# "A Systems Approach to Personalized Medicine"

**Talk and Discussion**

**NASA Ames**

**Mountain View, CA**

**March 28, 2013**

**Dr. Larry Smarr**

**Director, California Institute for Telecommunications and Information Technology**

**Harry E. Gruber Professor,**

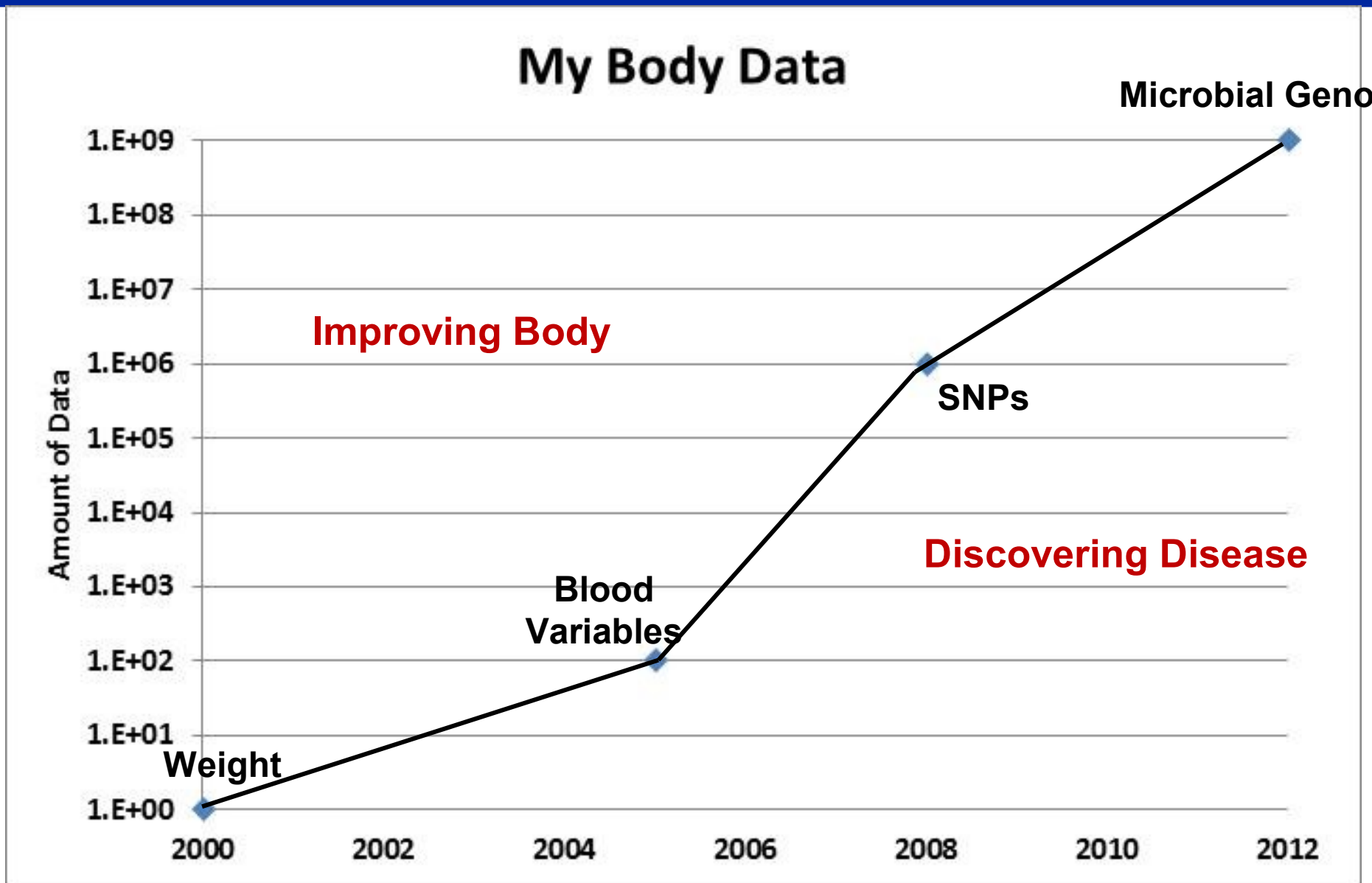
**Dept. of Computer Science and Engineering**

**Jacobs School of Engineering, UCSD**

UCSD

UCIrvine

# From One to a Billion Data Points Defining Me: The Exponential Rise in Body Data in Just One Decade!




UCSD
UCIrvine

# From Measuring Macro-Variables to Measuring Your Internal Variables




## The Patient of the Future

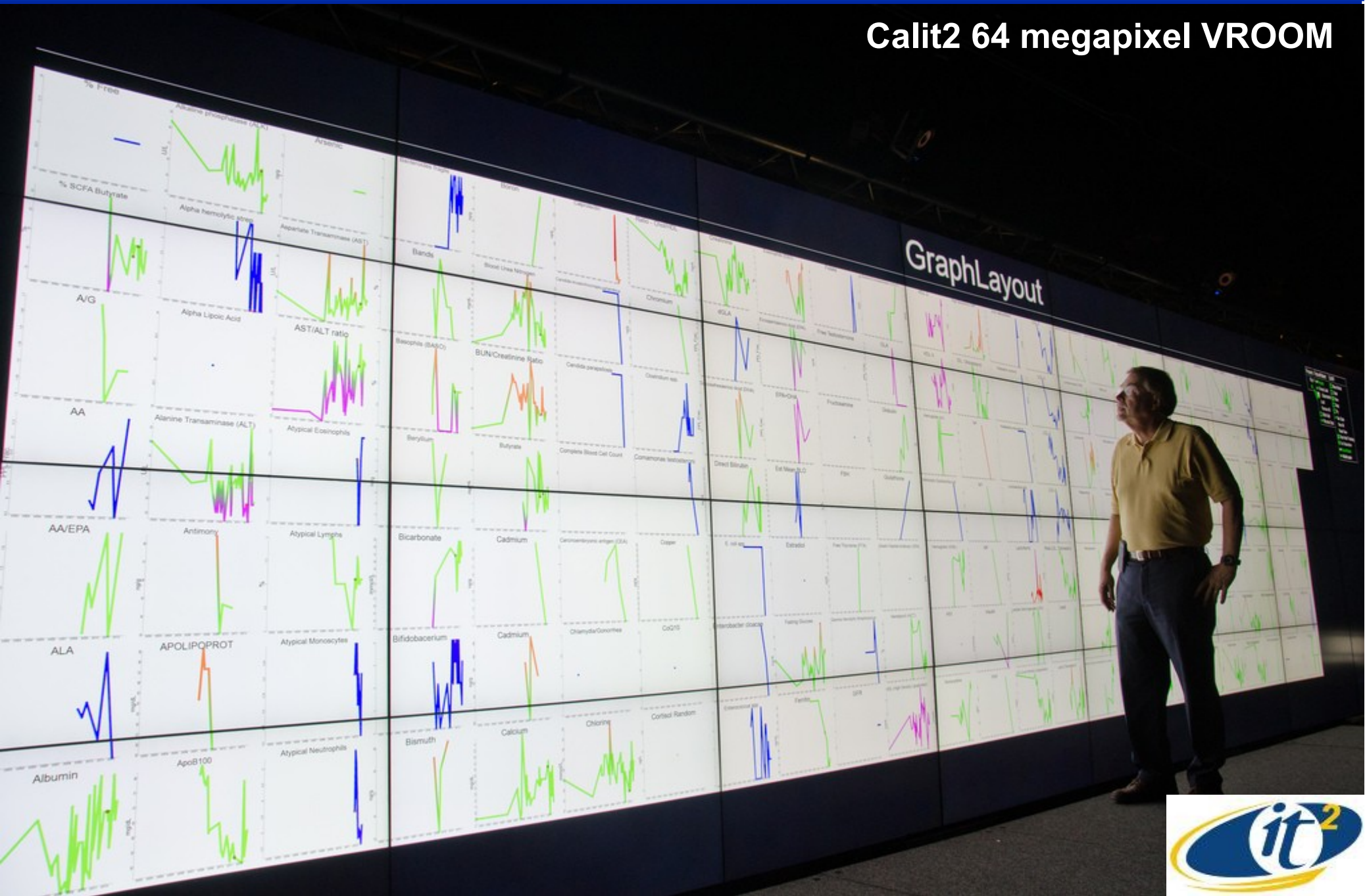
Internet pioneer Larry Smarr's quest to quantify everything about his health led him to a startling discovery, an unusual partnership with his doctor, and more control over his life.

