“Discovering the Other 90% of our Human Superorganism”

Remote Video Lecture to
The eResearch Australasia Conference 2014
Melbourne, Australia
October 28, 2014

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
http://lsmarr.calit2.net

Abstract

The human body is host to 100 trillion microorganisms, ten times the number of cells in the human body, and these microbes contain 100 times the number of DNA genes that our human DNA does. The microbial component of our “superorganism” is comprised of hundreds of species with immense biodiversity. Thanks to the National Institutes of Health’s Human Microbiome Program researchers have been discovering the states of the human microbiome in health and disease. To put a more personal face on the “patient of the future,” I have been collecting massive amounts of data from my own body over the last five years, which reveals detailed examples of the episodic evolution of this coupled immune-microbial system. An elaborate software pipeline, running on high performance computers, reveals the details of the microbial ecology and its genetic components. As a result of discovering the "missing" 90% of our bodies, we can look forward to revolutionary changes in medical practice over the next decade.

A Decade of eResearch Partnering With Australia

AUSTRALIAN AMERICAN LEADERSHIP DIALOGUE

aarnet Australia's Academic and Research Network



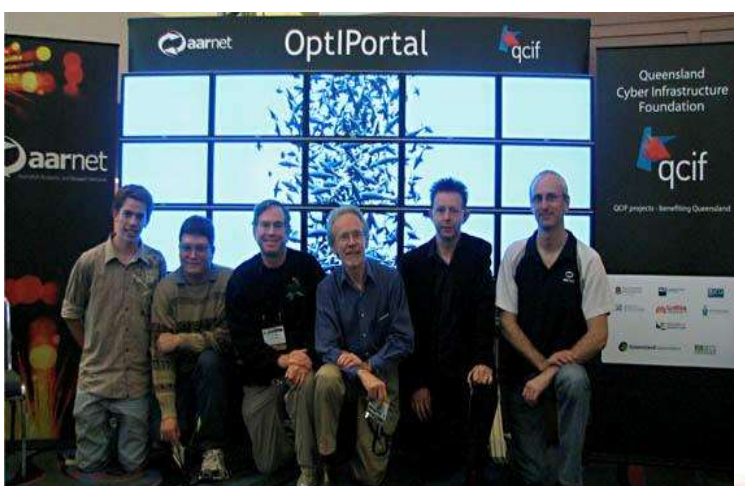
Phil Scanlan, AALD



University of Melbourne



David Abramson, Monash University

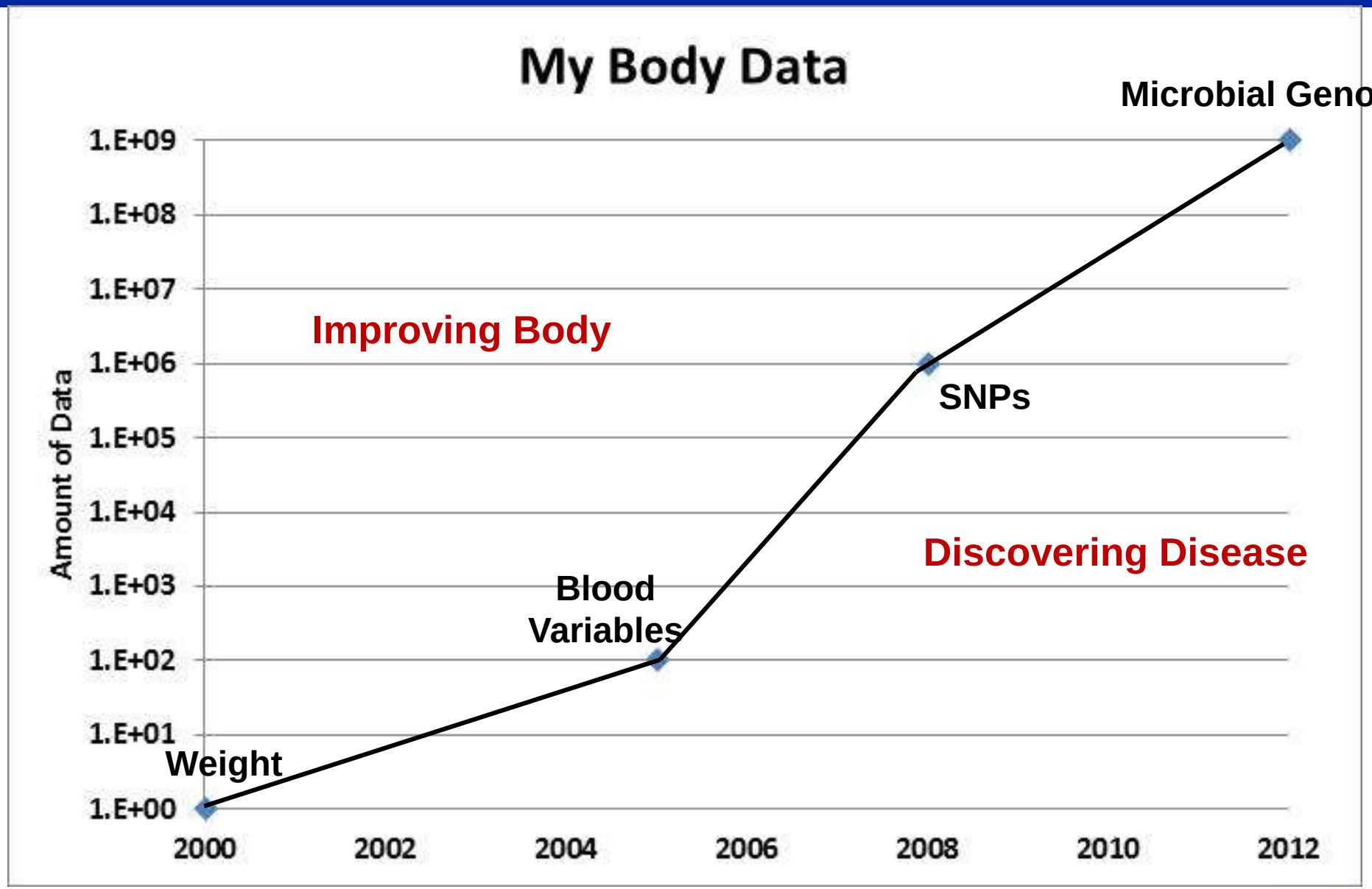


Bernard Pailthorpe, UQ

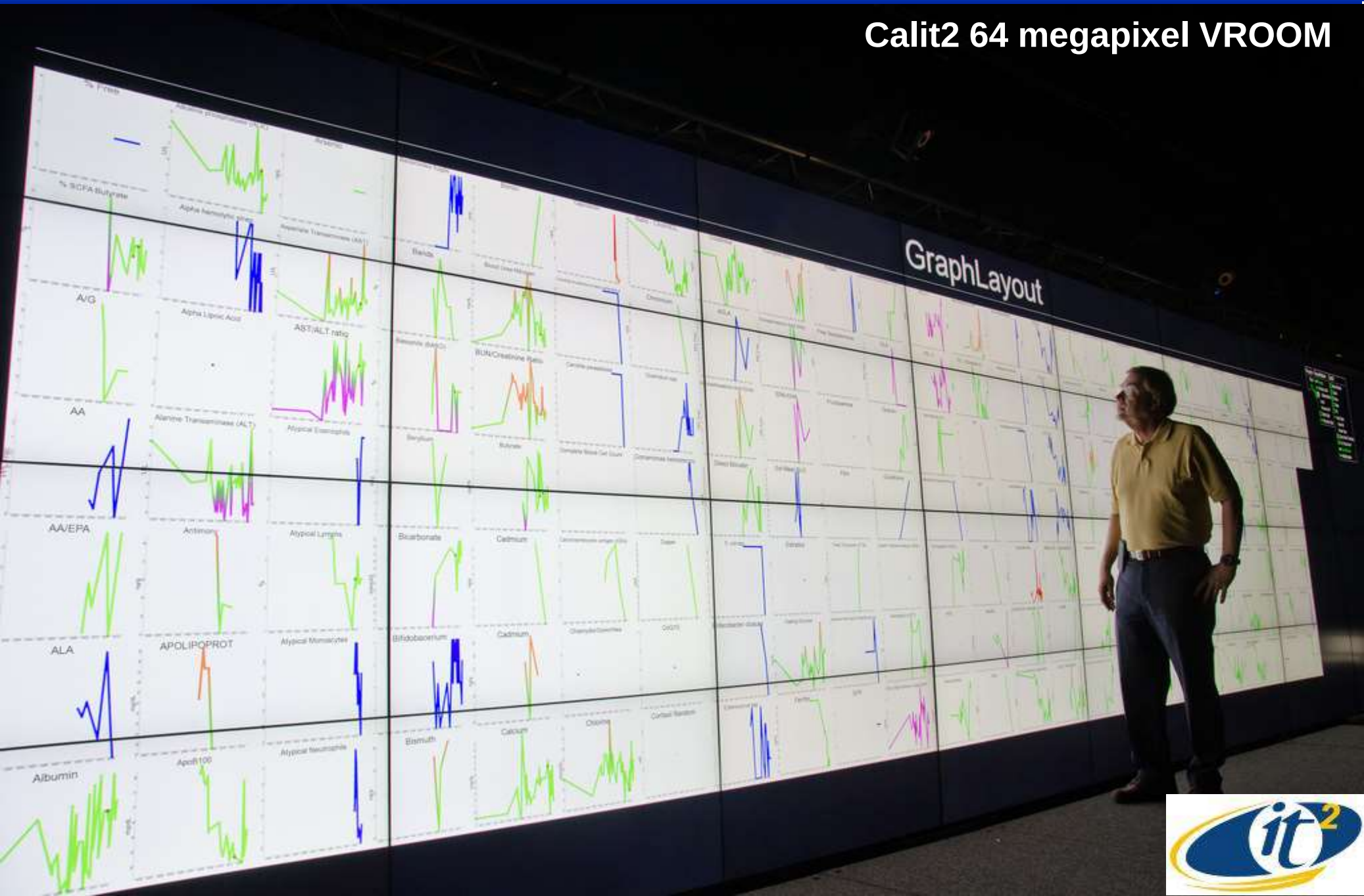
Chris Hancock, aarnet

it²

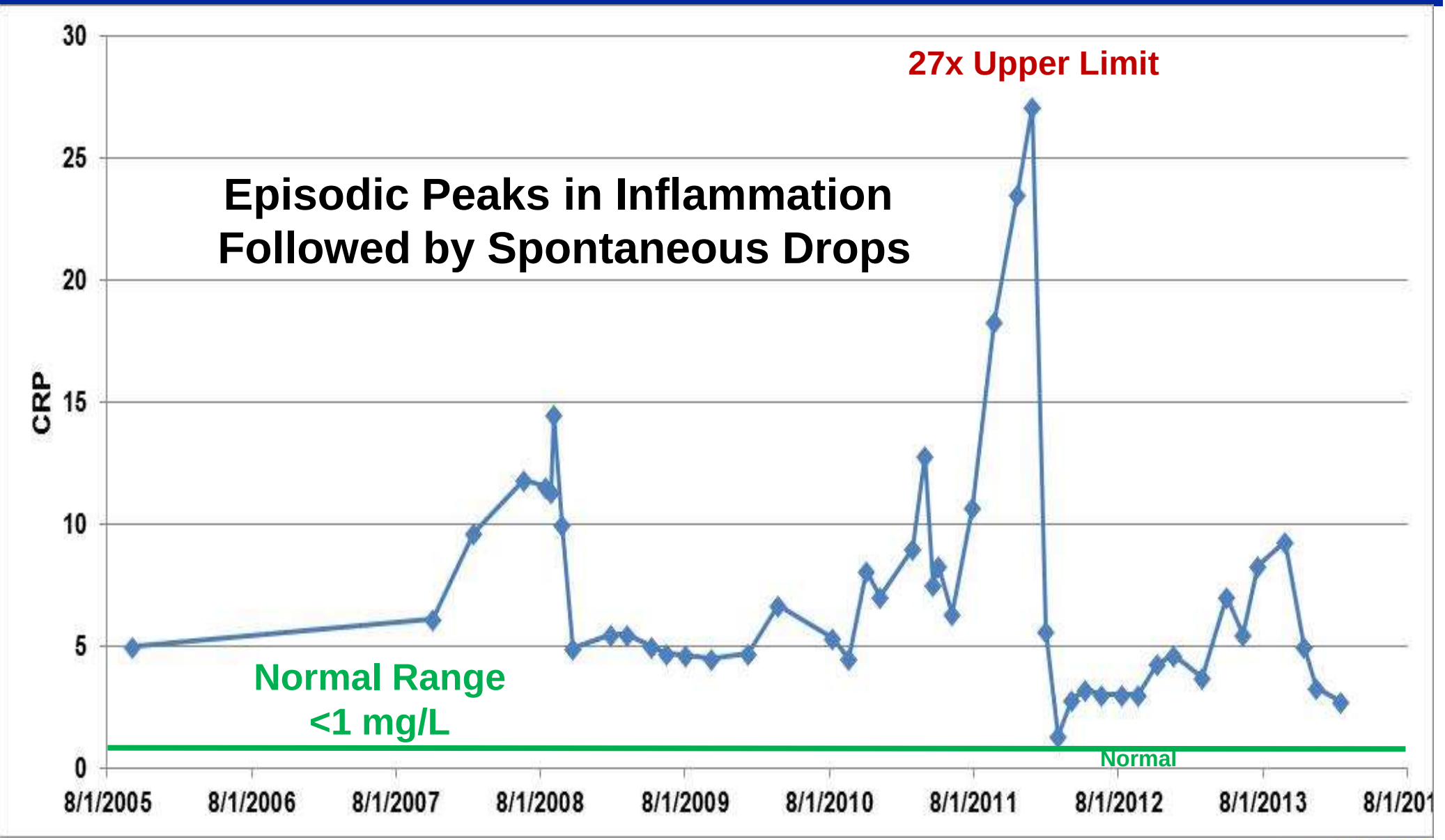
From One to a Billion Data Points Defining Me: The Exponential Rise in Body Data in Just One Decade



