P4 Medicine: Catalyzing a Revolution from Reactive to Proactive Medicine

Predictive, Personalized, Preventive and Participatory

Lee Hood

Institute for Systems Biology, Seattle

ISB's Approach to P4 Medicine

- Develop the tools and strategies for patient assays
- Bring P4 medicine to patients with the creation of the P4 Medical Institute (P4MI) in partnership with Ohio State Medical School

P4 Medicine Is Revolutionary

- P4 medicine is medicine of the present/near future.
- P4 medicine is driven by systems approaches to disease and emerging technologies
- P4 medicine will use measurements to quantify wellness and its transition into disease
- P4 medicine is revolutionary rather than evolutionary or incremental
- P4 medicine sees the patient (consumer) as the central focus of healthcare
- Pilot projects with informational assays in patient groups will be necessary to convince skeptics.
- P4 medicine will restructure the business plans of every sector of the healthcare industry—enormous economic opportunities
- P4 medicine will be effective, inexpensive and provide enormous economic benefits to economies—readily available to poor and rich.
- The national healthcare debate in the future should be reframed around P4 medicine rather than the old reactive medicine.

Medicine/Biology as an Information Science

- Two major types of information—digital information of the genome and environmental information
- Two informational structures connect the clash of digital and environmental information and phenotype—biological networks and molecular machines
- Biological information is hierarchical—DNA, RNA, proteins, networks, cells, tissues, individuals, etc—and these different types of information must be integrated

Essentials of Systems Biology

- Hypothesis-driven and hypothesis-generating
- Global data acquisition
- Integrate multi scalar data types
- Delineate biological network dynamics
- Formulate models that are predictive and actionable.
- Discovery science is key

INSTITUTE FOR Systems Biology

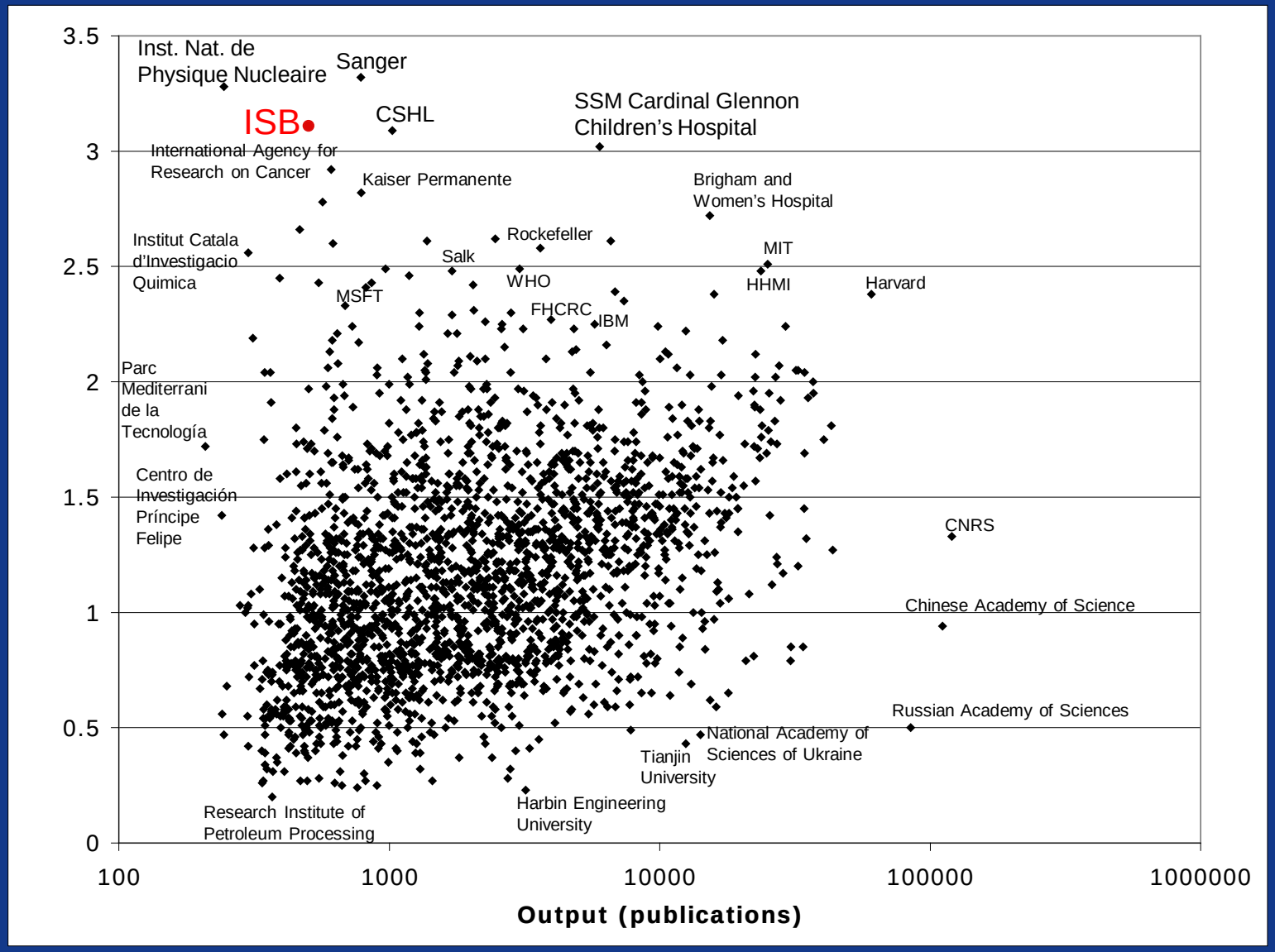
Institute for Systems Biology

Founded 2000—10th Anniversary

ISB has 13 faculty and 300 staff

Institute for Systems Biology

SCImago Institutions Rankings: http://www.scimagoir.com/



ISB 1st in US and 3rd in World for Impact of Papers

Institute for Systems Biology

A Systems Approach to a Neurodegenerative Disease (prion disease) in Mice—what are the disease-perturbed networks and how do they behave?

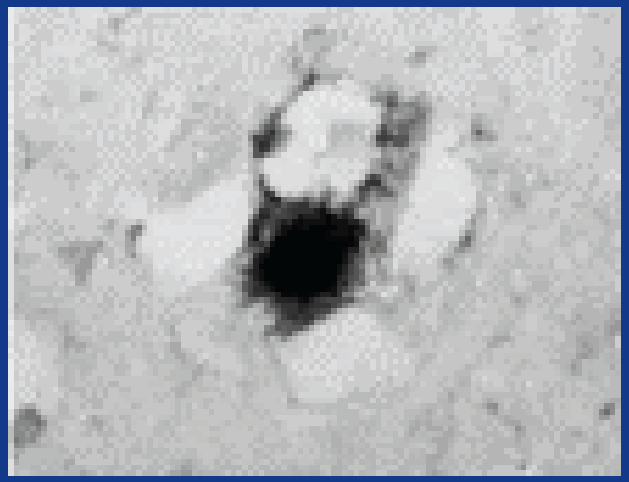
Prion Disease in Eight Mouse Strains: dealing with the signal to noise challenge employing subtractive biology

Group	Mouse	*Prnp* Genotype	Prion Strain	Incubation Time (d)
1	C57BL/6J	*a/a*	RML	~150
2	B6.I-1	*b/b*	301V	~120
3	FVB/NCr	*a/a*	RML	~150
4	B6.I-1	*b/b*	RML	~350
5	C57BL/6J	*a/a*	301V	~260
6	(FVB x FVB.129-*Prnp*^{tm1Zrch})	*a/0*	RML	~400
7	Tg(MoPrP-A)B4053	*30 x a*	RML	~60
8	FVB.129-*Prnp*^{tm1Zrch}	*0/0*	RML	No illness

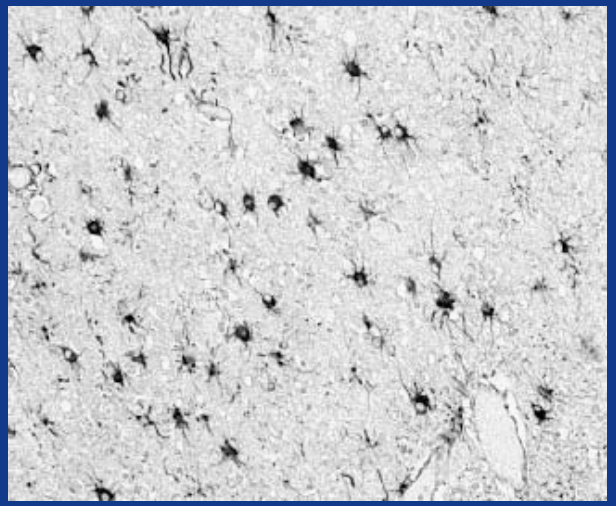
Differentially Expressed Genes--DEGs--7400 to 333

