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
Founded 2000—10th Anniversary

ISB has 13 faculty and 300 staff
ISB 1st in US and 3rd in World for Impact of Papers
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

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

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

No Clinical Sign

10 wks

PrPc Cleavage

Neuron

\( \alpha \)-cleavage

PrPc

\( \beta \)-cleavage

C2

ROS

Endocytosis

Caveolae

vesicle

1 Plasminogen

Complement Activation

Lipid acceptors

Signalizing

HDL

Liver

Microglia/ Astrocytes

CD68

CD84

CD14

CD4

Sphingolipid Metabolism

Androgen Metabolism

Cholesterol Homeostasis

Mitochondrion

GAG Metabolism

Lysosomal Proteases

2

3

6

Glycosylceramides

Galactosylceramides

Sphingomyelin
PrP accumulation and replication network—20 weeks
Sequential Disease-Perturbation of the Four Networks of Prion Disease

- Prion accumulation
- Glial Activation
- Synaptic Degeneration
- Neuronal Cell Death

Clinical Signs

0 wk
18~20 wk
22 wk

Cholesterol transport
Sphingolipid synthesis
Lysosome proteolysis

- Reactive Astrocytes
- Leukocyte extravasation
- Na\(^+\) channels
- Cargo transport
- Caspases

*Arachidonate metab./Ca\(^+\) sig.
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

Nerve cell death

18 wks

No Clinical Signs

2 wks

Clinical Signs

12 wks

20 wks
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-occurrences

Integrated Diagnostics—platform company for P4 medicine
Big New Projects at ISB

- The 2\textsuperscript{nd} 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
Samples
(DNA from blood lymphocytes)

Project flow

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

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

Deliverables
a) Reads
b) Coverage tables

[ ~20 T bytes of data ]
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
Recominational Genome Map from Miller’s Syndrome Children

65 crossovers in (2) male meioses (left)
104 crossovers in (2) female meioses (right)

- Both children inherited the same allele from both parents
- Each child inherited a different allele from each parent
- Children inherited the same allele from dad, different alleles from mom
- Children inherited the same allele from mom, different alleles from dad
**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 $8 \times 10^{-9}$ & $2.1 \times 10^{-8}$ per base per generation,

- The intergenerational mutation rate of the autosome is

  - $\sim 1.1 \times 10^{-8}$ /base/generation (~30% transversions)

*Science, 2010*
Simple recessive (SNPs)

Compound heterozygous (genes)

Disease Gene Candidates Reduced Analyzing Complete Family
DHODH  KIA0556
DNAH5  ZNF721

Miller’s gene

Ciliary dyskenesis gene
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
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

- TPPLC-MS/MS Data
- mZXML file format
- X! Tandem, SpectraST
- SEQUEST*, Mascot*
- pepXML file format
- XINTERACT
- XPRESS/AS
- APRatio
- Libra
- iProphet
- ProteinProphet
- protXML file format
- SBEAMS
- Gaggle
- Cytoscape
- PeptideAtlas

* Commercial software not part of TPP
TPP: Foundation for PeptideAtlas

Drives tool development and optimization

TransProteomic Pipeline

Advanced, uniform processing of all data
PeptideAtlas Datasets and Experiments

Number of Distinct Peptides

Cumulative Number of MS/MS spectra identified

NCI
Qian PNNL
HPPP I
Mallick
Novartis Microprot
Novartis Microprot2
Smireh PNNL
Glyco
Mallick
Paulovich FHCRC
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

Abundance (log₂ copies/cell)

41 copies/cell

1.3E6 copies/cell

<table>
<thead>
<tr>
<th>Protein</th>
<th>Weissman, copies/cell</th>
<th>Measured, copies/cell</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPO14</td>
<td>49</td>
<td>62</td>
</tr>
<tr>
<td>ENG1</td>
<td>64</td>
<td>198</td>
</tr>
<tr>
<td>ATC2 (Pep1)</td>
<td>98</td>
<td>89</td>
</tr>
<tr>
<td>ATC2 (Pep2)</td>
<td>98</td>
<td>140</td>
</tr>
<tr>
<td>SDS3</td>
<td>105</td>
<td>270</td>
</tr>
<tr>
<td>PRS7</td>
<td>125</td>
<td>2100</td>
</tr>
<tr>
<td>OSM1</td>
<td>432</td>
<td>3043</td>
</tr>
<tr>
<td>NFS1</td>
<td>504</td>
<td>479</td>
</tr>
<tr>
<td>MTNA</td>
<td>922</td>
<td>1127</td>
</tr>
<tr>
<td>CDC12</td>
<td>1169</td>
<td>1128</td>
</tr>
<tr>
<td>PDE1</td>
<td>1404</td>
<td>154</td>
</tr>
<tr>
<td>ATC6</td>
<td>1873</td>
<td>1063</td>
</tr>
<tr>
<td>ALDH2</td>
<td>2072</td>
<td>354</td>
</tr>
<tr>
<td>AROG</td>
<td>26272</td>
<td>5890</td>
</tr>
<tr>
<td>GLYC</td>
<td>67559</td>
<td>40178</td>
</tr>
<tr>
<td>6PGD1</td>
<td>101441</td>
<td>24988</td>
</tr>
<tr>
<td>EF3A</td>
<td>870578</td>
<td>51340</td>
</tr>
<tr>
<td>ALF (Pep1)</td>
<td>1018216</td>
<td>712630</td>
</tr>
<tr>
<td>ALF (Pep2)</td>
<td>1018216</td>
<td>749513</td>
</tr>
</tbody>
</table>
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
DEAL for *In vitro* molecular diagnostics: *Integrated nanotech/microfluidics platform*

Jim Heath, et al

300 nanoliters of plasma

5 minute measurement

Jim Heath, et al
Peptide Protein-Capture Agents

Jim Heath--Caltech
Antibody Displacement Technology—Heath--Caltech

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

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

20th Century Biomedicine • Analyzing one gene and one small problem at a time

ISB

21st Century Biomedicine

• 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
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
- Dahee 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