Visualizing Time Series of 150 LS Blood and Stool Variables, Each Over 5-10 Years

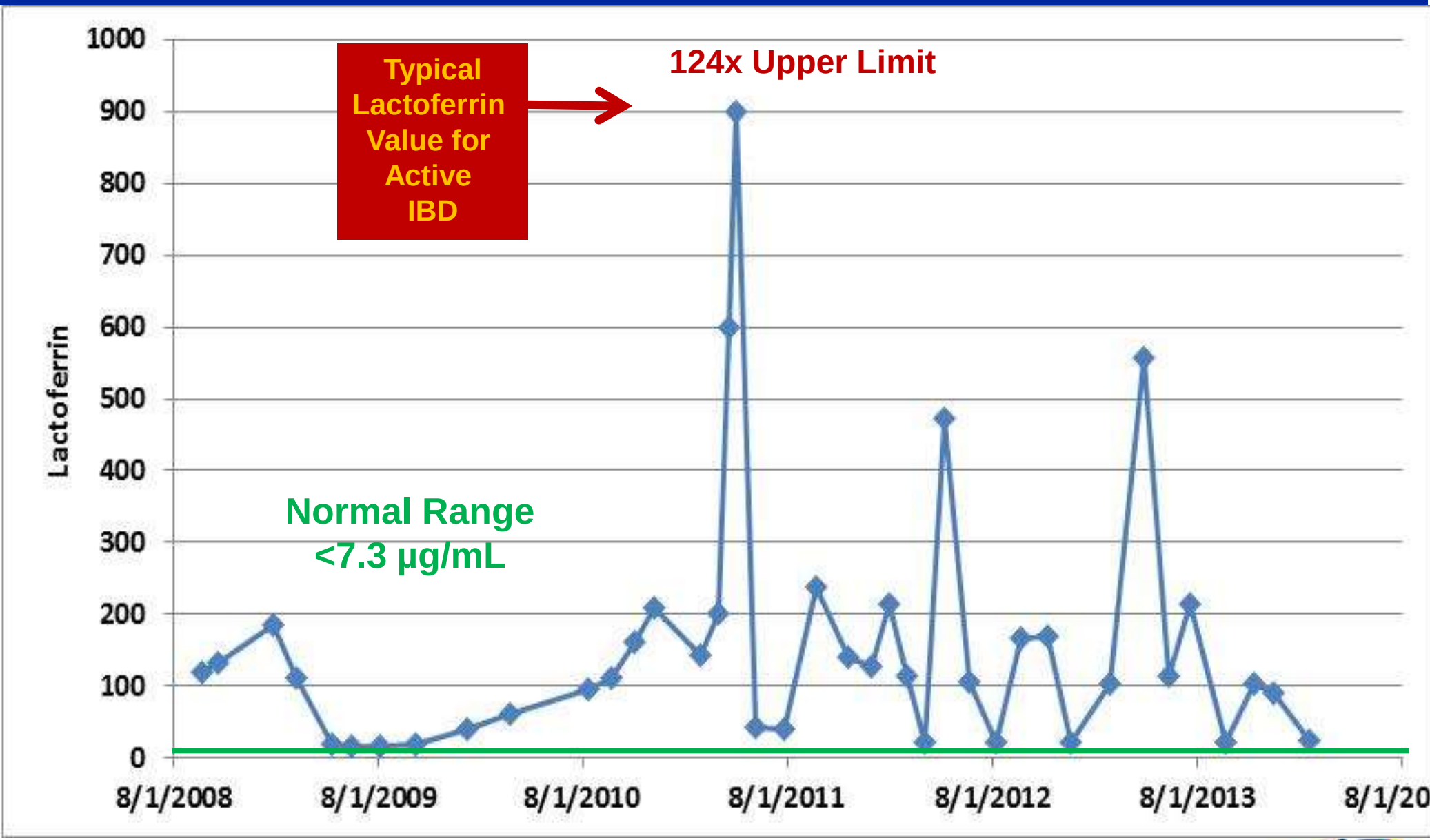


Only One of My Blood Measurements Was Far Out of Range--Indicating Chronic Inflammation



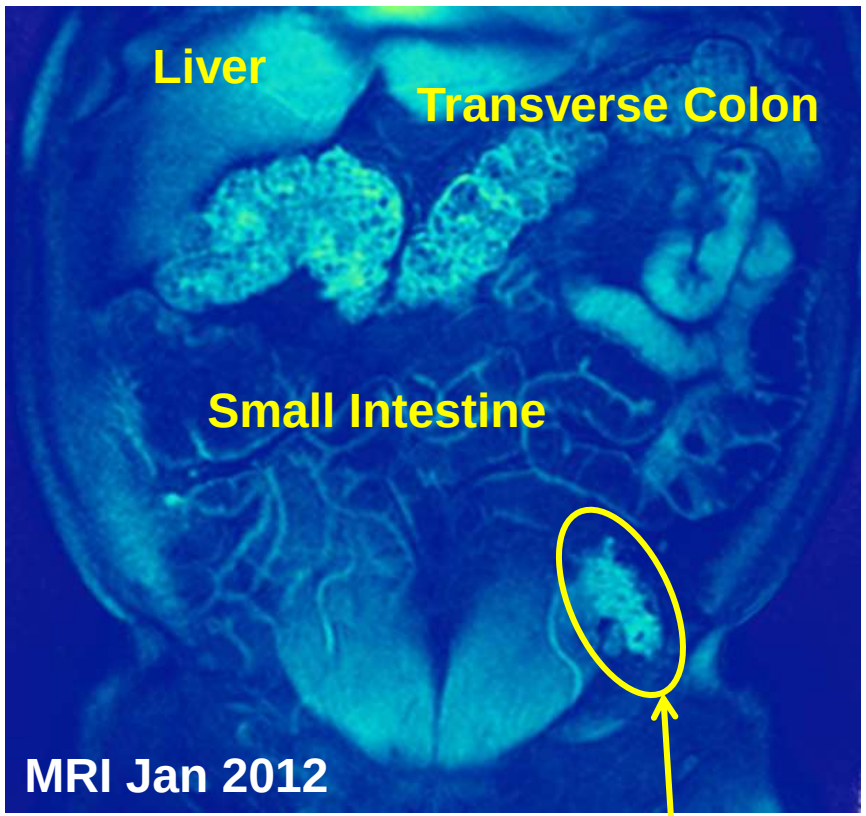
Complex Reactive Protein (CRP) is a Blood Biomarker for Detecting Presence of Inflammation

