"Towards Digitally Enabled Genomic Medicine"

Distinguished Lecture Series

Department of Computer Science and Engineering

UC San Diego

October 15, 2012

Dr. Larry Smarr

Director, California Institute for Telecommunications and Information Technology

Harry E. Gruber Professor,

Dept. of Computer Science and Engineering

Jacobs School of Engineering, UCSD

Abstract

Calit2 has, for over a decade, had a driving vision that healthcare is being transformed into “digitally enabled genomic medicine.” The global market for cell phones is driving down the cost of components needed for sensing many aspects of our body. Combined with advances in nanotechnology and MEMS, a new generation of body sensors is rapidly developing. As these real-time data streams are stored in the cloud, cross population comparisons becomes increasingly possible and the availability of biofeedback leads to behavior change toward wellness. To put a more personal face on the "patient of the future," I have been increasingly quantifying my own body over the last ten years. In addition to external markers I also currently track over 100 molecular and blood cell types in my blood and dozens of molecular and microbial variables in my stool. Through saliva I have obtained 1 million single nucleotide polymorphisms (SNPs) in my human DNA. My gut microbiome has been metagenomically sequenced, yielding 25 billion DNA bases. I will show how one can discover emerging disease states before they develop serious symptoms by graphing time series of these key variables and also will illustrate the power of multi-variant analysis across all these internal variables. Imagining a software system that can handle millions to billions of data points per person across billions of people leads to new challenges in computer science and engineering.

Calit2 Has Been Had a Vision of "the Digital Transformation of Health" for a Decade

- Next Step—Putting You On-Line!
  - Wireless Internet Transmission
  - Key Metabolic and Physical Variables
  - Model -- Dozens of Processors and 60 Sensors / Actuators Inside of our Cars

www.bodymedia.com



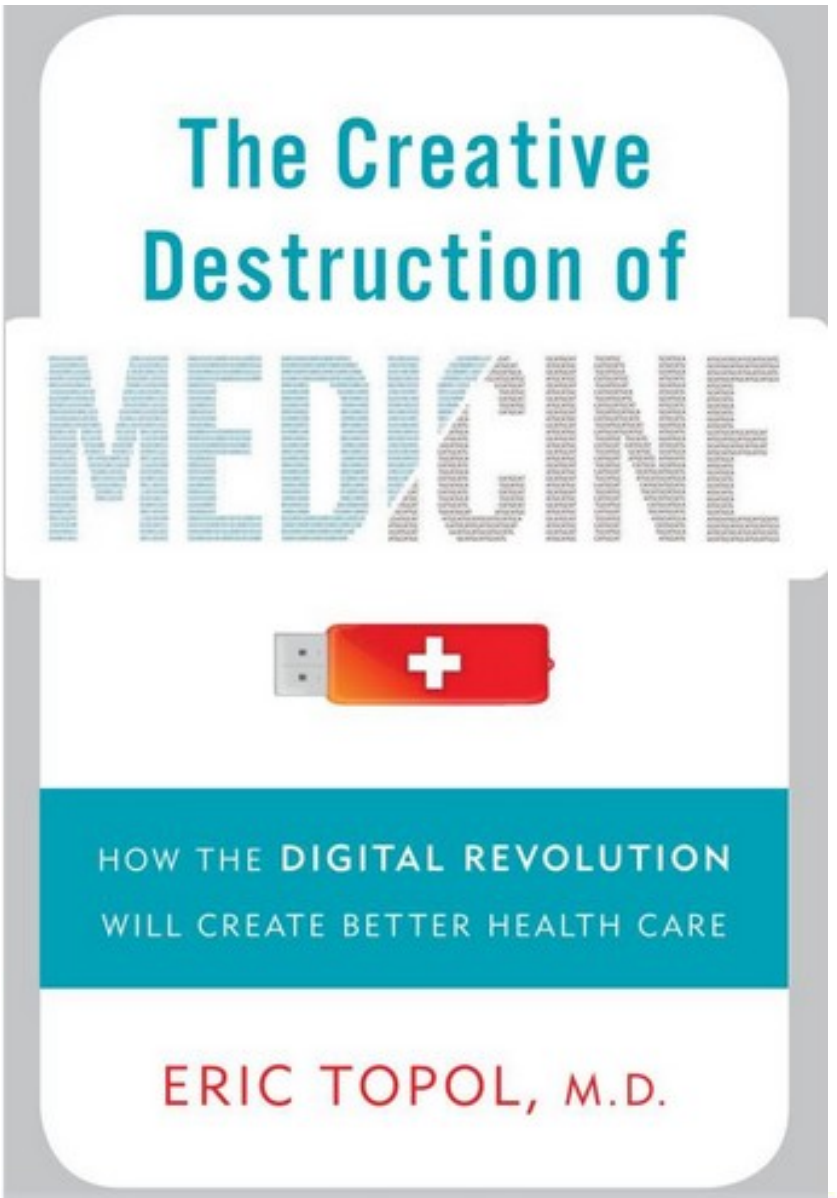
- Post-Genomic Individualized Medicine
  - Combine
    - Genetic Code
    - Body Data Flow
  - Use Powerful AI Data Mining Techniques

The Content of This Slide from 2001 Larry Smarr Calit2 Talk on Digitally Enabled Genomic Medicine

The Calit2 Vision of Digitally Enabled Genomic Medicine is an Emerging Reality



July/August 2011



February 2012

I Arrived in La Jolla in 2000 After 20 Years in the Midwest and Decided to Move Against the Obesity Trend

Age 51



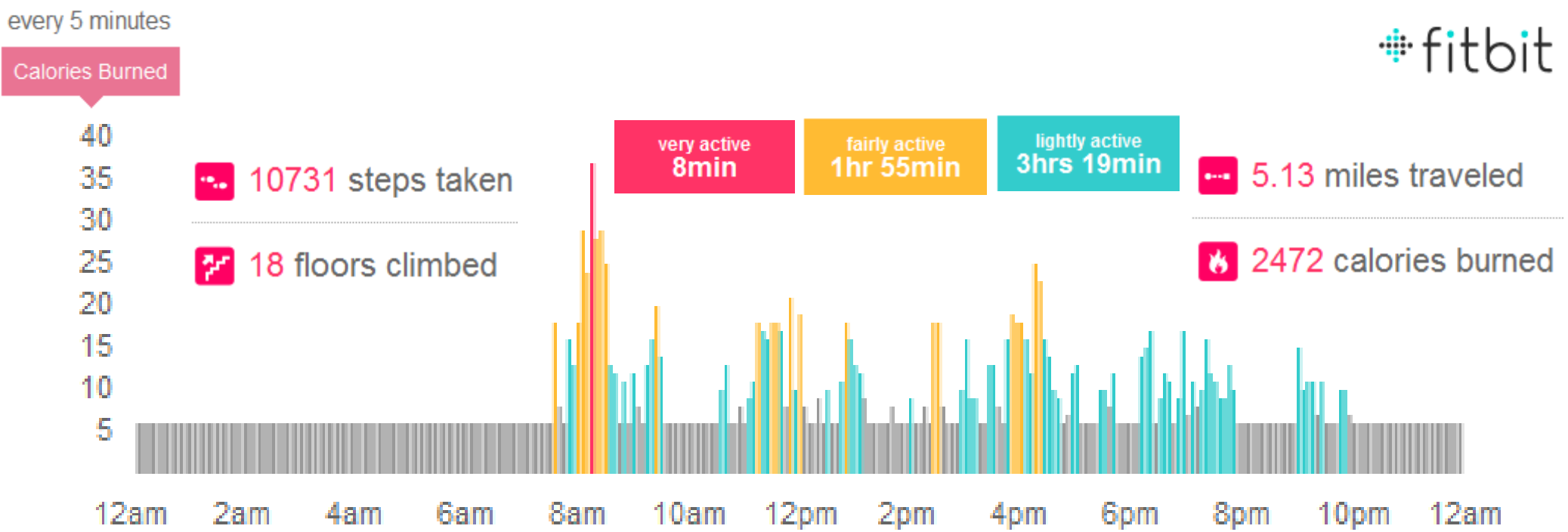


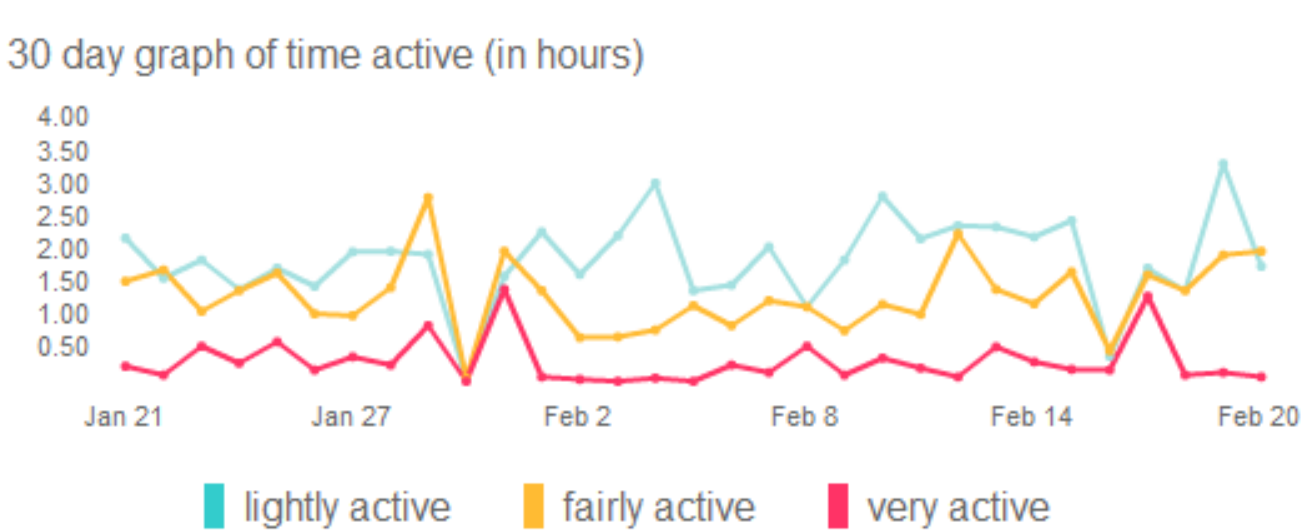
Age 61

I Reversed My Body's Decline By Altering My Nutrition and Exercise

See the full story at:
http://lsmarr.calit2.net/repository/092811_Special_Letter,_Smarr.final.pdf

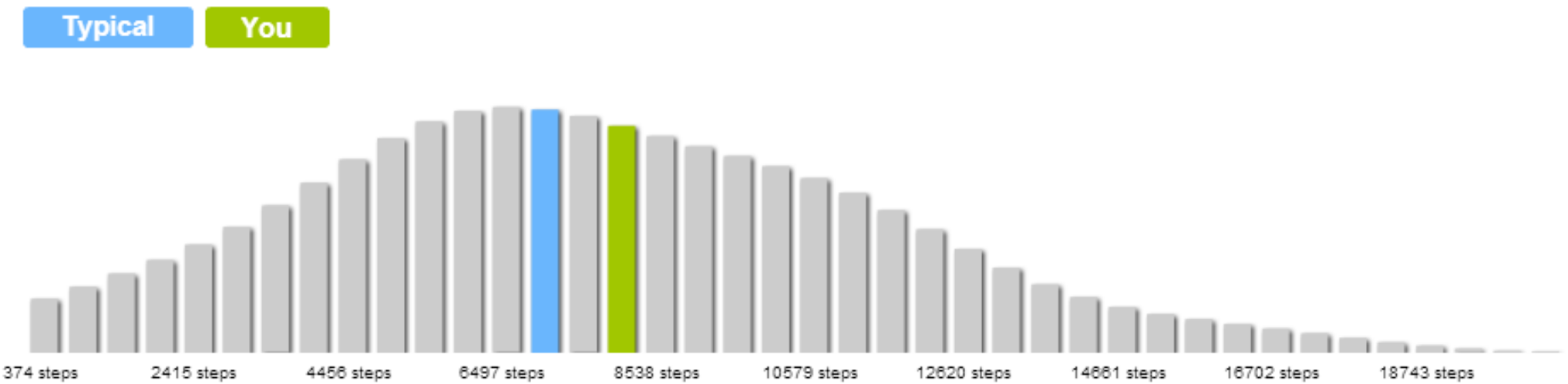
Wireless Monitoring
Helps Drive Exercise Goals







FitBit Compares Your Steps to Population of Your Age and Sex



This week you walked

8028 steps

15% more than the median for men 55 to 64 yrs who are normal or underweight

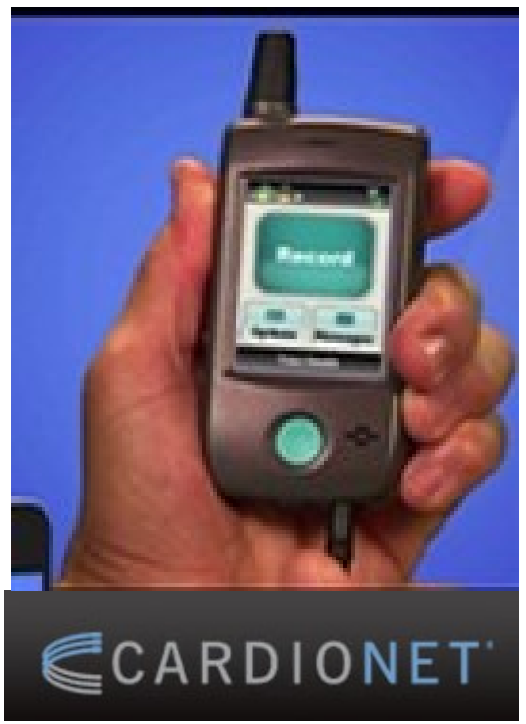
You are in the

60 percentile

of all men 55 to 64 yrs who are normal or underweight

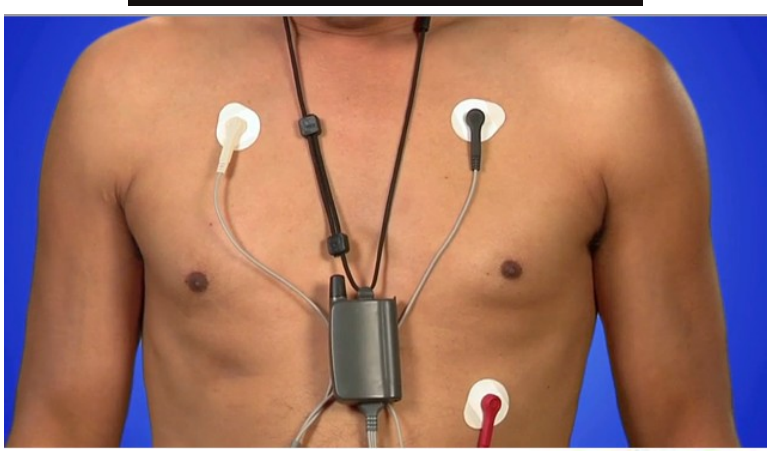
Calit2 is Using Several Heart Rate Wireless Monitors to Analyze Heart Rate Variability



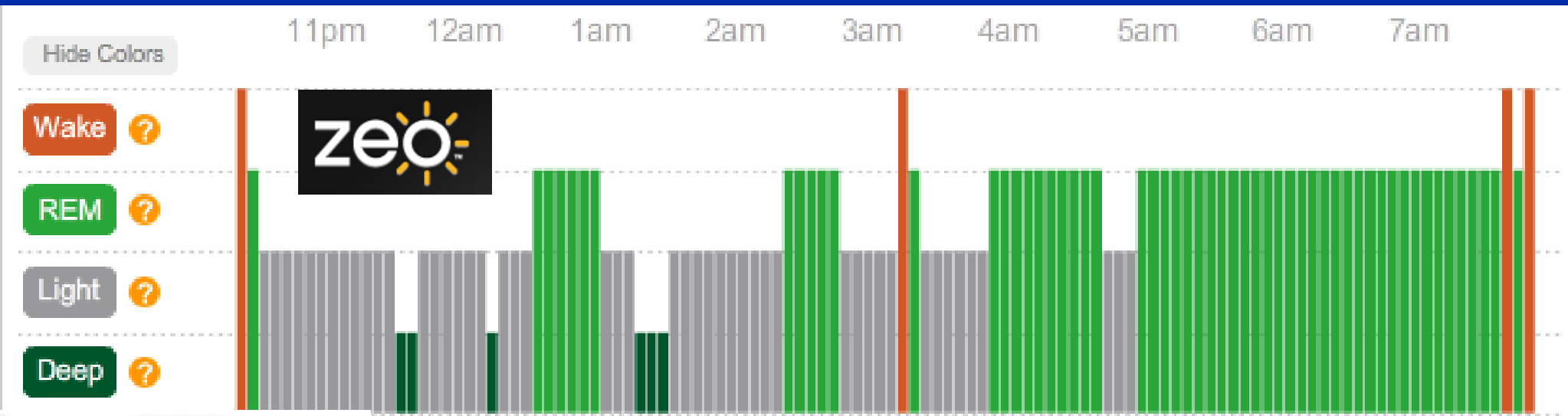


Zephyr technology





Quantifying My Sleep Pattern Using a Zeo - Surprisingly About Half My Sleep is REM!

