200 diverse aptamers to SARS-CoV-2 spike protein

- Five spike protein constructs, each done with five modified libraries
- We did twenty-five SELEX experiments and characterized individual synthetic SOMAmers for binding properties
- Goal is broad epitope coverage

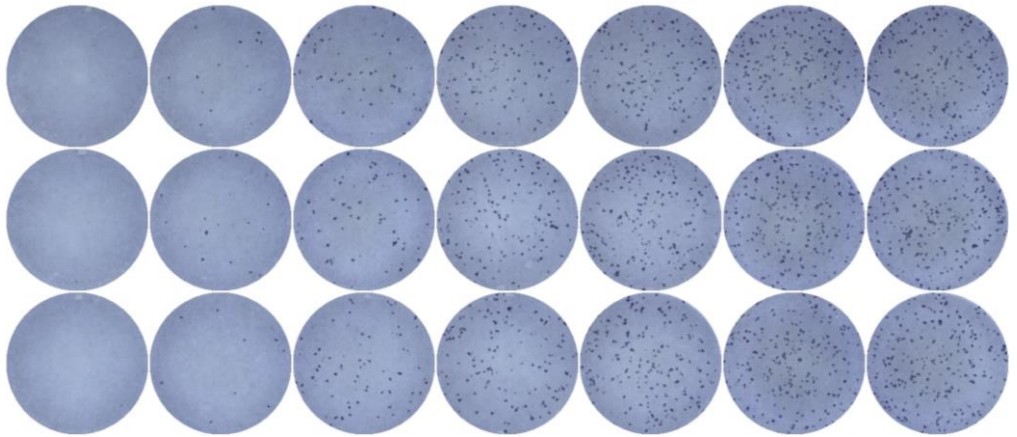


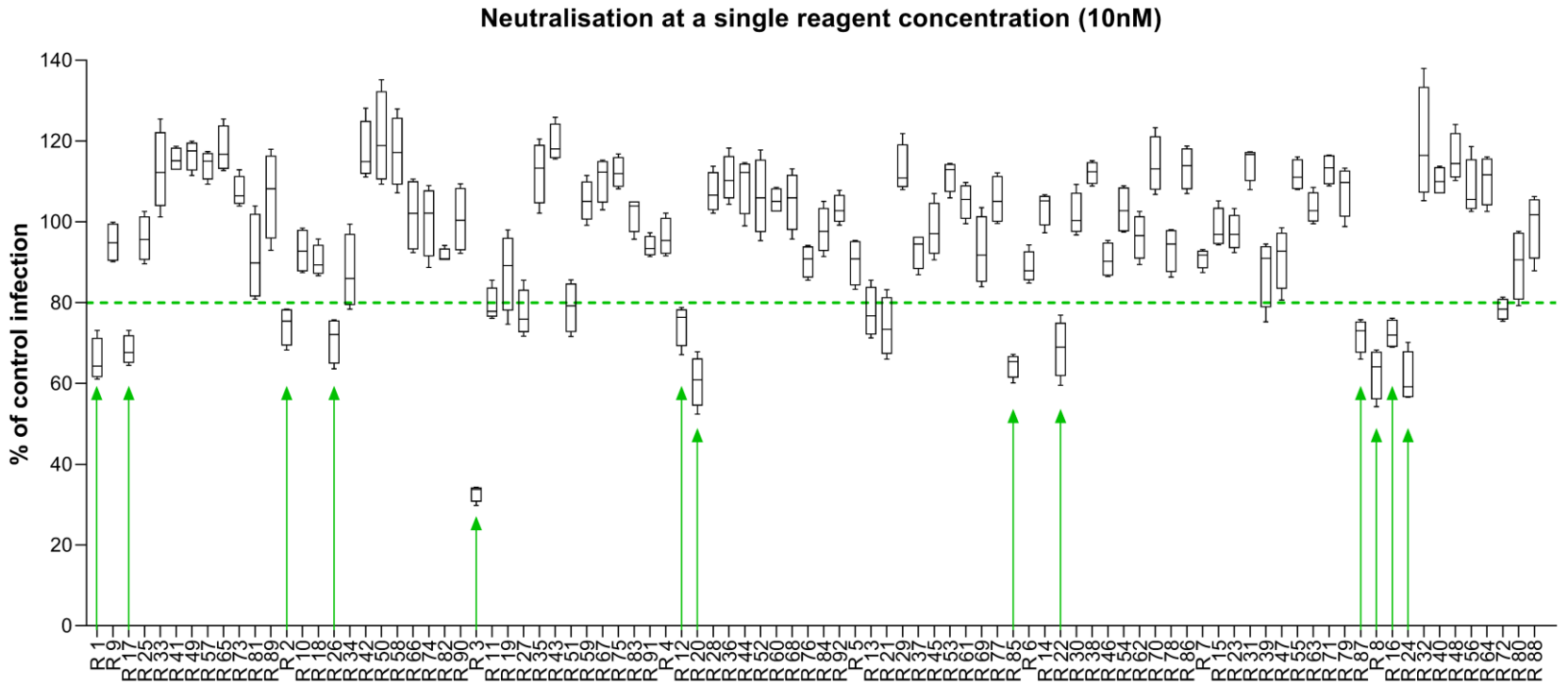
Some spike SOMAmers inhibit the virus in vitro

