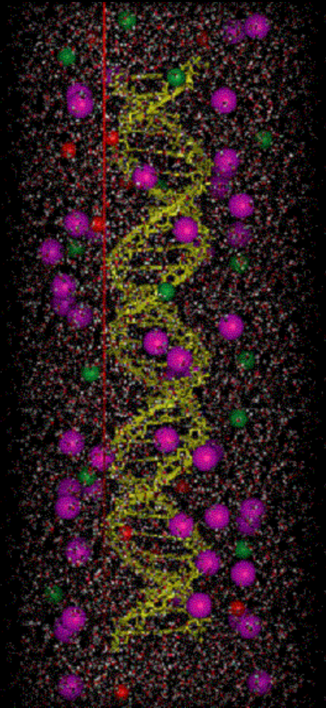
DNA

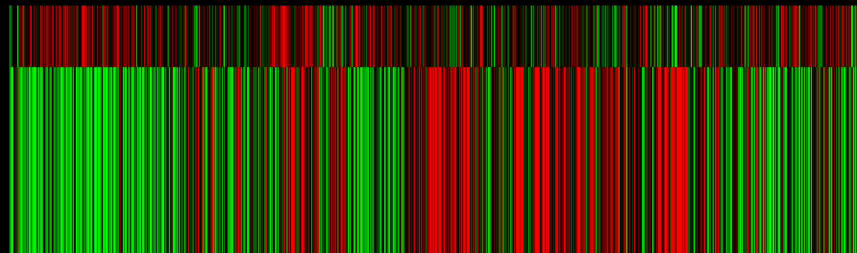
predicts certain possible future health issues



Entire genome sequence
~$50,000 today per patient
~$1k in 5 yrs or so
(next-generation, non-PRC technologies)

mRNA

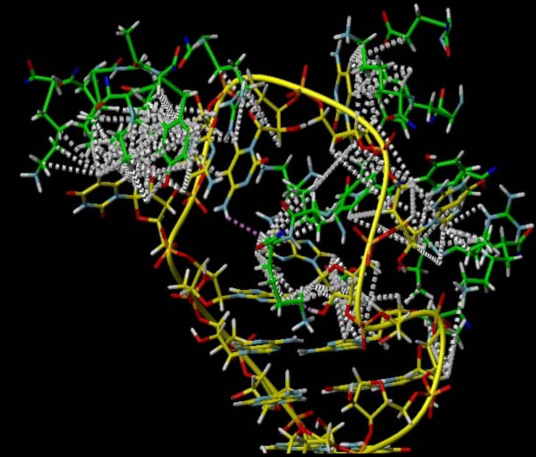
Limited info concerning evolving health;
advanced measurement technologies



100 element mRNA panels
(Genomic Health breast cancer)
~$2k for patient
Actual cost ~20 cents / mRNA (PCR based)
qPCR (quantitative) costs more

proteins

Potentially comprehensive information concerning evolving health;



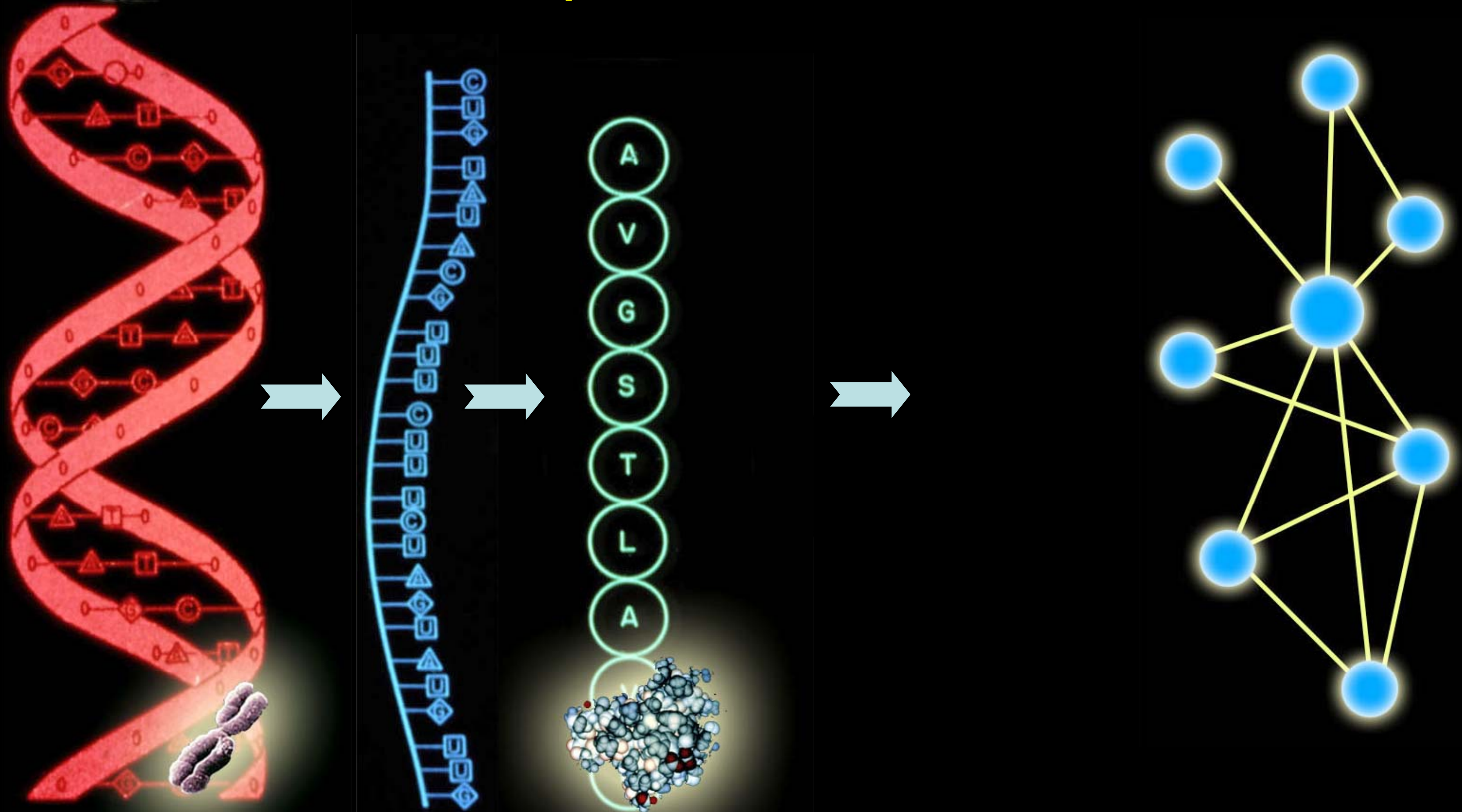
Pauciparameter
PSA, CA125, troponins
$50 per protein for patient
Actual cost ~$25
(antibody based)