Neuropathology Identifies 4 Networks

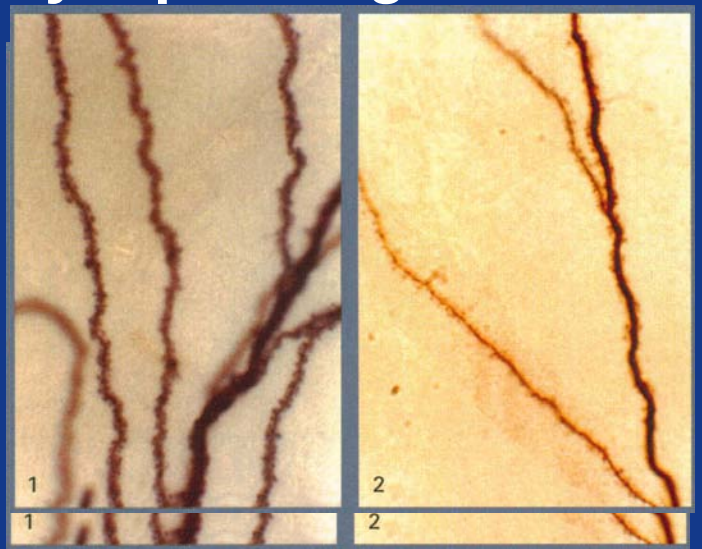
PrP accumulation



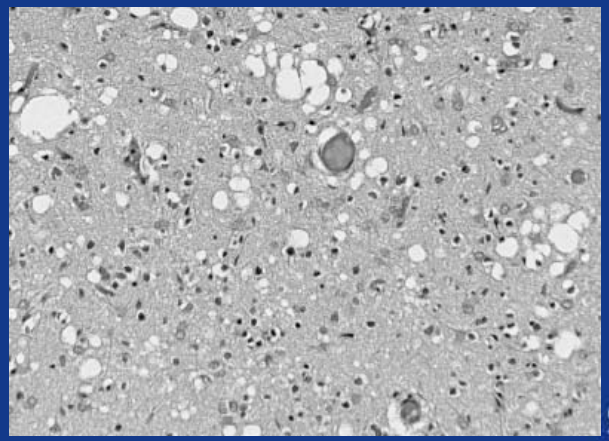
Microglia / Astrocyte activation



Synaptic Degeneration



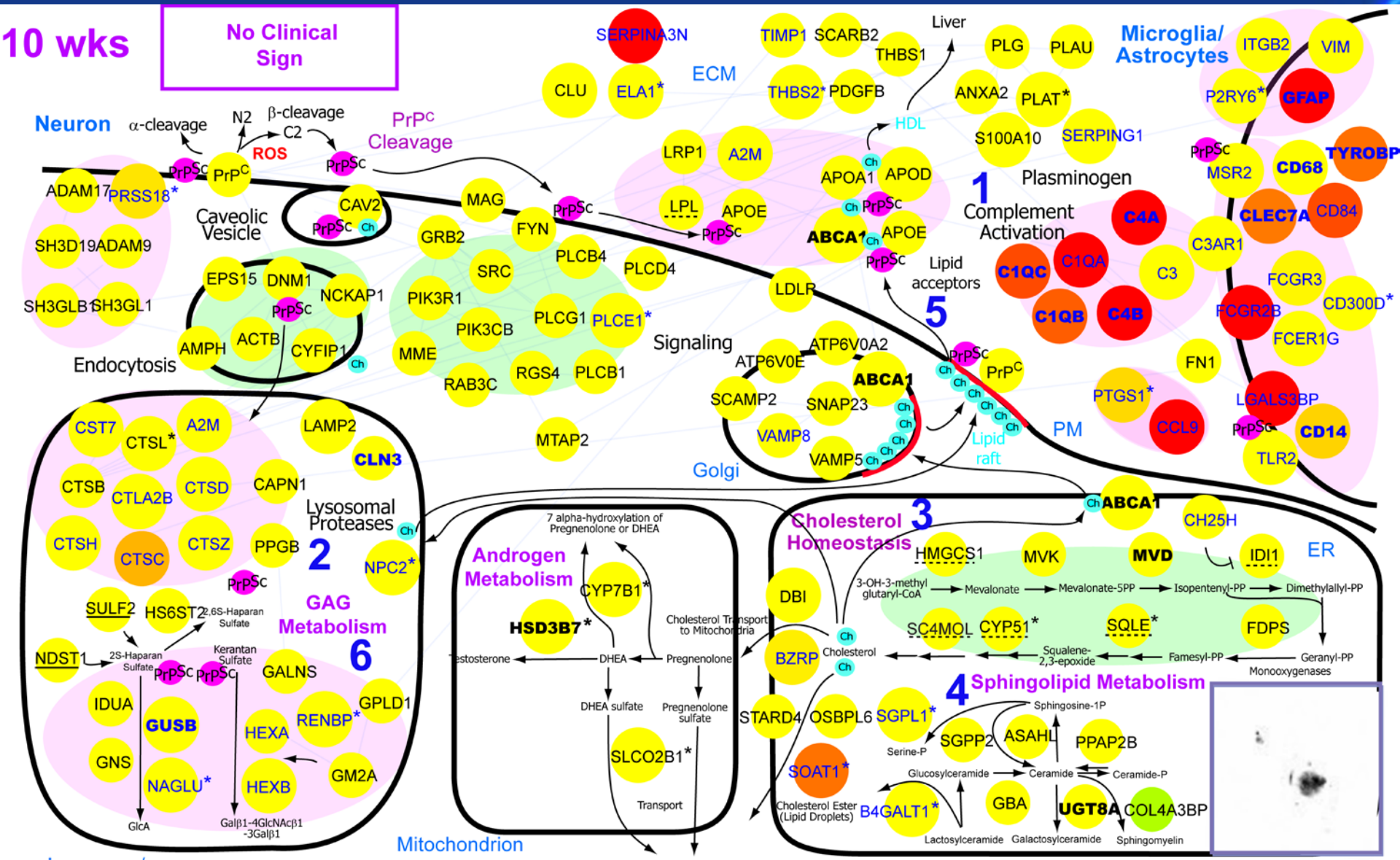
Nerve cell death



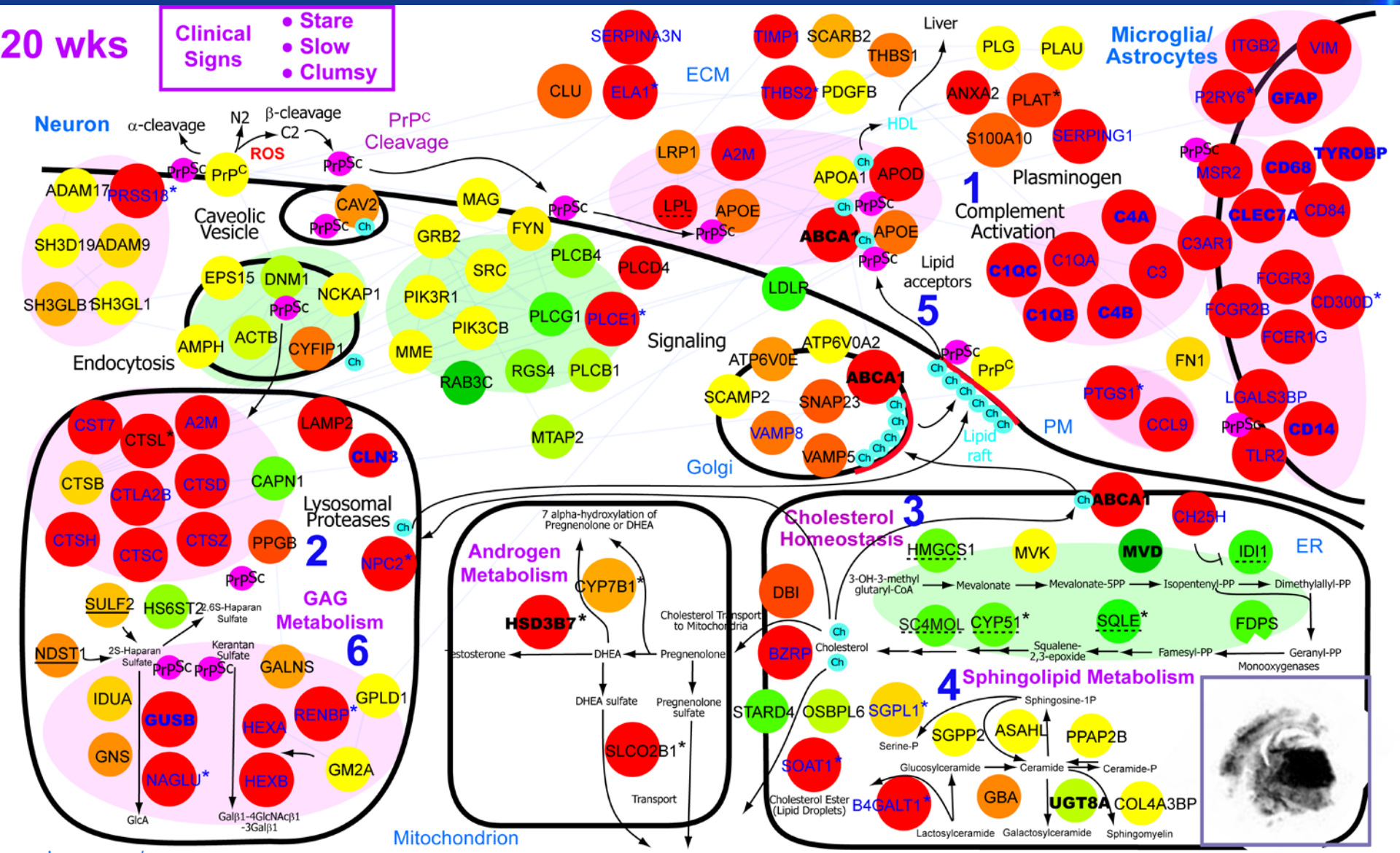
Integration of Six Data Types for Prion Disease Studies in Mice

- Deep brain transcriptome analyses at 10 time points across disease onset in 8 mouse strains
- Correlate with protein interaction data from known (histopathology) disease-perturbed networks
- Correlation with dynamical histopathological studies
- Correlation with clinical signs
- Distribution of infectious prion protein in the brains across disease progression
- Brain-specific blood protein concentration changes

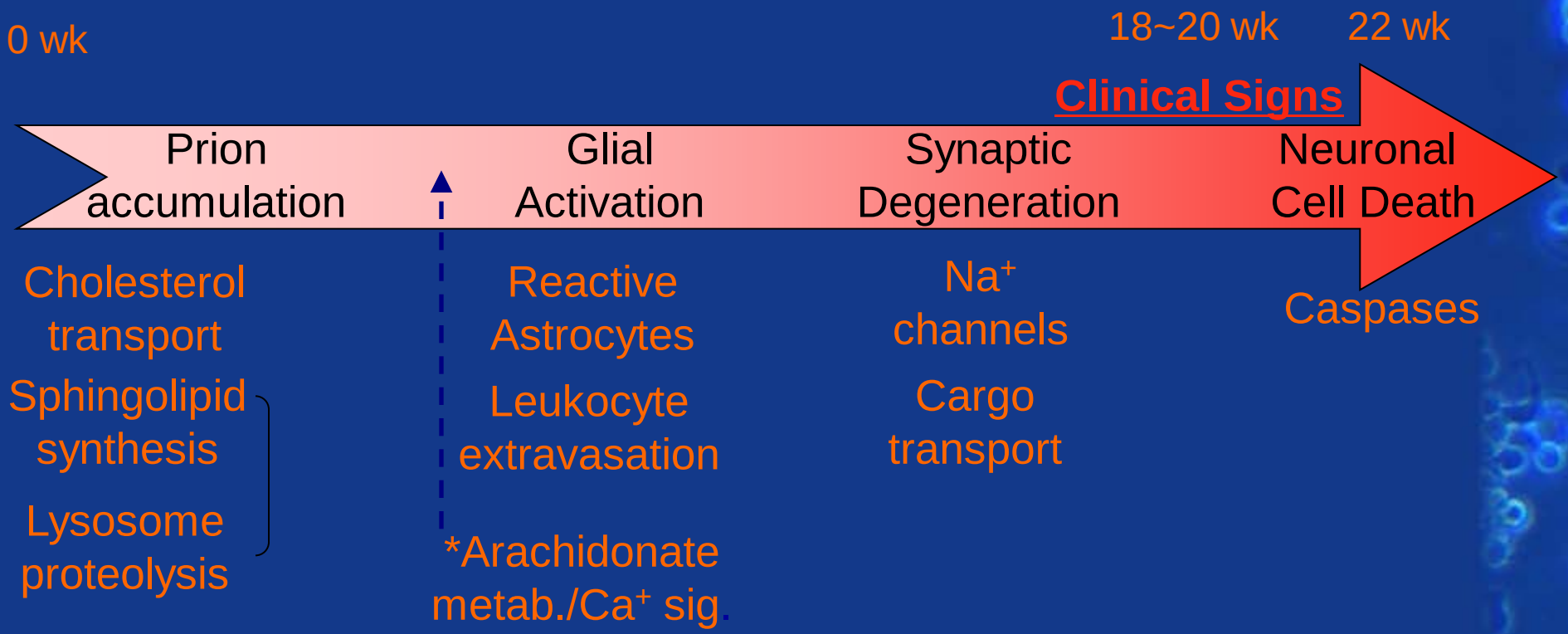
PrP accumulation and replication network—10 weeks