**www.technologyreview.com/biomedicine/39636**

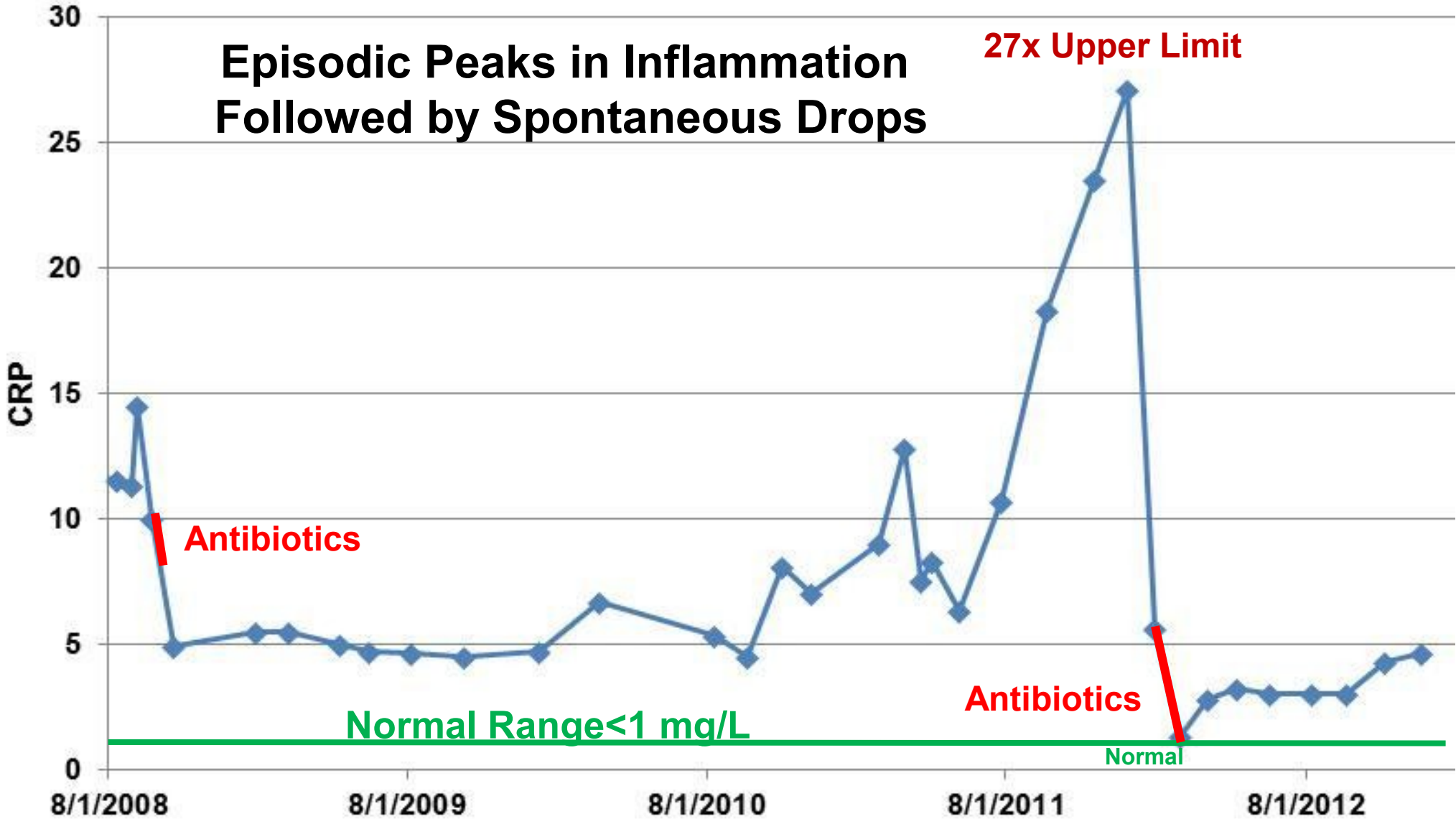
UCSD
UCIrvine

it²

# Visualizing Time Series of 150 LS Blood and Stool Variables, Each Over 5 Years



# Only One of My Blood Measurements Was Far Out of Range--Indicating Chronic Inflammation

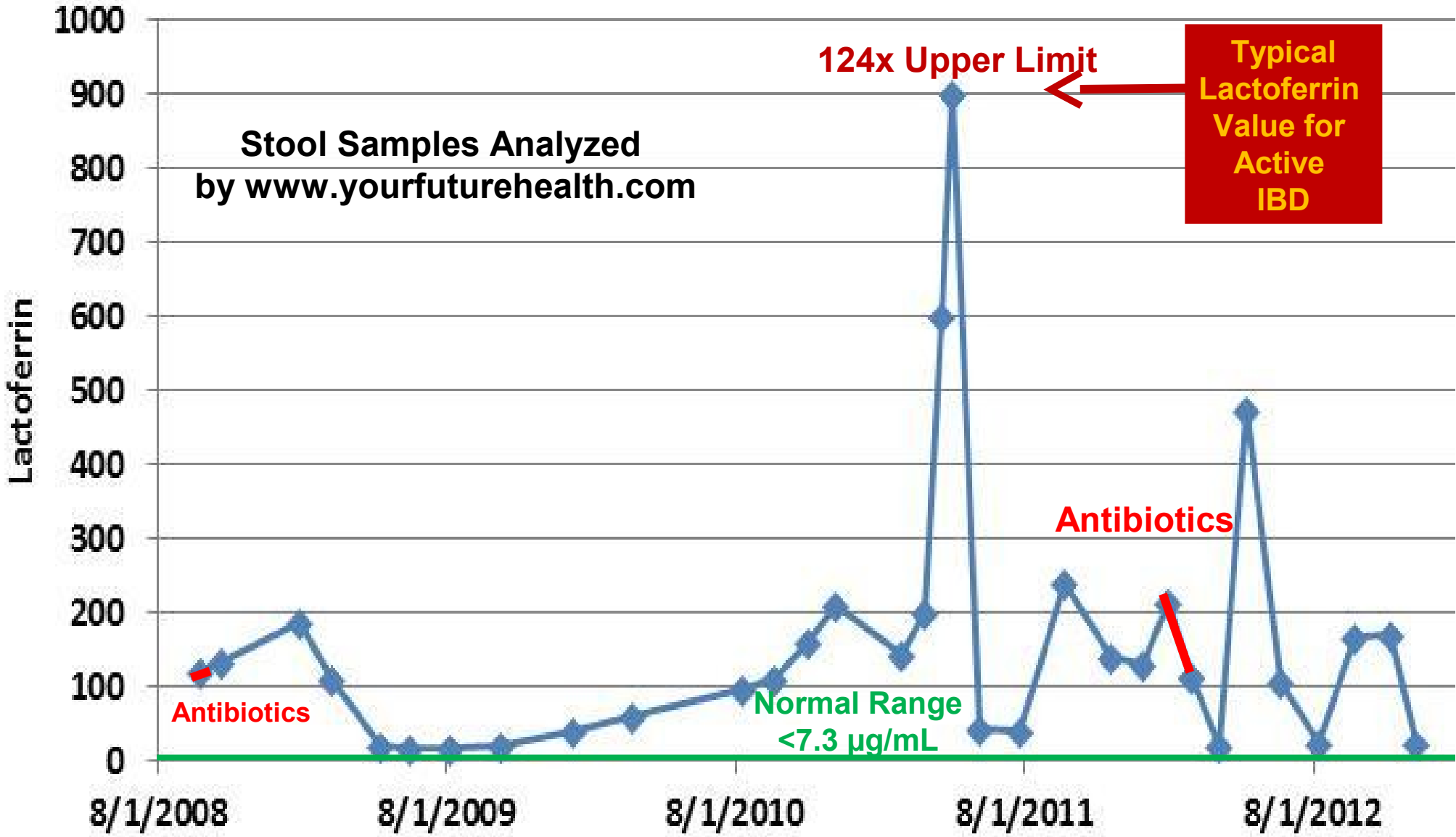



UCSD
UCIrvine

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

it²

# High Values of Lactoferrin (Shed from Neutrophils) From Stool Sample Suggested Inflammation in Colon



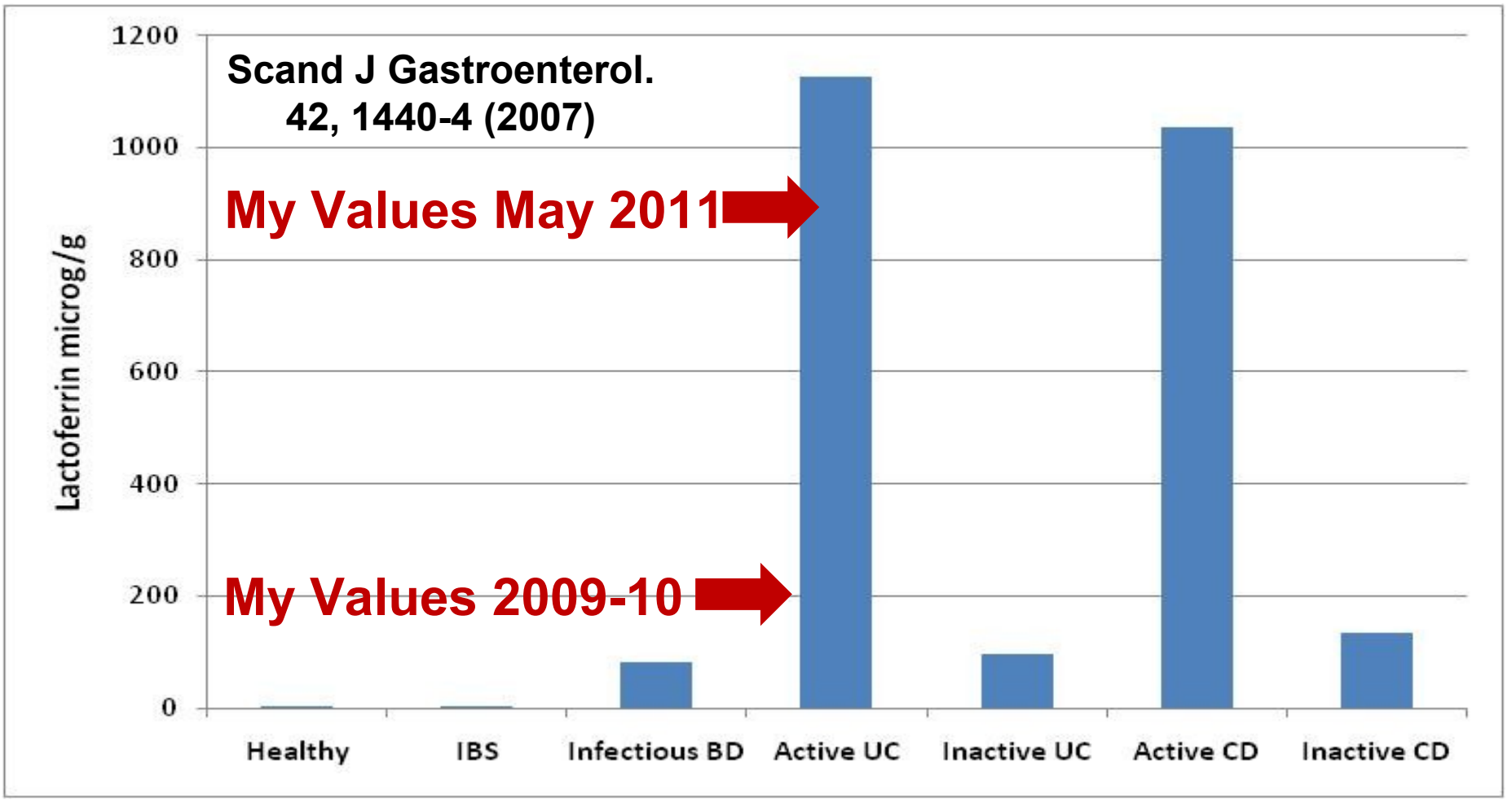

UCSD UCIrvine

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

it²

# High Lactoferrin Biomarker Led Me to Hypothesis I Had Inflammatory Bowel Disease (IBD)

**IBD is an Autoimmune Disease Which Comes in Two Subtypes: Crohn's and Ulcerative Colitis**

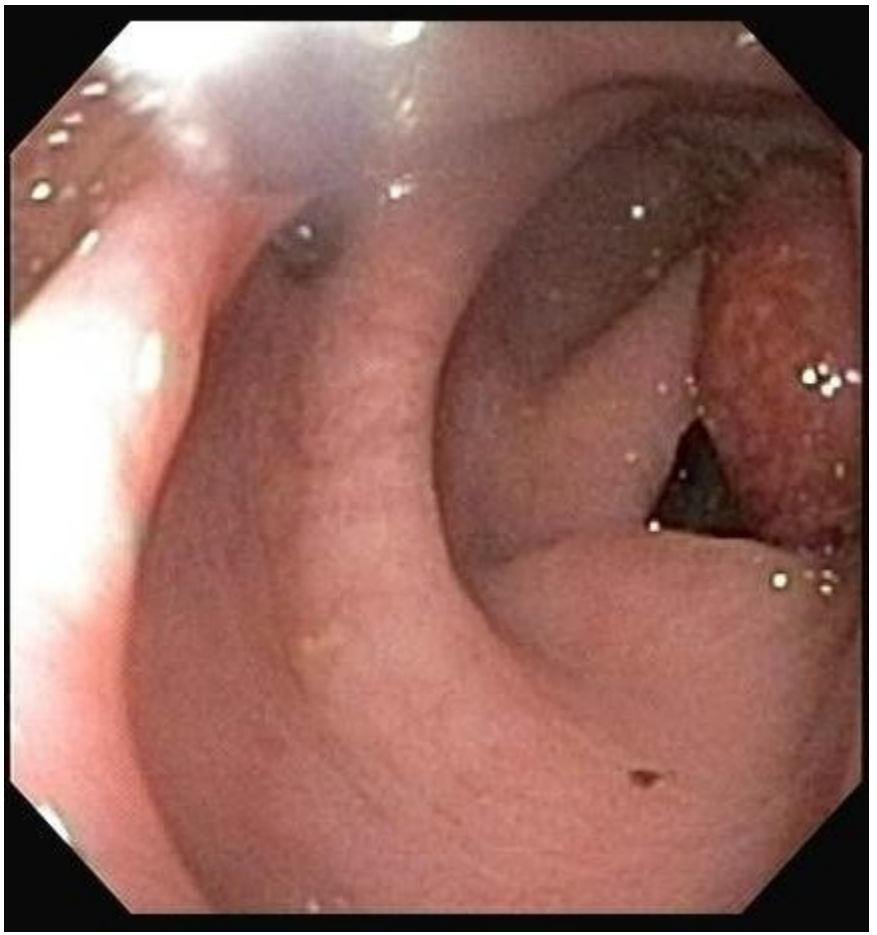



**Colonoscopy Revealed Inflamed Tissue**

# Colonoscopy Images Show Sigmoid Colon Inflammation



Dec 2010

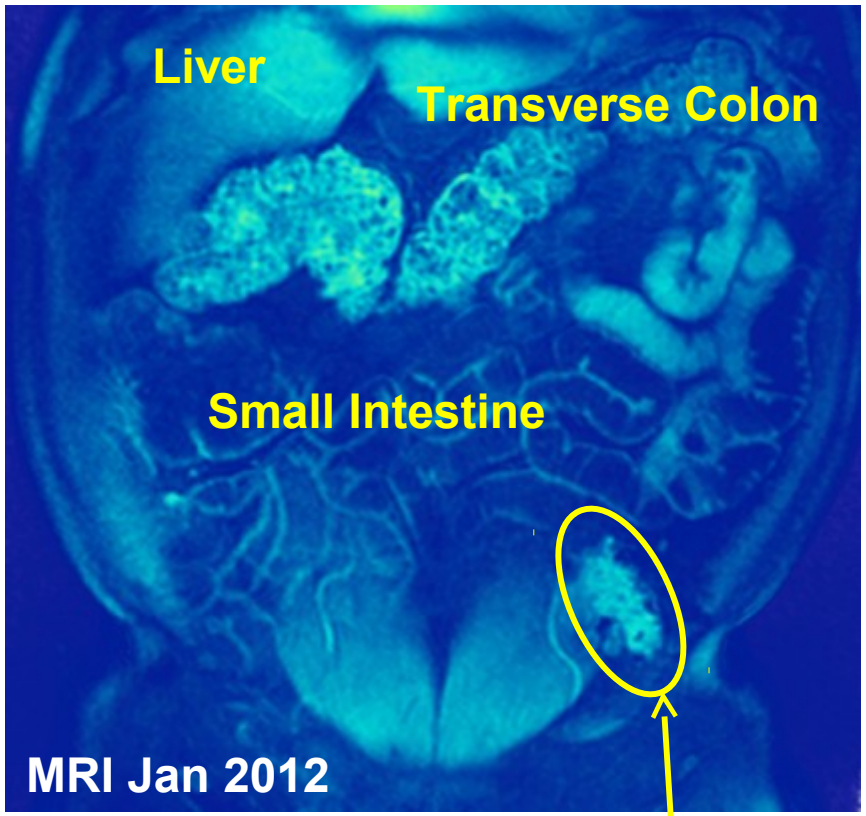


May 2011

UCSD
UCIrvine

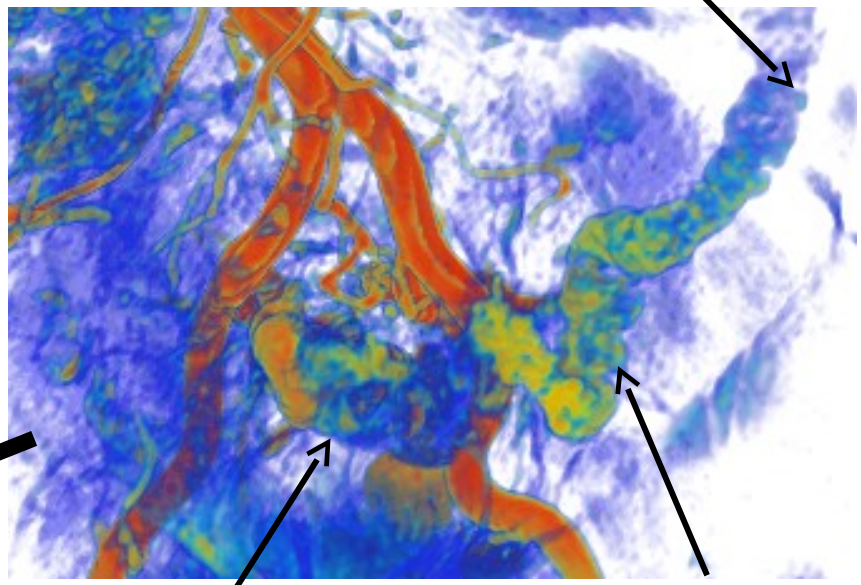
it²