Strategy # III:

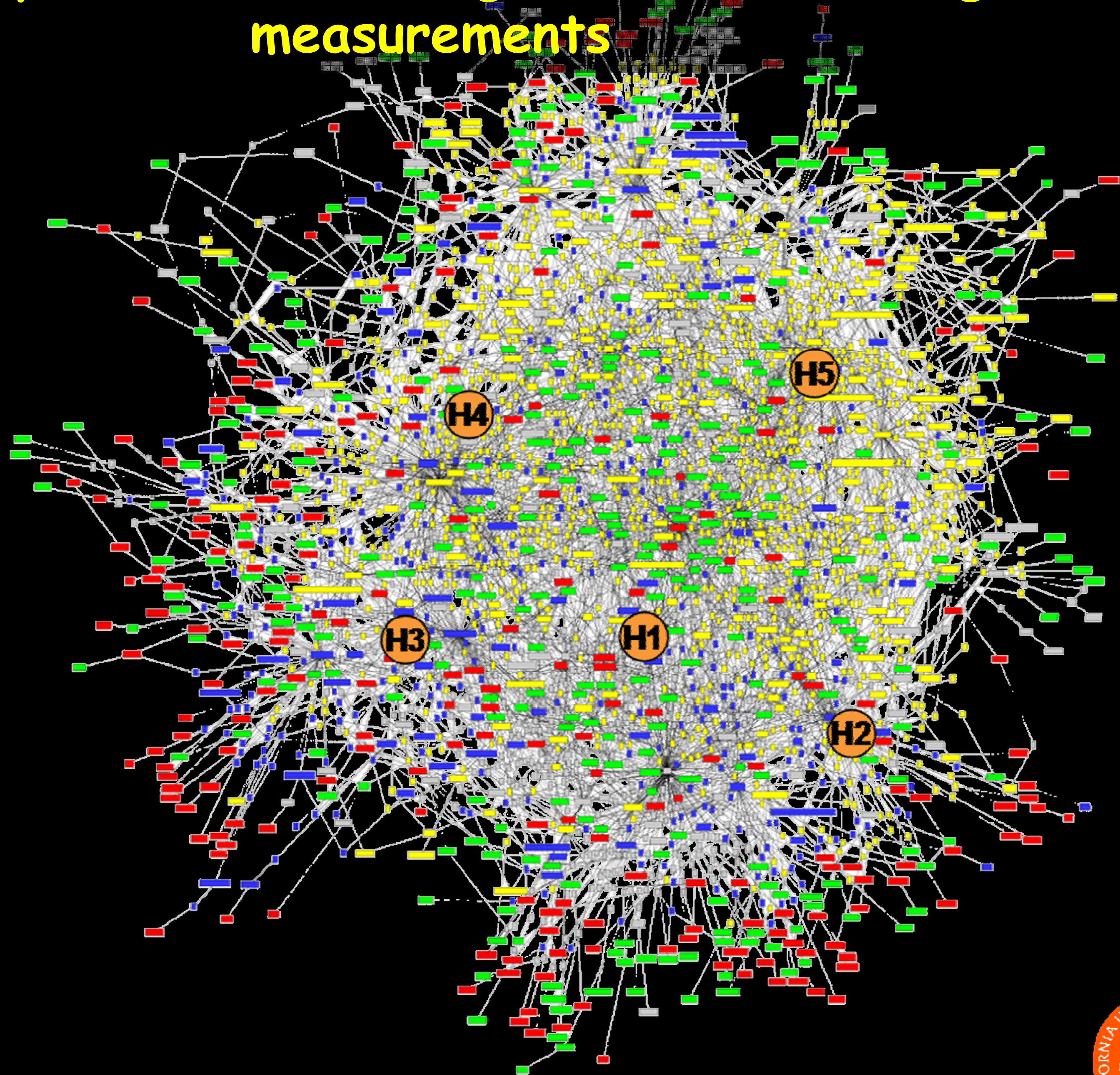
Measure Biological Function

(this is the hardest)

DNA mRNA protein Network



Network Hypotheses from large-scale mRNA & genomic measurements



Cytokines
IFNg, IL1b, IL6, Osm
CNTF OSMR IFI35 IFI30 IFIT1 PDIA3 TAPBP IL6ST IL6RA OSM IL1B PRTN3 IL6 IFNG KCNK6 IL13RA1 IL13
3b
TRADD TRAF3 TRAF2 FAS RIPK1 CRADD TNFRSF12A TNFRSF1A TNFAIP8L2 DAP TNF Tnfa
Nfkb activation
Chemokines
Tgfb SMAD3 DAXX CLU TGFBR1 TGFBR2 TGFB1 DAB2 SCARA3 MSR2
CD antigens
CD86 CD84 ANXA3 TLR2 CD14 LGALS1 CD28 SH2D1A PAP LGALS3BP CD8A SPP1 LGALS3 CD44 SELE 3d IL2RG IL4 CD53 IL2 CD9 TSPAN4 CCL6 PRG1 CCL3 CCL9
Glial Cells
Ccl6,9
Prp^Sc trapping (Complement)
C1QA C1QR1 C1QC C1QB C1S C4B C1R C3 C2
2b
MAPT APBB1 SNCA PRNP APP LAMA1
Amyloid
2a
ECM
RBPSUH DLL1 PSEN1 EDG1 MTAP2 HSPA5 PCAF NOTCH3 SKIIP ANXA4 GP2 EDG3 CAV1 PLEC1 MARK4
MAPK8 ATF2 PIK3R1 CSF3R HRAS1 MAPK8IP1 IRS1 PRKCM
Apoptosis
Fyn Signaling
3e 3c KCNIP3
BIRC3 CASP12 CSF2RB1 PTPN6 SH3BP2 FYN CSF1R GRAP2 PTPRC FCER1G LYN
BCL2 CASP3 CASP8 CREM CFLAR LCP2 INPP5D PLCG2 BLNK
BCL2A1A APAF1 BAX PYCARD SH3GLB1 BCL3 COPS2 IRF8 CASP1
CASP14 RXRA PRKDC GADD45A GADD45G CAPN2 CAPN1 CCND1 GADD45GIP1 CAPNS1 PCNA GADD45B
1c CSTA CST3 CSTB LAMP2 CTSH CTSB CTSL CST7 CTSS CTSC
Lysosomal Proteases
Neuron/ Astrocyte (?)
Ca2+ (ER)
TGM2 S100A10 S100A6 GAPDH ANXA2 S100A4 S100A11 S100B S100A1 S100A13
Endocytosis
S100s (Ca2+ binding protein)
1b 1a
STAT1 STAT3 NMI DNM1 HSPA8 TEC VAV1 DOCK2 DNM2
ADAMTS1 B2M A2M APOD COL2A1 SERPINA3N APOA2 LRP1 ELA1 APOE SPARC LRP2 PLAUR PLAT PLG SERPING1 PLAU PLTP KNG1 PAPPA THBS2 CLEC3B MMP9 TIMP1 MMP7 MMP1A
PrP^Sc transport (Lipid Metabolism)
3f
Integrins 1e
COL1A2 VCAM1 ITGA2 ITGB2 ICAM1 ITGB1 ITGAX CFH ITGAM ITGA6 IGTP NOV
Cytoskeleton reorganization
WASPIP MEN1 GC CFL1 WASF2 GFAP ACTC1 3a CNTN1 VIM ACTA1 ARHGDIB TSC1 DYNLL1 ANXA1 KRT2-8 L1CAM MSN VIL2 CFTR
Astrocyte Process
GABA receptors
GABRG2 SHANK2 GABARAP GABRB2 GABRA2 GABRB3 GABRG3 GABRQ GABRA1 GABRA3 GABRA4 GABRA6
Glutamate
GRIK2 GRIA1 GRIA4 DLGH4 CACNG2 PRKCC DLGH1 SDCBP SPNA2
SRF HMGA1 SLC17A6 SYT1 VAMP2 TTN CAPN3 STX1A SNAP23 CAMK4 CALM1 CAMK1D CAMKK2 CAMK1
Glutamate receptors
Ca2+
Synaptic Vesicular Transport
CaM kinase
Synapse