Adding Stool Tests Revealed Oscillatory Behavior in an Immune Variable



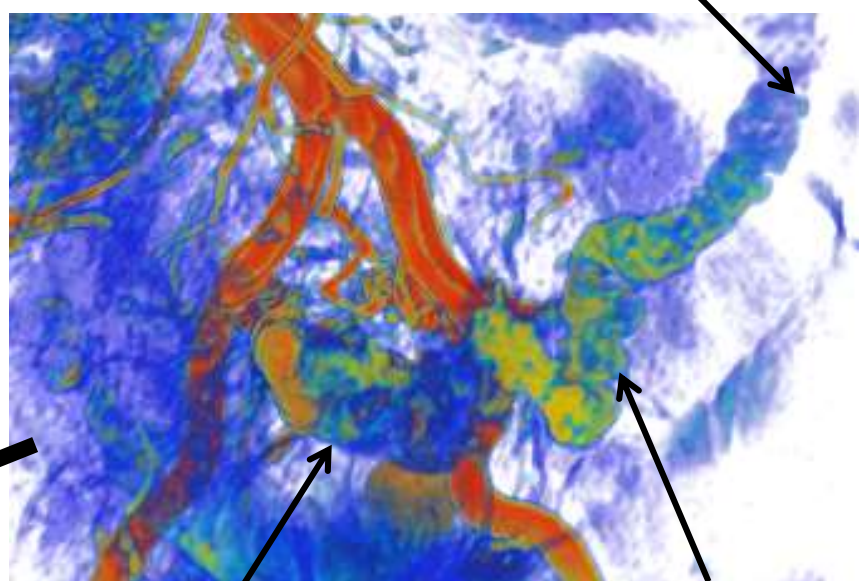
Lactoferrin is a Protein Shed from Neutrophils - An Antibacterial that Sequesters Iron

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

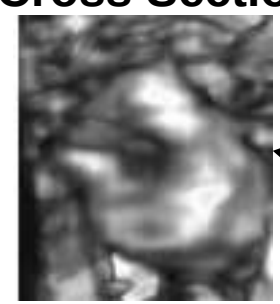
Descending Colon



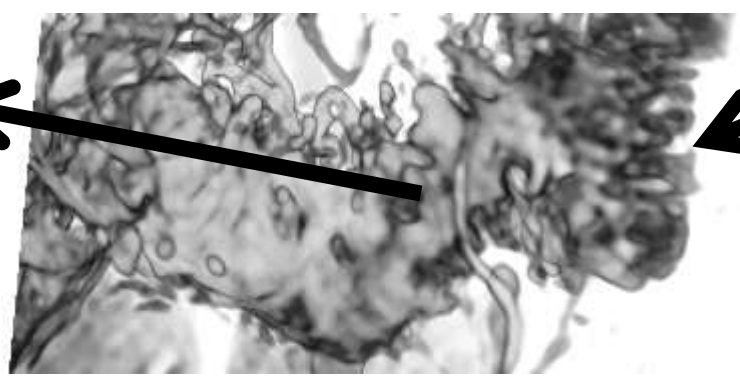
Major Kink

Sigmoid Colon Threading Iliac Arteries

Cross Section



Diseased Sigmoid Colon



Why Did I Have an Autoimmune Disease like IBD?