# 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 Calit2 Staff & DeskVOX Software**

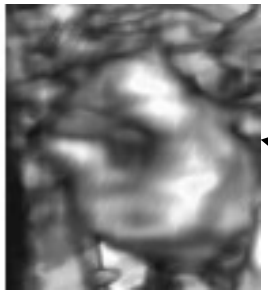
Descending Colon



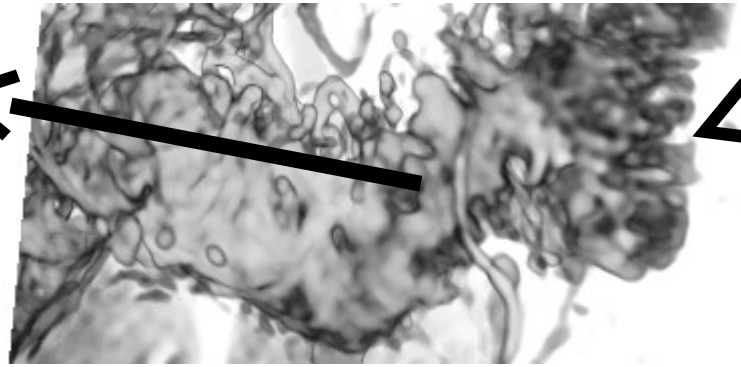
Major Kink

Sigmoid Colon Threading Iliac Arteries

Cross Section



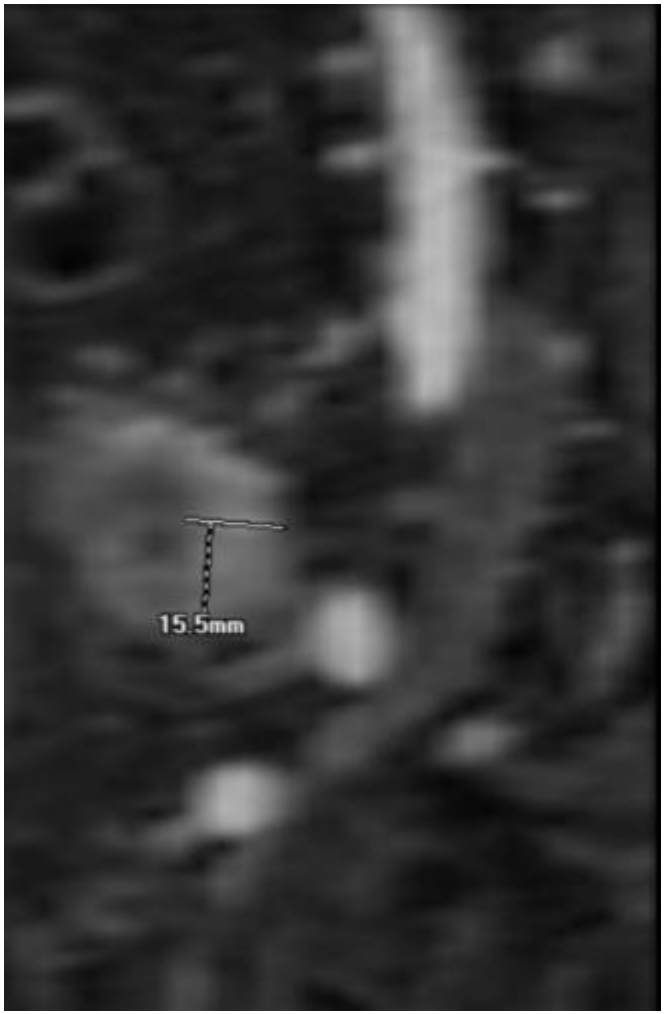
Diseased Sigmoid Colon

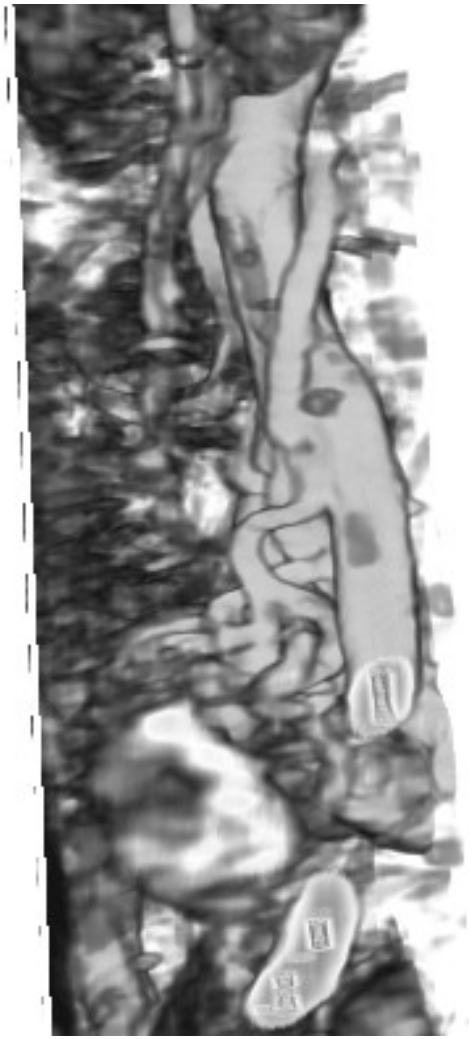


UCSD UCIrvine

it²

# Comparison of DeskVOX with Clinical MRI Slice Program



UCSD
UCIrvine

it²

# An MRI Shows Sigmoid Colon Wall Thickened Indicating Probable Diagnosis of Crohn's Disease



# 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!

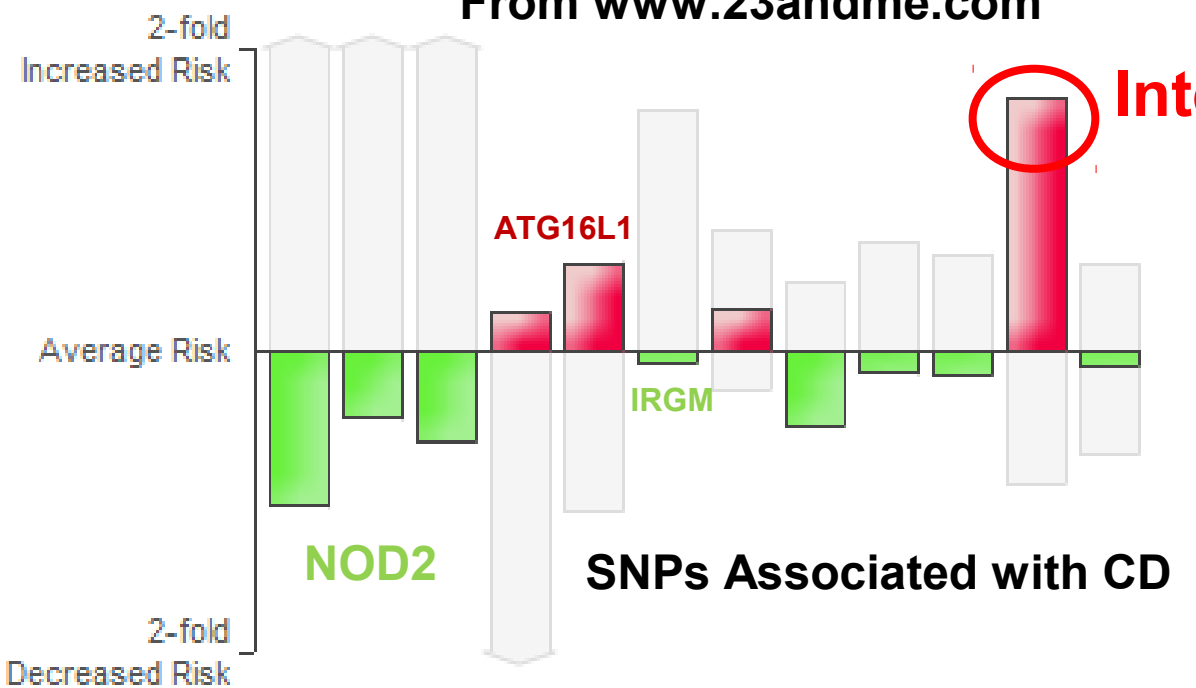
UCSD
UCIrvine

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

it²

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

From www.23andme.com




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

**Now Comparing 163 Known IBD SNPs with 23andme SNP Chip**

News and Views

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

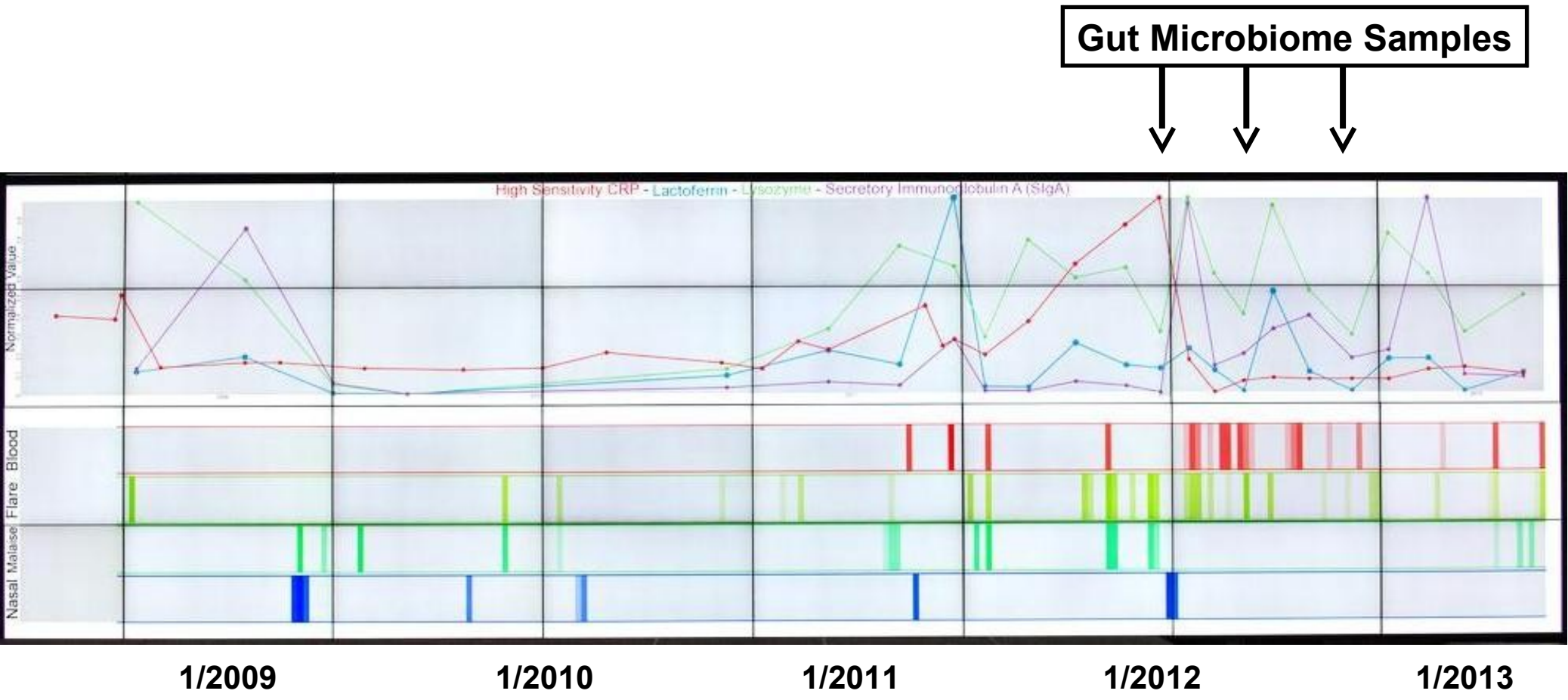
## IL-23: a master regulator in Crohn disease

Markus F Neurath[1]