PrP accumulation and replication network—20 weeks



Sequential Disease-Perturbation of the Four Networks of Prion Disease



Institute for Systems Biology

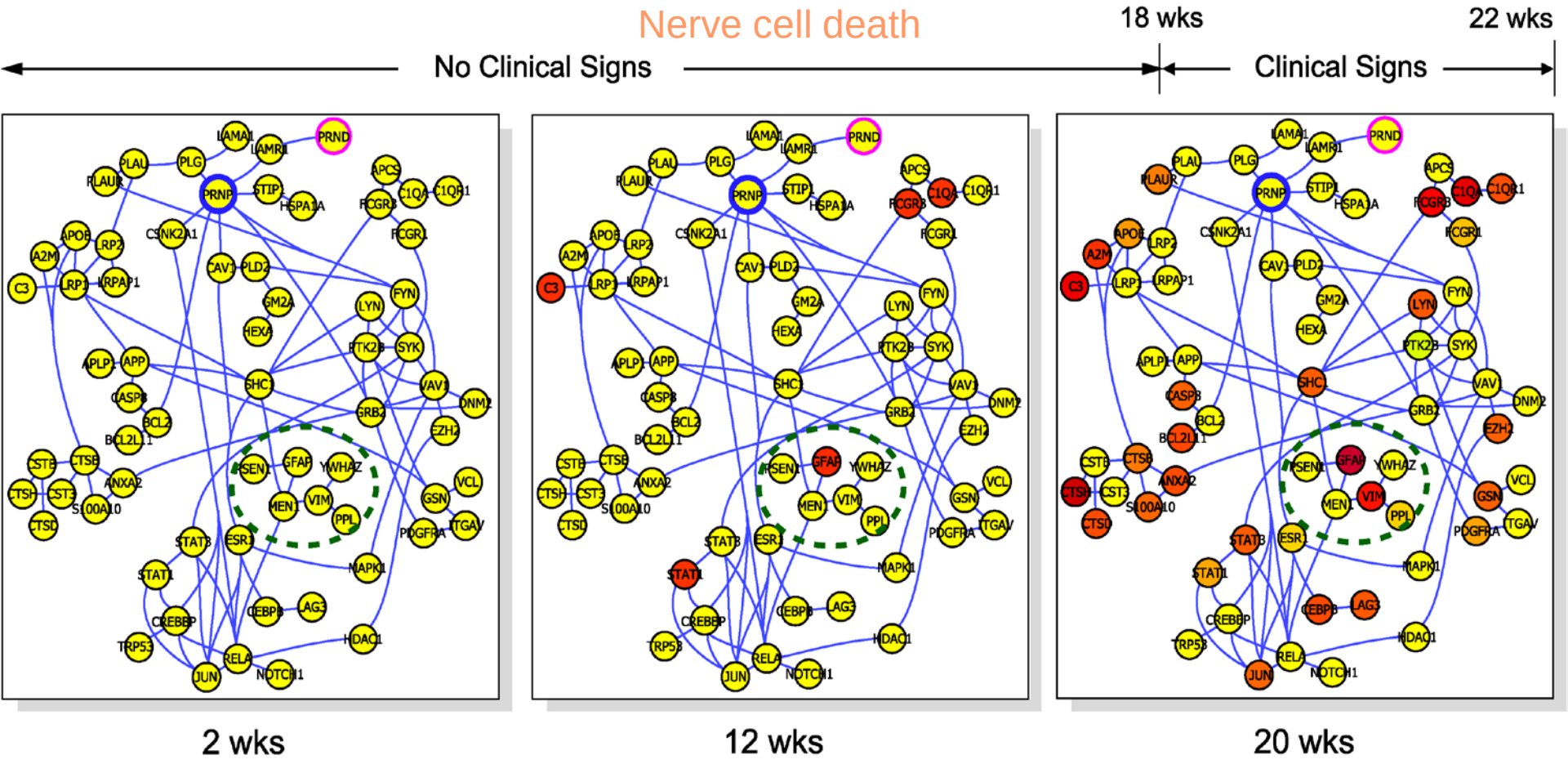
Differentially Expressed Genes (DEGs) Encoding Known and Novel Prion Disease Phenotypes

- About 300 DEGs encode core prion disease
- About 200/300 DEGs encode known disease pathogenic networks
- 100/300 DEGs encode novel pathogenic networks--the dark genes of prion disease
- Re-engineer disease-perturbed networks with drugs—new approach to drug target discovery and systems diagnosis

Making Blood a Window into Health and Disease: A Systems Approach to Blood Diagnostics

- Blood biomarkers that are chosen from dynamic network analyses—relevant to the biology of the disease
- Blood biomarkers that are organ specific

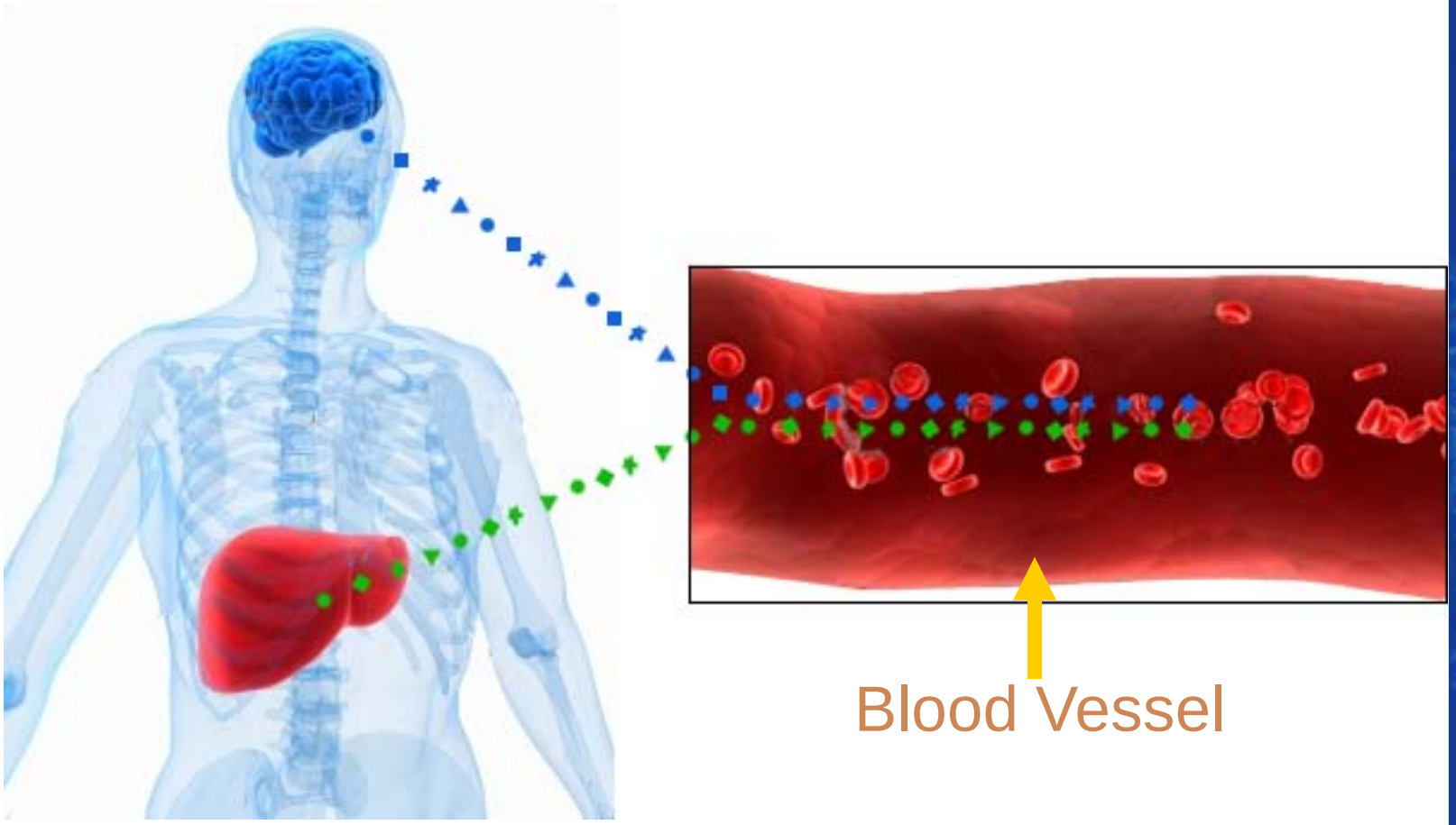
Dynamics of a Brain Network in Prion Disease in Mice



Organ-Specific Blood Fingerprints

Making Blood A Window Distinguishing Health and Disease

110 brain-specific blood proteins/80 liver-specific blood proteins



Why Blood Diagnostics Will Be the Key to P4 Medicine

- Early detection
- Disease stratification--prognosis
- Disease progression
- Follow therapy
- Assess re-occurances

Integrated Diagnostics—platform company for P4 medicine

Big New Projects at ISB

- The 2nd generation human genome project--focus on families where possible--capture, storage and comparative analyses of all human genomes and their relevant phenotypic data
- Human proteome project (ala the human genome project)
- Creating patient informational assays to explore new dimensions of data space—in conjunction with P4MI

Strategies and Technologies: Exploring New Dimensions of Patient Data Space

Genomics

Whole Genome Sequencing of Families: New Genomic Strategy

Project flow

Samples
(DNA from blood lymphocytes)

Complete Genomics
Powering large-scale human genome studies

Specifications:
~120 Gbase/genome
(mapped to Hg18 reference)

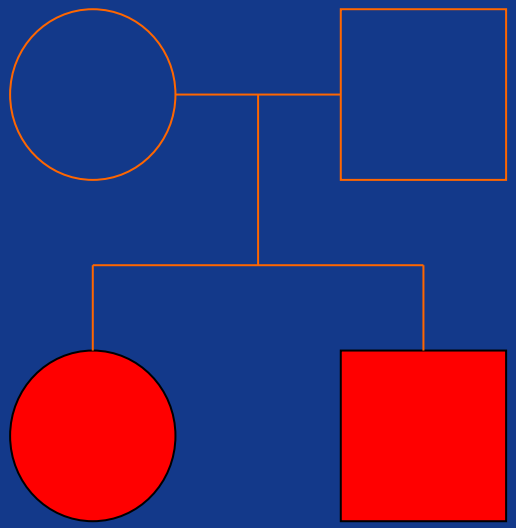
Deliverables
a) Reads
b) Coverage tables

[~20 T bytes of data]

Analyses
- Errors in family data set
- Variations in family
- Recombination of parental blocks
- Candidate gene identification
- de novo mutation rate

Institute for Systems Biology

Whole Genome Sequencing of Family of Four



Unaffected parents

Children with craniofacial Malformation (Miller Syndrome) and lung disease (ciliary dyskinesia)

Identify 70% of sequence errors using principles of Mendelian genetics—less than 1/100,000 error rate