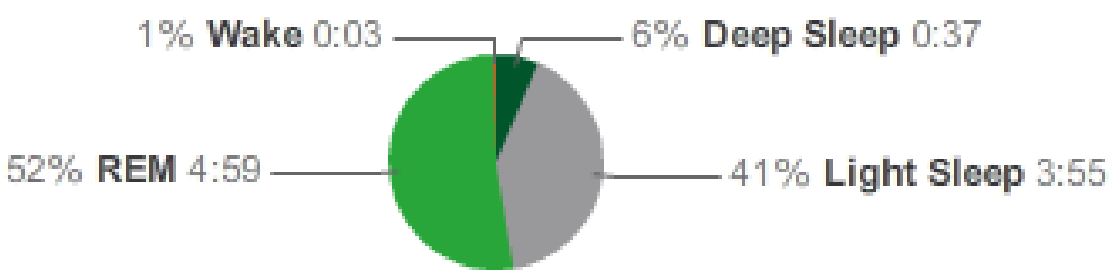




Zeo has database of ~10,000 users, over 200,000 nights

Total Z: 9 hr 31 min

Time to Z: 0 hr 2 min



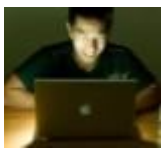
60 Year Old Male REM is Normally 20% of Sleep

Mine is Between 45-65% of Sleep

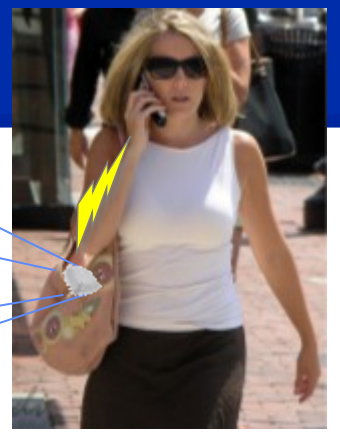
CitiSense –UCSD NSF Grant for Fine-Grained Environmental Sensing Using Cell Phones



4oz
30 compounds







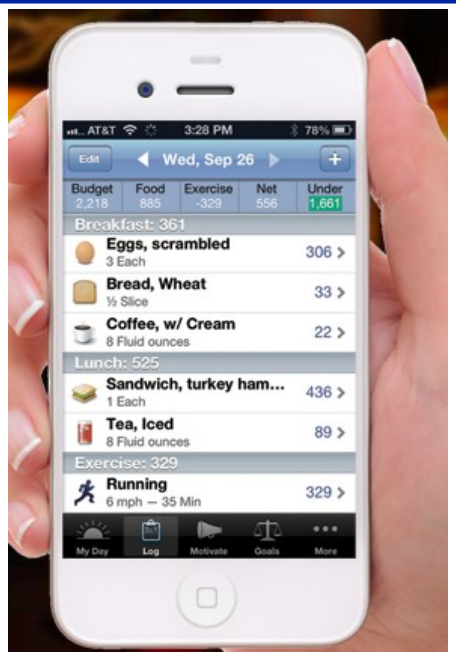






CitiSense Team
PI: Bill Griswold
Ingolf Krueger
Tajana Simunic Rosing
Sanjoy Dasgupta
Hovav Shacham
Kevin Patrick

Challenge-Develop Standards to Enable MashUps of Personal Sensor Data Across Private Clouds



Lose It-
Calories Ingested



Body Media-
Calories Burned



Withing/iPhone-
Blood Pressure





EM Wave PC-
Stress



Azumio-Heart Rate



Zeo-Sleep

From Measuring Macro-Variables to Measuring Your Internal Variables

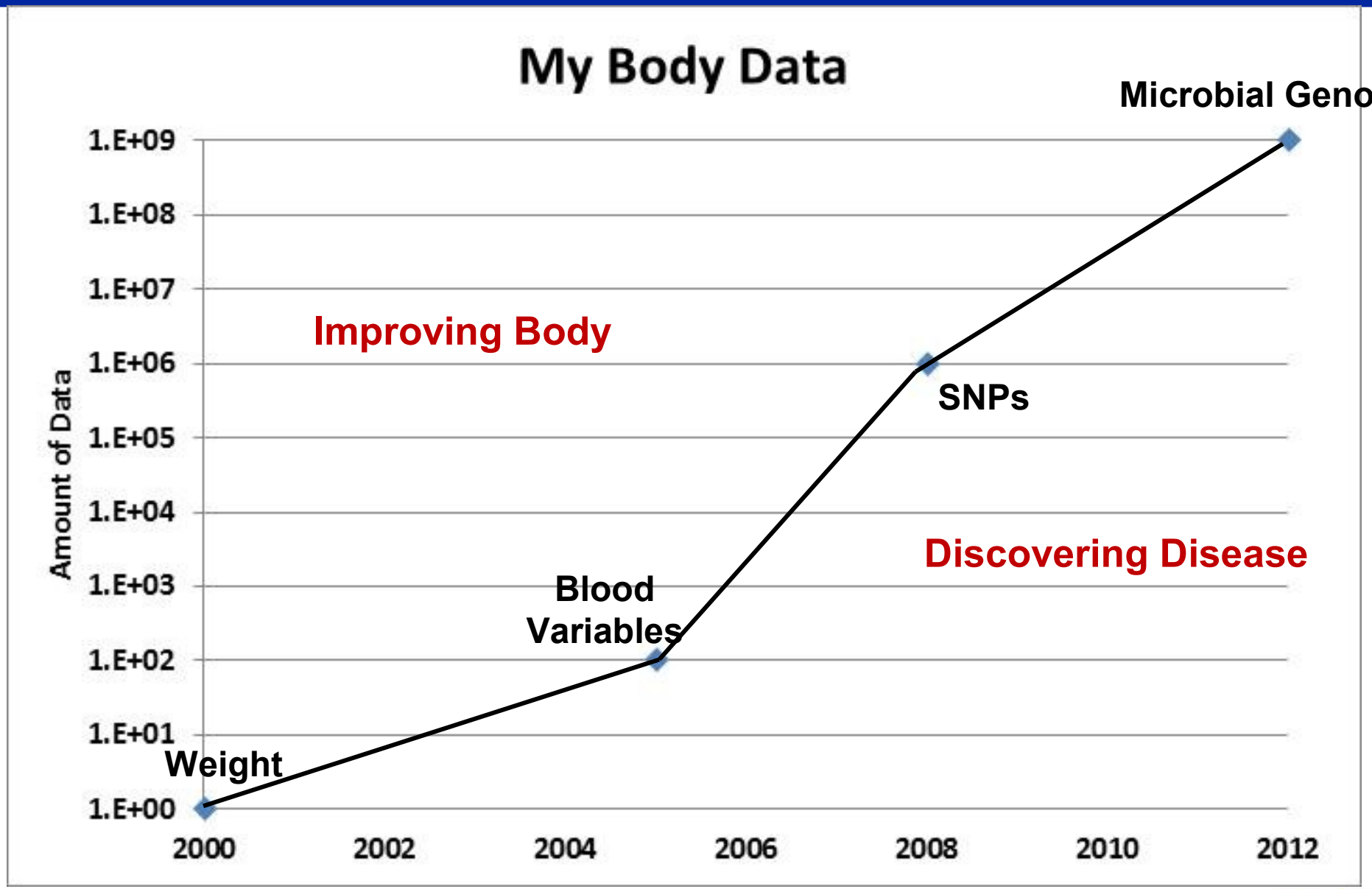


The Patient of the Future

Internet pioneer Larry Smarr's quest to quantify everything about his health led him to a startling discovery, an unusual partnership with his doctor, and more control over his life.

www.technologyreview.com/biomedicine/39636

Challenge: Creating a Population-Wide Software System: From One to Billions of Data Points Defining Me

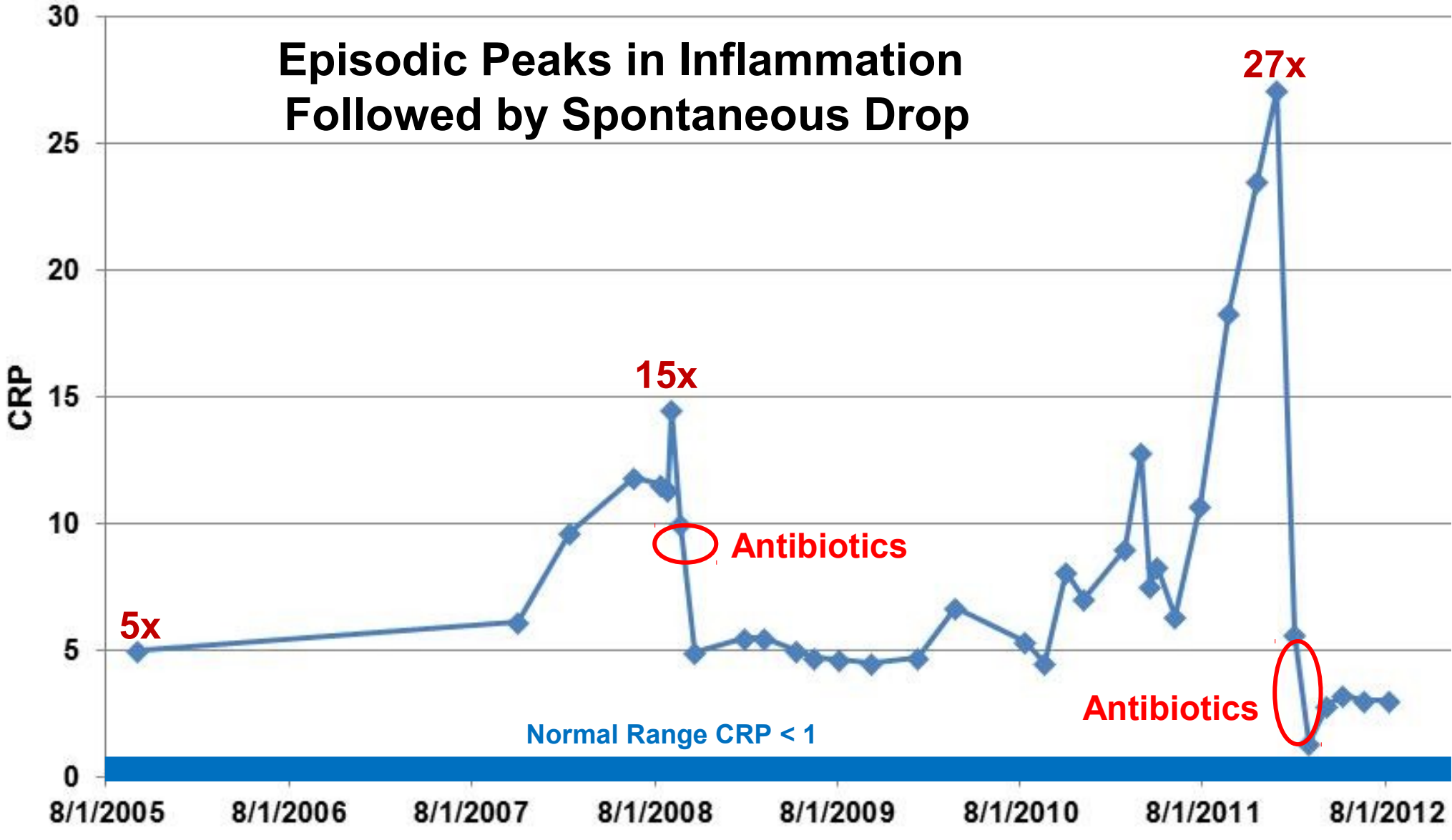


I Track 100 Variables in Blood Tests With Blood Samples Taken Monthly to Annually

- **Electrolytes**
  - Sodium, Potassium, Calcium, Magnesium, Phosphorus, Boron, Chlorine, CO_2
- **Micronutrients**
  - Arsenic, Chromium, Cobalt, Copper, Iron, Manganese, Molybdenum, Selenium, Zinc
- **Blood Sugar Cycle**
  - Glucose, Insulin, A1C Hemoglobin
- **Cardio Risk**
  - Complex Reactive Protein
  - Homocysteine
- **Kidneys**
  - Bun, Creatinine, Uric Acid
- **Protein**
  - Total Protein, Albumin, Globulin
- **Liver**
  - GGTP, SGOT, SGPT, LDH, Total Direct Bilirubin, Alkaline Phosphatase
- **Thyroid**
  - T3 Uptake, T4, Free Thyroxine Index, FT4, 2nd Gen TSH
- **Blood Cells**
  - Complete Blood Cell Count
  - Red Blood Cell Subtypes
  - White Blood Cell Subtypes
- **Cancer Screen**
  - CEA, Total PSA, % Free PSA
  - CA-19-9
- **Vitamins & Antioxidant Screen**
  - Vit D, E; Selenium, ALA, coQ10, Glutathione, Total Antioxidant Fn.