Glycosaminoglycan (GAG) Degradation
1d
PRNP CAV1 PLD2 GM2A HEXA HEXB NAGLU GUSB
Keratan Sulfate
Dermatan Sulfate
Heparan Sulfate
Amino-sugars

Dynamic Complement Activation
2a
Cell Lysis
C5B C3AR1 Phagocyte Recruitment
C1QA C1QB C1QC C1R C1S C4B C2 C3 PLG PLAT PLAU LRP1 SERPING1 PRNP A2M SERPINA3N

MHCII Antigen Presentation
2b
Endosome
PRNP IFI30 LGMN CTSB H2-D1 H2-K1 H2-A1
TCR Activation, Cytokine Production
Secretory Granule
CD74 LGMN CTSB MHCII CTSS CTSL SLIP CLIP MHCII

Synaptic Vesicle Pathway (SNARE)
3f
Vesicle
RAB14 LIN7C VAMP1 CPLX1 RPH3A VAMP8 SYT2 NSF SNAP29 CACNB4 CACNA2D2 SCN1A STX6 NLGN1 CACNA2D1
PM

Cytoskeleton reorganization 3a
CD antigens 3d
Gfap Vim Gadd45b Cflar Anti-apoptosis
Apoptosis 3e Fas Bcl2a1a Trp53
Glial cell Activation
Nfkb CD86 CD44 Lgals3 Ela1
PrP^Sc Trapping (Complement) 2a
Integrins 1e Col1a2 Icam1 Vcam1
Msr2/Scara3 (Glial cells)
PrP^Sc
Intracellular Signaling Pathways
Cytokines/ Chemokines
Ccl9 Ccl6 Cxcl10 C4b B2m S100a6 Plat Hmox1 Oxidative Stress Nfe2l2
Usf1/2 IL1b IL13 Osm IL6 Itgam Itgax
PrP^Sc accumulation (Lysosome/Endosome)
3b Cytokines/ Chemokines release Tnfa INFg Irf8 Tgfb1 A2m Casp8 Ets1
Creb1 1c Ctsb/c/h/l/s Fos/Jun Stat1 Stat3 Serpina3n Abca1 Timp1 Apoe 1b Tcfap2a
Nmi
PrP^Sc Accumulation (Lysosome/Endosome)
PrP^Sc Transport (Lipid Metabolism)
Cst7 C3
Sp1 Cebpa/b/d
2b Autocatalytic PrP^Sc replication
Prnp Lyzs Clu

Can we reduce representations like this into a few (<100) proteins worth measuring to achieve a diagnosis?

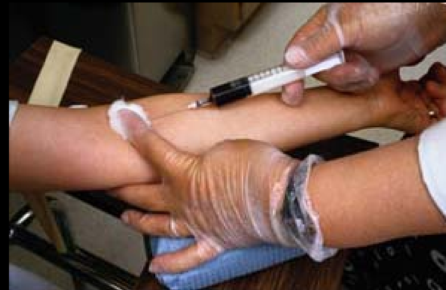
The blood proteome: The richest window into health & disease

~100,000 different proteins
(including post-translational modifications)
Concentrations range from
10^{-3}M to 10^{-17}M

How we use this is evolving into a very high technology, with design automation playing roles in many aspects

Conventional Blood protein measurement

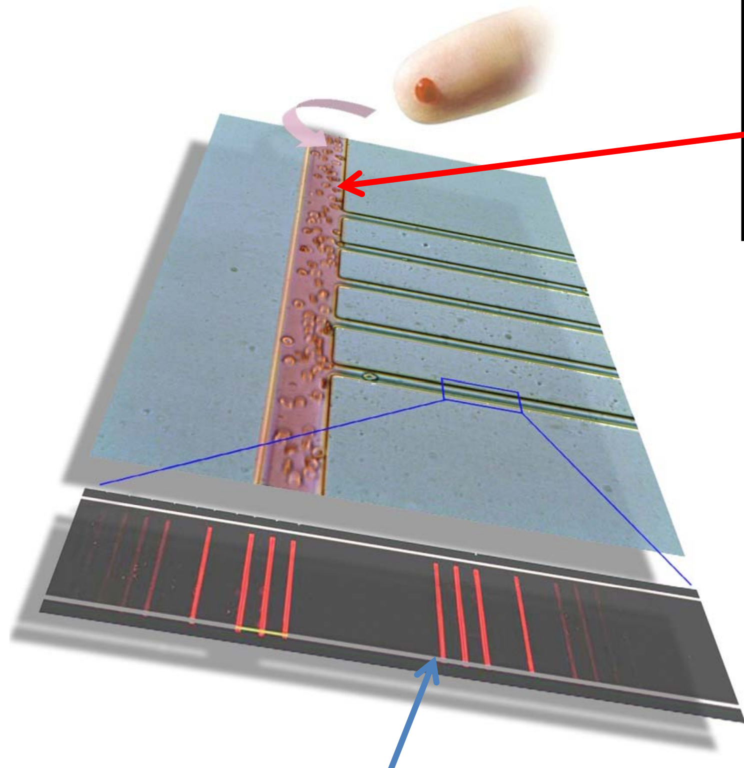
Extract ~5 ml blood



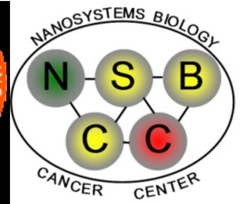
Centrifuge to separate plasma or serum

Measure proteins in 96 well plate

- *Slow (few hours);*
- *human intervention (costly)*
- *not comfortable for patient*
- *Doesn’t scale to lots of proteins*
- *Lacks sensitivity & dynamic range*