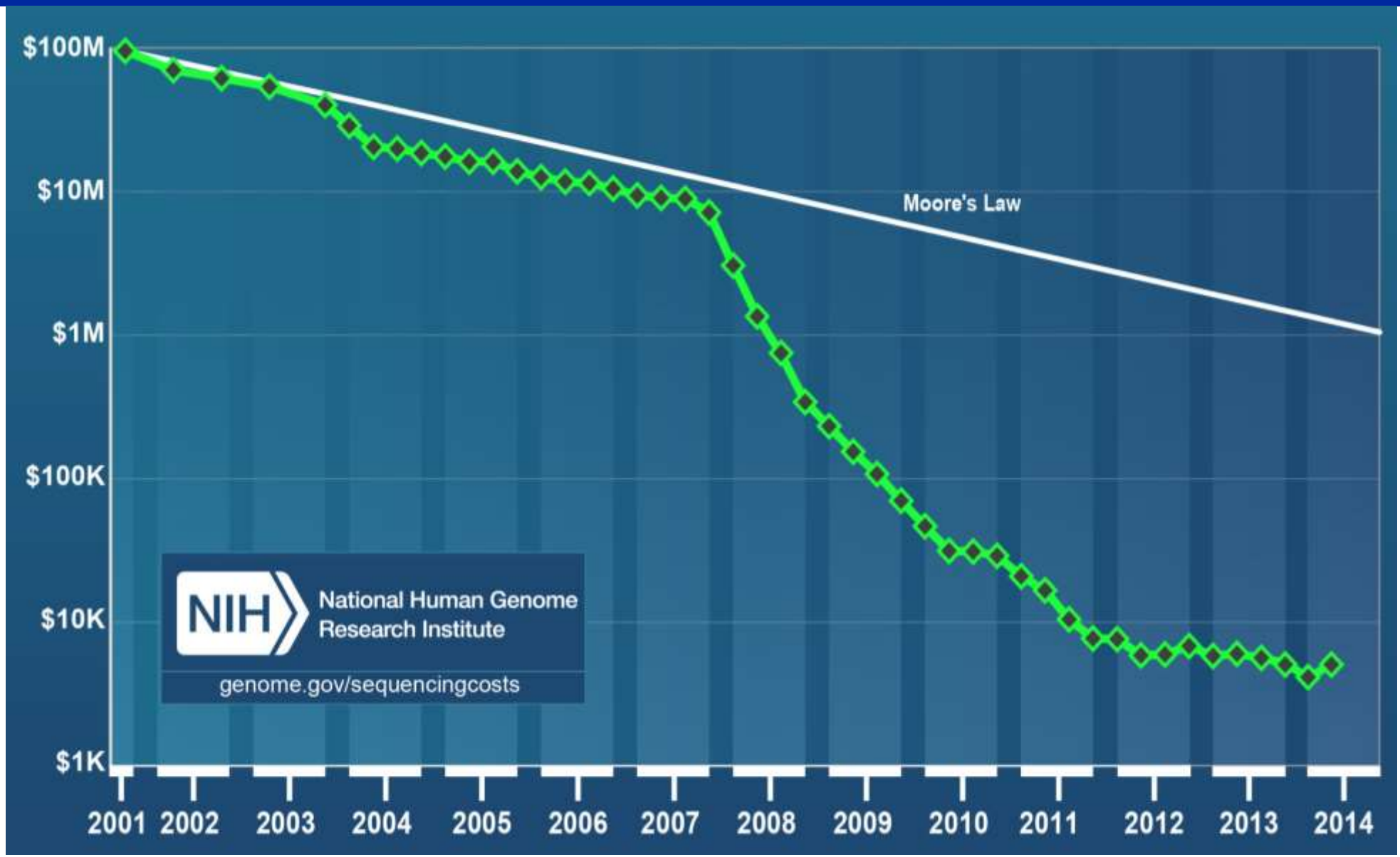
Despite decades of research,
the etiology of Crohn's disease
remains unknown.
Its pathogenesis may involve
a complex interplay between
host genetics,
immune dysfunction,