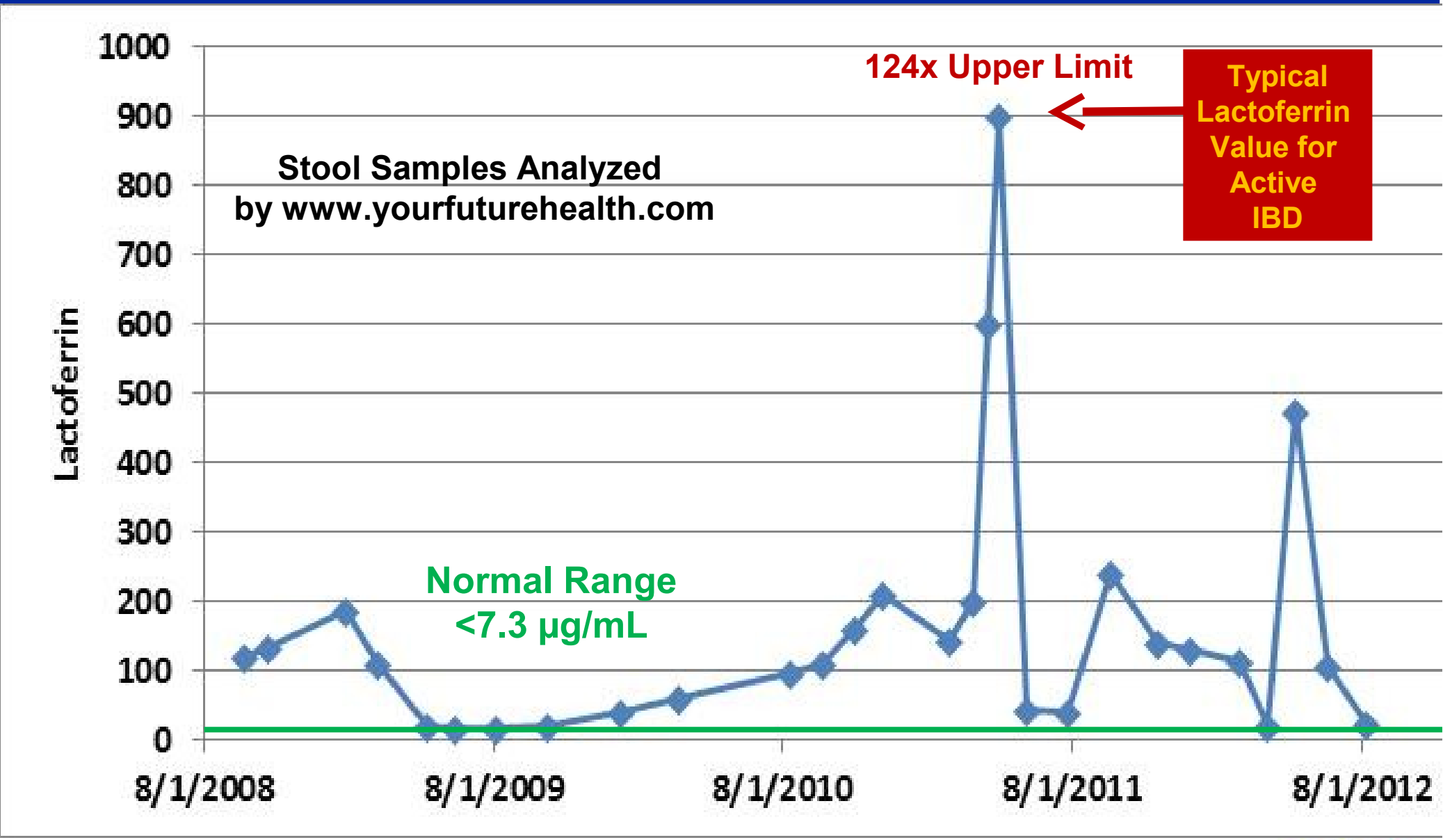
Only One of These Was Far Out of Normal Range

My Blood Measurements Revealed Chronic Inflammation



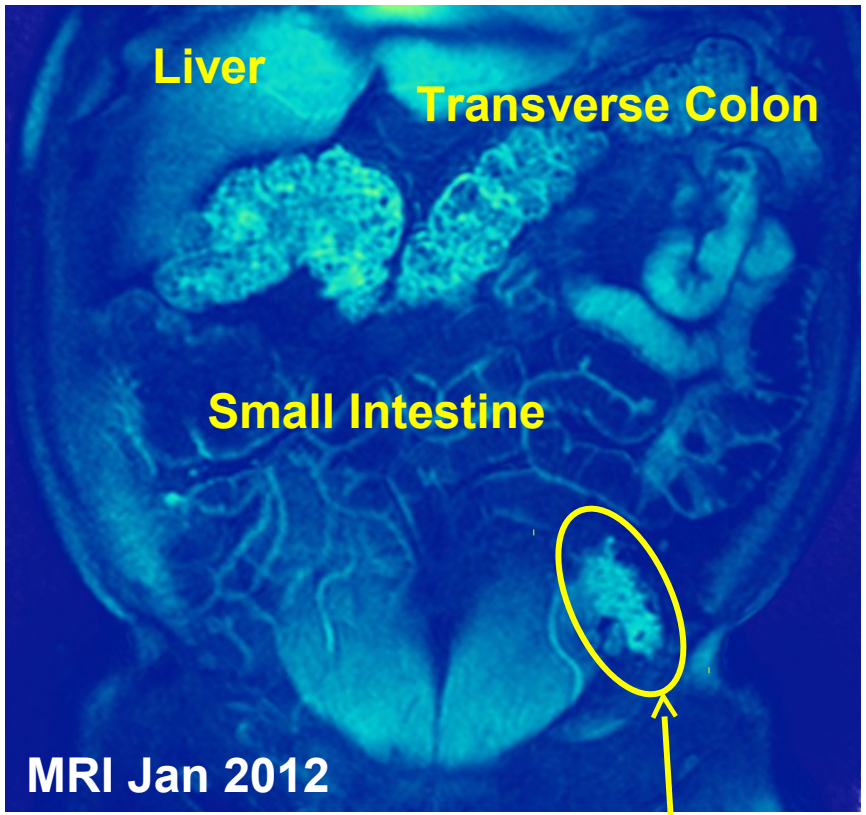
Complex Reactive Protein (CRP) is a Blood Biomarker for Detecting Presence of Inflammation

By Quantifying Stool Measurements Over Time I Discovered Source of Inflammation Was Likely in Colon



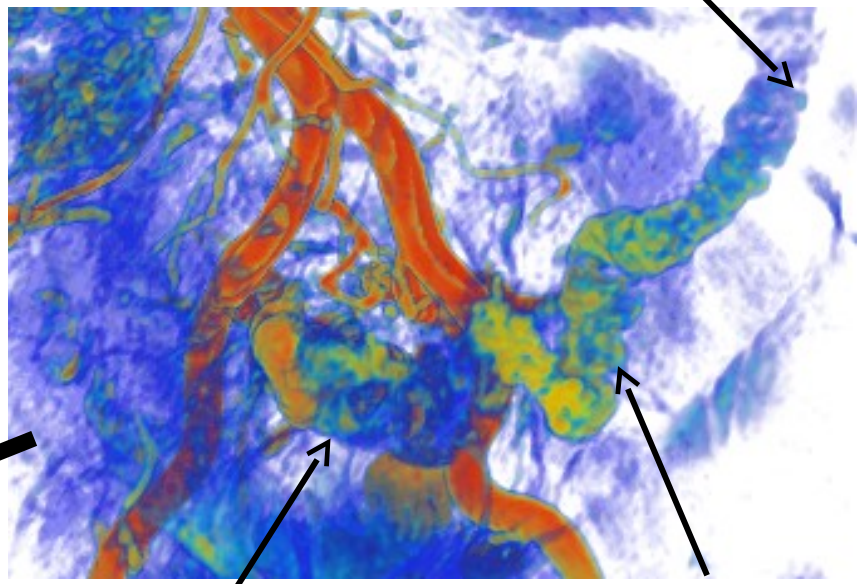
Lactoferrin is a Sensitive and Specific Biomarker for Detecting Presence of Inflammatory Bowel Disease (IBD)

Confirming the IBD (Crohn's) Hypothesis: Finding the "Smoking Gun" with MRI Imaging



I Obtained the MRI Slices From UCSD Medical Services and Converted to Interactive 3D Working With Jurgen Schulze's DeskVOX Software

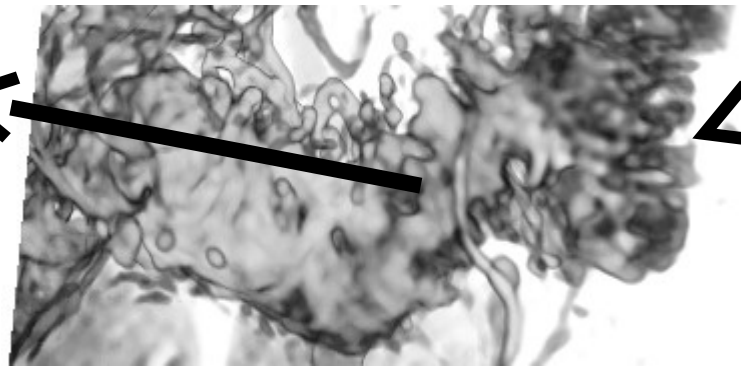
Descending Colon



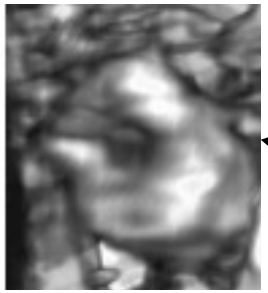
Major Kink

Sigmoid Colon
Threading Iliac Arteries

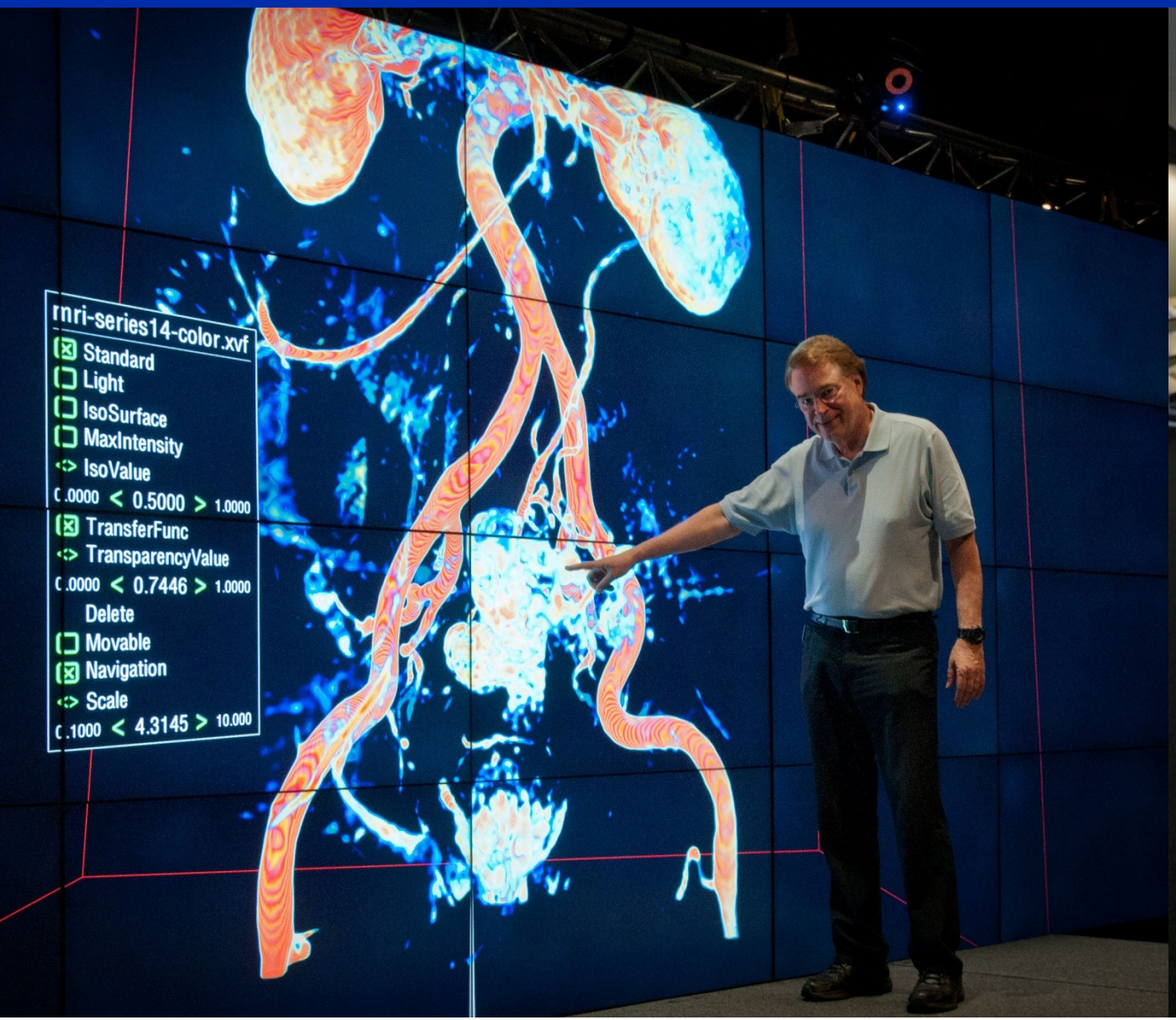
Diseased Sigmoid Colon



Cross Section



Interactive Visualization and 3D Hard Copy from LS MRI Data





Research: Calit2 FutureHealth Team

Challenge: Is it Possible for Software to Intercompare Digital Human Bodies?

- **Videos of Me Giving Tours of My Insides:**
  - **http://www.youtube.com/watch?v=9c4DtJ_L_Ps**
  - **www.theatlantic.com/magazine/archive/2012/07/the-measured-man/309018/**