- Vero cells exposed to virus in the presence of SOMAmer (92 tested)
- Viscous overlay prevents distant spread of progeny virus, producing a “focus” of infected cells, approximating an individual infectious event
- With William James, Oxford



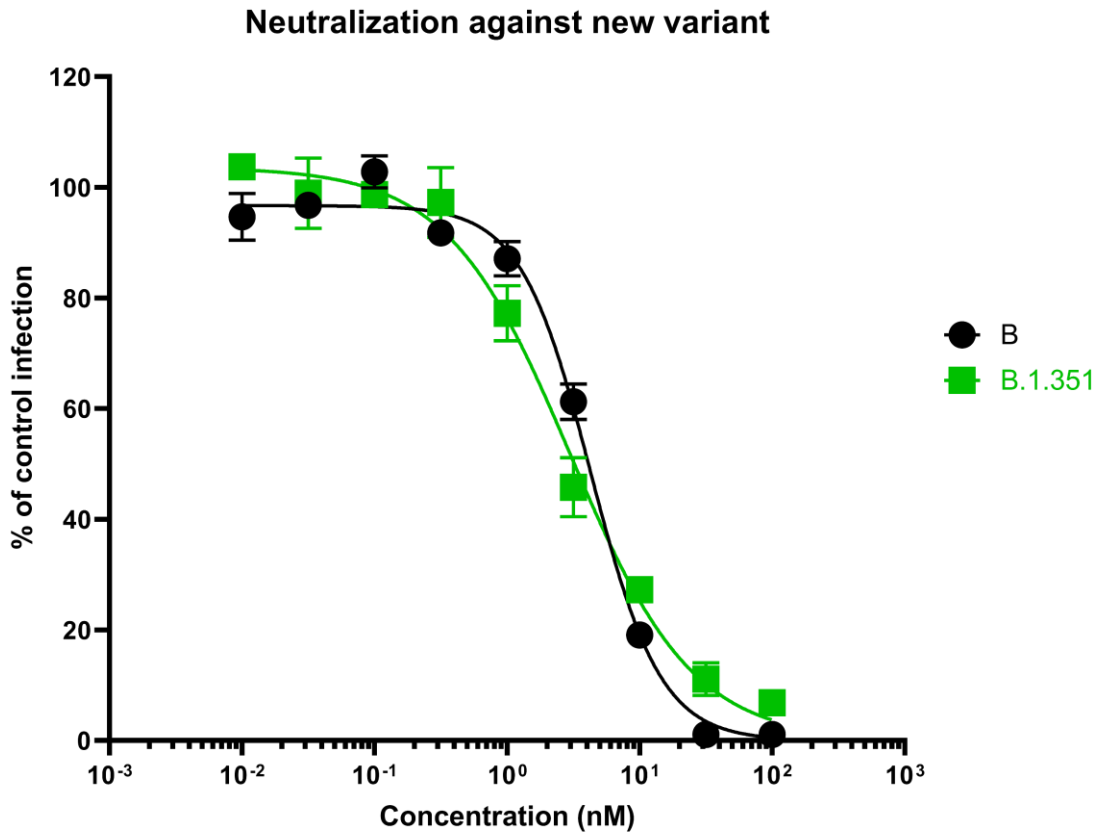
SOMAmer concentration





SOMAmer reagents show antiviral activity, including against new variant

- The minimally truncated (29-mer) form of the most potent SOMAmer 26874-29_20 showed no loss of antiviral inhibitory potency against the South African B.1.351 variant
- This variant is known to escape neutralization by many RBD-targeted monoclonal antibodies