Whole-blood



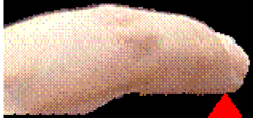
plasma

Antibody-barcodes

Heath group;
Nature Biotech; 2008

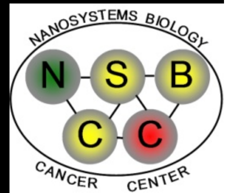
Separating Plasma from whole blood

Dr. Brian Yen & Ophir Vermesh



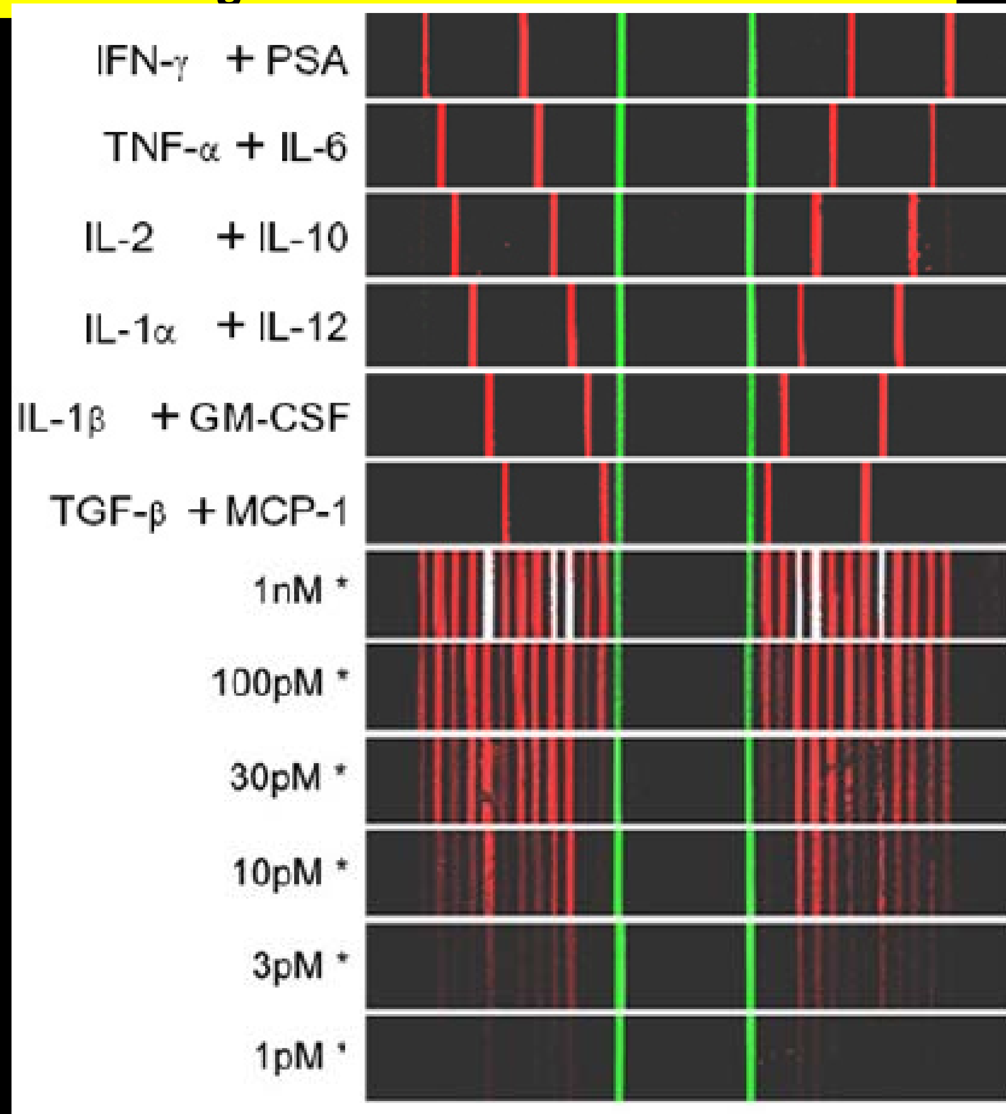
Blood
& tissue
handling

say region

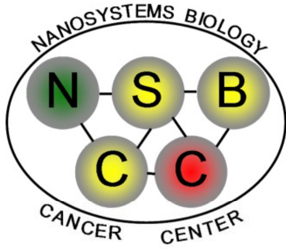


Technology must be simple, robust, quantitative and accurate to 10% on a log scale

Required for commercialization
AND for using devices in clinical trials
AND for using devices to learn new science



Nature Biotech, 2008



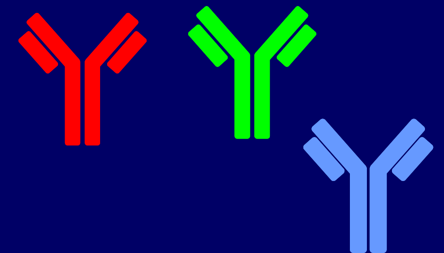
Robotics for chip manufacture

2nd installations (one at UCLA to support clinical trials)

Habib Amad

40 chips per day
6 fingerpricks per chip
20 proteins per fingerprick
$500 total cost
Or 10 cents/protein