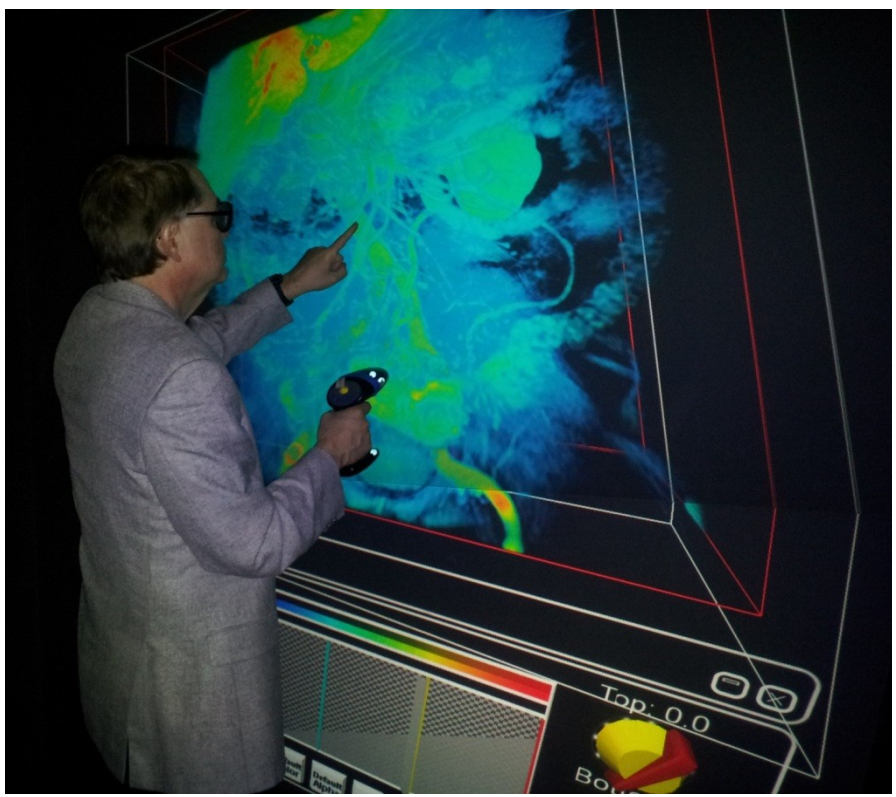


Photo & DeskVOX Software Courtesy of Jurgen Schulze, Calit2

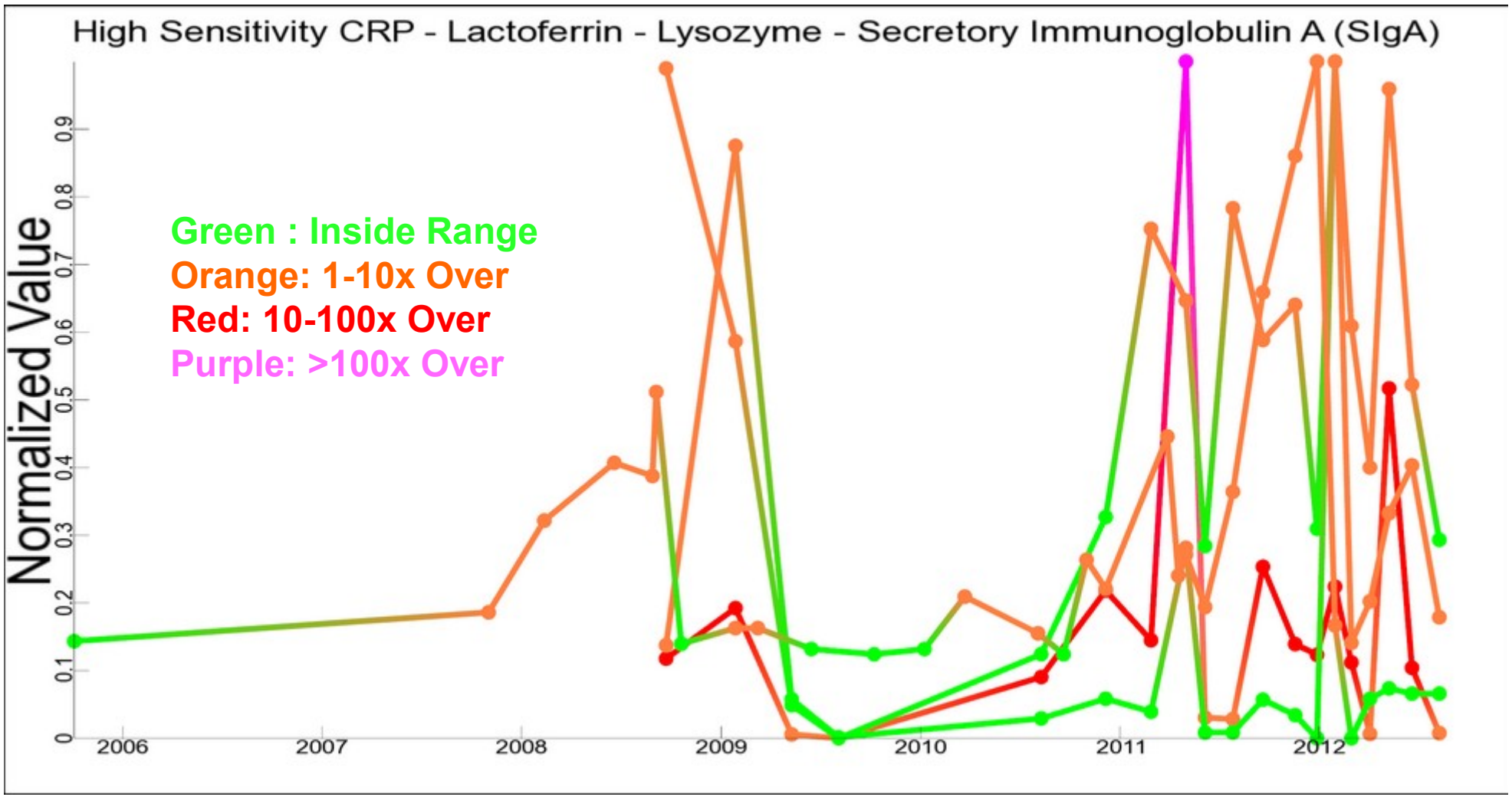
Why Did I Have an Autoimmune Disease like IBD?

Despite decades of research,
the etiology of Crohn's disease
remains unknown.
Its pathogenesis may involve
a complex interplay between
host genetics,
immune dysfunction,
and microbial or environmental factors.
--*The Role of Microbes in Crohn's Disease*

So I Set Out to Quantify All Three!

Paul B. Eckburg & David A. Relman
Clin Infect Dis. 44:256-262 (2007)

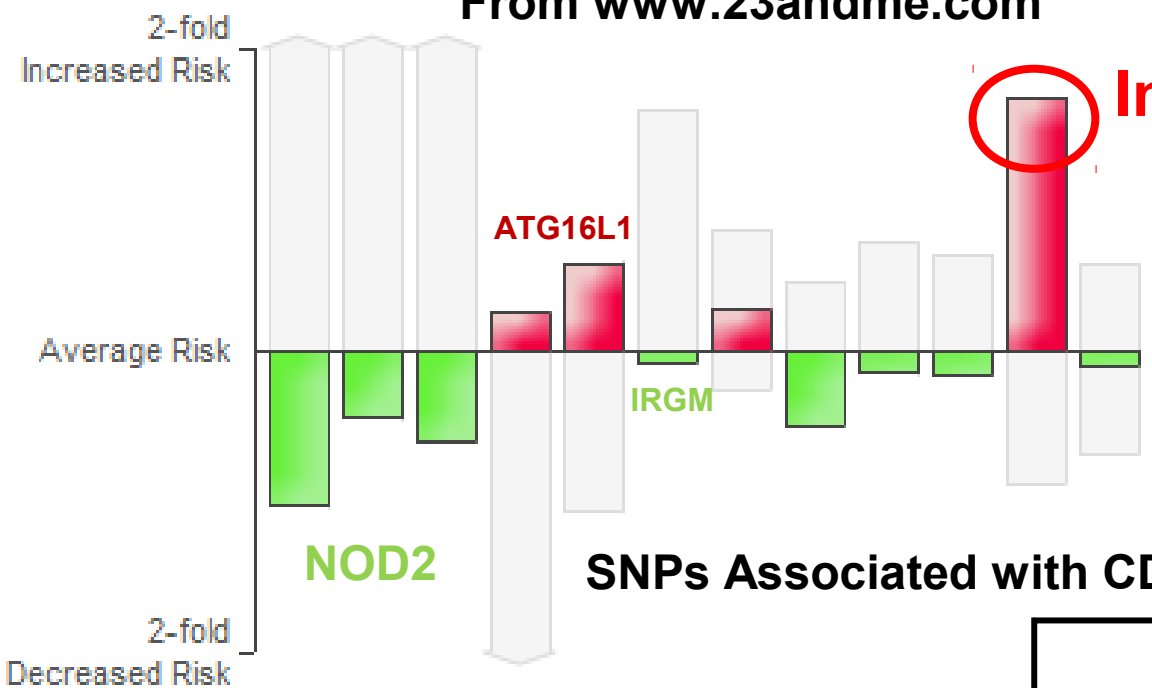
Putting Multiple Immunological Biomarker Time Series Together, Reveals Major Immune Dysfunction



Source: Calit2 Future Health Expedition Team

I Wondered if Crohn's is an Autoimmune Disease, Did I Have a Personal Genomic Polymorphism?

From www.23andme.com



SNPs Associated with CD

Polymorphism in Interleukin-23 Receptor Gene — 80% Higher Risk of Pro-inflammatory Immune Response

~ 1 Million Single Nucleotide Polymorphisms (SNPs) Make Up About 90% of All Human Genetic Variation

News and Views

Nature Medicine **13**, 26 - 28 (2007)
doi:10.1038/nm0107-26

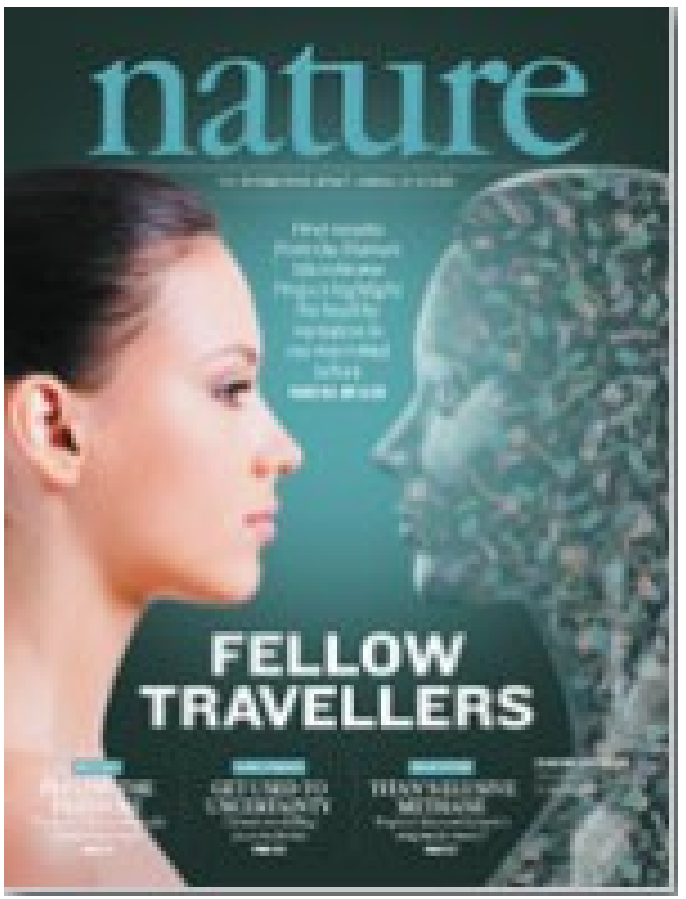
IL-23: a master regulator in Crohn disease

Markus F Neurath[1]

Intense Scientific Research is Underway on Understanding the Human Microbiome



June 8, 2012



June 14, 2012

Determining My Gut Microbes and Their Time Variation



Shipped Stool Sample
December 28, 2011

I Received
a Disk Drive April 3, 2012
With 35 GB FASTQ Files

Weizhong Li, UCSD
NGS Pipeline:
230M Reads
Only 0.2% Human

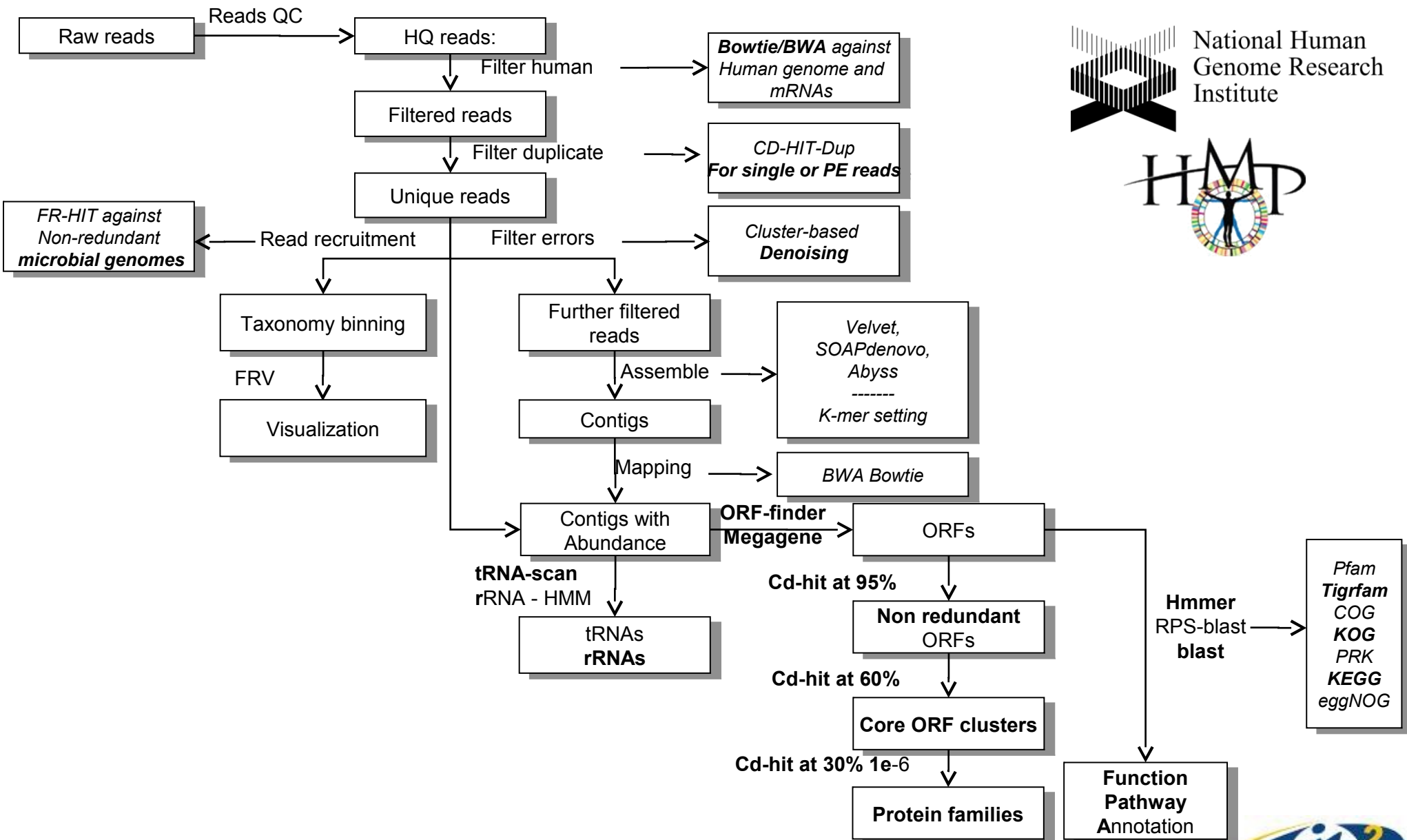
Required 1/2 cpu-yr
Per Person Analyzed!