UCSD
UCIrvine

it²

# Four Immune Biomarkers Over Time Compared with Four Signs/Symptoms



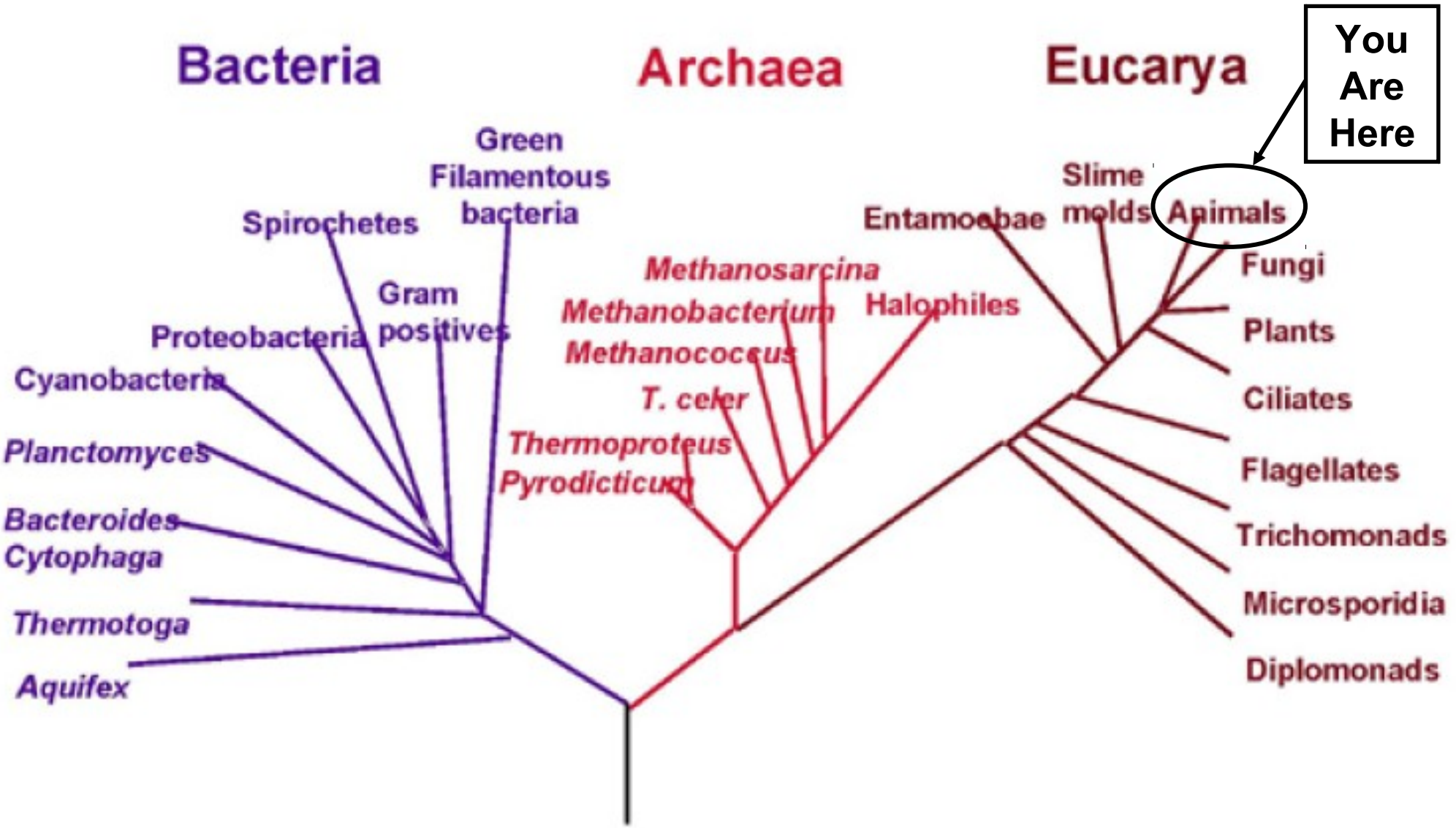

**Here Immune biomarkers are normalized 0 to 1, with 1 being the highest value in five years**

UCSD
UCIrvine

Source: Photo of Calit2 64-megapixel VROOM

it²

# However, Most Biological Diversity on Earth is in the Microbial World



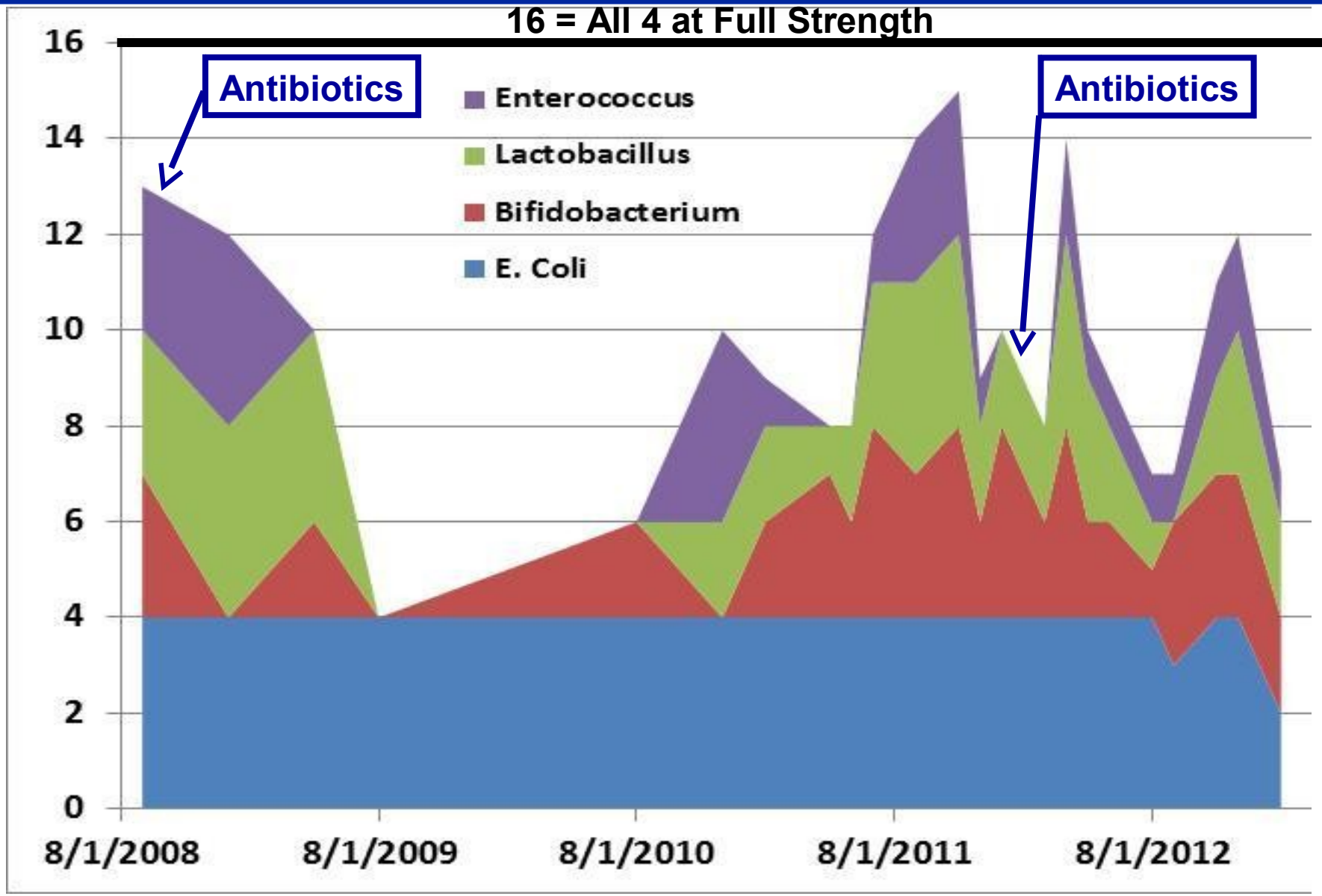

UCSD
UCIrvine

**So You Have Many Phyla of Microbes Within You!**

**Source: Carl Woese, et al**

it²

# Cultured Bacteria From Stool Tests Showed Large Time Variations in Gut Microbiome




UCSD
UCIrvine

**Antibiotics: Levaquin & Metronidaloze**

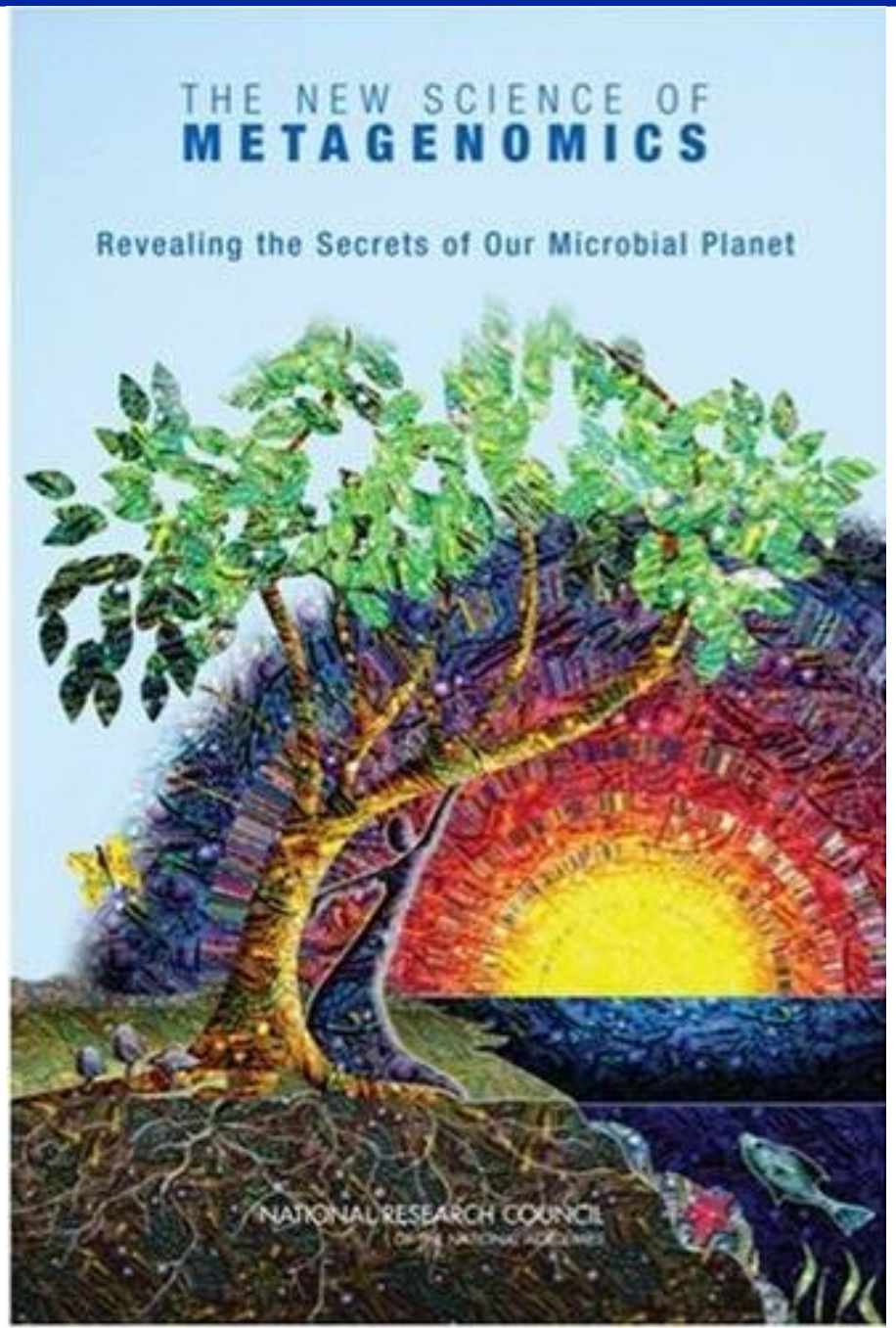
**Values From www.yourfuturehealth.com stool test**

it²

# But How Can You Determine Which Microbes Are Within You?

**NRC Report:**

**Metagenomic data should be made publicly available in international archives as rapidly as possible.**




**"The emerging field of metagenomics, where the DNA of entire communities of microbes is studied simultaneously, presents the greatest opportunity -- perhaps since the invention of the microscope – to revolutionize understanding of the microbial world." –**

**National Research Council**
**March 27, 2007**

UCSD
UCIrvine

it²

# Intense Scientific Research is Underway on Understanding the Human Microbiome




June 8, 2012



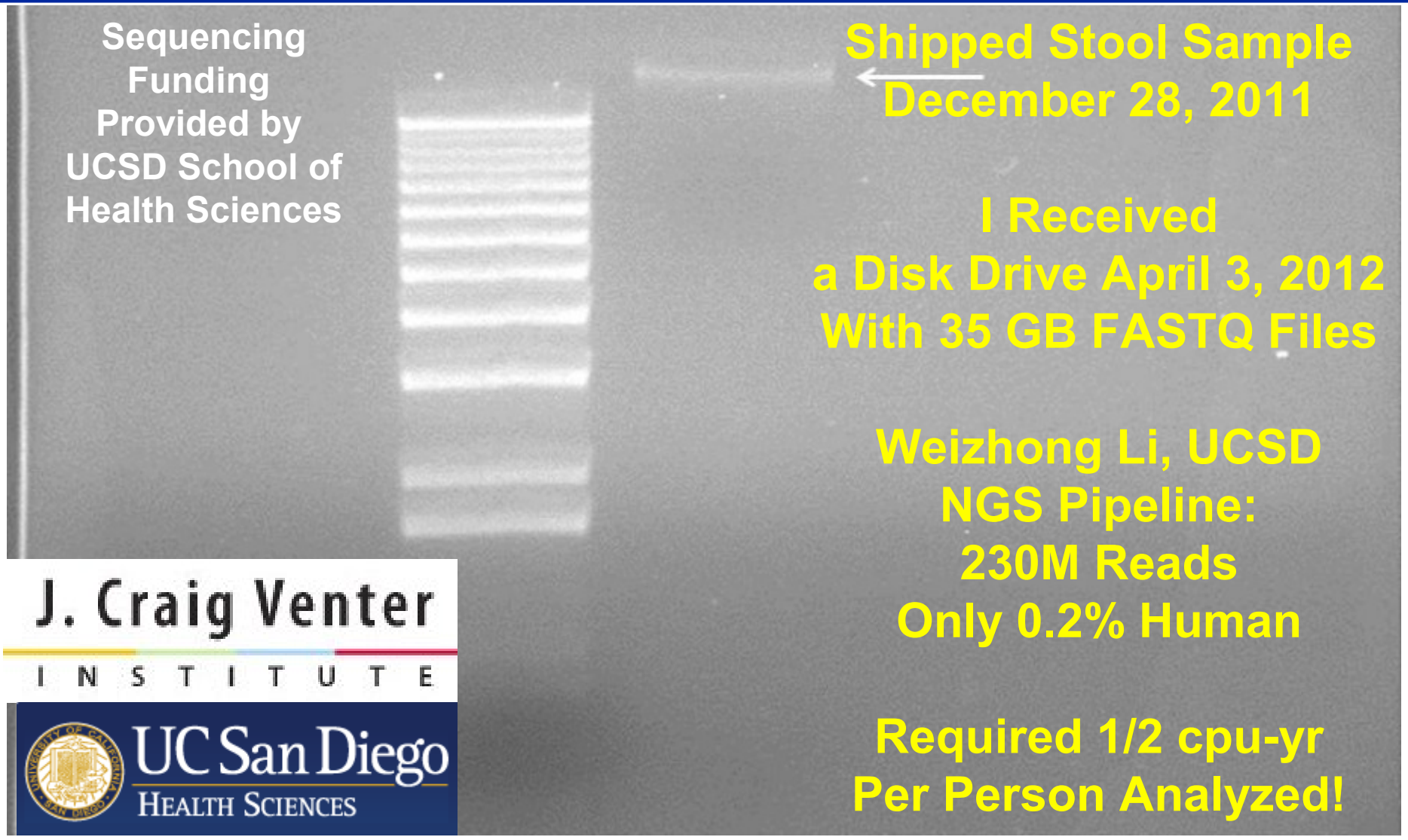

June 14, 2012

UCSD
UCIrvine

**From Culturing Bacteria to Sequencing Them**

it²

# To Map My Gut Microbes, I Sent a Stool Sample to the Venter Institute for Metagenomic Sequencing



Sequencing Funding Provided by UCSD School of Health Sciences

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!

J. Craig Venter INSTITUTE

UC San Diego HEALTH SCIENCES

Gel Image of Extract from Smarr Sample-Next is Library Construction

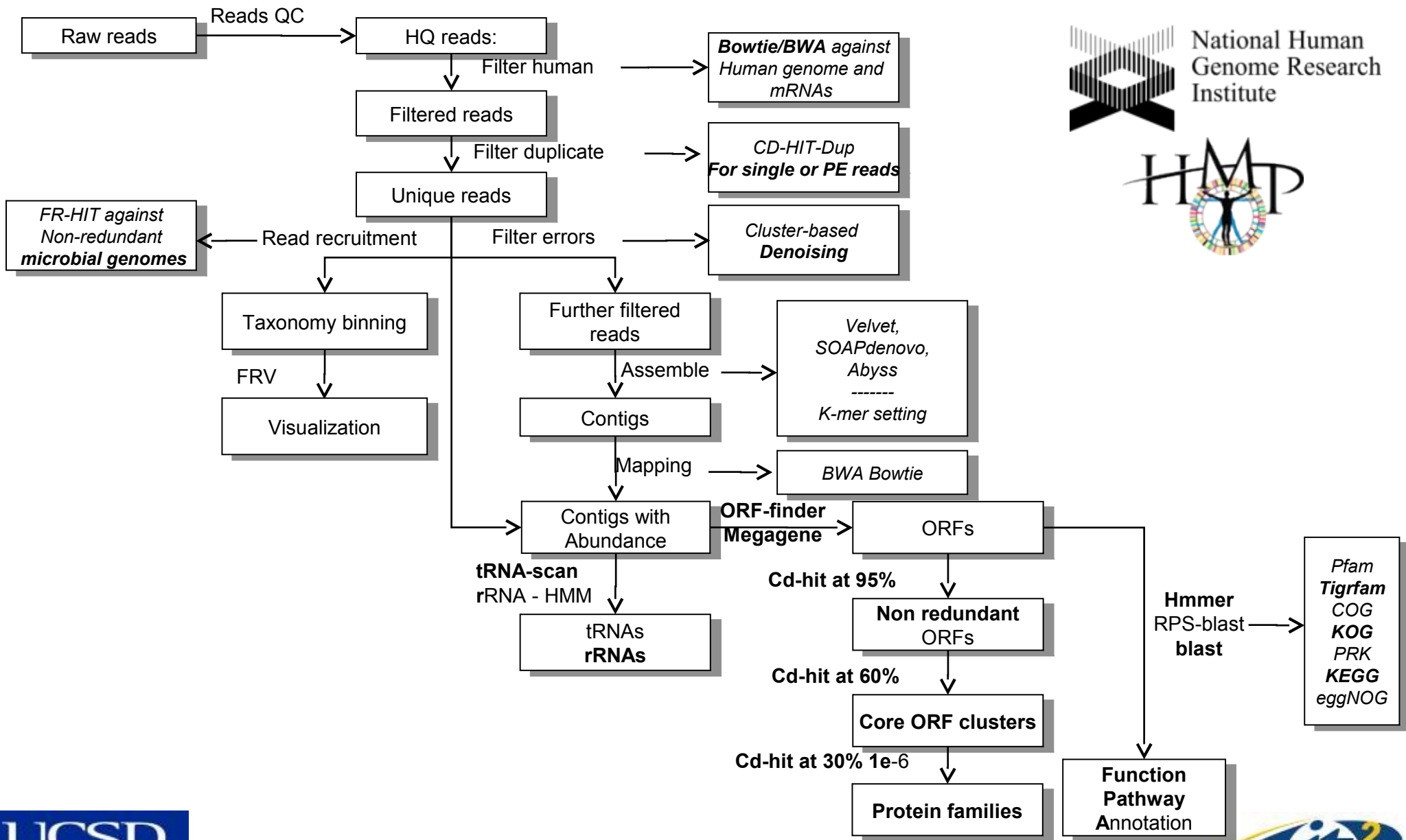
Manny Torralba, Project Lead - Human Genomic Medicine
J Craig Venter Institute