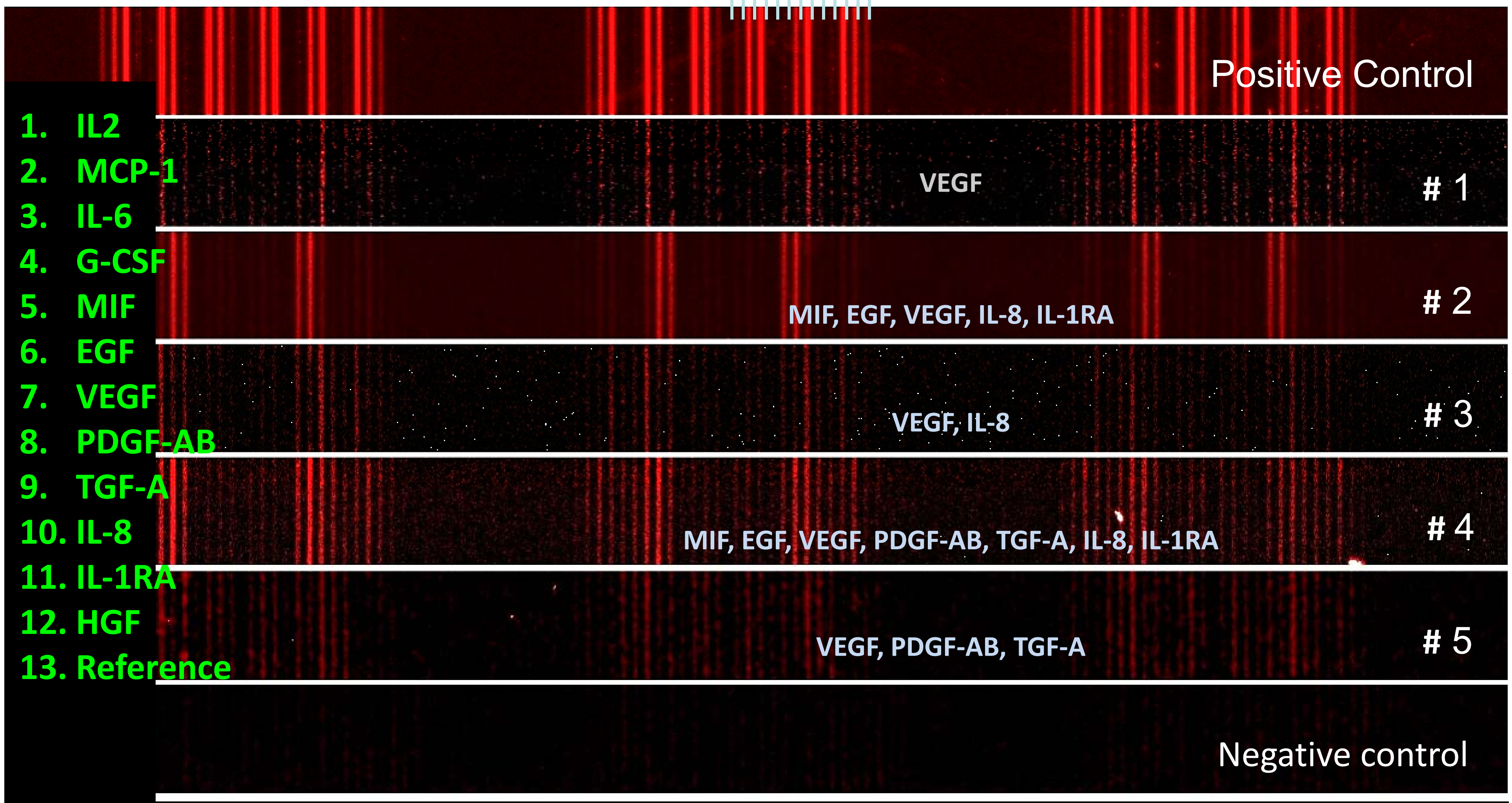
Cost is limited by antibodies



Glioblastoma Patient Trial

Biomarkers 1 to 12

Patient ID



Electronic Design Automation

What is the basis for the panels of biomarkers?

1. Literature: provides 4 or 5 potential protein biomarkers
2. Deep transcriptome analysis to identify genes that are expressed only in the brain: provides ~100 protein biomarkers
3. As many (or more than) 100,000 measurements carried out on specific patient's tissue (surgically resected): provides ~ 20 protein biomarkers

Measurements carried out as a function of time, cell type, molecular (drug) perturbation, etc., on proteins, mRNAs, genes, etc.

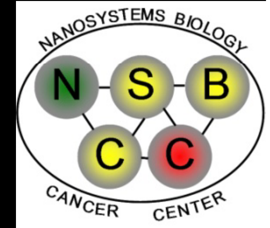
The ideal panel may vary from patient to patient, and putting it together can be beyond an individual's capacity to mine data.

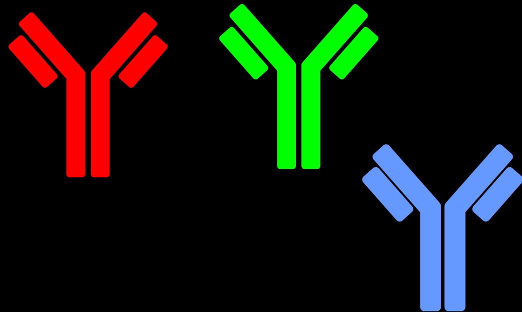
Examples of such experiments on cells derived from a glioblastoma patient's tumor

We need algorithms that can take many (perhaps 10^8), diverse experimental measurements and utilize them to back out:

- A hypothesis for how the system works
- How the system has been perturbed by disease
- A few measurements we can make that will reflect the state of the system

The biggest protein-measurement bottleneck: Protein Capture Agents





Antibodies can cost ~$500 per milligram

They are chemically, biochemically, and physically unstable

Can cost ~$\$10^4$-$\$10^5$ to develop

Keeping a panel of ~20 antibody pairs stable for a 20 protein blood assay can cost as much as the antibodies themselves

A 100 protein (antibody) assay would be almost impossibly expensive to maintain

Pasadena test for a Protein Capture Agent

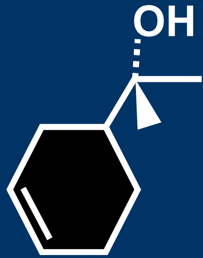
Rosemary Rohde & Heather Agnew

Store, as a powder, in your car trunk on an August day in Pasadena
Retrieve one year later
Capture agent still exhibits antibody-like selectivity and sensitivity

Technology must be adaptable to high throughput manufacturing

Protein Capture Agents

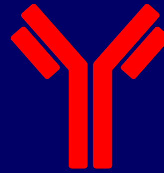
Chemically prepared libraries



chemical space & molecular size are trade-offs – e.g. a comprehensive 6-mer (short) peptide library constructed from 18 artificial amino acids is >30M compounds – a barely manageable number