UC San Diego
HEALTH SCIENCES

J. Craig Venter
INSTITUTE

it²

We Used Weizhong Li Group's Metagenomic Computational NextGen Sequencing Pipeline



National Human Genome Research Institute

HMP

PI: (Weizhong Li, UCSD):
NIH R01HG005978 (2010-2013, $1.1M)

We Used SDSC's Gordon Data-Intensive Supercomputer to Analyze JCVI Sequences of LS Gut Microbiome

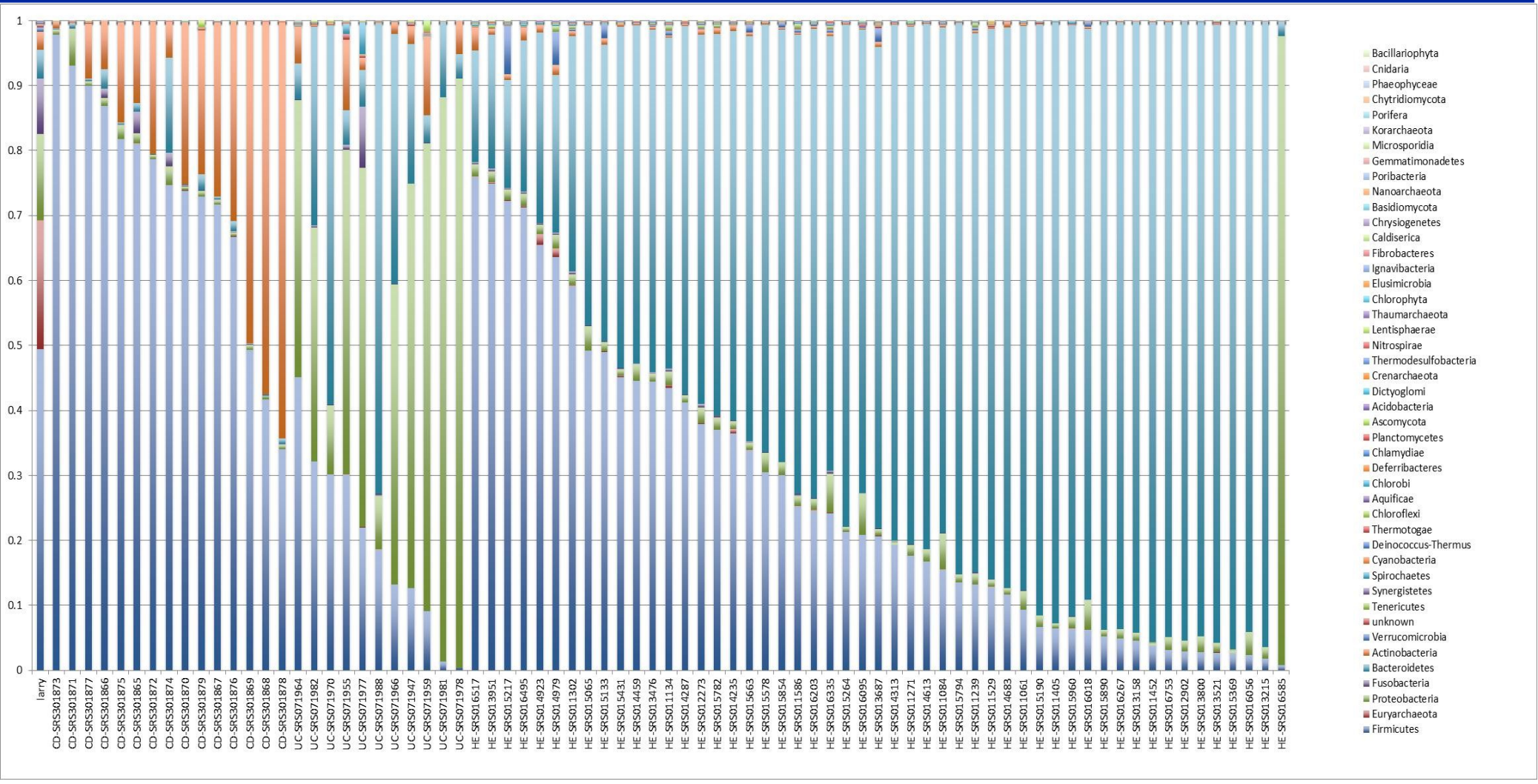
- **Analyzed Healthy and IBD Patients:**
  - LS, 13 Crohn's Disease & 11 Ulcerative Colitis Patients, + 150 HMP Healthy Subjects
- **Gordon Compute Time**
  - ~1/2 CPU-Year Per Sample
  - > 200,000 CPU-Hours so far
- **Gordon RAM Required**
  - 64GB RAM for Most Steps
  - 192GB RAM for Assembly
- **Gordon Disk Required**
  - 8TB for All Subjects
  - Input, Intermediate and Final Results

Venter Sequencing of LS Gut Microbiome:
230 M Reads
101 Bases Per Read
23 Billion DNA Bases

Enabled by a Grant of Time on Gordon from SDSC Director Mike Norman

SDSC

Metagenomic Sequencing of Gut Bacteria: Phyla Distribution Detects Different IBD Types



LS **Crohn's** **Ulcerative Colitis** **Healthy**

J. Craig Venter INSTITUTE

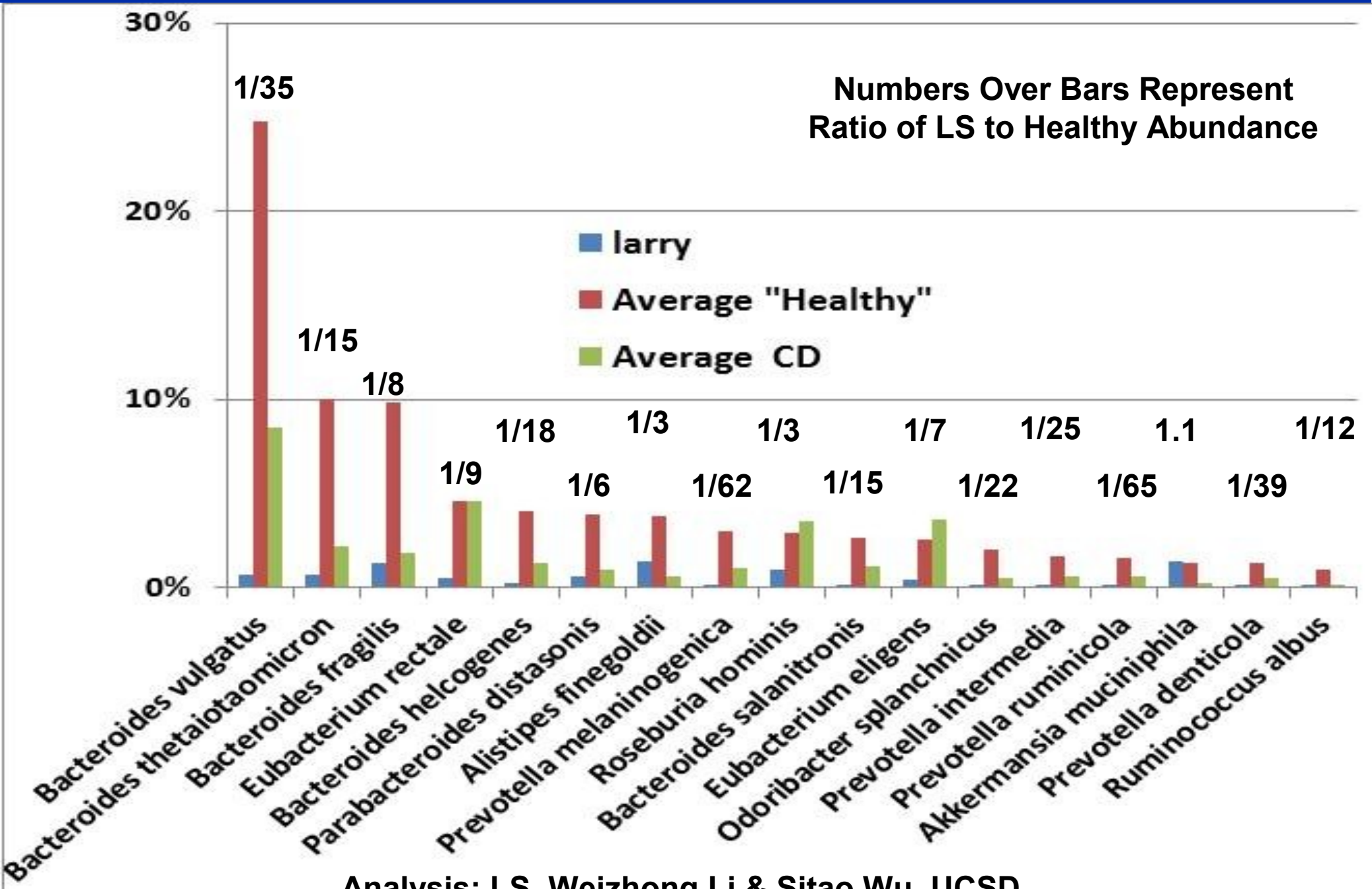
UC San Diego Health Sciences

SDSC

it²

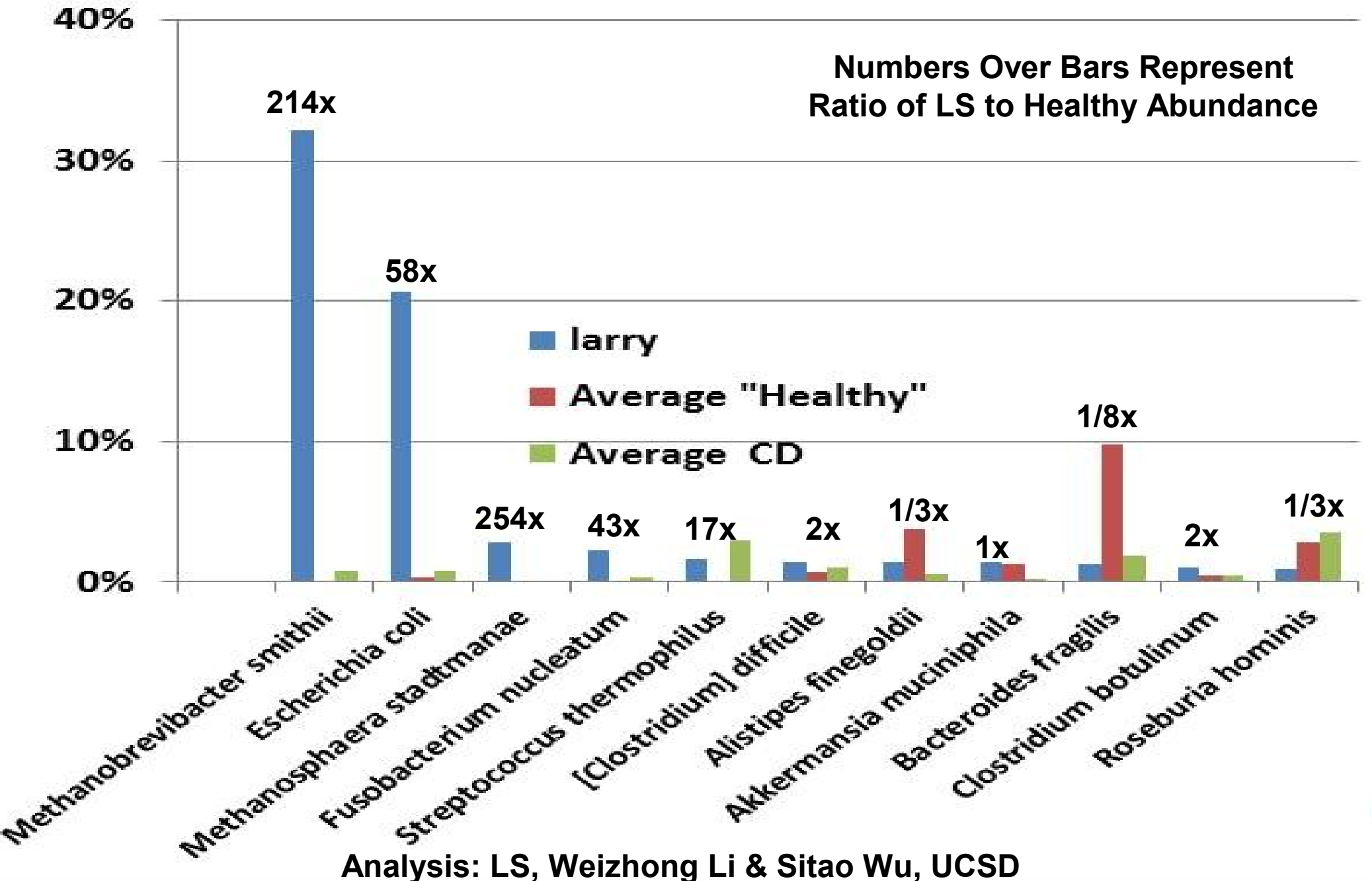
Analysis: Weizhong Li & Sitao Wu, UCSD

Almost All Abundant Species (≥1%) in Healthy Subjects Are *Severely Depleted* in LS Gut



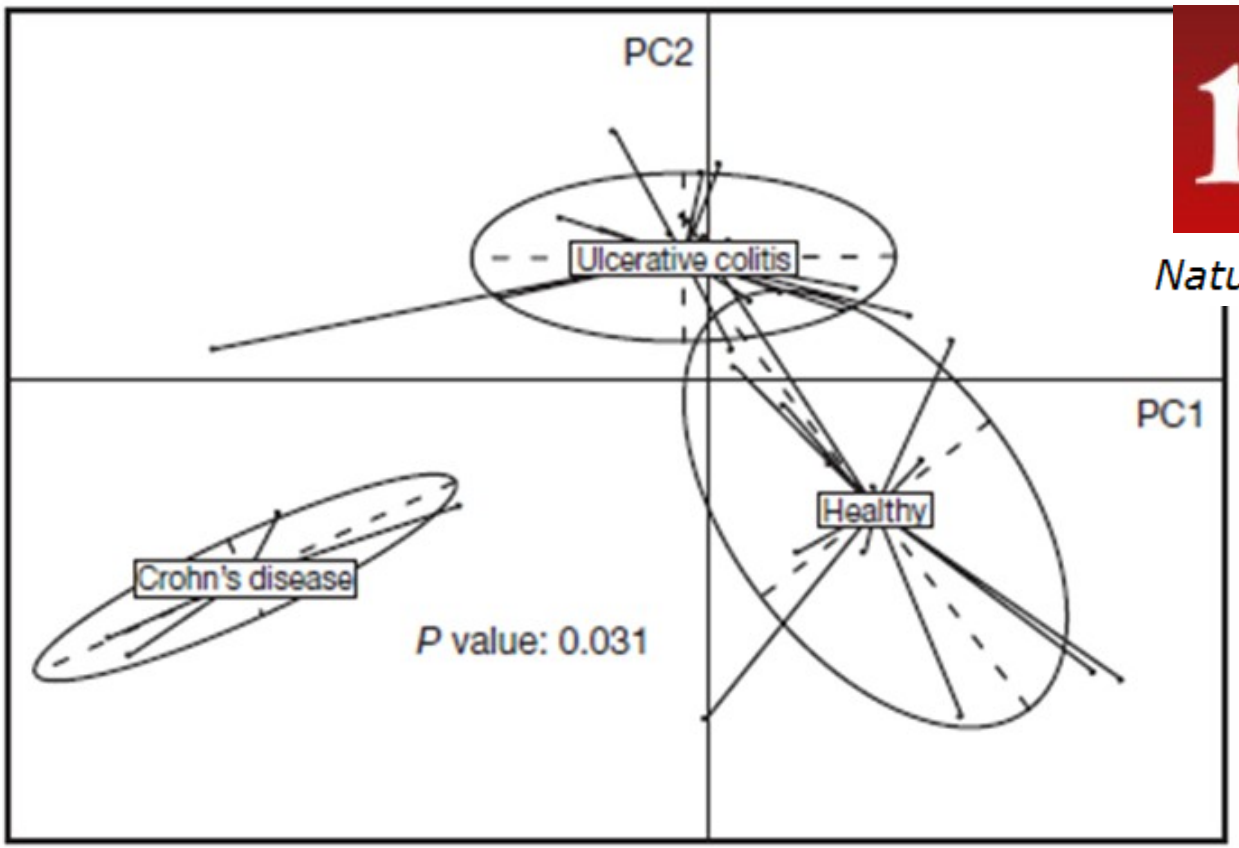
Analysis: LS, Weizhong Li & Sitao Wu, UCSD

LS Abundant Microbe Species (≥1%) Are *Dominated by Rare Species* in Healthy Subjects



Analysis: LS, Weizhong Li & Sitao Wu, UCSD

Microbial Metagenomics Can Diagnose Disease States



nature

Nature **464**, 59-65 (4 March 2010)