and microbial or environmental factors.

--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)

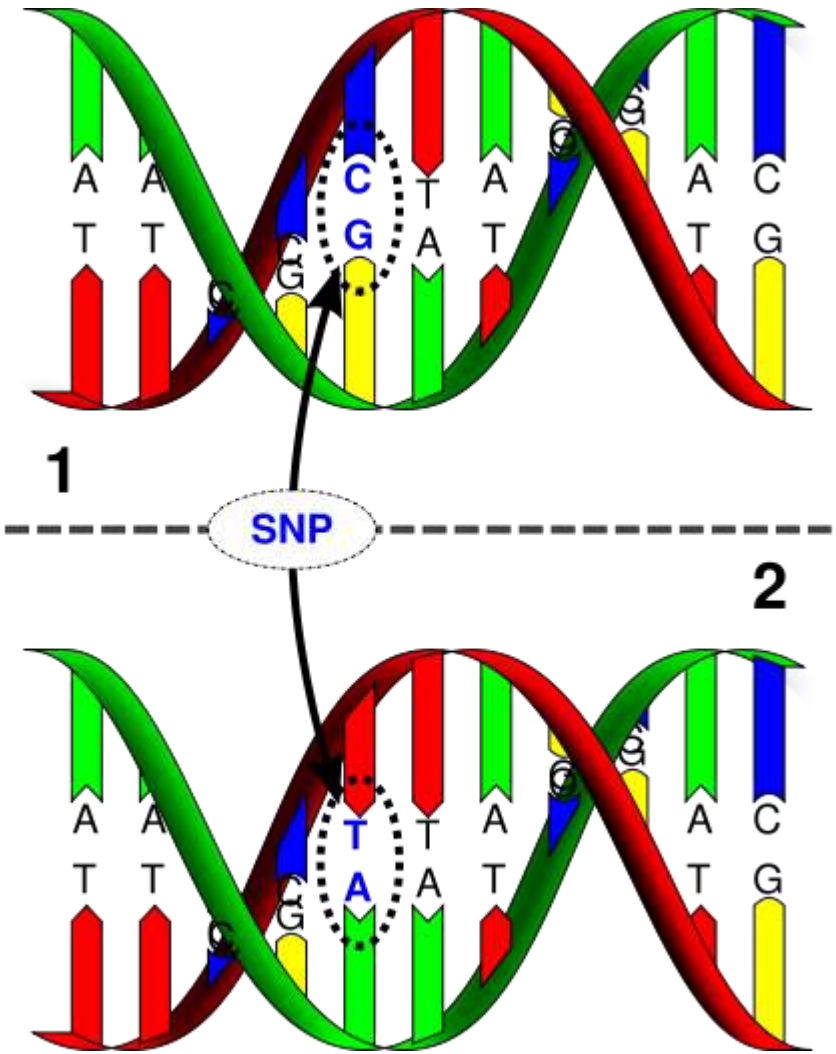
The Cost of Sequencing a Human Genome Has Fallen Over 10,000x in the Last Ten Years