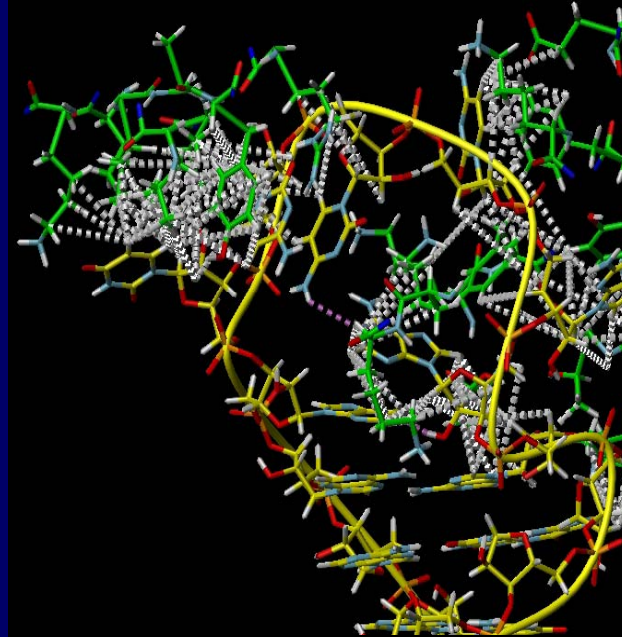
Stability, solubility, etc., can be built in

Biologics



chemical space & molecular size are both achievable

Stability, solubility, etc., are generally not achieved



Antibody-like affinities and selectivities (from artificial peptide-like capture agents) requires the sampling of comprehensive chemical space for a 25-30-mer peptide constructed from 18-22 amino acids, over multiple generations

Manufacturable & Stable Protein Capture Agents

Requirements of a good strategy

- Simple & robust chemistry
- Comprehensive chemical space & high molecular weight
- Capture agent stability built-in at the start
- Prior knowledge about protein target IS NOT required
- Entire scheme may be automated

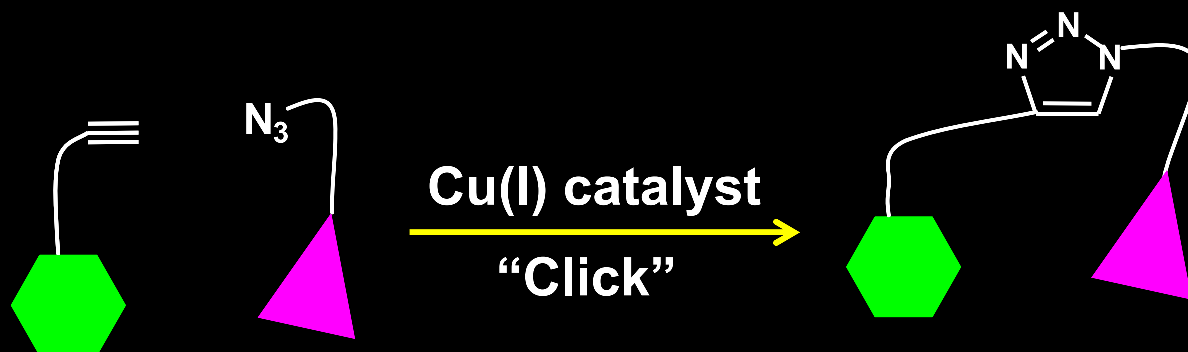
And..

Antibodies: start → finish 24 – 36 weeks
Capture agents: start → finish 2-3 weeks

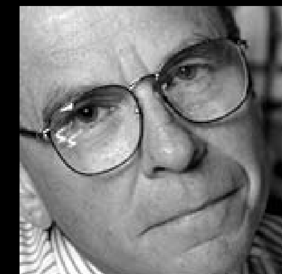
A novel approach to Small Molecule Inhibitors

Very reliable chemistry (Huisgen 1,3-dipolar cycloaddition)

R. Huisgen, G. Szeimies, L. Möbius, *Chem. Ber.* **1967**, *100*, 2494–2507.

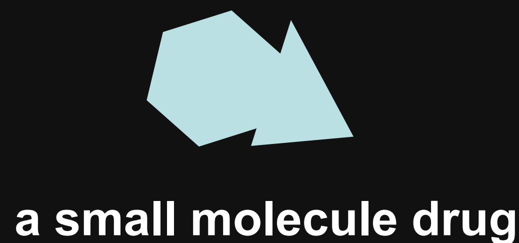


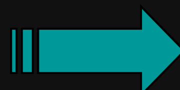
K. Barry Sharpless

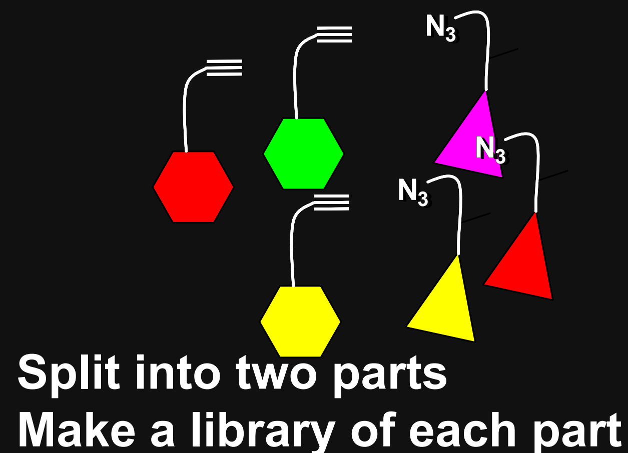


H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chem. Int. Ed.* **2001**, *40*, 2004–2021.

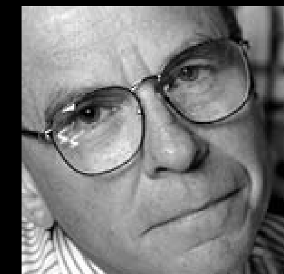
A novel approach to Small Molecule Inhibitors