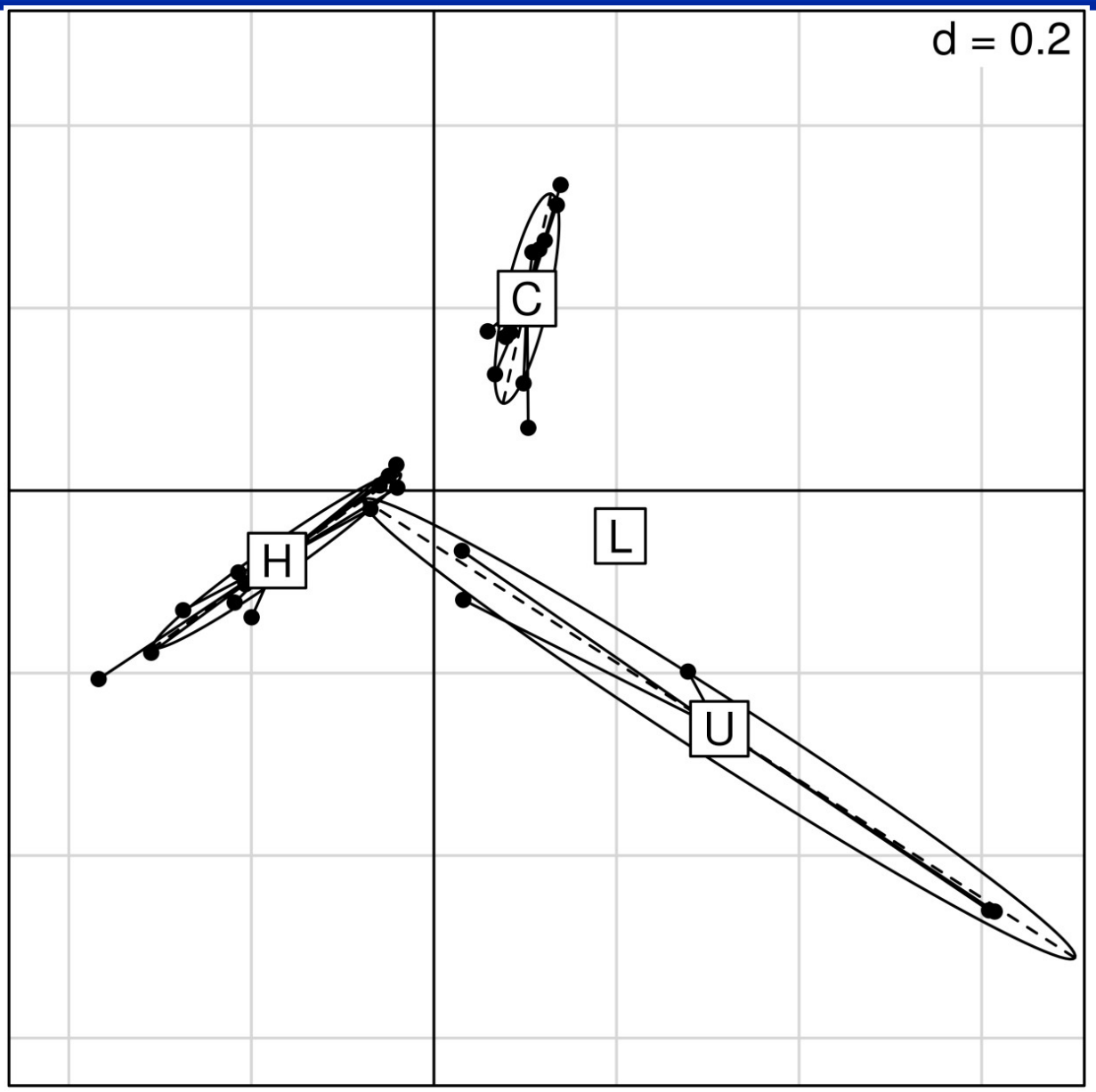
IBD Patients Harbored, on Average, *25% Fewer Microbial Genes* than the Individuals Not Suffering from IBD.

Figure 4 | Bacterial species abundance differentiates IBD patients and healthy individuals. Principal component analysis with health status as

A human gut microbial gene catalogue established by metagenomic sequencing

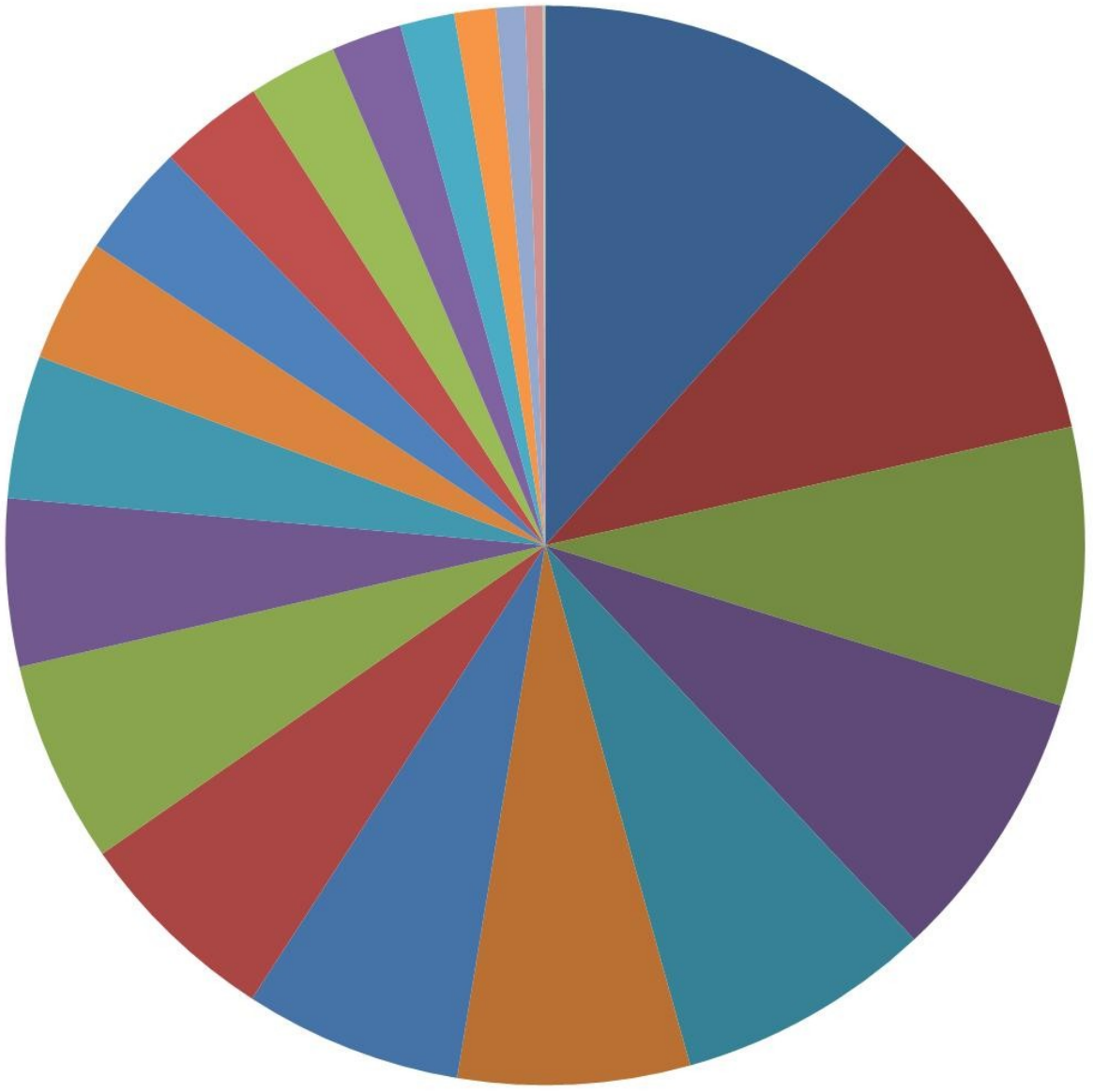
Junjie Qin[1]*, Ruiqiang Li[1]*, Jeroen Raes[2,3], Manimozhiyan Arumugam[2], Kristoffer Solvsten Burgdorf[4],

Our Principal Component Analysis Based On Microbial Species Abundance



Analysis: Weizhong Li & Sitao Wu, UCSD

Analysis of Clusters of Orthologous Groups (COGs) - Gene Family Distribution in LS Gut Microbiome



- General function prediction only
- Function unknown
- Translation, ribosomal structure and biogenesis
- Replication, recombination and repair
- Amino acid transport and metabolism
- Transcription
- Carbohydrate transport and metabolism
- Cell wall/membrane/envelope biogenesis
- Energy production and conversion
- Inorganic ion transport and metabolism
- Coenzyme transport and metabolism
- Nucleotide transport and metabolism
- Posttranslational modification, protein turnover, chaperones
- Signal transduction mechanisms
- Defense mechanisms
- Lipid transport and metabolism
- Intracellular trafficking, secretion, and vesicular transport
- Cell cycle control, cell division, chromosome partitioning
- Secondary metabolites biosynthesis, transport and catabolism
- Cell motility
- Chromatin structure and dynamics
- RNA processing and modification
- Cytoskeleton
- Extracellular structures

J. Craig Venter INSTITUTE

UC San Diego Health Sciences

SDSC

it2

Analysis: Weizhong Li & Sitao Wu, UCSD

Where I Believe We are Headed: Predictive, Personalized, Preventive, & Participatory Medicine

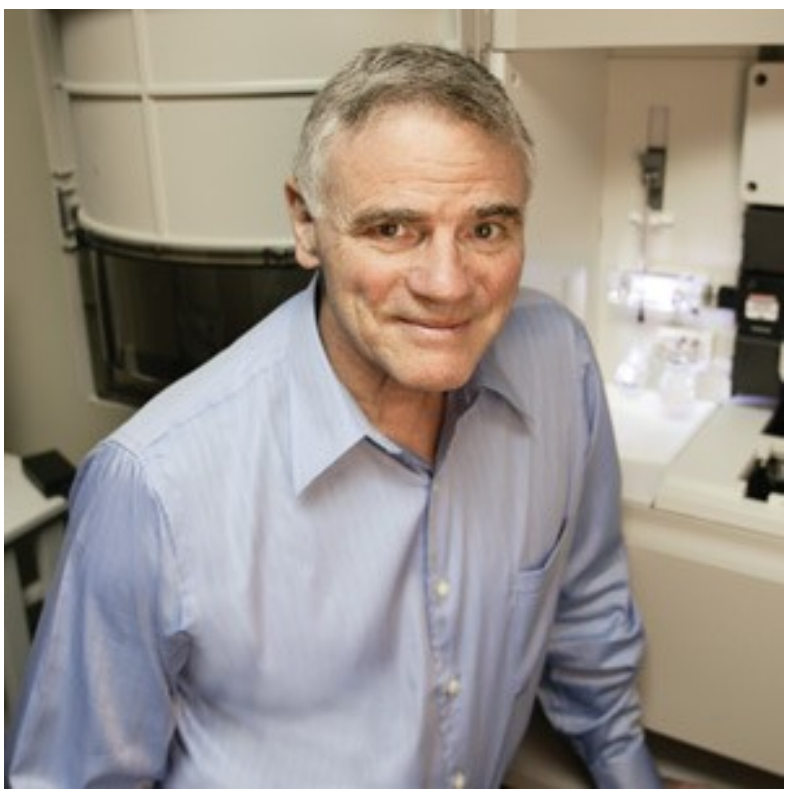
Newsweek

I am Leroy Hood's Lab Rat!

A Doctor's Vision of the Future of Medicine

Leroy Hood
NEWSWEEK

From the magazine issue dated Jul 13, 2009



Using a "LifeChip"
Quantify ~2500 Blood Proteins,
50 Each from 50 Organs or Cell Types
from a Single Drop of Blood
To Create a Time Series

www.newsweek.com/2009/06/26/a-doctor-s-vision-of-the-future-of-medicine.html

Invited Paper for Focus Issue of Biotechnology Journal, Edited by Profs. Leroy Hood and Charles Auffray.

Perspective

Quantifying your body: A how-to guide from a systems biology perspective

Larry Smarr

University of California, San Diego, CA, USA

Download Pdfs from my Portal:

http://lsmarr.calit2.net/repository/Biotech_J._LS_published_article.pdf

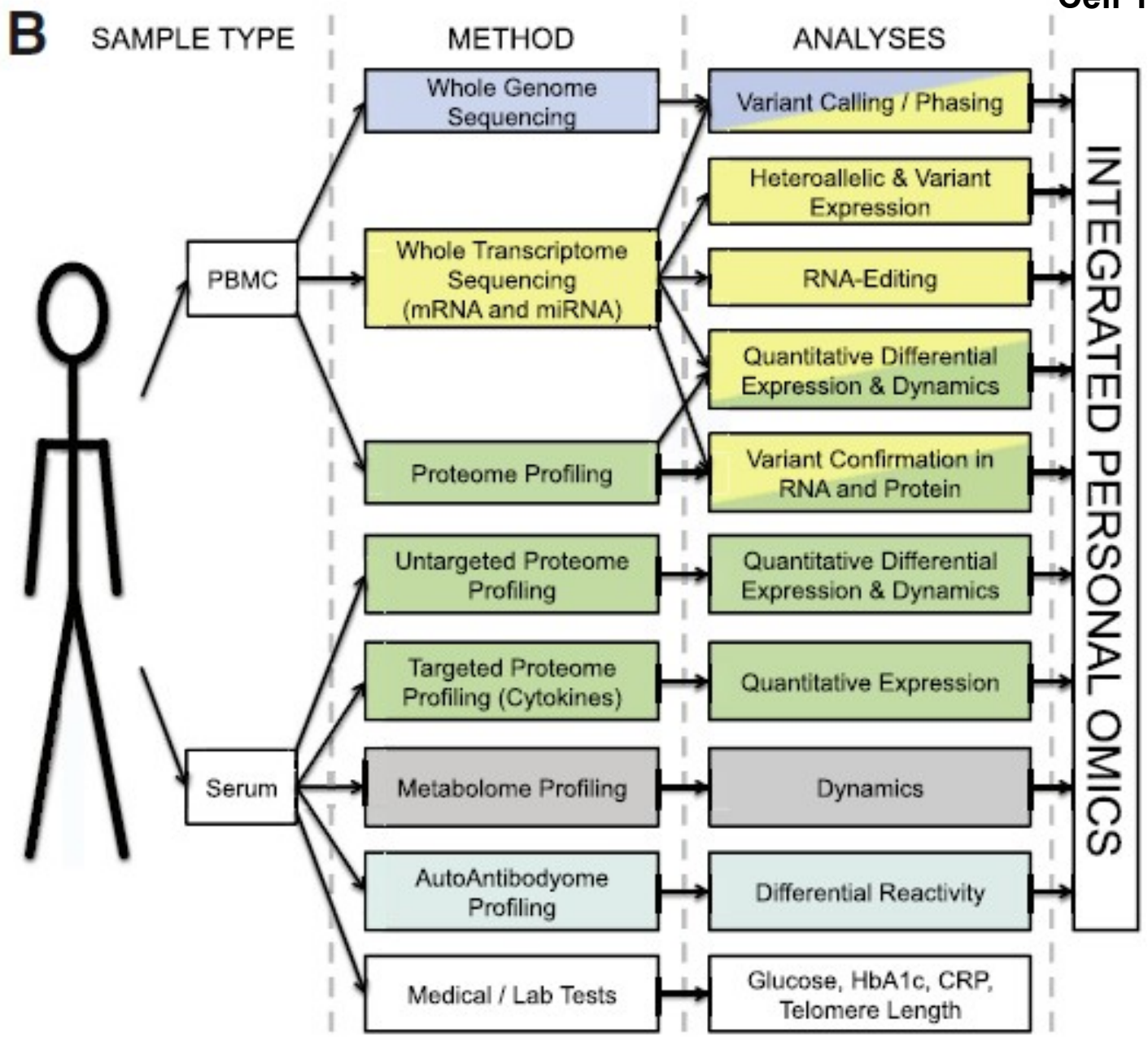
http://lsmarr.calit2.net/repository/Biotech_J._Supporting_Info_published.pdf