January 25, 2012

UCSD UCIrvine

it²

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




National Human Genome Research Institute

HMP

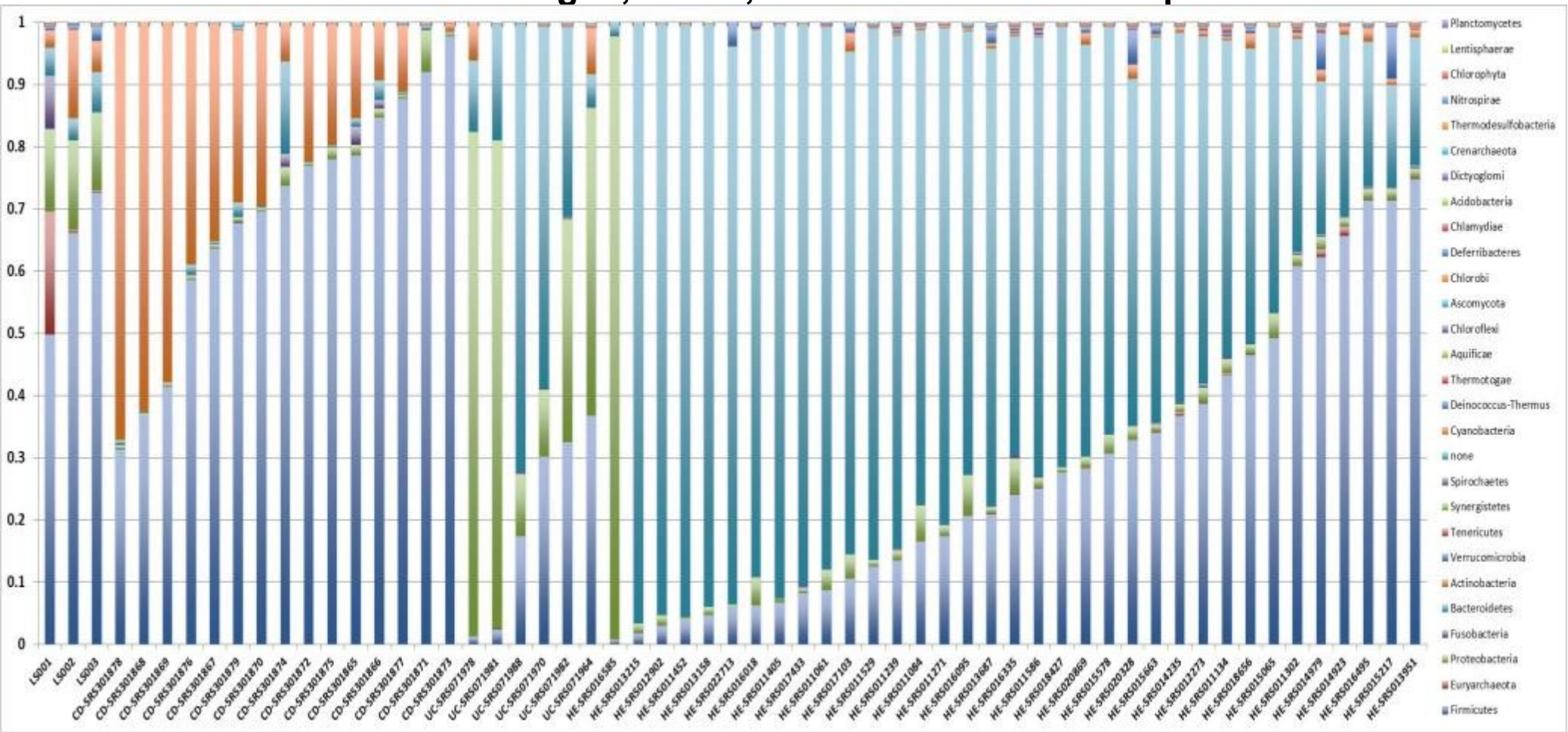
UCSD UCIrvine

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

it²

# Computations Reveal Gut Microbial Phyla Abundance: LS, Crohn's, UC, and Healthy Subjects

**Source: Weizhong Li, UCSD; Calit2 FuturePatient Expedition**




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

**Bacterial Phyla**

UCSD
UCIrvine

## Toward Noninvasive Microbial Ecology Diagnostics

# 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**

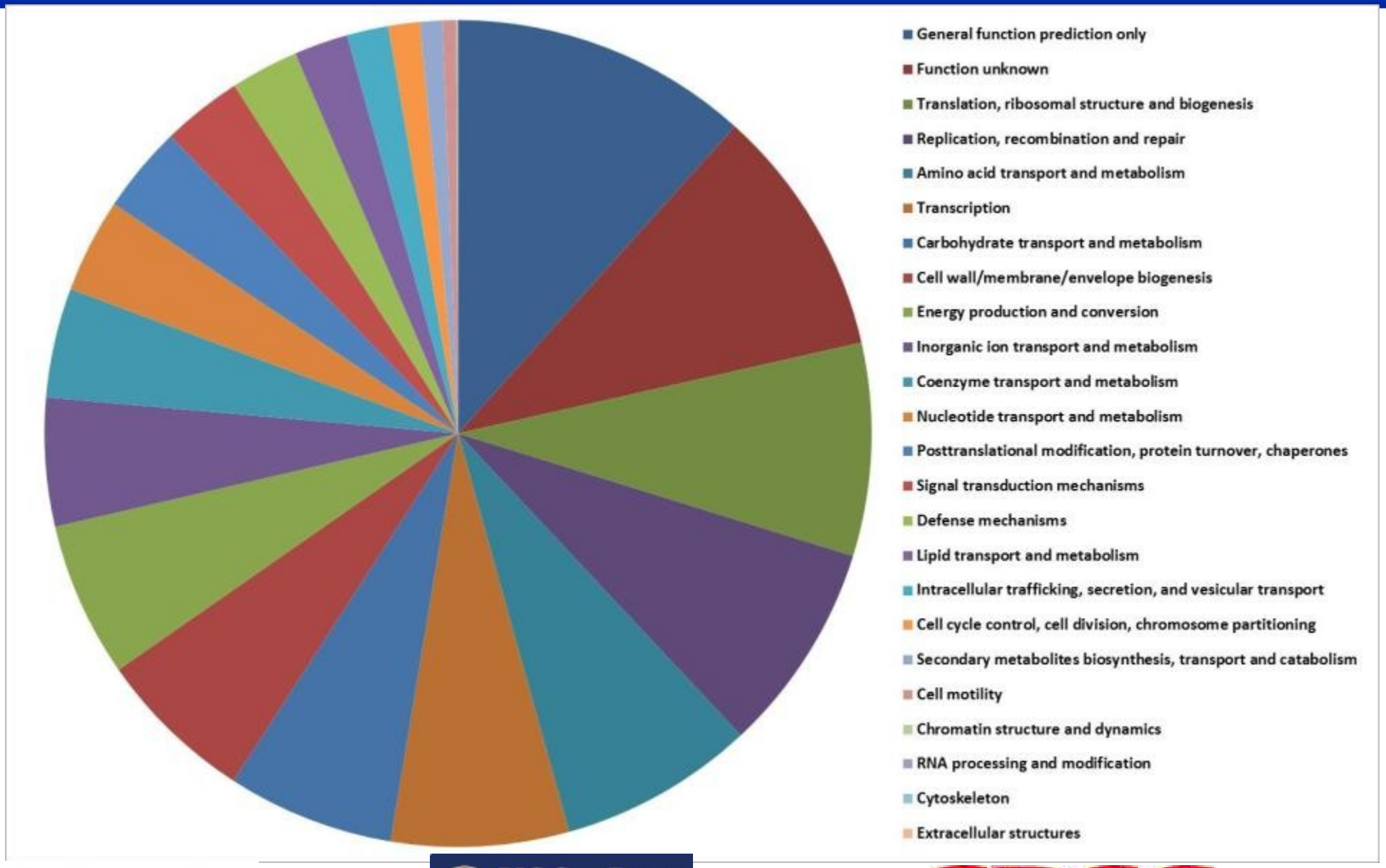
**Enabled by a Grant of Time on Gordon from SDSC Director Mike Norman**

SDSC

UCSD
UCIrvine

it²

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




J. Craig Venter INSTITUTE

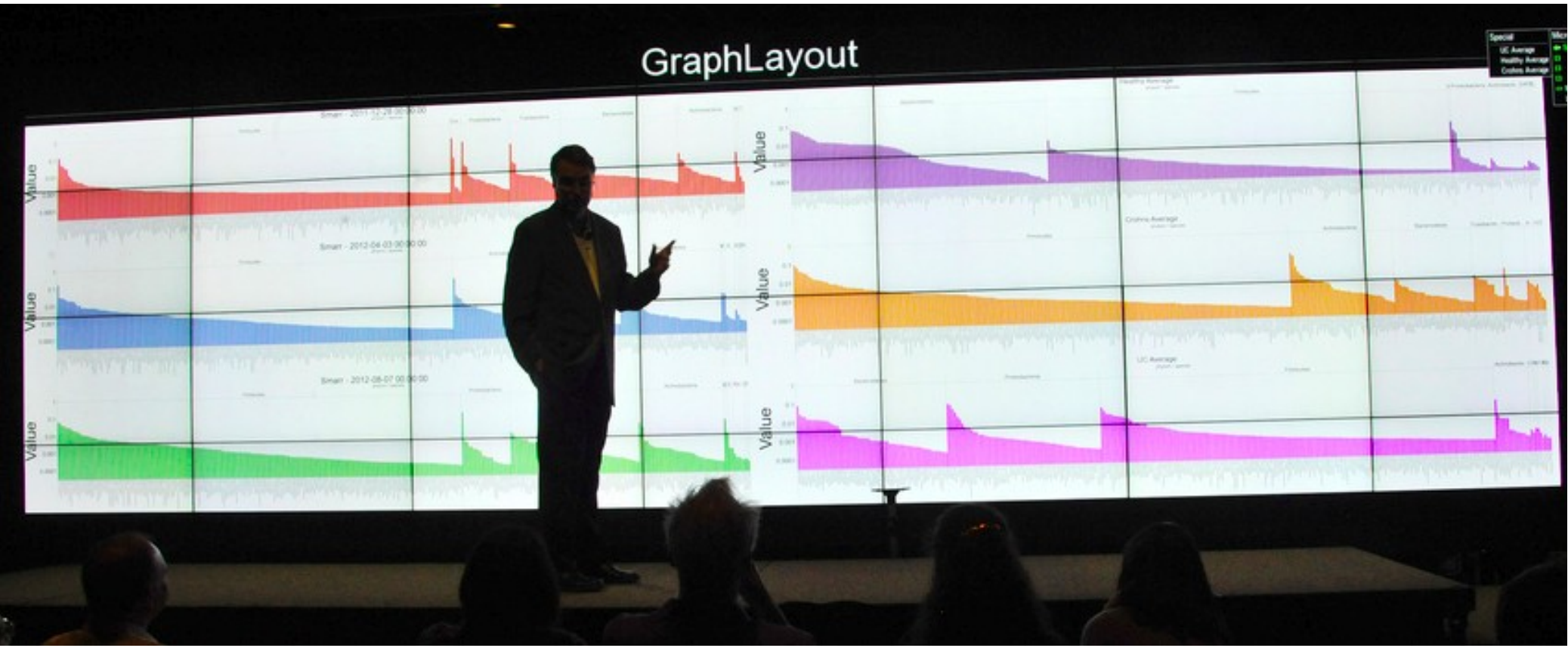
UC San Diego Health Sciences

SDSC

it²

Analysis: Weizhong Li & Sitao Wu, UCSD

# Using Calit2's 64 Megapixel Tiled Display Wall To Analyze Human Microbiome Complexity




**Comparing 3 LS Time Snapshots (Left)**
**with Healthy, Crohn's, UC (Right Top to Bottom)**

UCSD
UCIrvine

**Calit2 VROOM-FuturePatient Expedition**

it²

# LS Gut Microbe Species 12/28/11 (red) compared to Average of Healthy Subjects (blue)

**Species are Organized by Microbial Phyla**
**Each Species is a Bar,**
**Height is Logarithmic Abundance,**



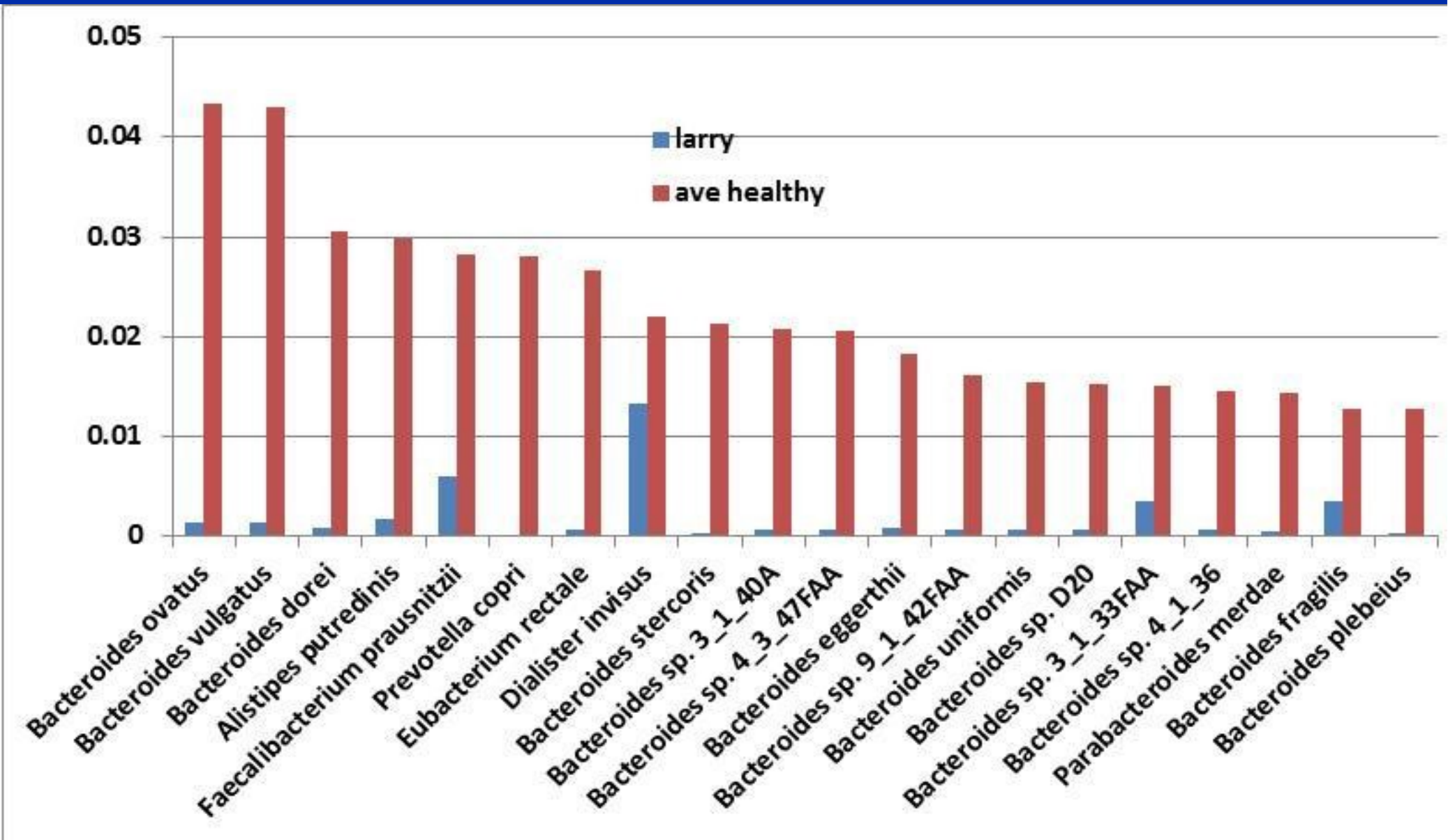

**Derived from metagenomic sequencing of LS stool sample.**

UCSD
UCIrvine

**Source: Photo of Calit2 64-megapixel VROOM**

it²

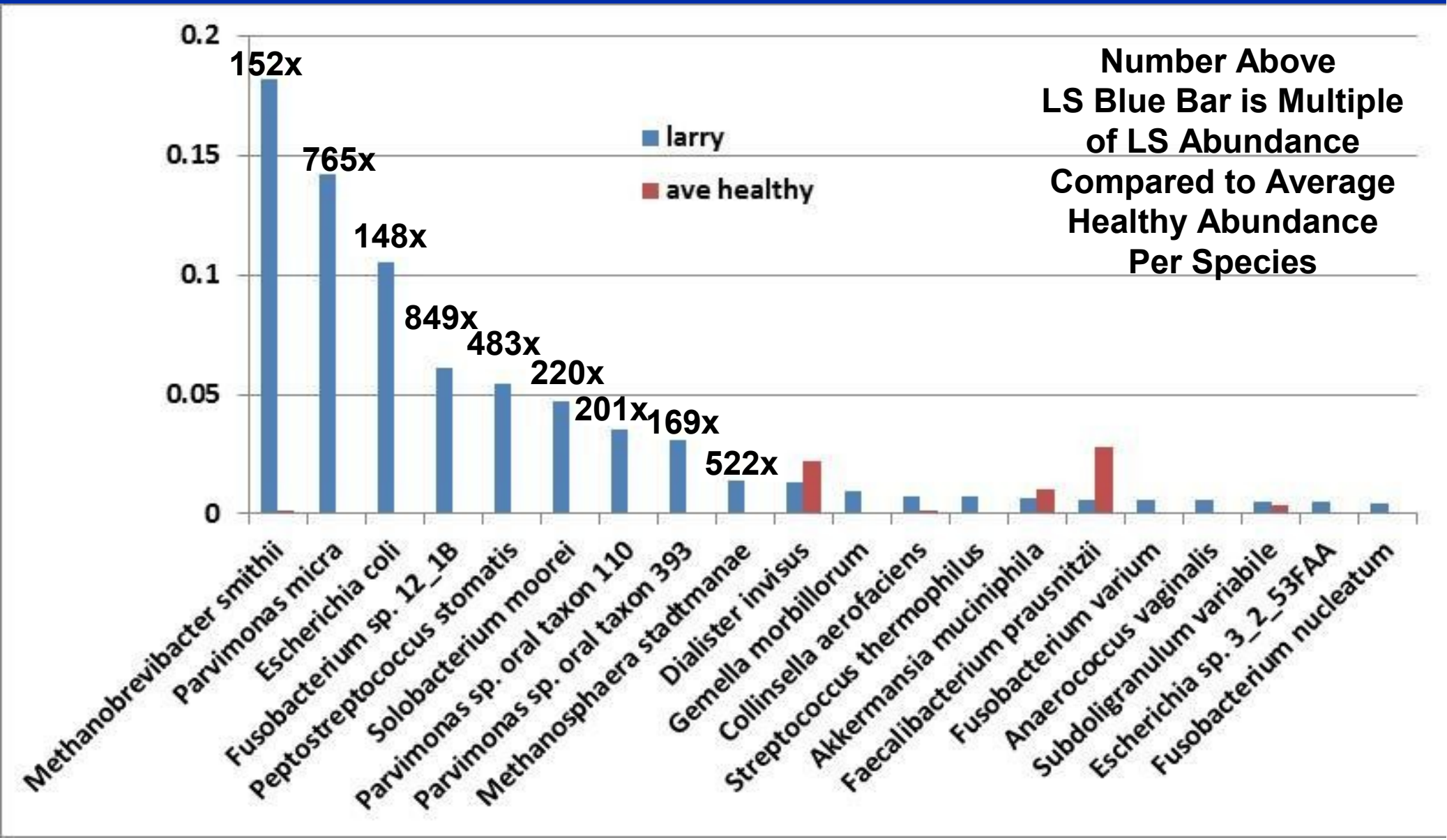
# Almost All Abundant Species (≥1%) in Healthy Subjects Are *Severely Depleted* in LS Gut