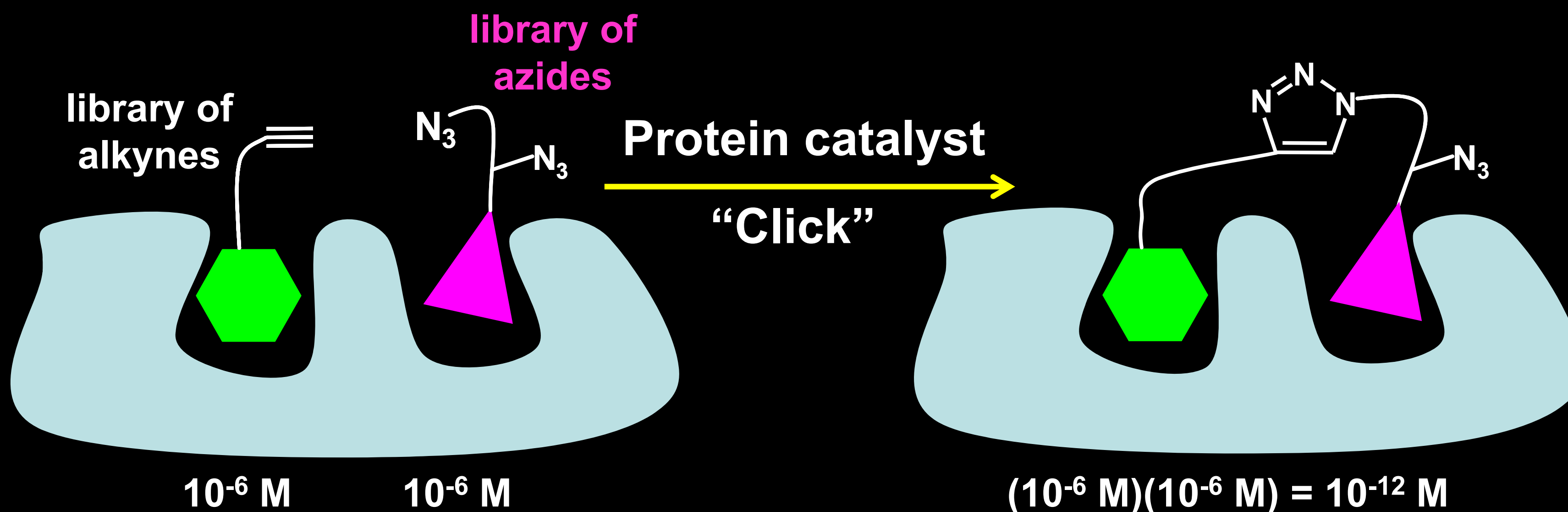






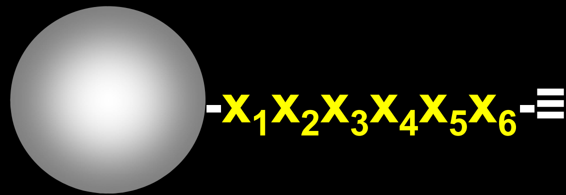
K. Barry Sharpless





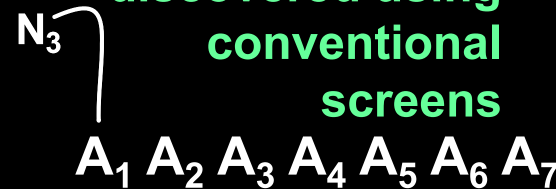
H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chem. Int. Ed.* **2001**, *40*, 2004–2021.

10^8 **element** **bead-based 6-mer peptide library built from artificial and non-natural amino acids**



x_i = artificial or non-natural amino acid
$i=1\text{-}18$

An azide terminated 7-mer peptide anchor ligand discovered using conventional screens