Biotechnol. J. 2012, 7, 980–991



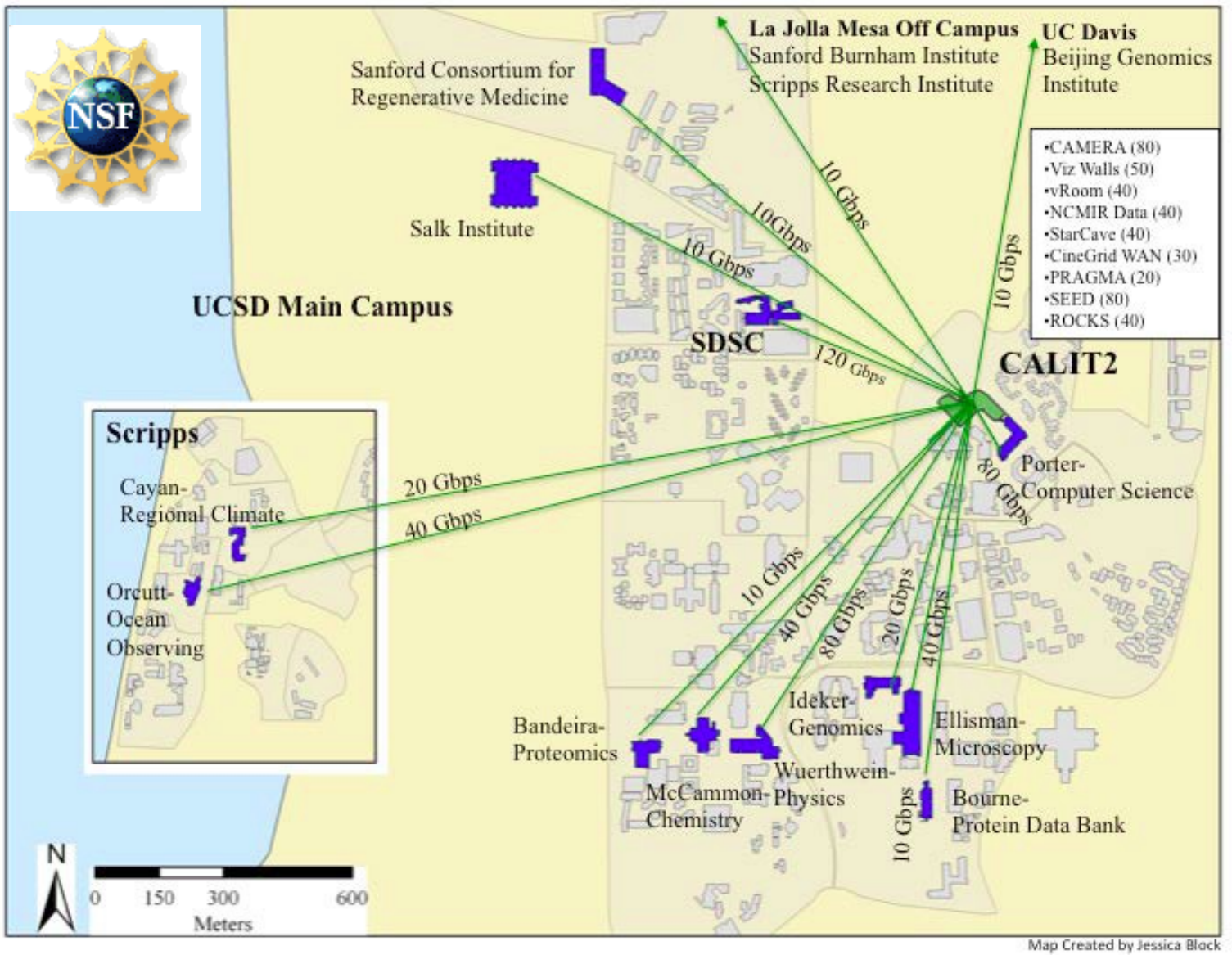
Integrative Personal Omics Profiling: 1000x the Data I Have Taken

Cell 148, 1293–1307, March 16, 2012



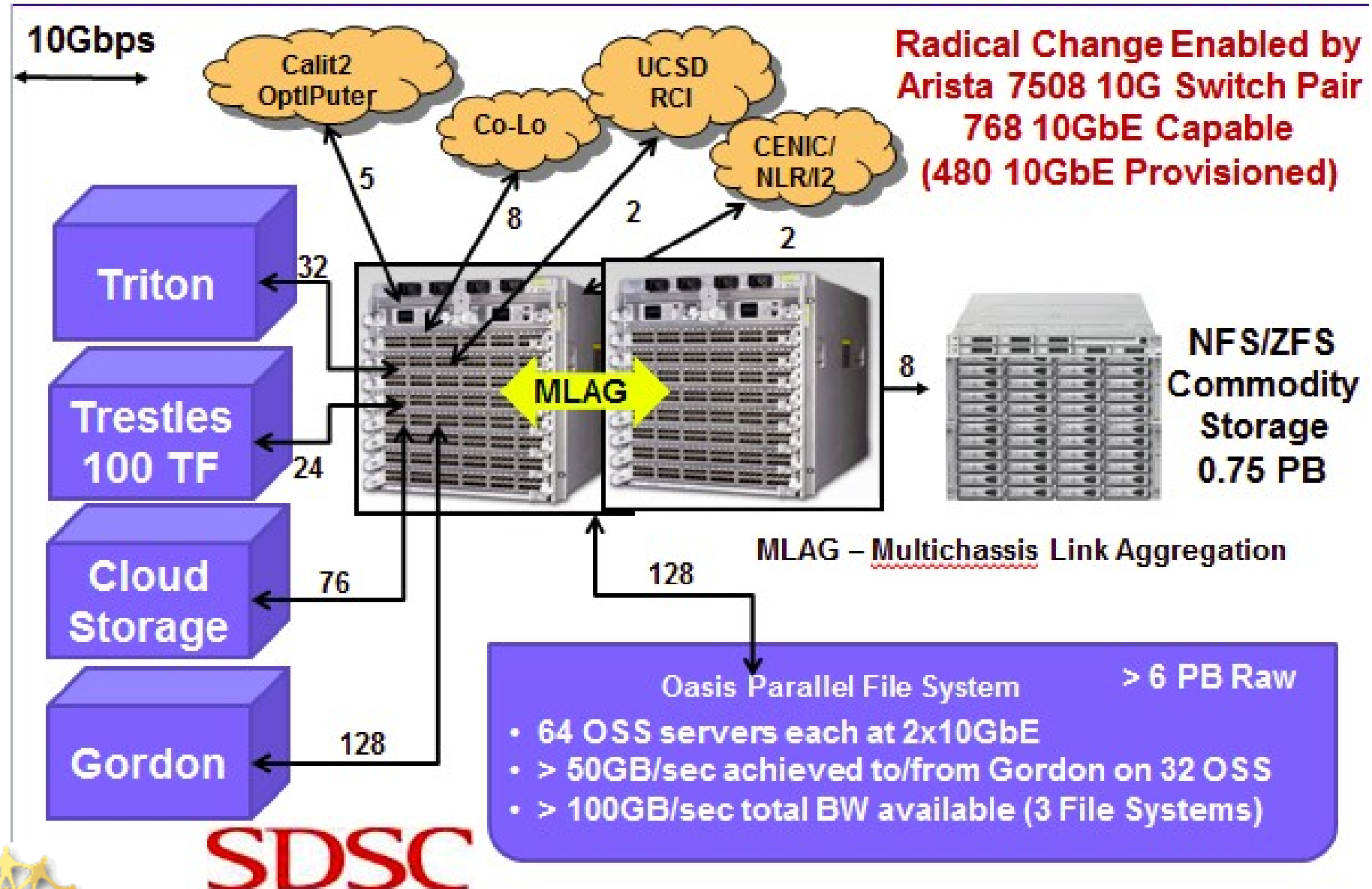
- **Michael Snyder, Chair of Genomics Stanford Univ.**
- **Genome 140x Coverage**
- **Blood Tests 20 Times in 14 Months**
  - **tracked nearly 20,000 distinct transcripts coding for 12,000 genes**
  - **measured the relative levels of more than 6,000 proteins and 1,000 metabolites in Snyder's blood**

Creating a Big Data Freeway System: NSF Has Awarded Prism@UCSD Optical Switch



Phil Papadopoulos, SDSC, Calit2, PI

Arista Enables SDSC's Massive Parallel 10G Switched Data Analysis Resource



New NIH Center for Biomedical Computing: integrating Data for Analysis, Anonymization, and SHaring (iDASH)

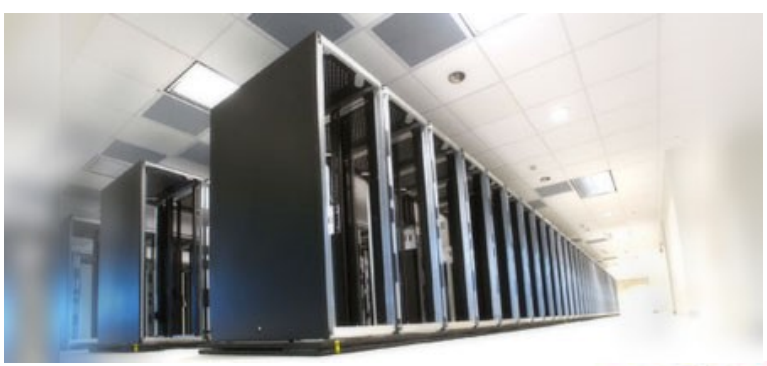
- **Data Exported for Computation Elsewhere**
  - Users download data from iDASH



- **Computation Comes to the Data**
  - Users access data in iDASH
  - Users upload algorithms into iDASH

- **iDASH Exportable Cyberinfrastructure**
  - Users download infrastructure

Private Cloud at SD Supercomputer Center
Medical Center Data Hosting
HIPAA certified facility

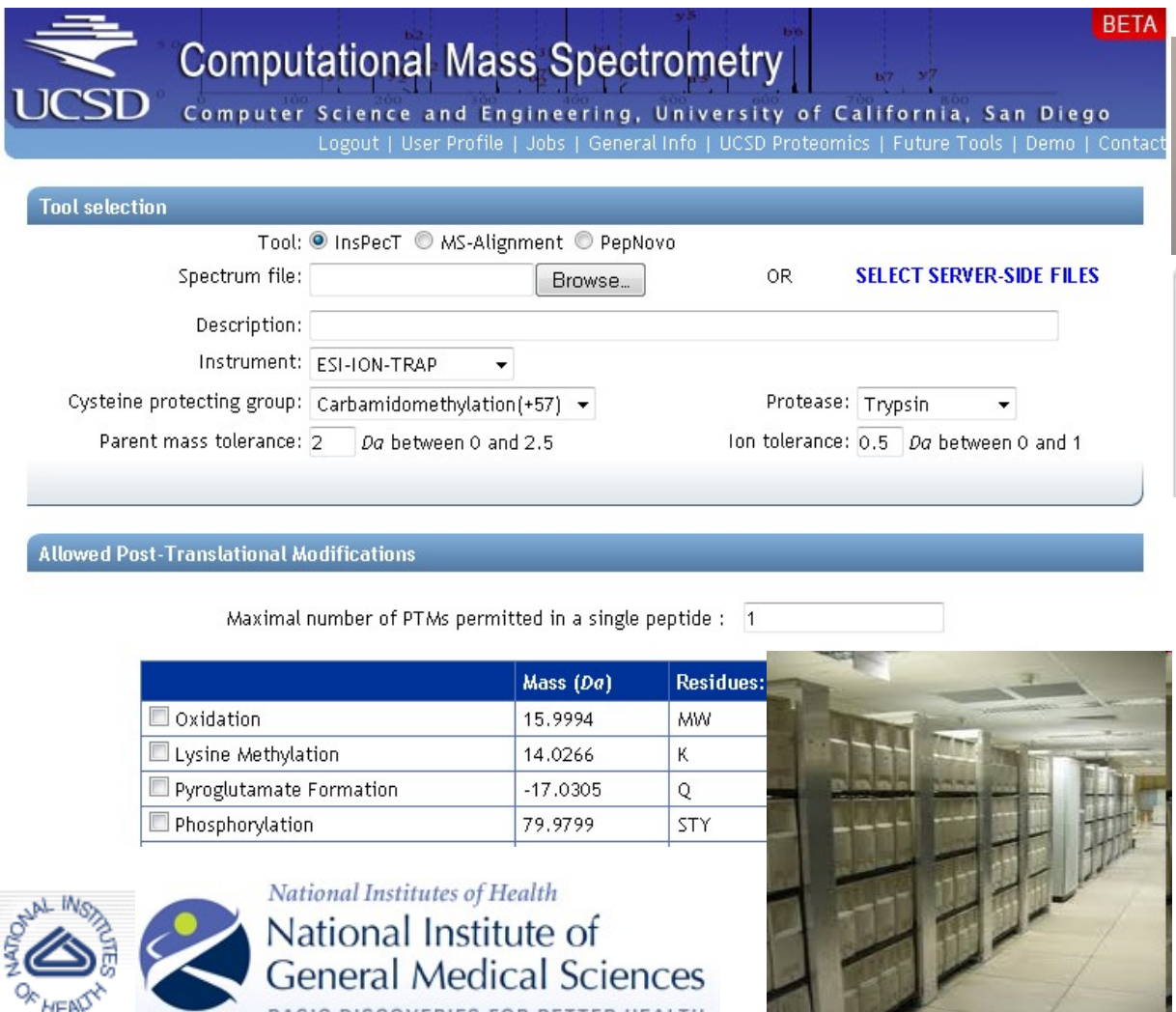


Source: Lucila Ohno-Machado, UCSD SOM

funded by NIH U54HL108460

UCSD Center for Computational Mass Spectrometry Becoming Global MS Repository

ProteoSAFe: Compute-intensive discovery MS at the click of a button



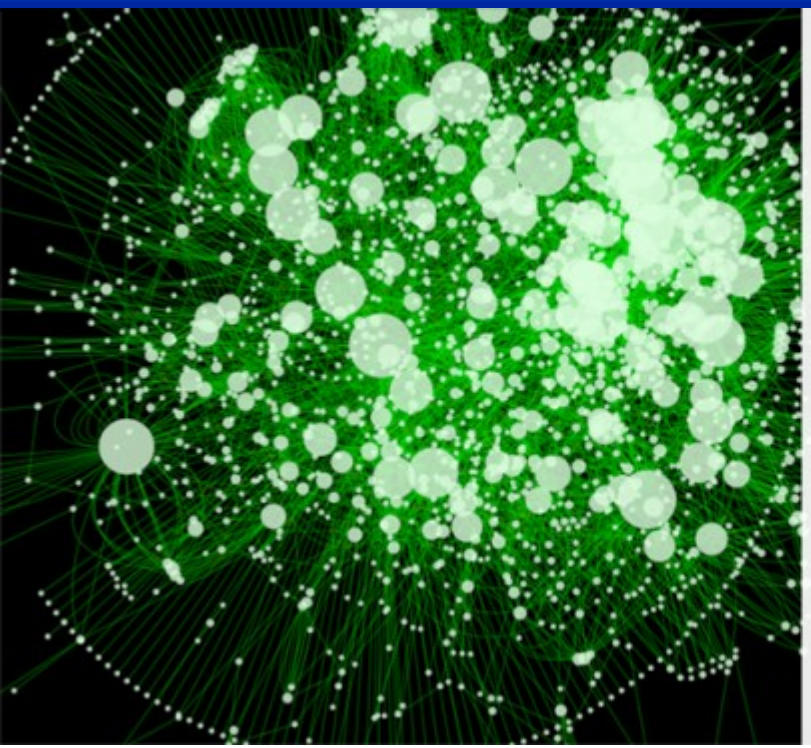
MassIVE: repository and identification platform for all MS data in the world