UCSD
UCIrvine

it²

# Top 20 Most Abundant Microbial Species In LS vs. Average Healthy Subject

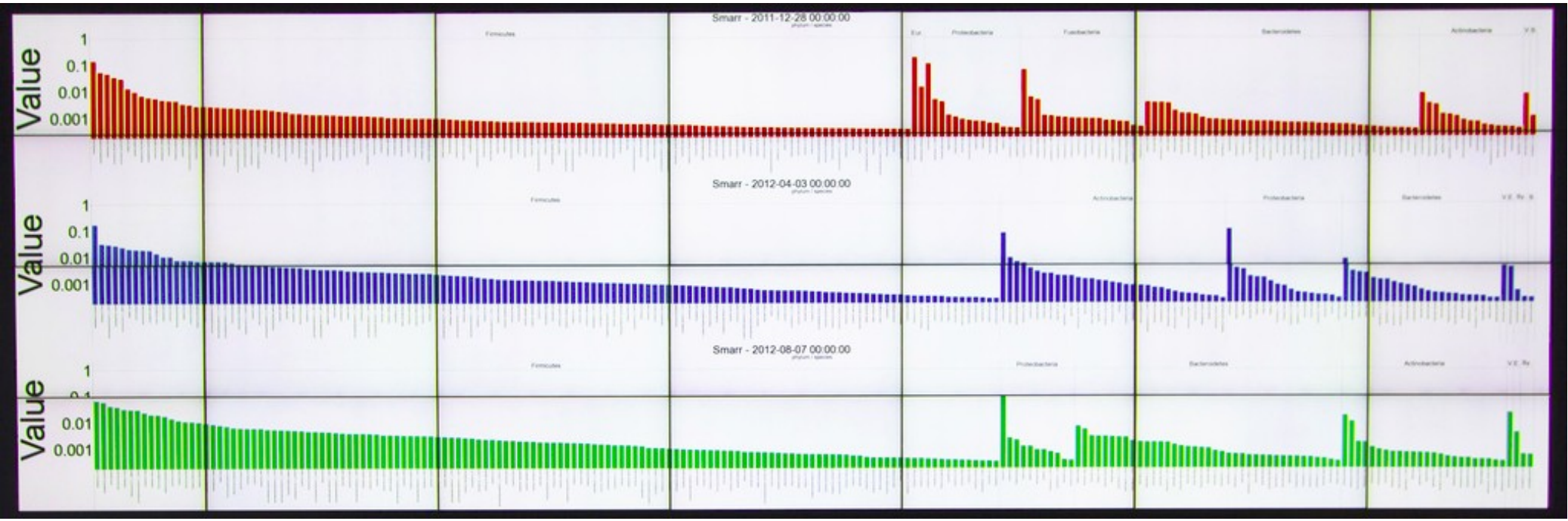



UCSD
UCIrvine

Source: Sequencing JCVI; Analysis Weizhong Li, UCSD
LS December 28, 2011 Stool Sample

it²

# 200 LS Gut Microbe Species at 3 Times 12/28/11, 4/3/12, 8/7/12




**Red is at Highest Value of CRP**
**Blue is the Day After End of Antibiotic/Prednisone Therapy**
**Green is Four Months Later**

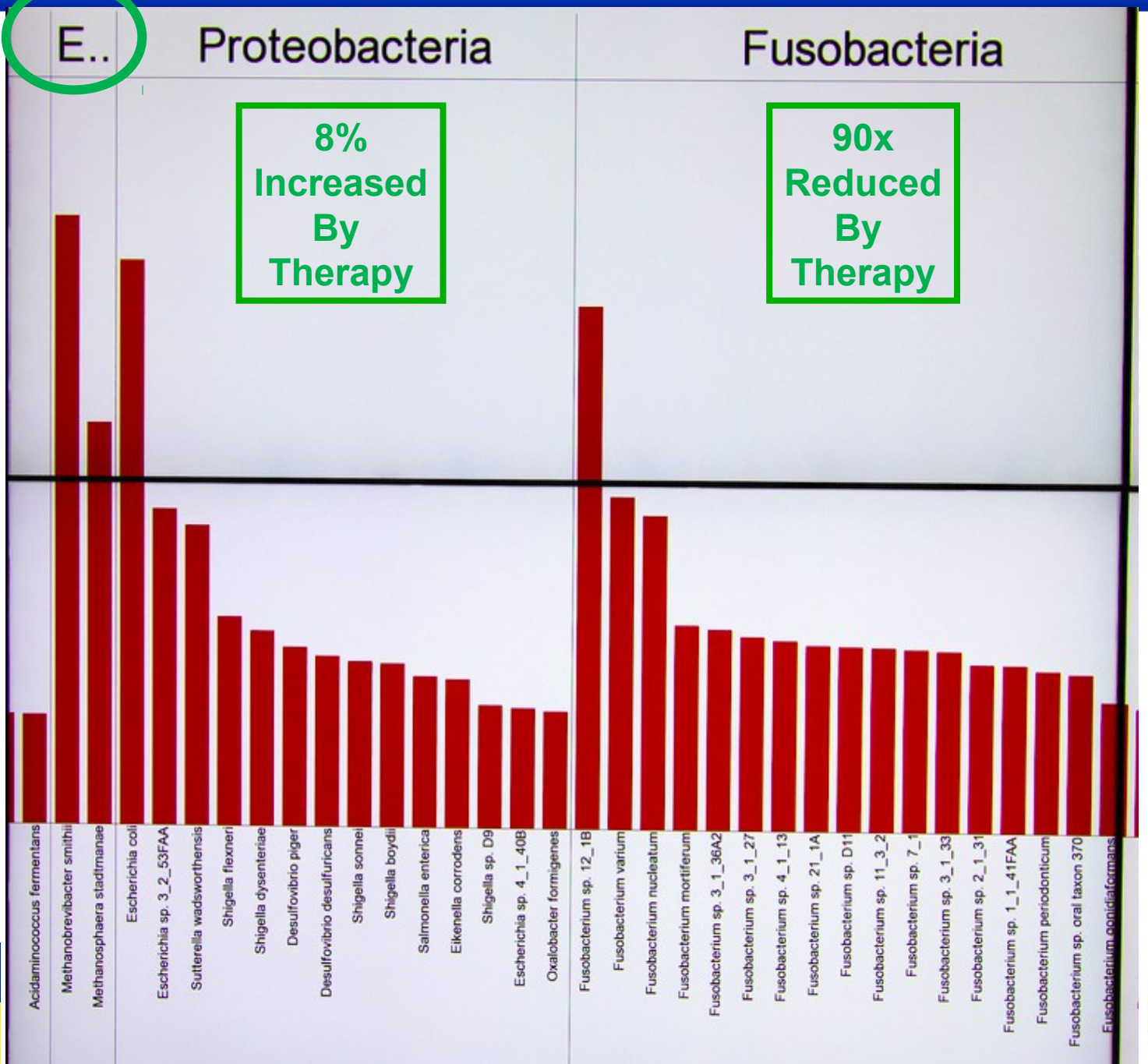
UCSD
UCIrvine

**Source: Photo of Calit2 64-megapixel VROOM**

it²

# Closeup of Uncommon LS Microbes 12/28/11 Stool Sample

**45x Reduced By Therapy**




Two separate research teams have found strikingly high concentrations of *Fusobacterium* in tumor samples collected from colorectal cancer patients.
October 18, 2011