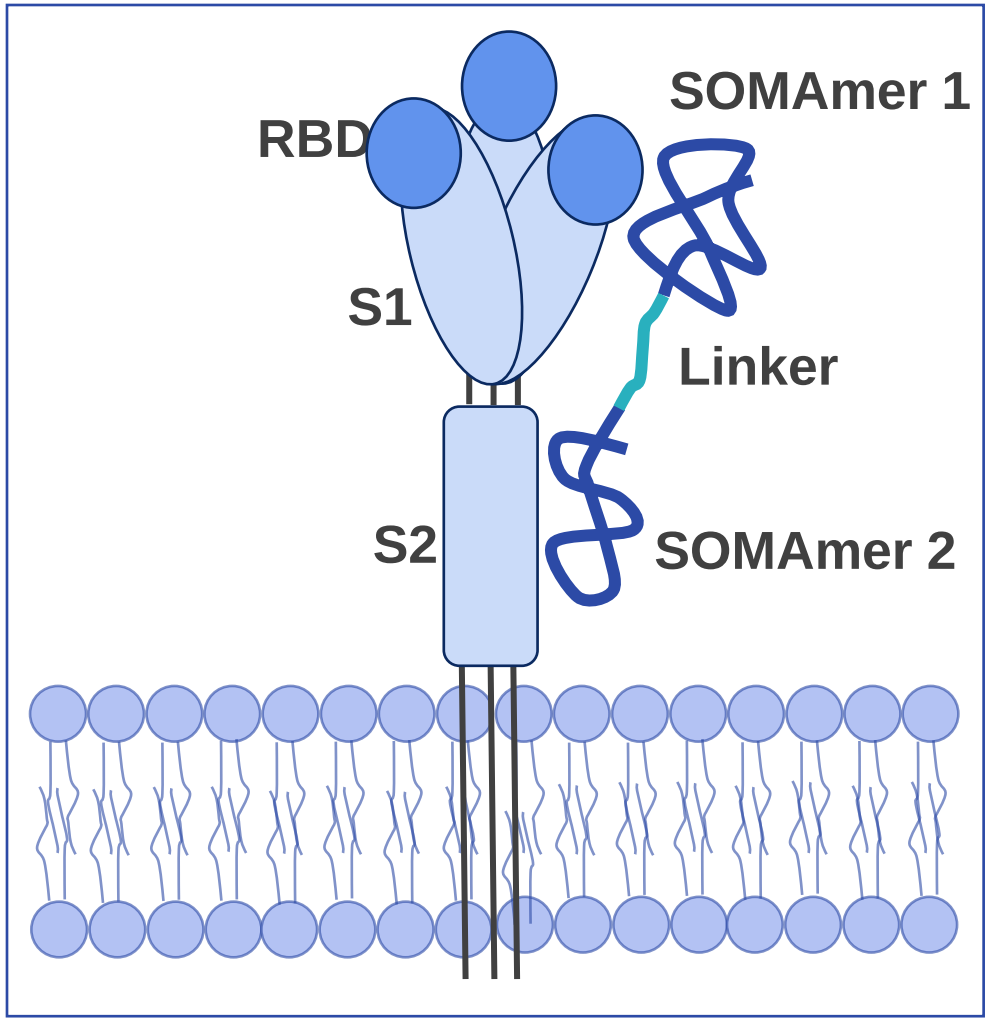
Lead reagents bind spike variants

		Reagent					
		26874-29_20	26860-75_18	26876-3_18	26860-16_18	26876-13_18	Seq ID
		RBD	RBD	S2	RBD	S2	Target
Variant	Source	29-mer	44-mer	28-mer	35-mer	28-mer	Length
wt S1/S2 monomer	Sino Biologicals	0.4 nM	0.4 nM	0.6 nM	0.5 nM	0.6 nM	K_d (wt)
S1/S2 "wt" stable trimer (mutations in S2 to stabilize trimer) F817P, A892P, A899P, A942P, K986P, V987P, R683A and R685A	Acro Biosystems	✓	✓	✓	✓	NB	
UK B1.1.7 monomer HV69-70 del, Y144 del, N501Y, A570D, D614G, P681H, T716I, S982A, D1118H	Sino Biologicals	✓	✓	✓	✓	✓	
P.1 Brazil stable trimer L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, D614G, H655Y, T1027I, V1176F	Acro Biosystems	✓	✓	✓	✓	NB	
S. African B.1.351 stable trimer L18F, D80A, D215G, LAL242-244del, R246I, K417N, E484K, N501Y, D614G, A701V	Acro Biosystems	✓	✓	✓	✓	NB	

✓ Indicates affinity to variant is within 3-fold of wild type

NB No binding detected up to 50 nM protein

SOMAmer multimers for enhanced binding affinity



- Linking two SOMAmers has several potential advantages:
  - Dramatic increase in potency
    - Theoretical best case: $K_d^{Dimer} = K_{d1} \times K_{d2}$
    - Even partial success results in super-potent ligands