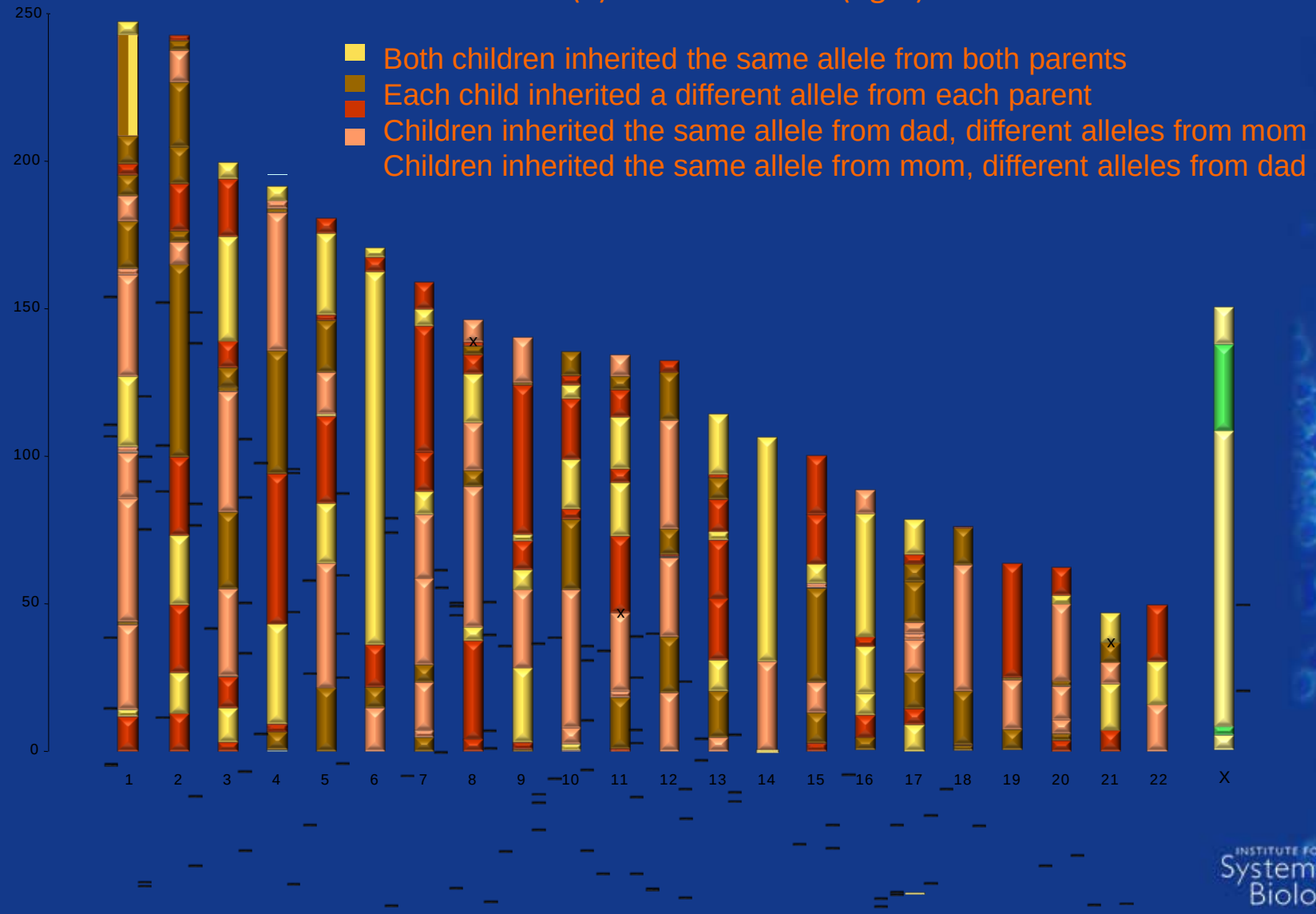
Discovery of about 230,000 rare variants in family—confirmed by identification in two or more family members

Institute for Systems Biology

Recominational Genome Map from Miller's Syndrome Children

65 crossovers in (2) male meioses (left)

104 crossovers in (2) female meioses (right)



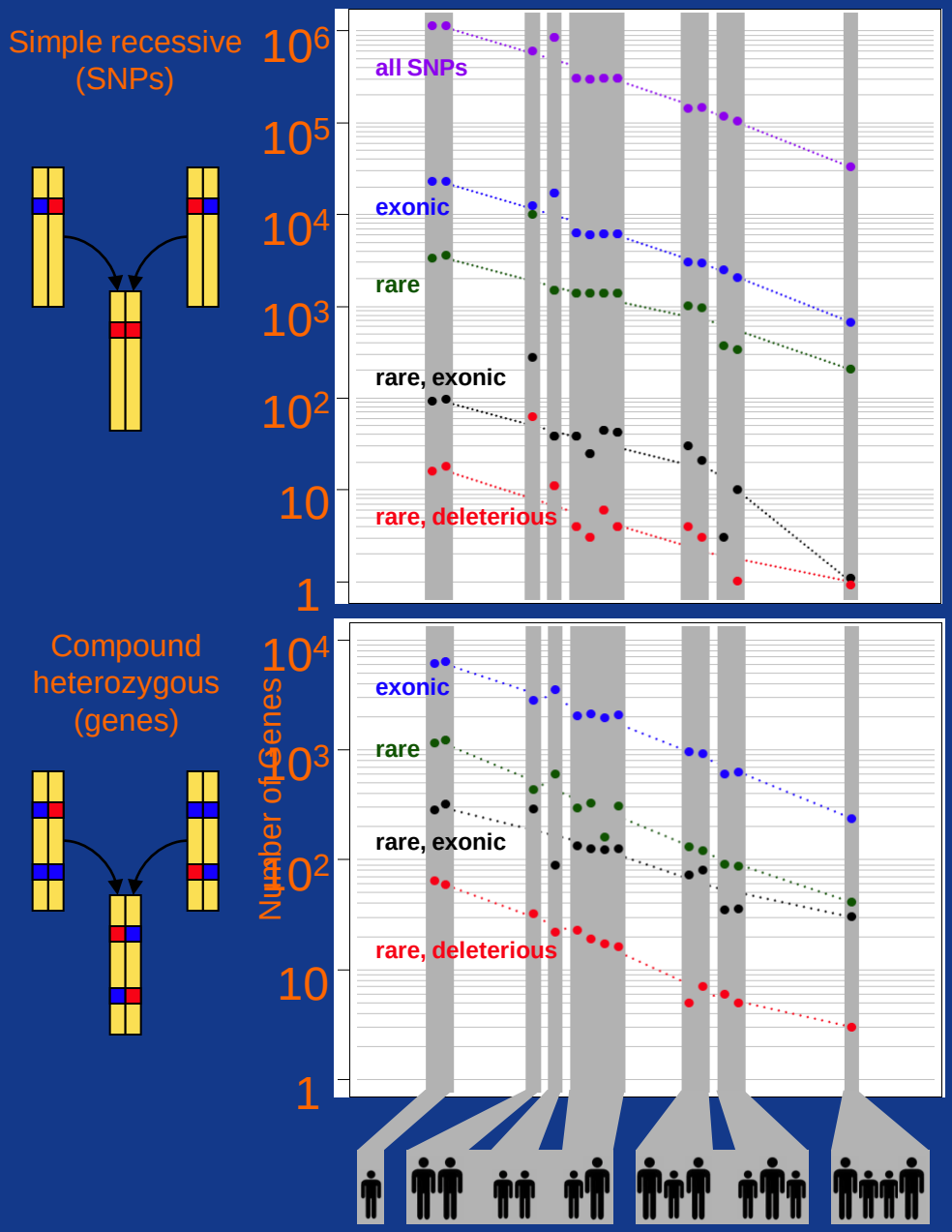
Institute for Systems Biology

Inter-generational base-change mutations

- "Genetic errors": Genetically impossible SNPs in kids will most likely be sequencing errors, but some (perhaps ~1/1000) will be new mutations. We can find these!
- New mutations, germline nucleotide substitutions, have never been directly measured before.
- Low error rate & family makes this approach work
- We trapped by Agilent hybridization selection and resequenced ~60,000 sites
- Indirect, phylogenetic estimate is between 8×10^{-9} & 2.1×10^{-8} per base per generation,
- The intergenerational mutation rate of the autosome is
  - **~1.1×10^{-8} /base/generation (~30% transversions)**