Source:
Nuno Bandeira,
Vineet Bafna,
Pavel Pevzner,
Ingolf Krueger,
UCSD
proteomics.ucsd.edu

National Institutes of Health
National Institute of General Medical Sciences
BASIC DISCOVERIES FOR BETTER HEALTH

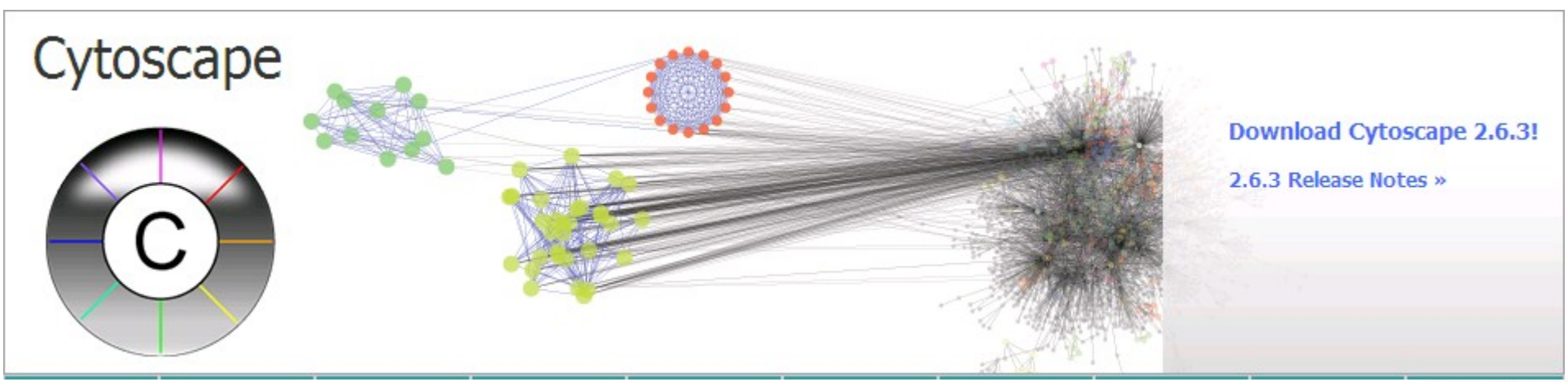
Integrating Systems Biology Data: Cytoscape



Cytoscape

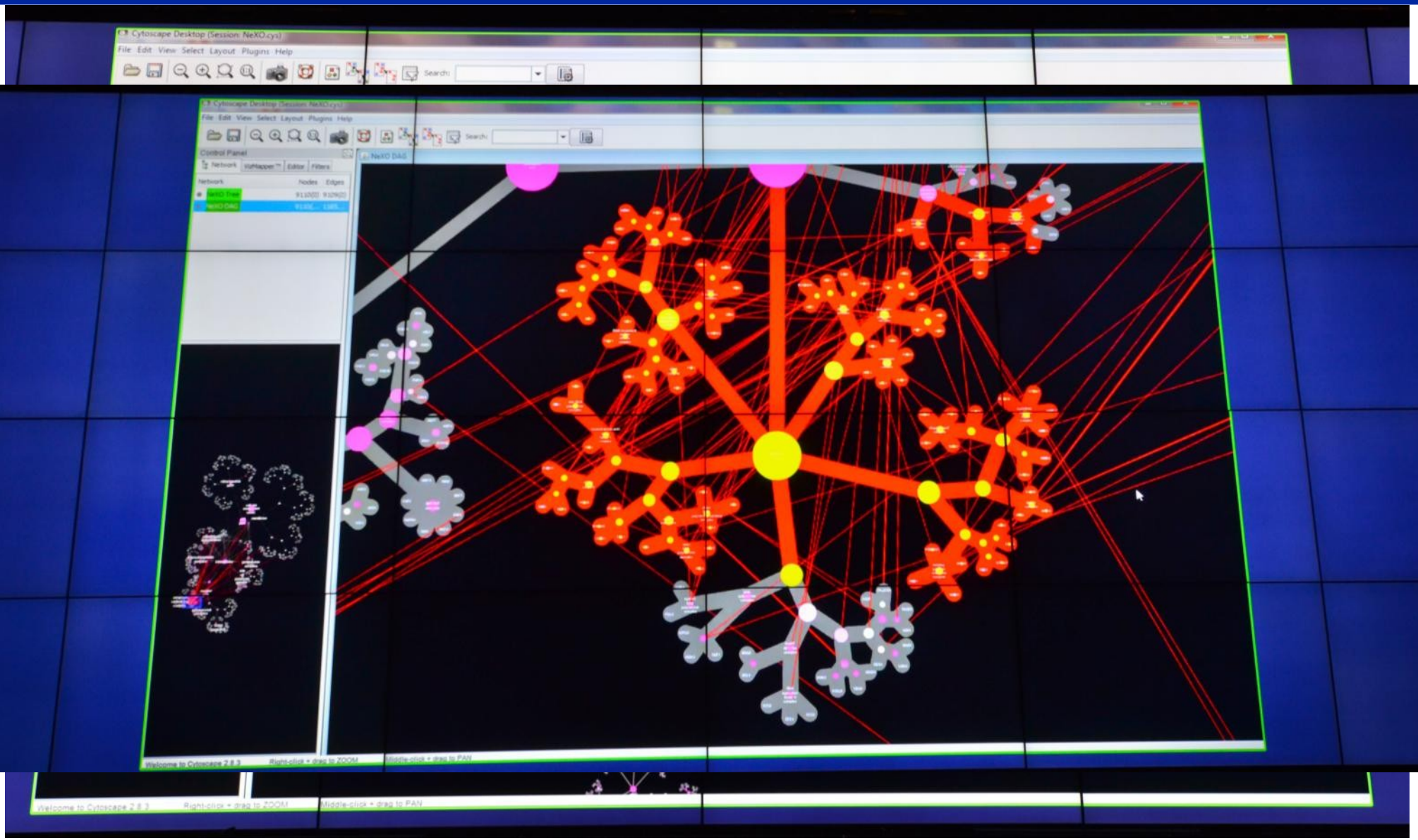
An Open Source Platform for Network Analysis and Visualization

- **OPEN SOURCE Java Platform for Integration of Systems Biology Data**
- **Layout and Query of Interaction Networks (Physical And Genetic)**
- **Visual and Programmatic Integration of Molecular State Data (Attributes)**



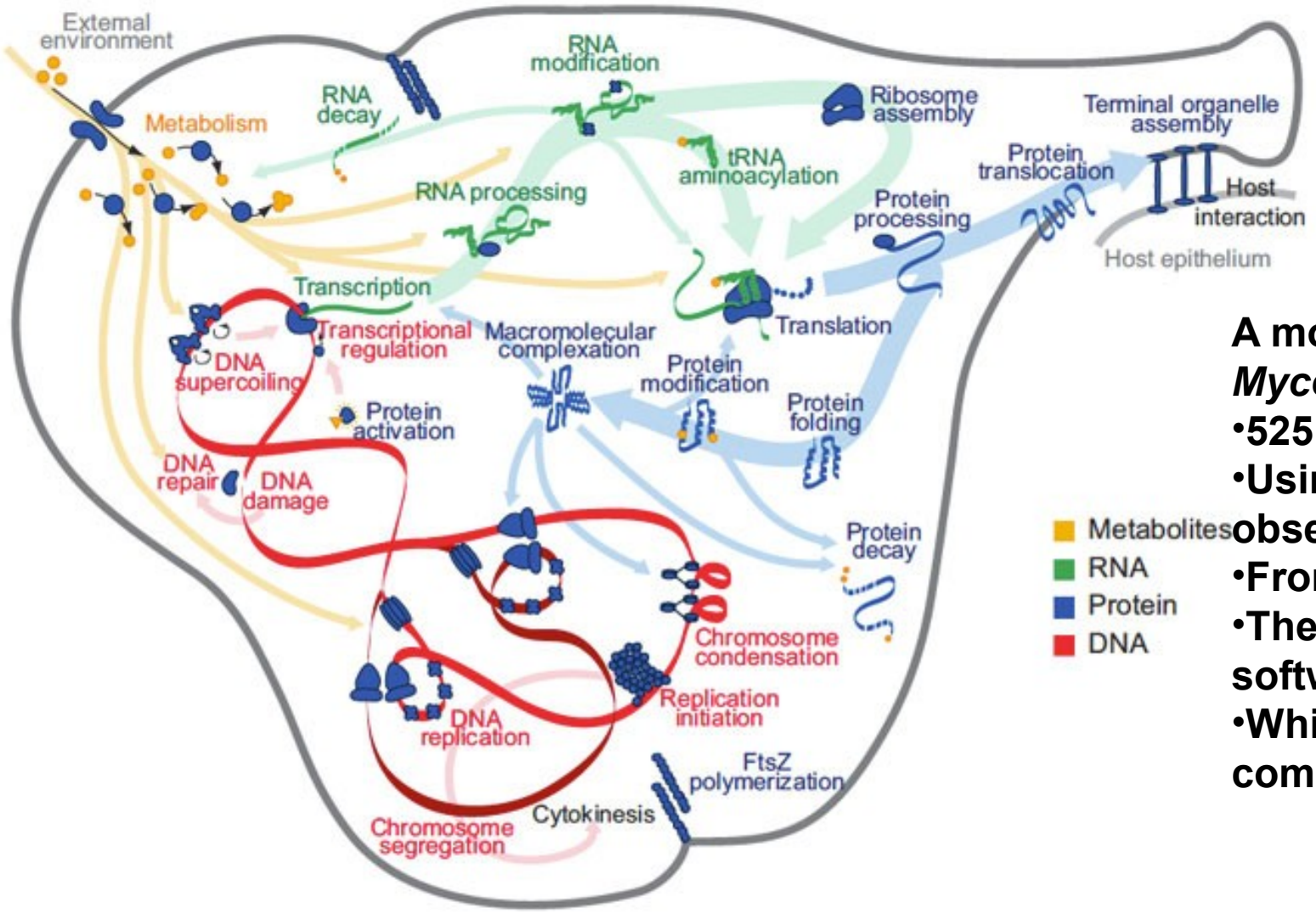
www.cytoscape.org

Cytoscape Genetic Networks
On Vroom-64MPixels Connected at 50Gbps



Calit2 Collaboration with Trey Idekar Group

“A Whole-Cell Computational Model Predicts Phenotype from Genotype”



Cell

Cell *150*, 389–401, July 20, 2012

A model of *Mycoplasma genitalium,*

- **525 genes**
- **Using 1,900 experimental observations**
- **From 900 studies,**
- **They created the software model,**
- **Which requires 128 computers to run**

Jonathan R. Karr,[1,4] Jayodita C. Sanghvi,[2,4] Derek N. Macklin,[2] Miriam V. Gutschow,[2] Jared M. Jacobs,[2] Benjamin Bolival, Jr.,[2] Nacyra Assad-Garcia,[3] John I. Glass,[3] and Markus W. Covert[2,*]

[1]Graduate Program in Biophysics

[2]Department of Bioengineering

Stanford University, Stanford, CA 94305, USA

[3]J. Craig Venter Institute, Rockville, MD 20850, USA

The Stanford/JCVI Paper Was Hailed as a Historic Breakthrough

Cell

Leading Edge
Previews

The Dawn of Virtual Cell Biology

Peter L. Freddolino[1] and Saeed Tavazoie[1,*]
[1]Department of Biochemistry and Molecular Biophysics and Initiative in Systems Biology, Columbia University

In First, Software Emulates Lifespan of Entire Organism

By JOHN MARKOFF
Published: July 20, 2012

The New York Times

Science

the Atlantic

To Model the Simplest Microbe in the World, You Need 128 Computers

JUL 23 2012, 1:05 PM ET 25

Share 16 Tweet 116 +1 42 Recommend 244

What a pathbreaking computer simulation tells us about the future of biotechnology.

Alexis Madrigal is a senior editor at *The Atlantic*.

Early Attempts at Modeling the Systems Biology of the Gut Microbiome and the Human Immune System

Trends in Biotechnology June 2011, Vol. 29, No. 6

Prospects for systems biology and modeling of the gut microbiome

Fredrik H. Karlsson, Intawat Nookaew, Dina Petranovic and Jens Nielsen

Department of Chemical and Biological Engineering, Chalmers University of Technology, Kemivägen 10, SE-412 96 Göteborg, Sweden

2012 IEEE 26th International Parallel and Distributed Processing Symposium

High-Performance Interaction-Based Simulation of Gut Immunopathologies with ENteric Immunity SImulator (ENISI)

Keith Bisset*, Md. Maksudul Alam*, Josep Bassaganya-Riera*, Adria Carbo*, Stephen Eubank*, Raquel Hontecillas*, Stefan Hoops*, Yongguo Mei*, Katherine Wendelsdorf*, Dawen Xie*, Jae-Seung Yeom*†, and Madhav Marathe*†

**Virginia Bioinformatics Institute*

†Department of Computer Science

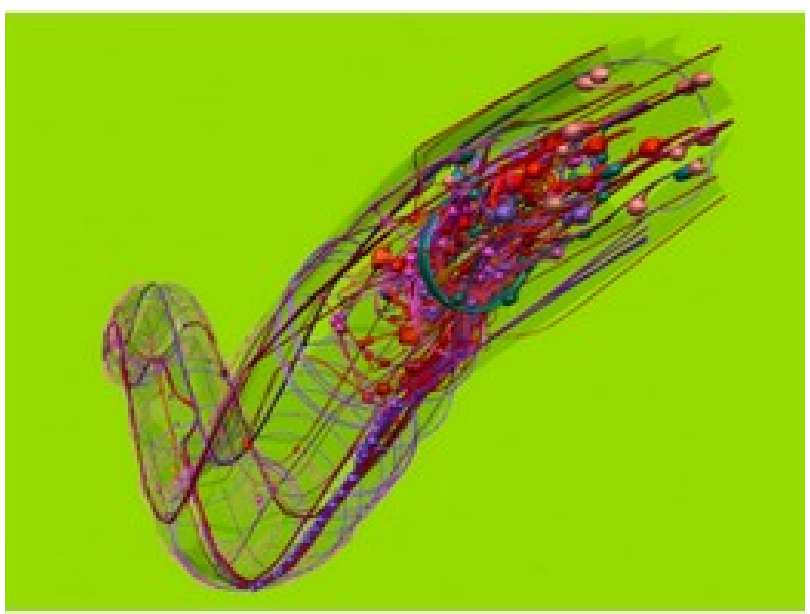
Virginia Tech, Blacksburg VA 24061

Next Challenge:
Building a Multi-Cellular Organism Simulation

OpenWorm is an attempt to build a complete cellular-level simulation of the nematode worm Caenorhabditis elegans. Of the 959 cells in the hermaphrodite, 302 are neurons and 95 are muscle cells.

The simulation will model electrical activity in all the muscles and neurons. An integrated soft-body physics simulation will also model body movement and physical forces within the worm and from its environment.

- Latest news / current status
- Biology of C. elegans
- The nervous system
- Simulation approach
- Software stack
- Computer hardware
- 3D WebGL model



A 3D rendering of the C. elegans nervous system.

www.artificialbrains.com/openworm

A Vision for Healthcare in the Coming Decades

Using this data, the planetary computer will be able to build a computational model of your body and compare your sensor stream with millions of others. Besides providing early detection of internal changes that could lead to disease, cloud-powered voice-recognition wellness coaches could provide continual personalized support on lifestyle choices, potentially staving off disease and making health care affordable for everyone.

ESSAY
An Evolution Toward a Programmable Universe
By LARRY SMARR
Published: December 5, 2011

The New York Times

Science