    - We have done this with other targets
    - Goal is essentially irreversible binding
  - Reduced chance of mutations leading to loss of binding affinity
    - Mutations at both sites are needed
    - Probability of that event is the product of individual probabilities

Acknowledgments

SELEX and Structures
Dan Schneider
Sheela Waugh
Shashi Gupta
Dan Drolet
Bharat Gawande
Chi Zhang
Thale Jarvis
Amy Gelinas

Beryllium (now USB)
Doug Davies
Tom Edwards

Yale
Anna Marie Pyle
Xiaoming Ren

Chemistry
John Rohloff
Jeff Carter
Cate Fowler
Brian Ream
Greg Wardle
Matt Otis
Wes Mayfield
Steve Wolk

Bioinformatics
Dom Zichi
Alan Williams

Medical Group
Rachel Osteroff
Steve Williams

Otsuka Pharmaceutical Co.
Miyuki Futakuchi
Akihiko Hisaminato
Tomoki Suzuki
Haretsugu Hishigaki
Masao Hirota
Yuichi Ishikawa

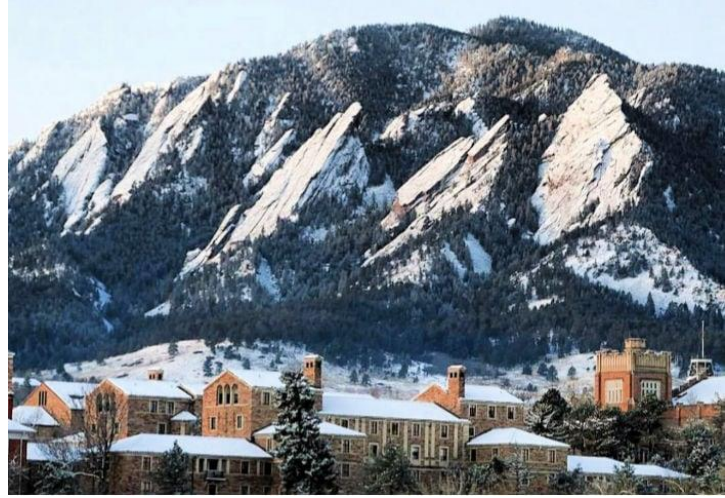
Linda Crnic Institute for Down Syndrome
Joaquin Espinosa
Tom Blumenthal

Discovery Sciences
Sheri Wilcox
Eric Zimmerman

University of Cambridge
John Danesh
James Peters
Adam Butterworth
Ben Sun

Max Planck Institute, Munich
Ralf Jungmann
Sebastian Strauss
Philipp Nickels

University of Oxford
Javier Gilbert-Jaramillo
Michael Knight
Maeva Dupont
William James



Direction and vision
Larry Gold