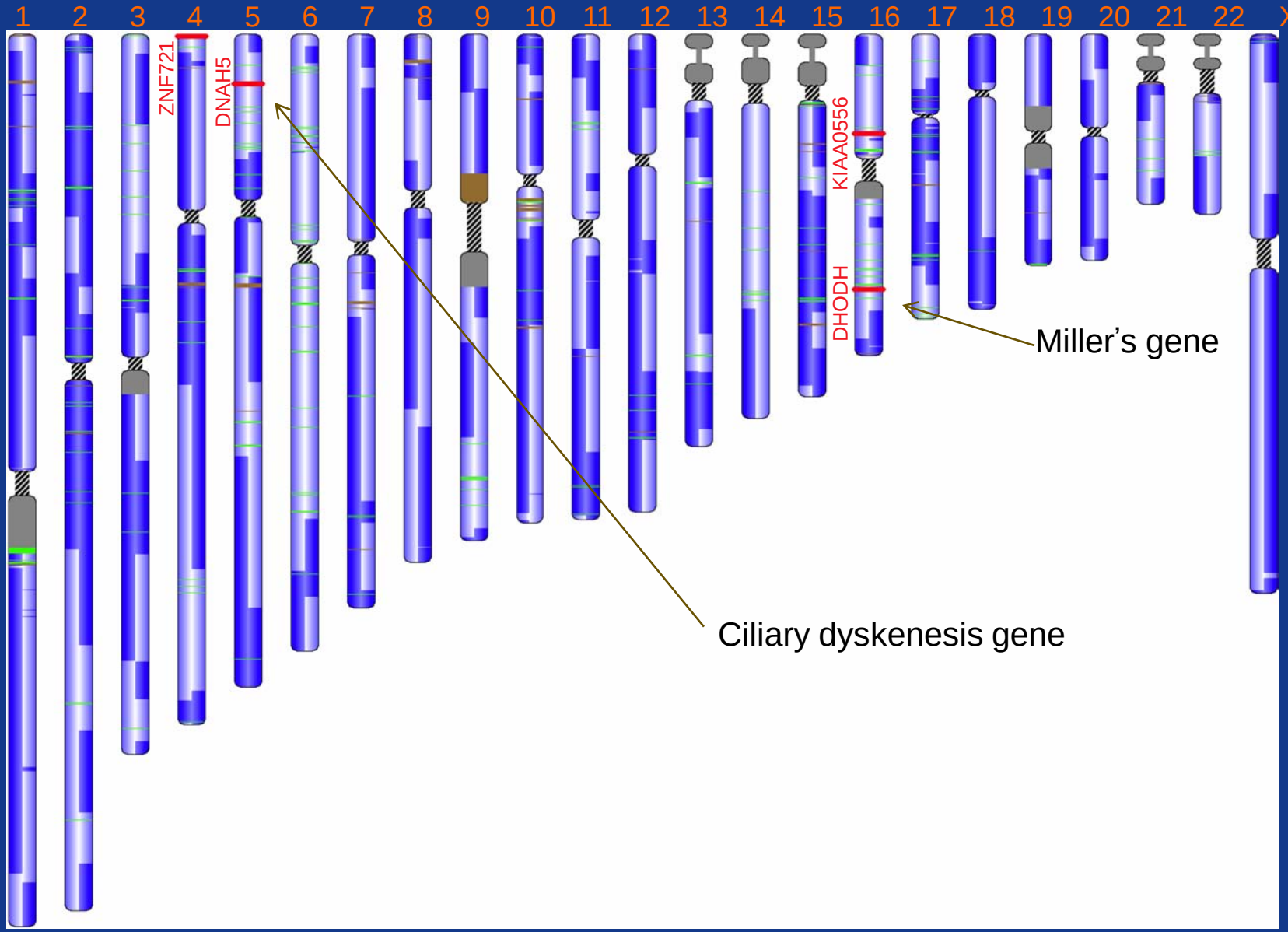
Science, 2010

Institute for Systems Biology



Disease Gene Candidates Reduced Analyzing Complete Family

Institute for Systems Biology



Center for Complete Human Genome Sequence Analyses

- Collect all available complete genome sequences—maintaining family and phenotypic relationships but removing names—cloud storage
- Develop software to evaluate their quality—base call and assembly—cloud analyses
- Develop software to annotate and reduce data storage
- Store phenotype data for each genome—molecular, higher phenotype, classic medical records—use to stratify for relevant phenotypes
- All by all and relevant subsets of genome comparisons of genomes within stratified types

Proteomics

The Human Proteome Project: Creation of a Human MRM Proteome Atlas

An SRM/MRM assay for every human protein

R. Moritz, L. Hood R. Aebersold

Collaboration with Agilent and OriGene

Institute for Systems Biology

Why Did the Genome Project Transform Biology?

- Provided a complete parts list of genes (and proteins)—key for global systems approaches
- Made all genes (and all other potentially interesting regions of the genome) available to all scientists
- High throughput sequencing and other technologies drove the development of high throughput data generation platforms
- Drove the development of sophisticated computational and mathematical approaches to biology
- Enable mass-spectrometry-based proteomics
- Instituted the vision of immediate open data access
- Provided genomic access to plants, animals and microbes
- Transformed medical diagnostics—pharmacogenomics and disease diagnosis
- Transformed our understanding of evolution

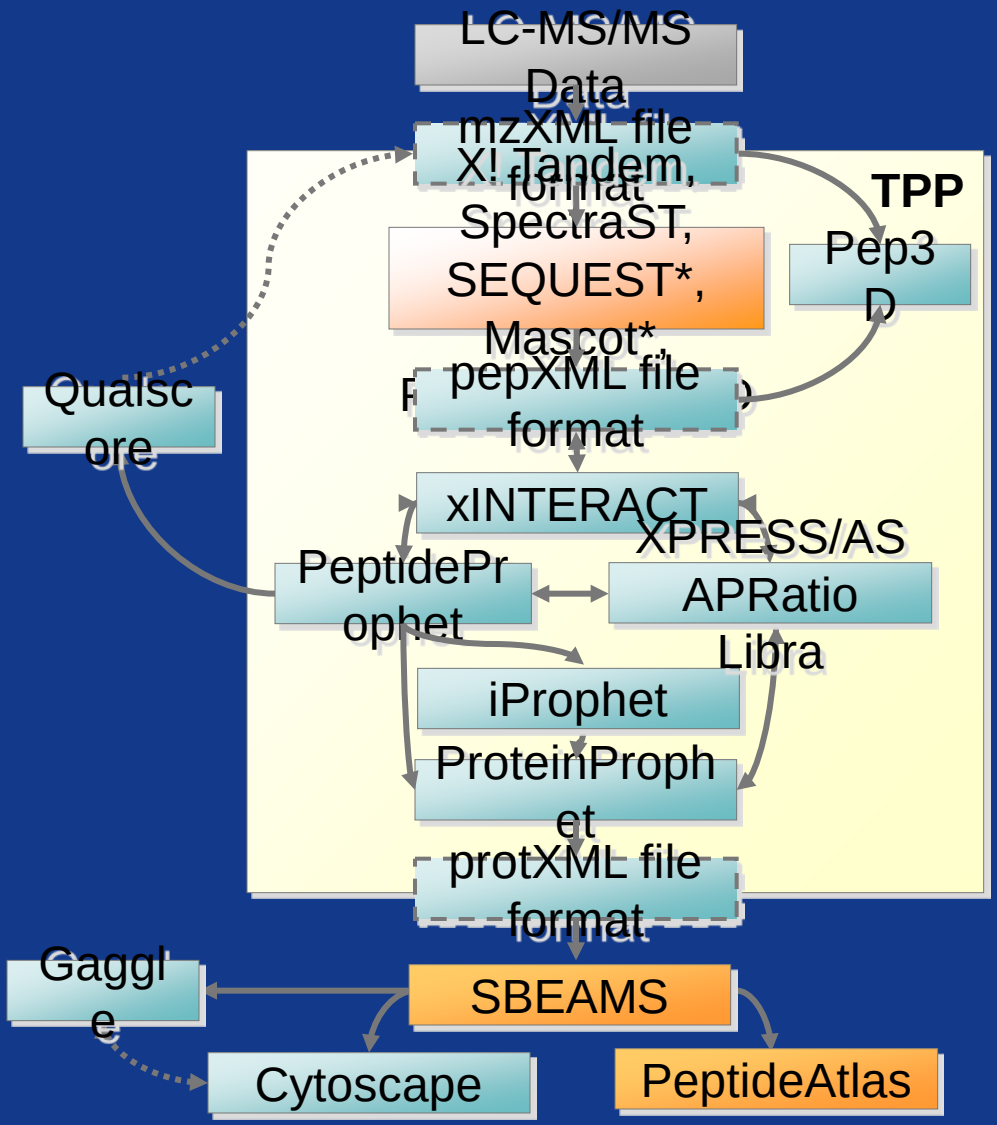
Proteomics Is Very Complex

Global Characterization of Proteins

- Identification
- Quantification
- Modification
- Interaction
- Half-Life
- Compartmentalization
- Three-Dimensional Structures and Dynamics
- Assigning Function