UCSD
UCIrvine

it²

# DIY Systems Biology - Toward P4 Healthcare

Perspective

**Over 1000 Downloads So Far**

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

*Larry Smarr*

University of California, San Diego, CA, USA

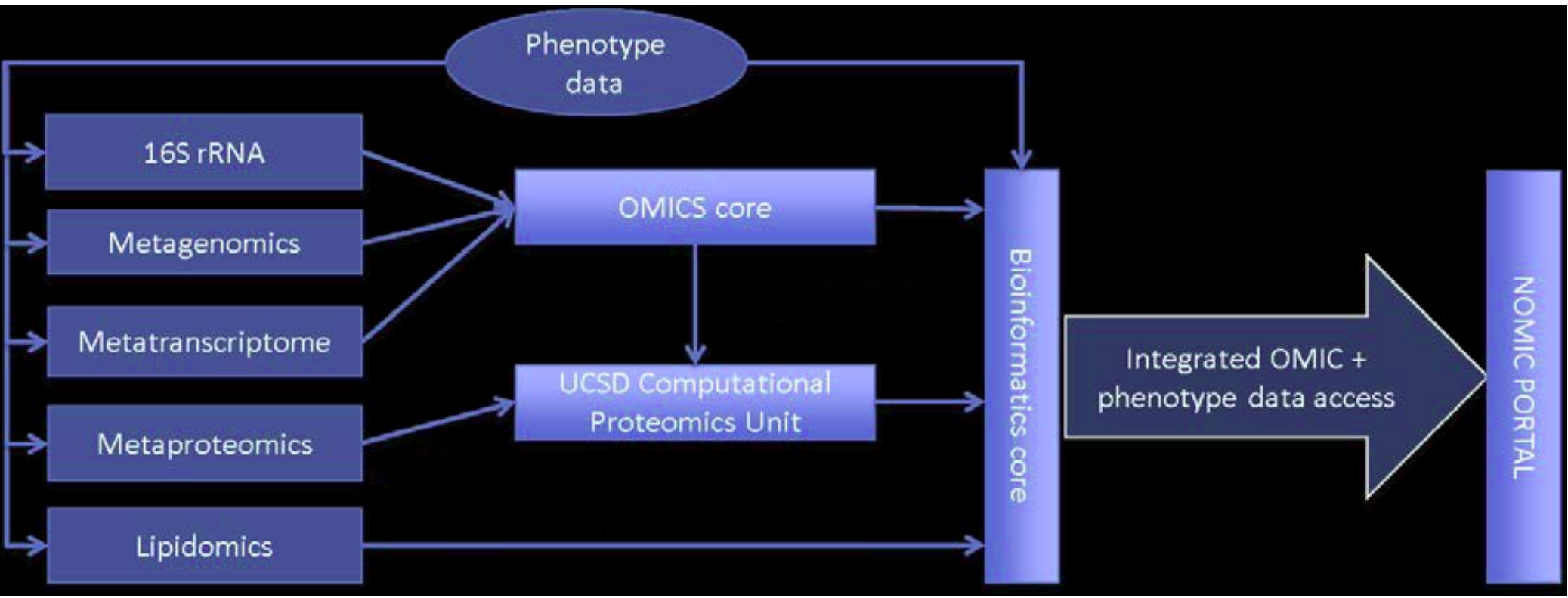
**Download pdfs from Journal:**

**http://onlinelibrary.wiley.com/doi/10.1002/biot.201100495/full**




Biotechnol. J. 2012, 7, 980–991

# Proposed UCSD Integrated Omics Pipeline

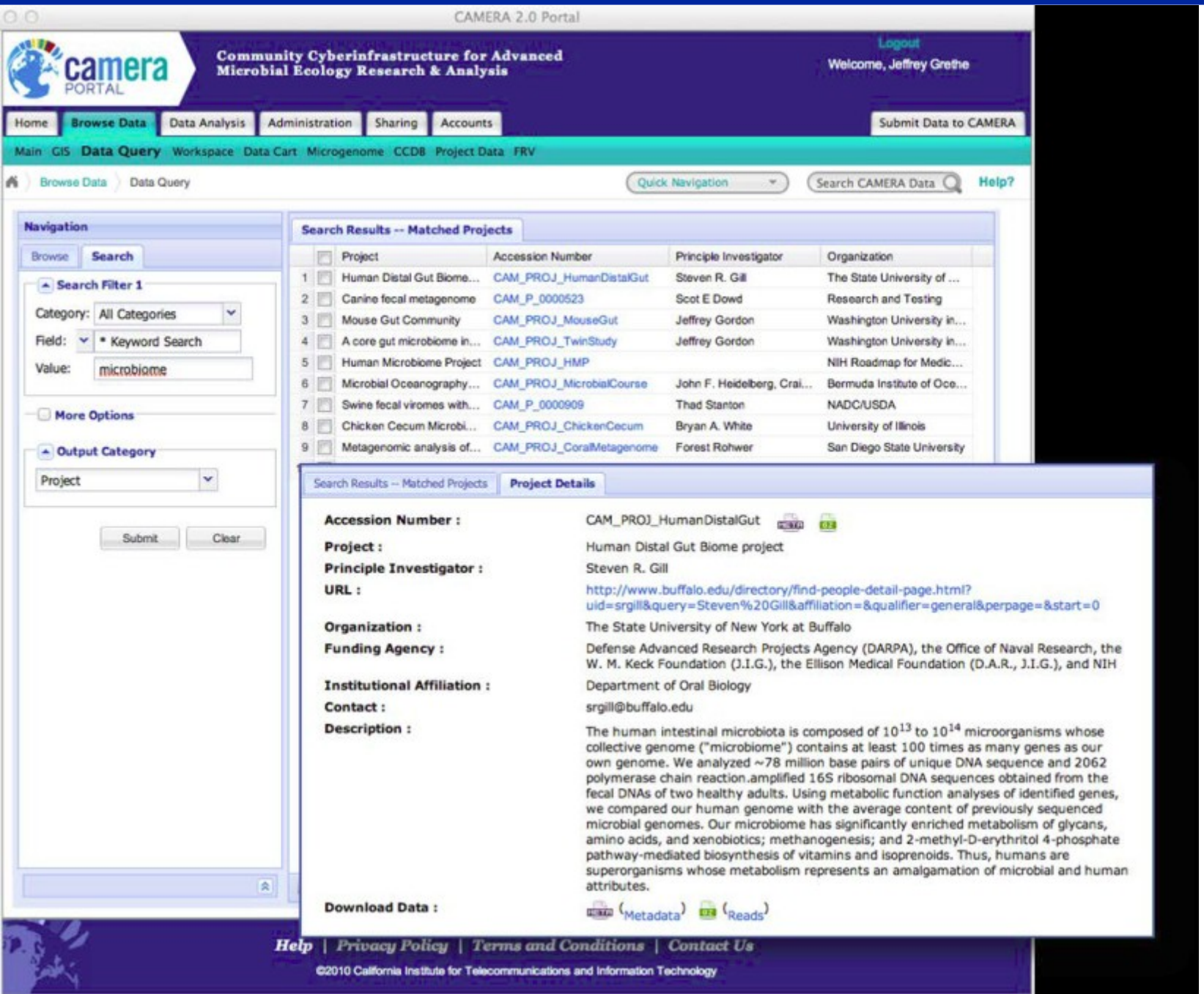



UCSD
UCIrvine

Source: Nuno Bandiera, UCSD

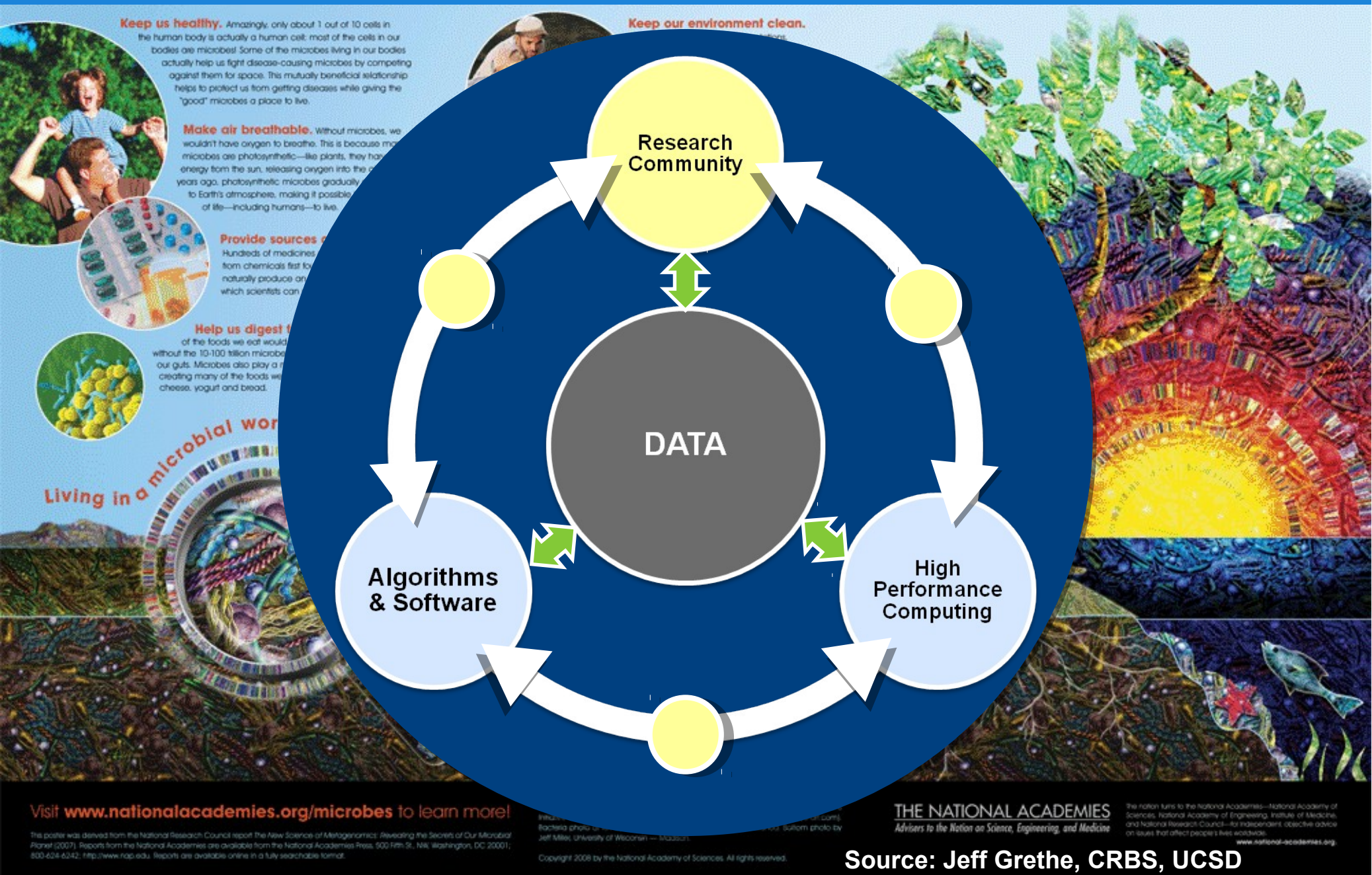
it²

# CAMERA as an Example for the NOMIC Portal Query/Hierarchy System




**Source:**
**Jeff Grethe,**
**CRBS, UCSD**

# Ecosystem to Amplify Understanding of Microbial Community Structure & Function




Source: Jeff Grethe, CRBS, UCSD

# Access to Computing Resources Tailored by User's Requirements and Resources










UCSD Triton



NSF/SDSC Gordon



NSF/SDSC Trestles



NSF/TACC Lonestar



NSF/TACC Ranger




NSF/RCAC Steele

**Infrastructure Services Extend CAMERA Computations to 3rd Party Compute Resources**

Windows Azure

amazon web services

UCSD

UCIrvine

Source:
Jeff Grethe,
CRBS, UCSD

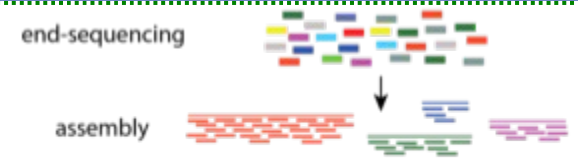
EAGER: Multi-Domain, Workflow-Driven Computation System for Microbial Ecology Research and Analysis

# PhyloMETARE

## *Explore, Analyze & Compare Transcriptomes*

D

***Data***




***Data Analysis***

Analyze Sequence Data




phylogeny | metabolic reconstruction | comparative genomics | population genetics | compare environments

**Source:**
**Jeff Grethe,**
**CRBS, UCSD**

METAREP
JCVI Metagenomics Reports

website http://www.jcvi.org/phylo-metarep
source code http://github.com/jcvi/METAREP
blog http://blogs.jcvi.org/tag/metarep
contact metarep-support@jcvi.org

**Dash Board**

Phylo-Metagenomics Reports (v1.3.1-beta)

JCVI supports the Gordon and Betty Moore Foundation's (GBMF) **Marine Microbiology Initiative (MMI)** with Phylo-Metagenomics Reports (**Phylo-METAREP** ), a dedicated online resource to make large volumes of annotated marine transcriptomes amenable for exploratory data analysis.

Phylo-METAREP provides a suite of **high-performance** web based tools to **view, query, browse** and **compare** annotated transcriptomes in realtime.

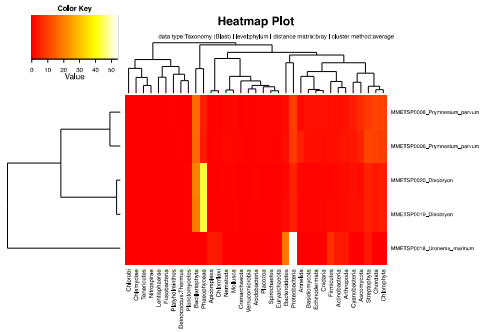
It enables scientists to **slice and dice** data and **compare multiple datasets/conditions** at various functional and taxonomic levels applying statistical tests as well as hierarchical clustering, multidimensional scaling and heatmaps.
Additional information:

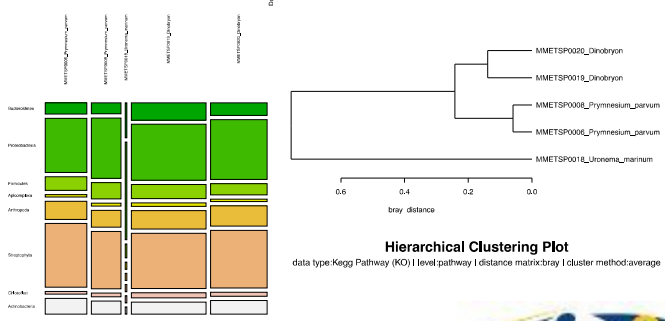
Flyer | Manual | Open Source | Publication | Open Virtualization Format

camera
Marine Microbial Ecology

Login
Username: moore54
Password: •••••
Remember me for 2 weeks
Forgot password?
Login

## Diverse Analysis Functions




UCSD
UCIrvine

GORDON AND BETTY MOORE FOUNDATION

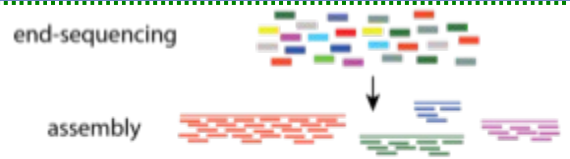
**A new community resource for comparing complex microbial gene expression patterns**

it2

# VIROME

## *Explore, Analyze &Compare Viral Genomes/Metagenomes*

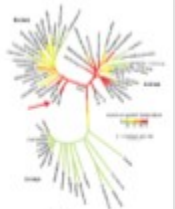
***Data***



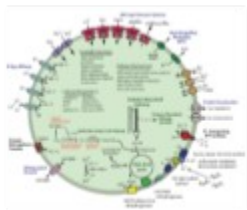

camera
Microbial Ecology

***Data Analysis***

Analyze Sequence Data



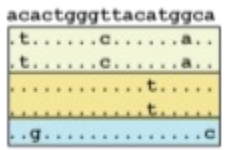
phylogeny



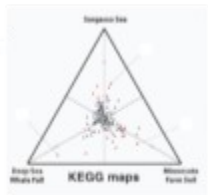
metabolic reconstruction



comparative genomics




population genetics



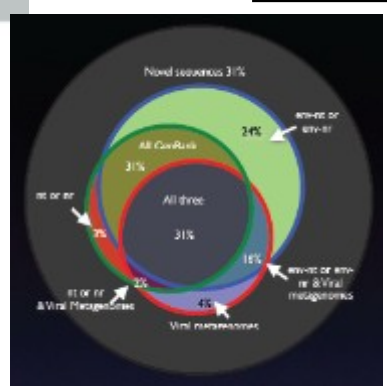

compare environments

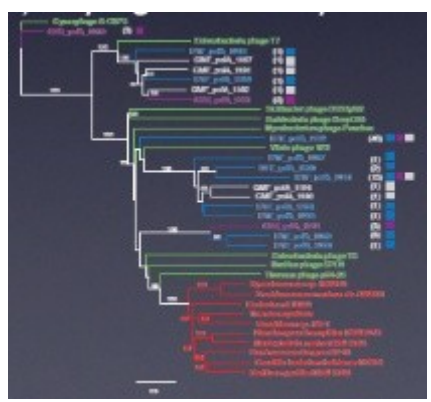
**Resource for analysis of viral metagenomes**

**Source: Jeff Grethe, CRBS, UCSD**




**Diverse Analysis Functions**

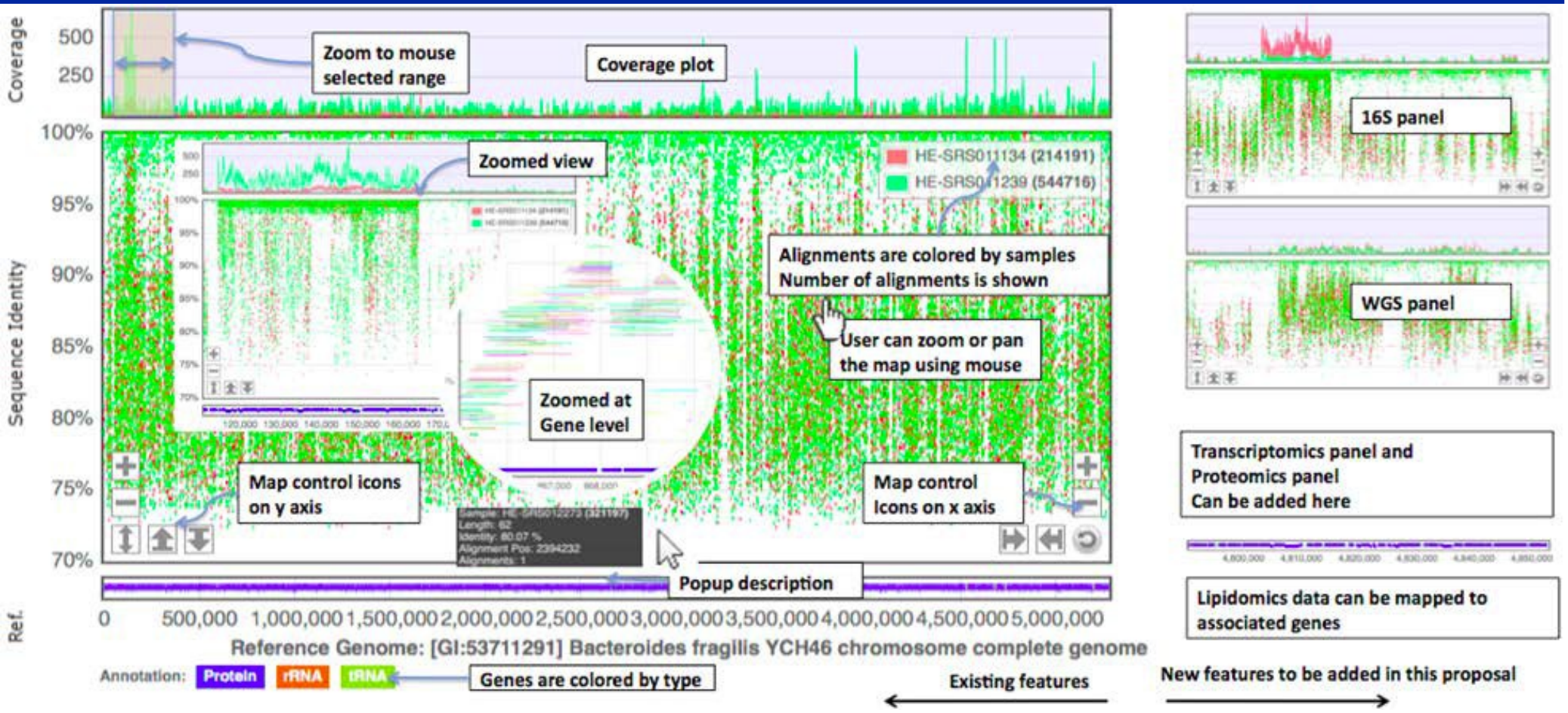


GORDON AND BETTY MOORE FOUNDATION

it2

# Fragment Recruitment Viewer (FRV) Interface

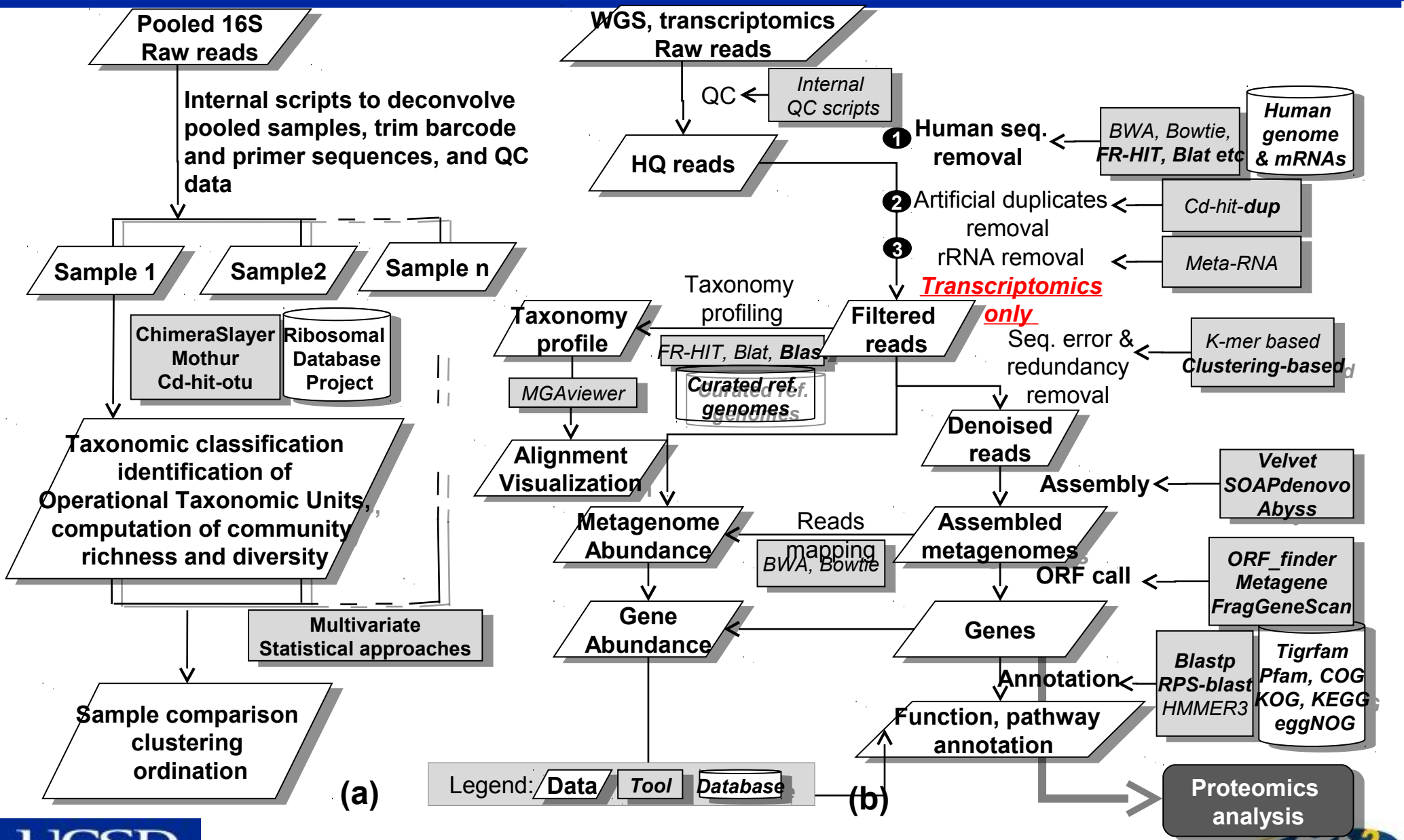



**X-axis is the genome coordinate, and y-axis is alignment identity (%). The top is genome coverage. The bottom shows genes or other genomic features. Users can zoom, resize, and pan the plot by mouse or using icons at corners in a similar way as Google Maps. Right illustrates new functions and interface to be implemented in order to handle multiple integrated omics data types by using multiple synchronized FRV panels.**