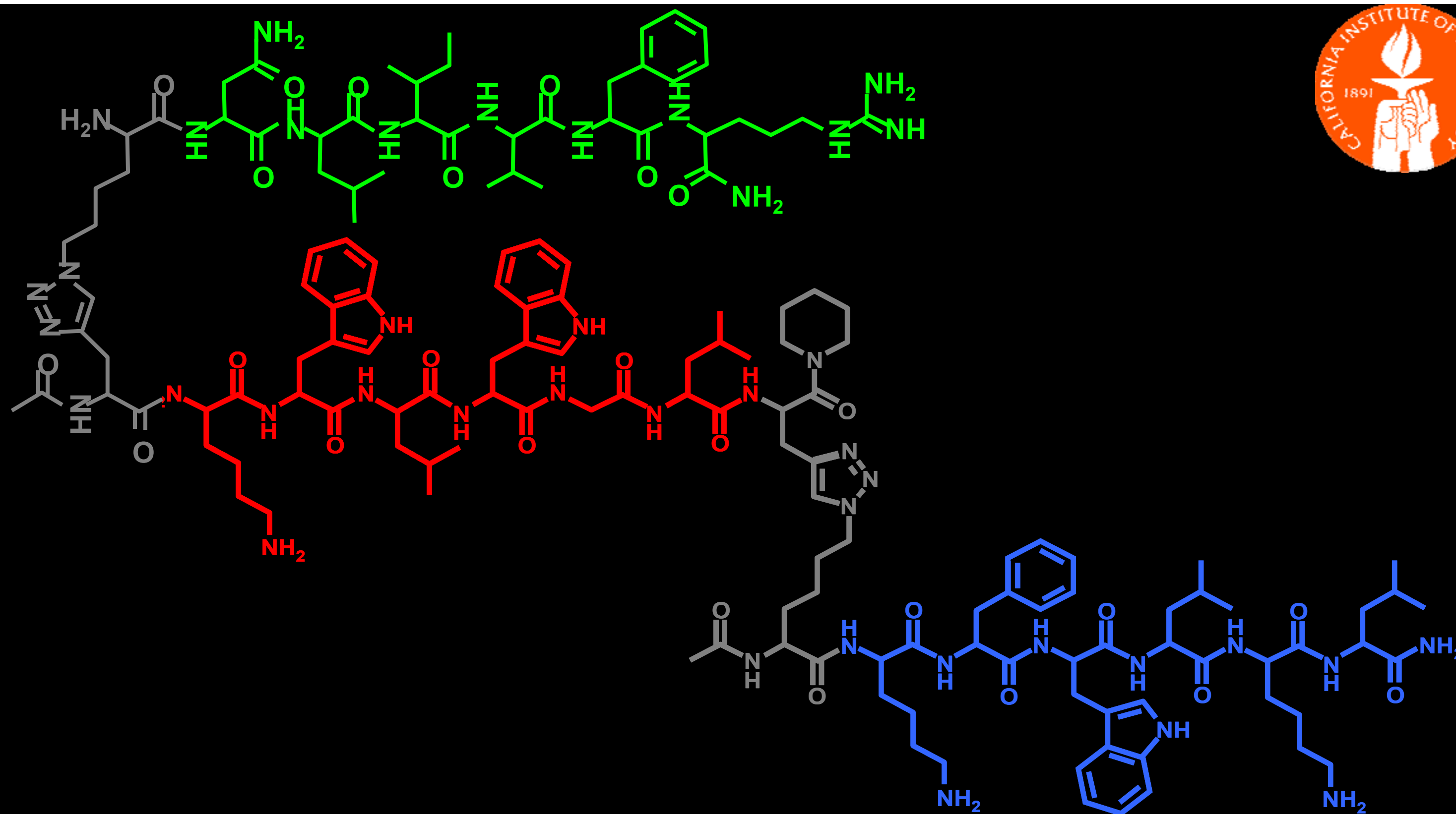


Protein target

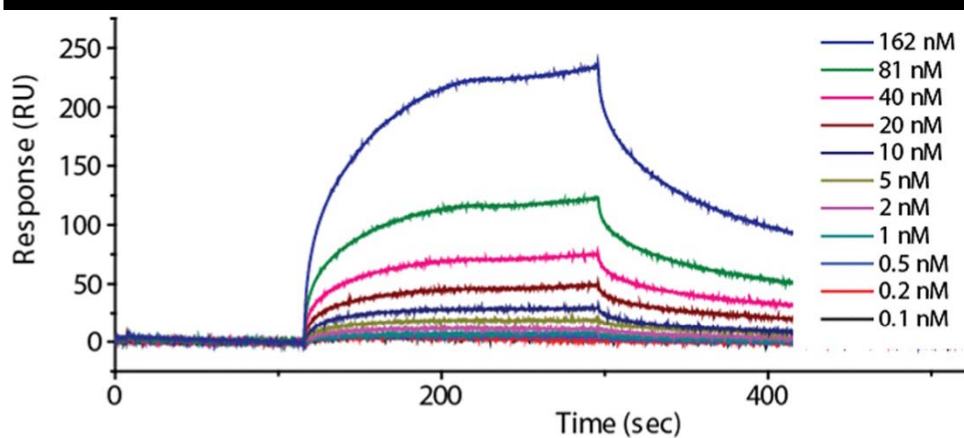
Protein + anchor ligand incubated with large peptide (bead) library

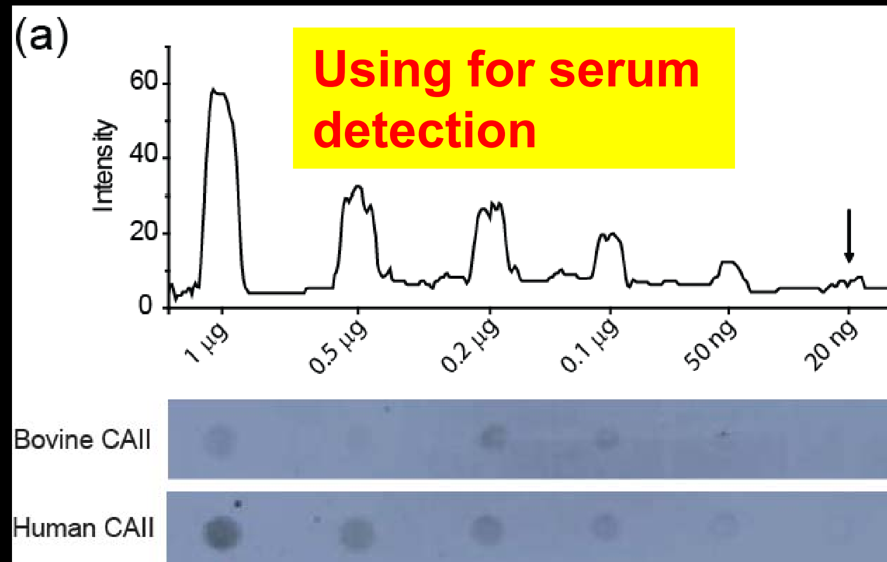
Protein couples best library peptides with anchor ligand by catalyzing formation of triazole

A biligand is formed. That biligand may be used to form a triligand, which can be used to form a tetraligand, etc…

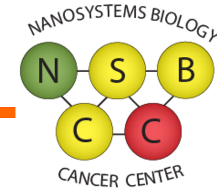


Human CAII (40 nM affinity)



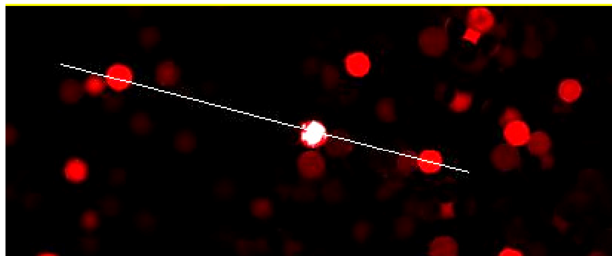


Making the approach high throughput

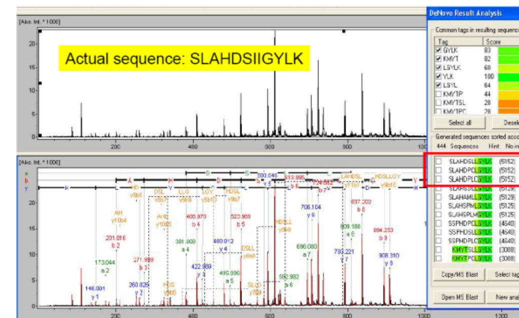


1. Make 34 million peptides, one peptide per bead
24 hr = 1 library (only make once)

2. Incubate with fluorescently labeled protein: 4 hours



3. Identify 100 hit beads: 4 hours



4. Single bead peptide sequencing to identify hits: MALDI TOF/TOF
~4 hours

5. Make focused peptide library, and repeat 2-4
26 hrs

~2-3 days to identify an anchor peptide

protein → multi-ligand → 1-2 weeks

In Singapore: Jaehong Lim Su Seong Lee Junhoe Cha Sylvia Tan Shi Yun Yeo

Making the approach high throughput

NANOSYSTEMS BIOLOGY CANCER CENTER
CALIFORNIA INSTITUTE OF TECHNOLOGY 1891
Institute of Bioengineering and Nanotechnology A*STAR

1. Make 34 million acetylated peptides, one peptide per bead
24 hr = 1 library; only do once

2. Incubate with fluorescently labeled protein and anchor ligand: 4 hours

3. Identify 40 hit beads: 4 hours

4. Single bead peptide sequencing to identify hits: MALDI TOF/TOF
~3 hours

Actual sequence: SLAHDSIIGYLK

5. Make focused peptide library, and repeat 2-4
26 hrs

~2-3 days to identify a biligand capture agent

Ligand → biligand → triligand → tetraligand → pentaligand

2-3 days per step; tri- and tetra-ligands ≡ good antibodies

Design Automation

1.

2.

3.

5.

In principle, each step may be fully automated based upon results of the previous step – human intervention is not necessary – the final capture agent may be fully generated via design automation of the manufacturing process.

Ligand → biligand → triligand → tetraligand → pentaligand

2-3 days per step; tri- and tetra-ligands ≡ good antibodies