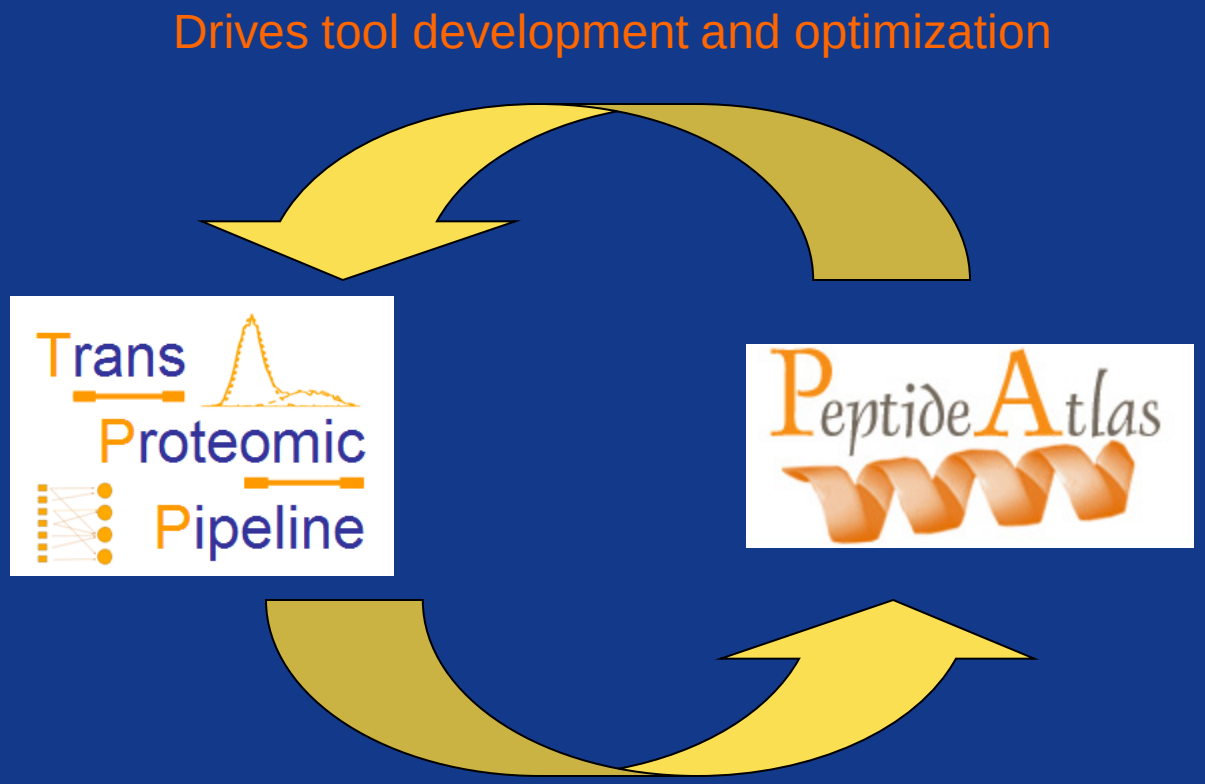
ISB Proteomics Pipeline

Trans Proteomic Pipeline (TPP) components

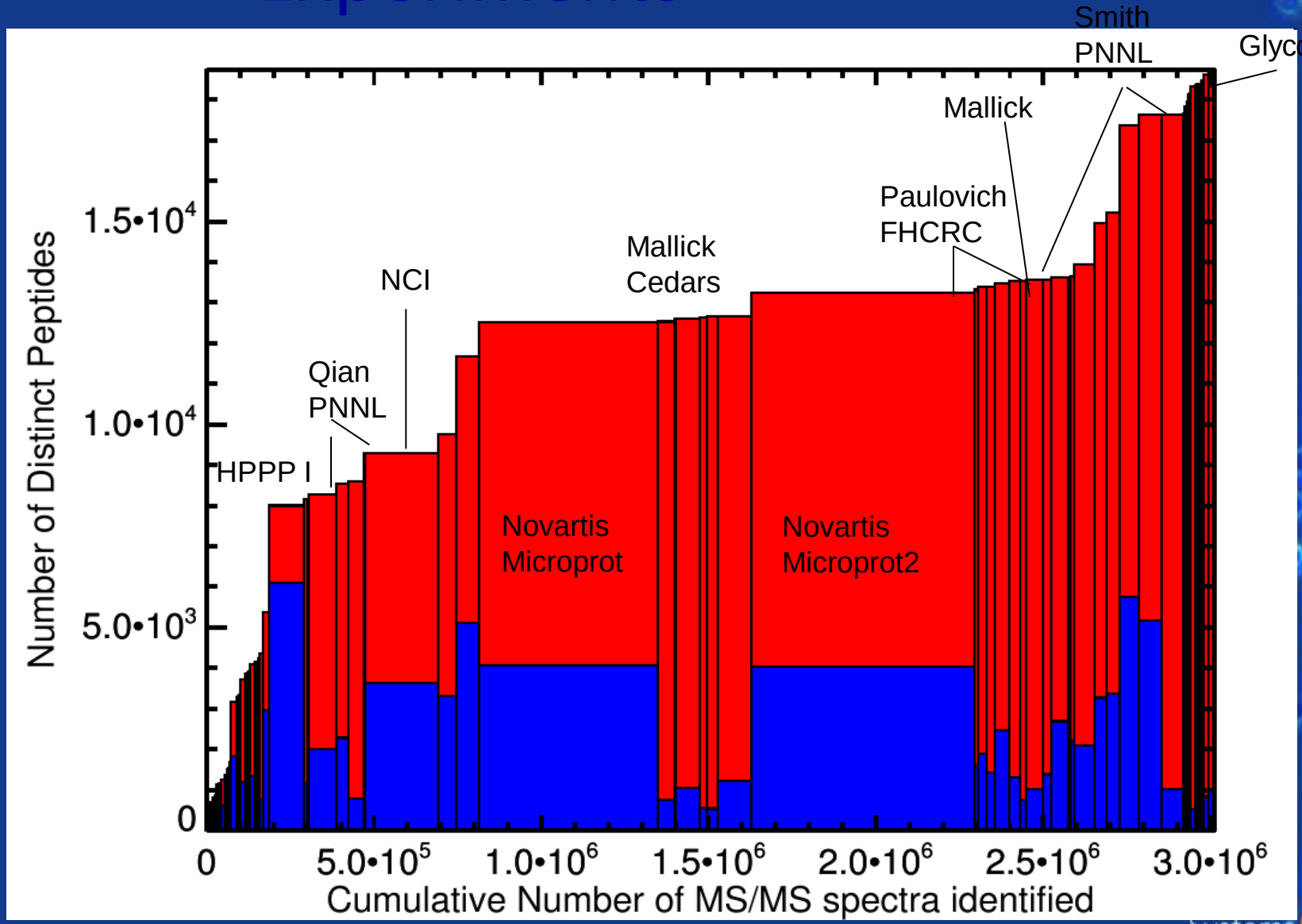


* Commercial software not part of TPP

TPP: Foundation for PeptideAtlas



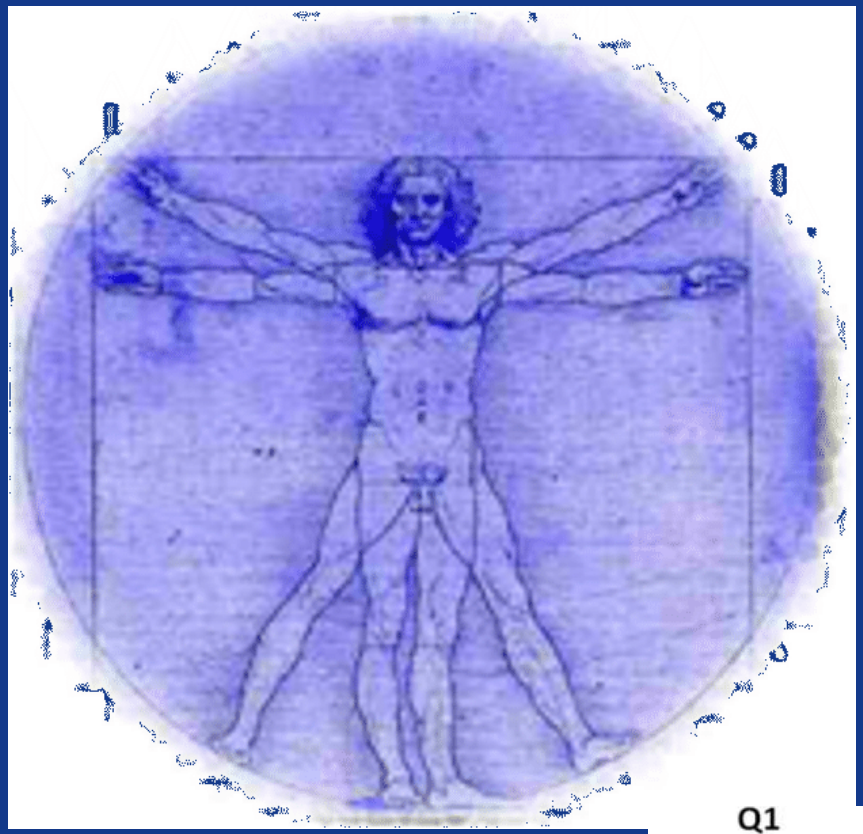
PeptideAtlas Datasets and Experiments



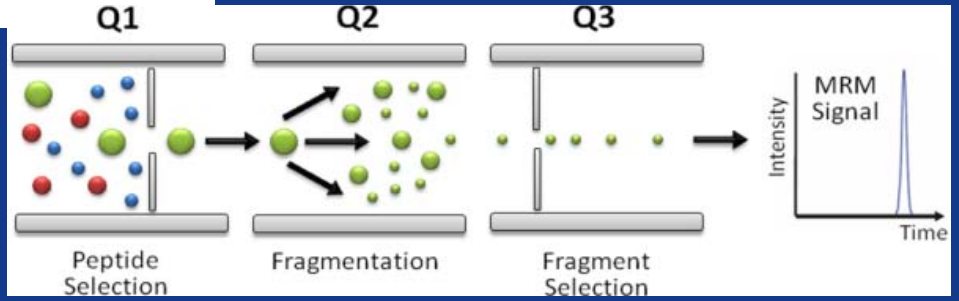
Systems Biology

Targeted Proteomics

Human MRMAtlas—in next 3 years



20,333 proteins (20,328)
32,562 proteins incl. isoforms
~500K distinct peptides (7-30aa)



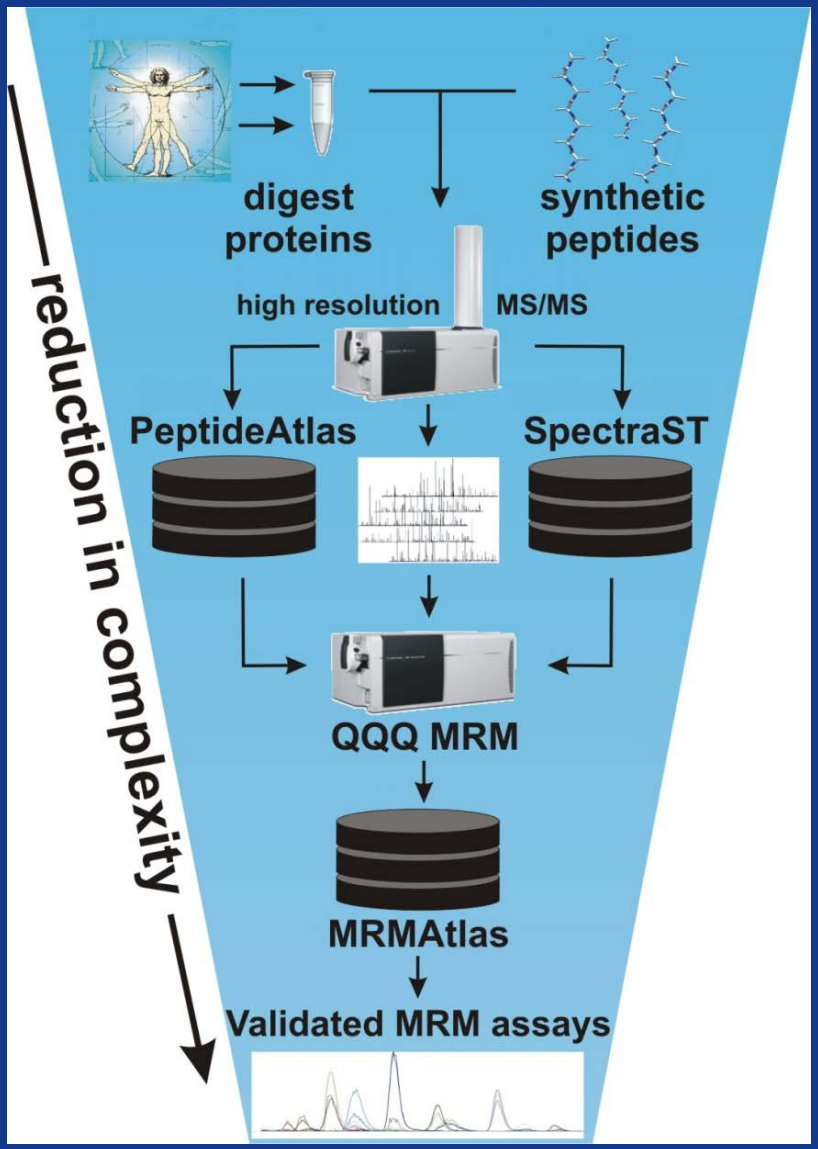
R. Moritz ISB/R. Aebersold ETH
Agilent and OriGene

Developments at ISB - MRMAtlas

Human proteins
(from natural source or synthetic)

Develop Human PeptideAtlas
(from tryptic digests or synthetic peptides)

Develop Human MRMAtlas
(verified quantitative assays)

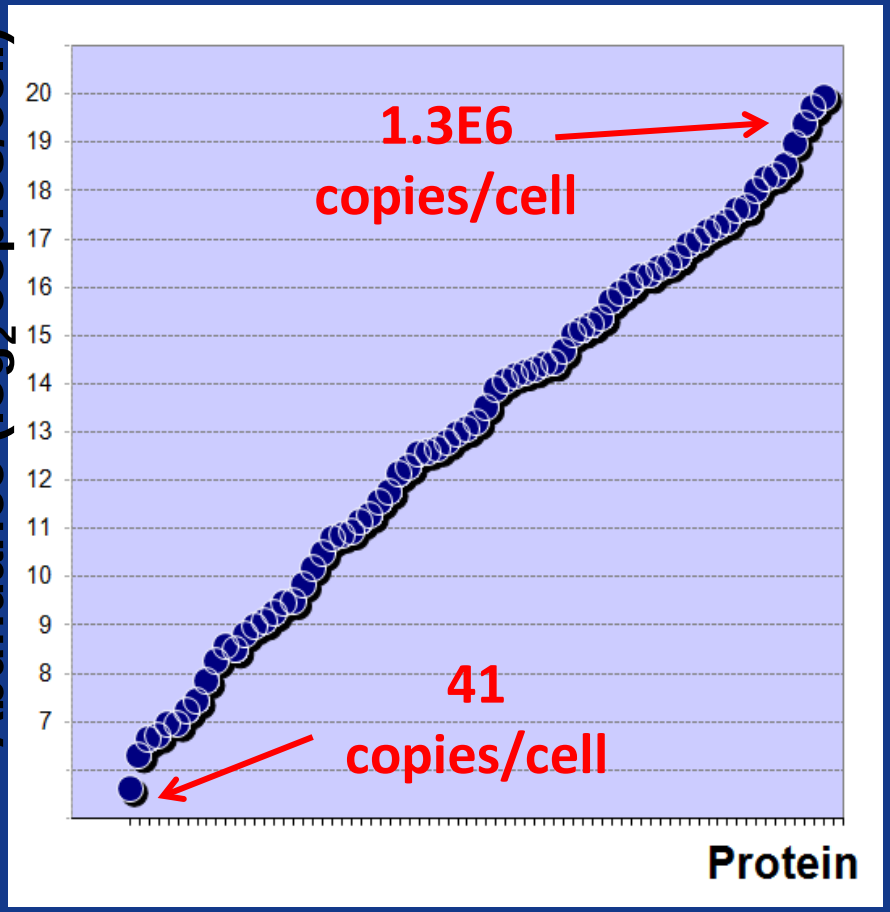


Synthetic "proteotypic" peptide
(from inexpensive synthesis without purification)