UCSD
UCIrvine

Source: Weizhong Li, UCSD

# Combined 16S, Metagenomics and Metatranscriptomics Pipeline

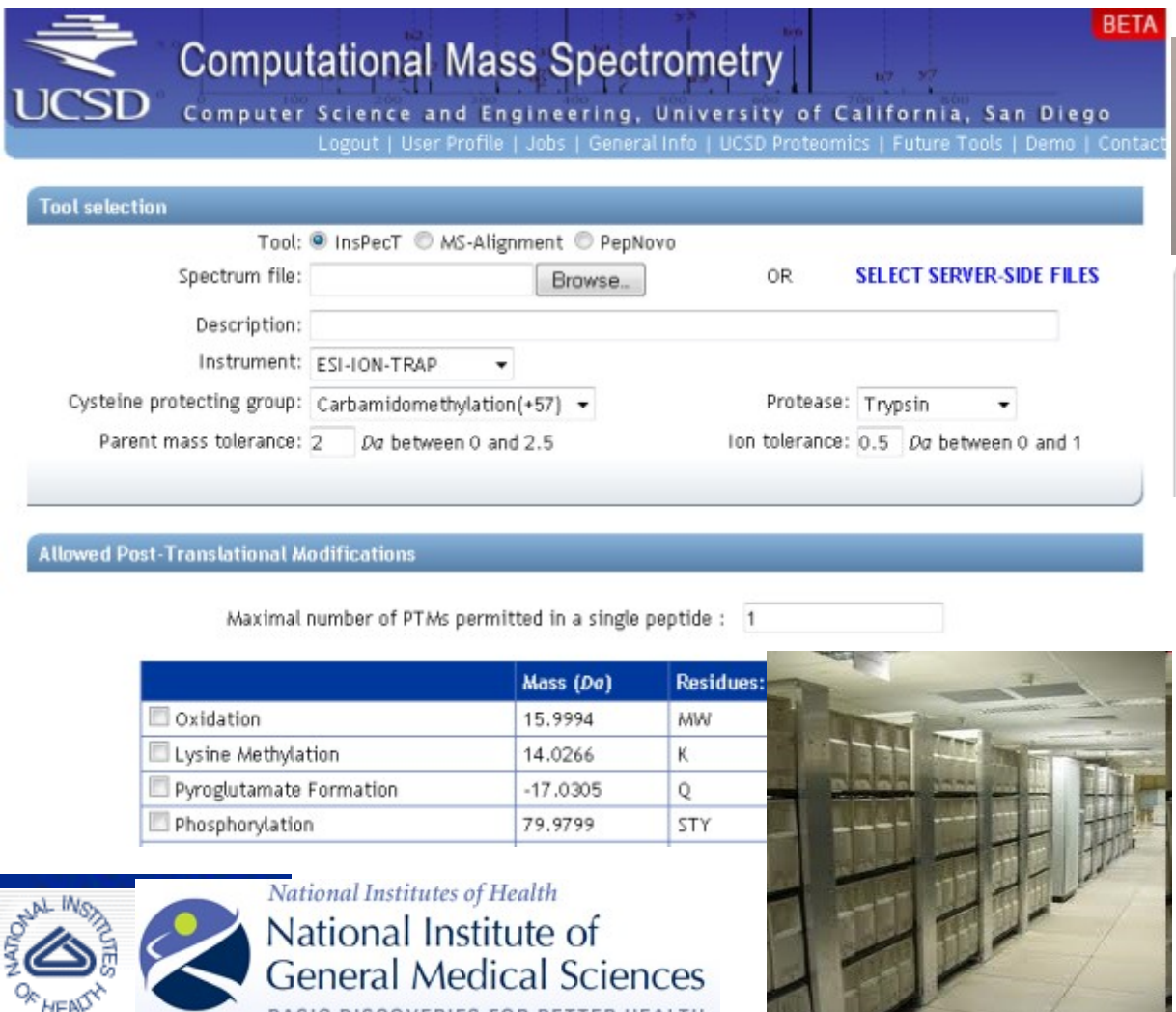



Source: Weizhong Li, UCSD

# 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**







**Statistics**

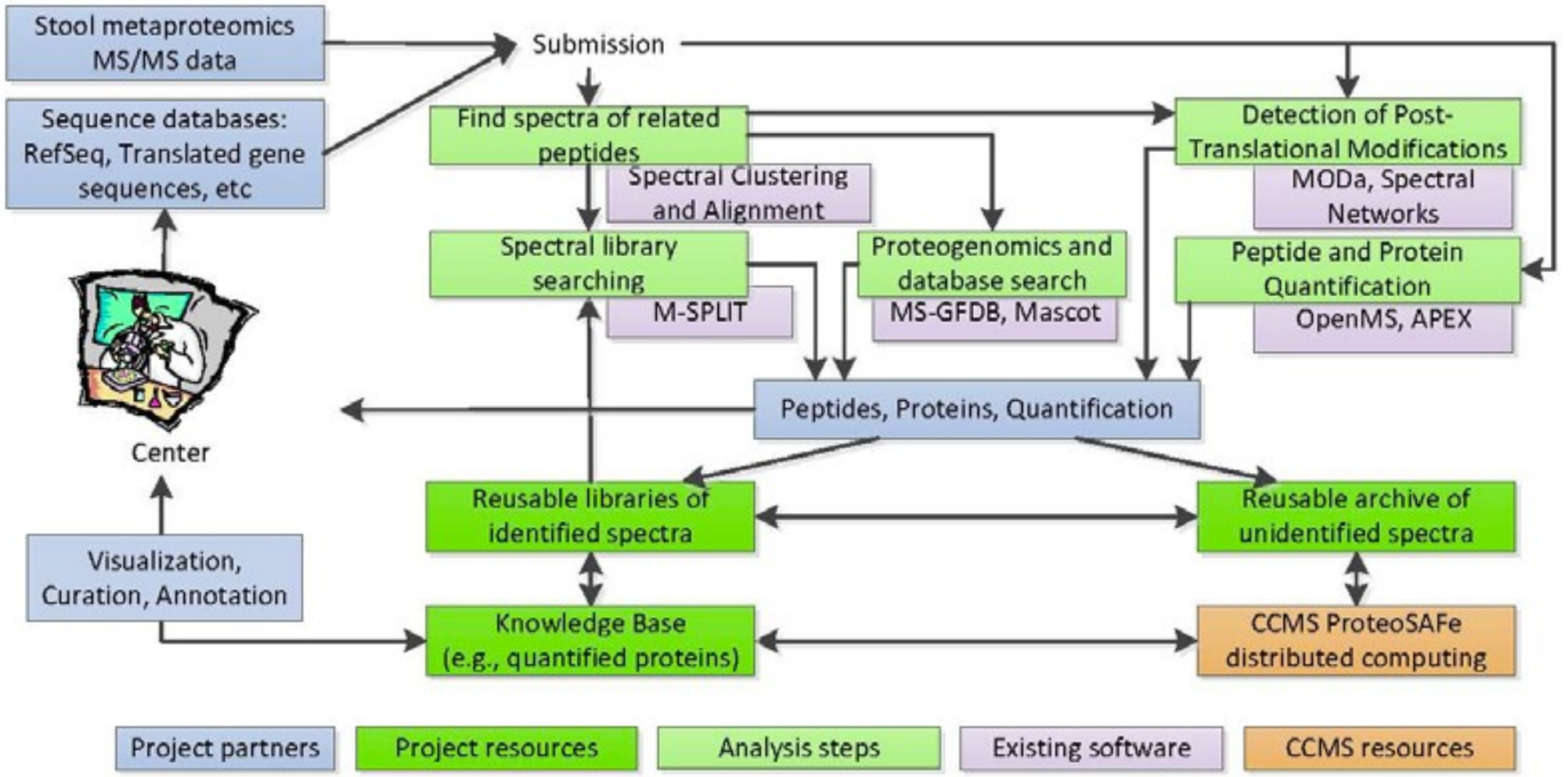
| | |
|---|---|
| Members | 1544 |
| Groups & Projects | 355 |
| Data Size (uncompressed) | 23.3 TB |
| Data Files | 13,264,659 |



**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

# Metaproteomics Analyses Work Flow

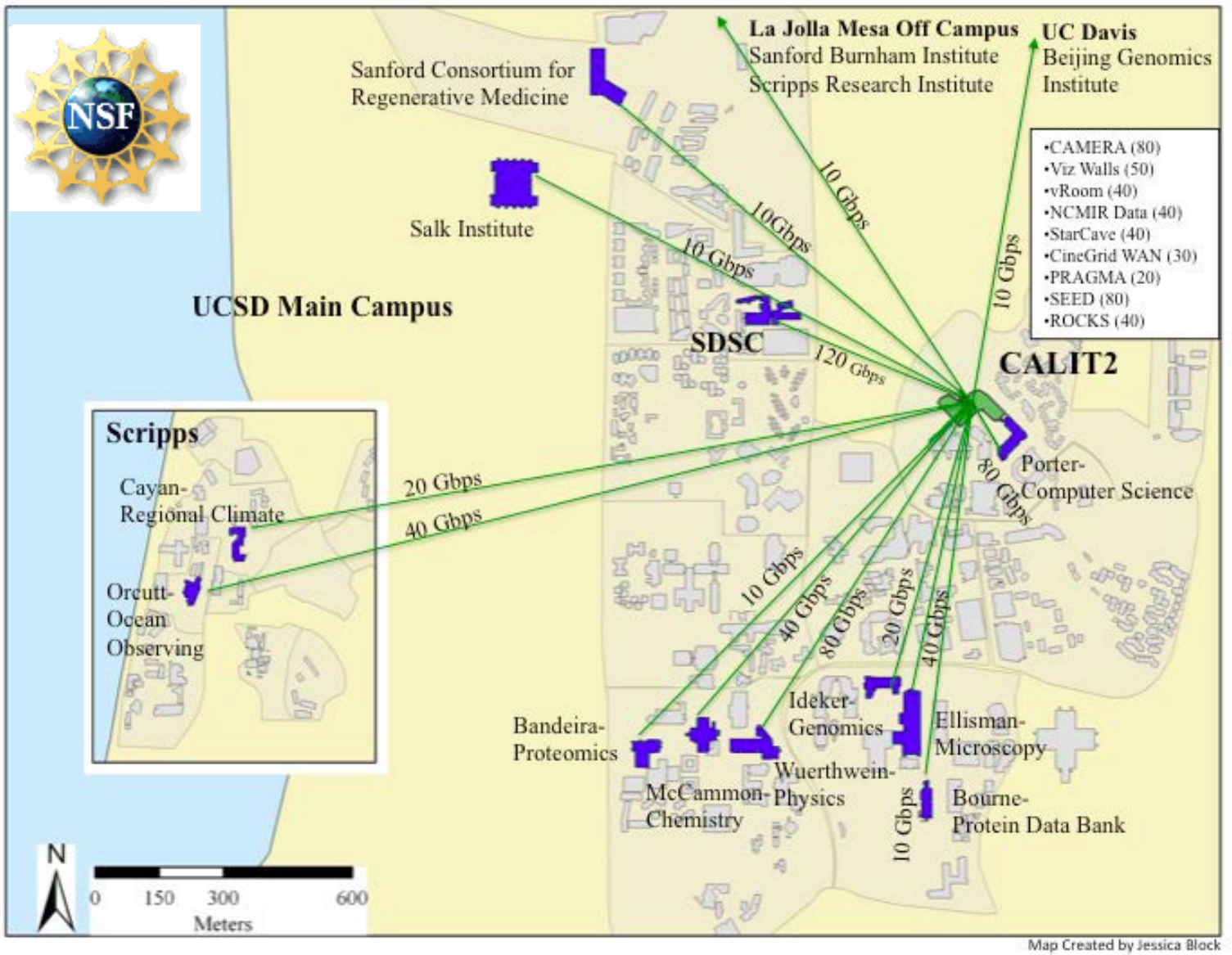



UCSD
UCIrvine

Source: Nuno Bandeira, UCSD

it²

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

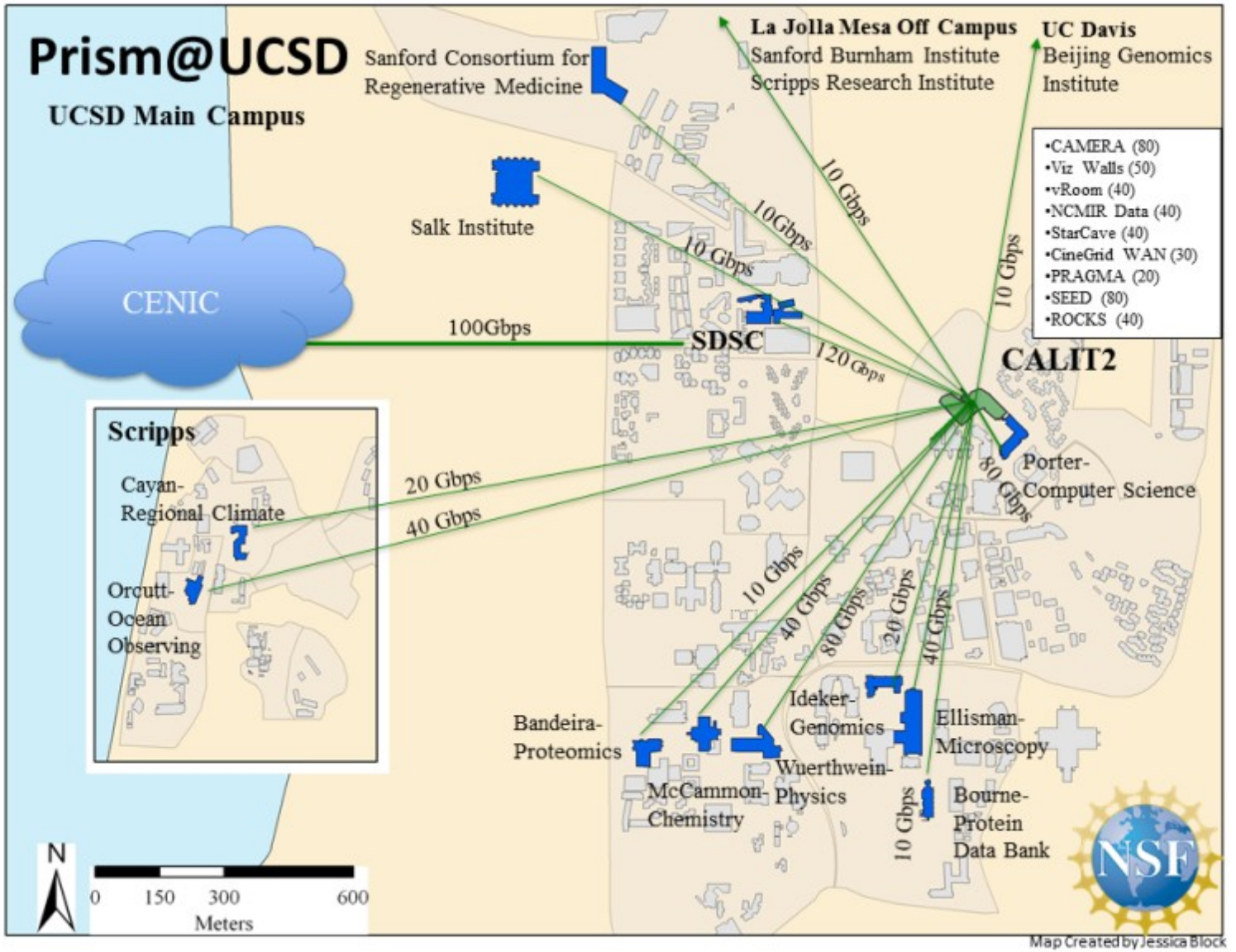



UCSD UCIrvine

**Phil Papadopoulos, SDSC, Calit2, PI**

SDSC it²

# PRISM@UCSD Enables
## Connection to Remote Campus Compute & Storage Clusters




UCSD
UCIrvine