Develop optimized transitions from PeptideAtlas

INSTITUTE FOR Systems Biology

MRM Assays for 97% Yeast Proteins



Protein	Weissman, copies/cell	Measured, copies/cell
SPO14	49	62
ENG1	64	198
ATC2 (Pep1)	98	89
ATC2 (Pep2)	98	140
SDS3	105	270
PRS7	125	2100
OSM1	432	3043
NFS1	504	479
MTNA	922	1127
CDC12	1169	1128
PDE1	1404	154
ATC6	1873	1063
ALDH2	2072	354
AROG	26272	5890
GLYC	67559	40178
6PGD1	101441	24988
EF3A	870578	51340
ALF (Pep1)	1018216	712630
ALF (Pep2)	1018216	749513

Proposal for a Human Proteome Project

- Create targeted assays (SRM/MRM) for all human proteins
- Do the same for model organisms (mouse, rat, fly, nematode, etc.)
- Develop technologies to increase the power of proteome analyses
- Develop the computational and mathematical tools necessary for proteome analyses
- Develop the software to make all biological networks/molecular machines and their proteins accessible to biologists (protein chemists)

Microfluidic Protein Chip:

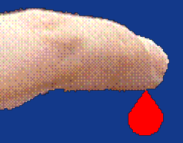
Assay 2500 Organ-Specific Blood Proteins from Millions of Patients Using a Drop of Blood

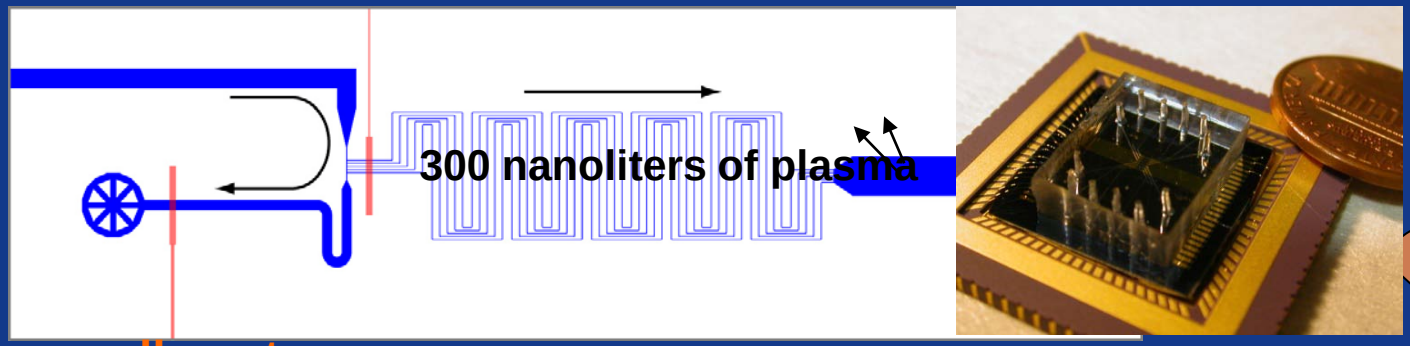
- Jim Heath--Caltech

Institute for Systems Biology

DEAL for *In vitro* molecular diagnostics:

Integrated nanotech/microfluidics platform





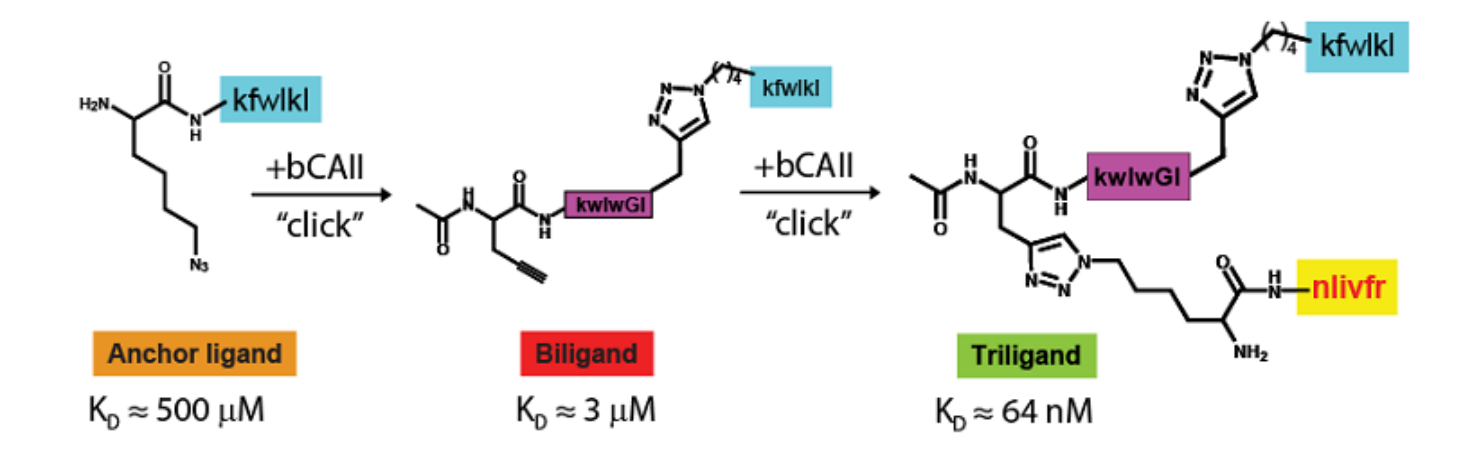
5 minute measurement

Jim Heath, et al

Peptide Protein-Capture Agents

Jim Heath--Caltech

Antibody Displacement *Technology—Heath--Caltech*



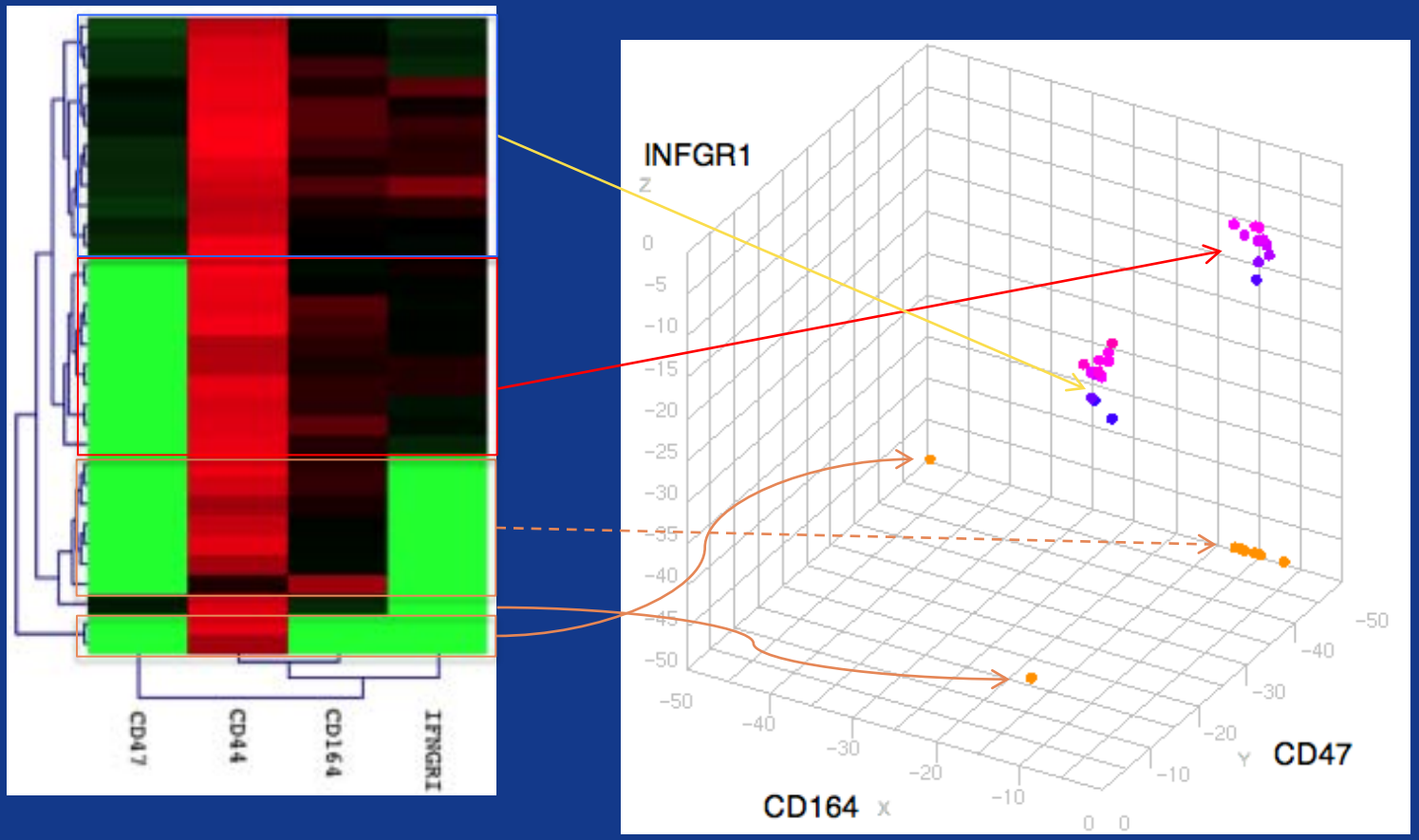
An example of a triligand PCC agent for bovine carbonic anhydrase II

Protein Catalyzed Capture Agents:

- triligands determined by repeated screening of target protein across synthetic bead-bound peptide libraries
- anchor peptide is selected on the first screen
- protein catalyzes the formation of second ligand to anchor ligand on second screen
- protein catalyzes the formation of the third ligand to the anchor and second ligand on third screen
- high affinity, stable and easily manufactured triligand capture agents

Single-Cell Analysis

Quantitative transcriptome clustering of single cells from the human glioblastoma cell line U87



Individual Patient Information-Based Assays of the Present/ Future (I)

- Genomics
  - Complete individual genome sequences—predictive health history—will be done sequencing families
  - Complete individual cell genome sequences—cancer.
  - Complete MHC chromosomal sequence in families—autoimmune disease and allegies
  - 200 Actionable SNPs—pharmacogenetics-related and disease-related genes
  - Sequence 1000 transcriptomes simultaneously in one DNA sequencing run from single cancer cells to identify quantized cells states and dissect cancer
  - Analyze aging transcriptome profiles
- Proteomics
  - 2500 blood organ-specific blood proteins from 300 nanoliters of blood in 5 minutes—twice per year (50 proteins from 50 organs)—wellness assessment.
  - Array of 13,000 human proteins—against autoimmune or allergic sera--stratify.
  - Single molecule protein analyses—blood organ-specific proteins

Individual Patient Information-Based Assays of the Present/ Future (II)

- Single cells
  - Analyze 10,000 B cells and 10,000 T cells for the functional regions of their immune receptors—past and present immune responsiveness—follow vaccinations—identify autoimmune antibodies.
  - Analyze individual blood macrophages—inflammation, etc.
  - Use pore technology to separate epithelial cells from blood cells--cancer
- iPS (stem) cells
  - Analyze individual stem (iPS) cells from each individual differentiated to relevant tissues to get important phenotypic information—molecular, imaging and higher level phenotypic measurements.

Stratification of Complex Genetic Diseases—e.g. Alzheimer's

- Collect families of patients with the relevant disease (families will stratify disease to certain extent)
- Create iPS cells from each individual
- Differentiate these cells to neurons
- Probe the neurons with single cell analyses to identify the degree of heterogeneity
- Probe these neurons (individually or collectively) with ligands, drugs and relevant RNAi's
- Analyze their transcriptome, miRNAome and proteome responses
- Global comparisons across and within families of the molecular data—for stratification

A New Approach to Analyzing Genomic Variability in Cancer

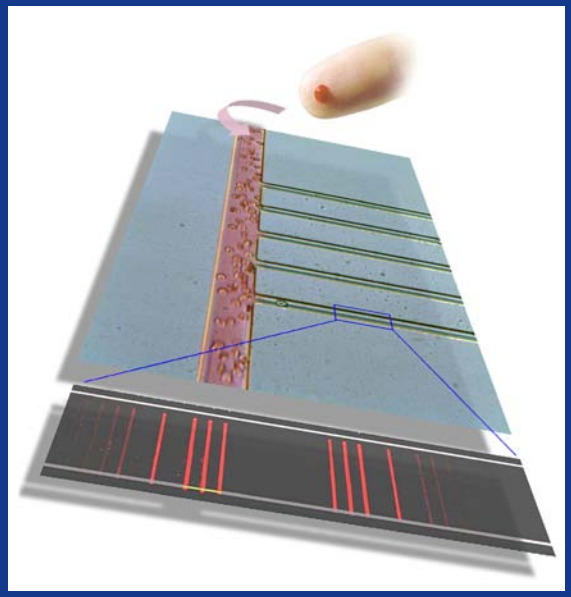
- Analyze at the single-cell level 1000 cells from each of 5 individual human glioblastomas—to quantize individual cell types
- Analyze the complete genome sequences of the family of one of these individuals with glioblastoma—to obtain high accurate family sequence data
- Cell sort the quantized cell populations from this individual's tumors using CD markers
- Analyze the complete genome sequences of the diffferent quantized cell populations from the individual's glioblastoma as well as their transcriptomes/miRNAs

New Approaches to Autoimmune Disease

- Sequence the 4 megabase MHC locus in families with autoimmune disease so that all MHC genes and their cis/trans relationships may be delineated.
- Use a protein chip with 13,000 human proteins to identify autoimmune antibodies in sera from autoimmune patients.
- Characterize 1000 individual T and 1000 individual B cells from the sera of autoimmune patients

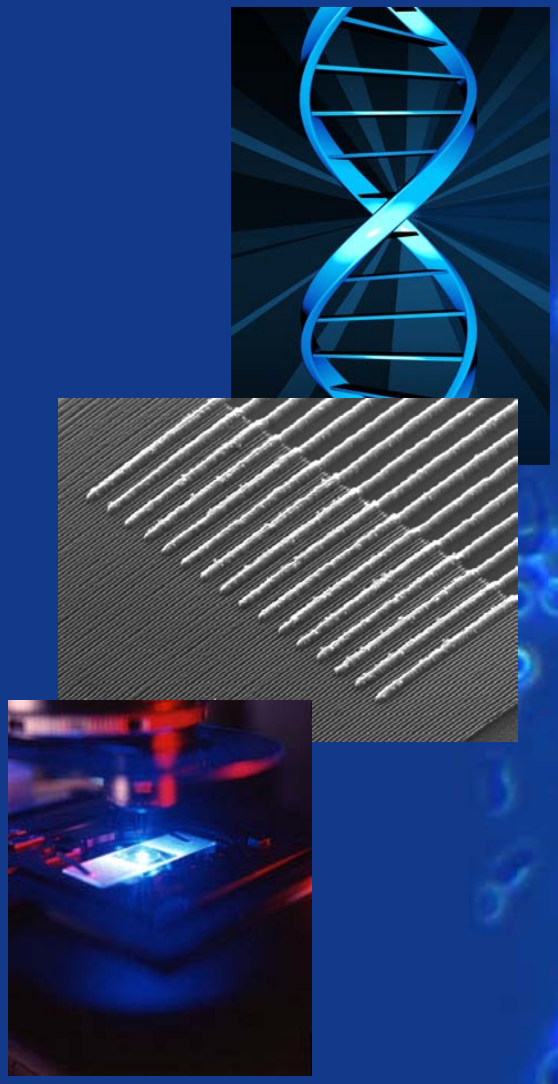
Predictive, Personalized, Preventive and Participatory (P4) Medicine

- Driven by systems approaches to disease, new measurement (nanotechnology) and visualization technologies and powerful new computational tools, P4 medicine will emerge over the next 10-20 years



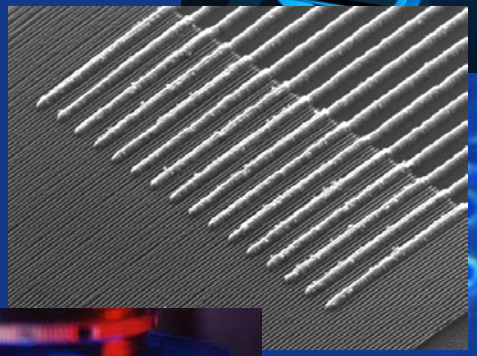
P4 Medicine

- Predictive:
  - Probabilistic health history--DNA sequence
  - Biannual multi-parameter blood protein measurements
  - In vivo molecular imaging



P4 Medicine

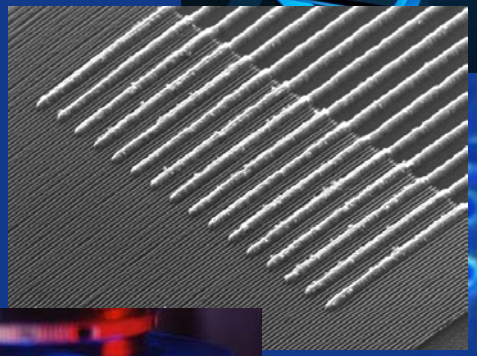
- Personalized:
  - Unique individual human genetic variation mandates individual treatment
  - Patient is his or her own control—longitudinal data
  - Billions of data points on each individual
  - 100s millions patients with billions data points



Institute for Systems Biology

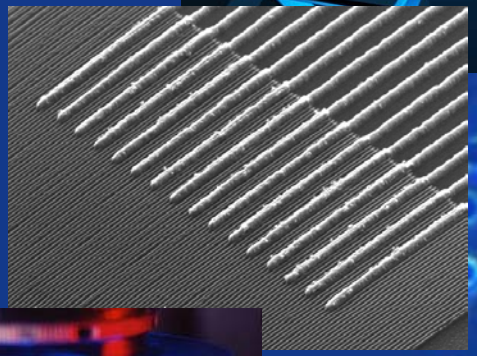
P4 Medicine

- Preventive:
  - Design of therapeutic and preventive drugs via systems approaches
  - Systems approaches to vaccines will transform prevention of infectious diseases
  - Transition to wellness assessment



P4 Medicine

- Participatory:
  - Patient understands and participates in medical choices
  - Physicians trained before P4 will have to understand it
  - Medical community—interconnected and educated
  - Create IT for healthcare to handle billions of patients, each with billions of data points



Inventing the Future

- Analyzing one gene and one small problem at a time

- Systems analysis of biology and medicine--e.g., predictive, preventive, personalized and participatory (P4) medicine
- Technology development
- Pioneer computational tools
- Transferring knowledge to society--joining academics and industry--changing K-12 science education--P4 medicine and society
- Strategic partnerships--for hard scientific problems--P4 medicine--industrial, academic, government, international

Institute for Systems Biology

Accelerating the Realization of P4 Medicine: ISB Strategic Partnerships

- ISB/Luxembourg—develop the strategies and tools for P4 medicine--attack two fundamental problems of P4 medicine--$100 million/5 years

- ISB/Ohio State University Medical School—P4 Medicine Institute—bring P4 medicine to patients—55,000 employee population where OSU is payer/provides—two pilot projects

The P4 Medicine Institute (*http://www.P4MI.org*)

- Non-profit 501c3
- ISB and Ohio State founding members
- Committed to bringing P4 medicine to patients—initially through two pilot projects—wellness and lung cancer
- Seeking academic and industrial partners who share the P4 vision and have complementary skills/resources
- Bringing on consultants to analyze the societal challenges of P4 medicine—ethics, security, confidentiality, policy, regulation, economics

P4 Medicine Is Personalized Medicine and Far More!

- P4 medicine is medicine of the present/near future.
- P4 medicine is driven by systems approaches to disease and emerging technologies
- P4 medicine will use measurements to quantify wellness and its transition into disease
- P4 medicine is revolutionary rather than evolutionary or incremental
- P4 medicine sees the patient (consumer) as the central focus of healthcare
- Pilot projects with informational assays in patient groups will be necessary to convince skeptics.
- P4 medicine will restructure the business plans of every sector of the healthcare industry—enormous economic opportunities
- P4 medicine will be effective, inexpensive and provide enormous economic benefits to economies—readily available to poor and rich.
- The national healthcare debate in the future should be reframed around P4 medicine rather than the old reactive medicine.

Acknowledgements

Prion--Institute for Systems Biology

Daehee Hwang
Inyoul Lee
Hyuntae Yoo
Eugene Yi (proteomics core facility)
Bruz Marzolf (Affymetrix core facility)

Nanotechnology—protein chips, protein-capture agents--Jim Heath, Caltech

MRM protein assays and Human Proteome—R Moritz, R Aebersold, Origene and Agilant

Single-cell analyses—Leslie Chen and Qiang Tian

Luxemburg Strategic Partnership—David Galas, Diane Isonaka, Rudi Balling (Lux)

Prion--McLaughlin Research Institute
Great Falls, Montana

Ranjit Giri
Douglas Spicer
Rajeev Kumar
Rose Pitstick
Rebecca Young
George A. Carlson

Family genome project—ISB/UW/Utah/Complete Genomics—David Galas

P4MI Institute—Fred Lee, Clay Marsh (OSU)

Single protein analysis—Chris Laustead

INSTITUTE FOR Systems Biology