This Has Enabled Sequencing of Both Human and Microbial Genomes

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

Person A



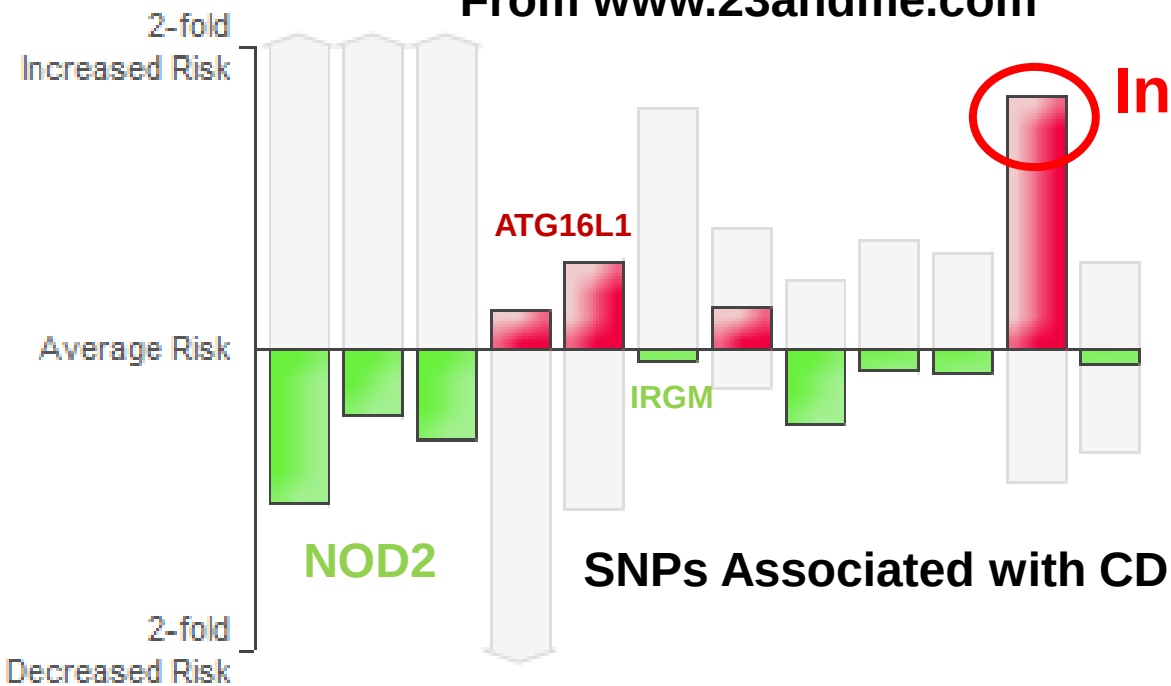
Person B

SNPs Occur Every 100 to 300 Bases Along Human DNA

www.23andme.com Tracks One Million SNPs

I Found I Had One of the Earliest Known SNPs Associated with Crohn's Disease

From www.23andme.com



Polymorphism in Interleukin-23 Receptor Gene — 80% Higher Risk of Pro-inflammatory Immune Response rs1004819

23andme is Collecting 10,000 IBD Patient's SNPs to Map Into the 163 Known SNPs Associated with IBD

News and Views

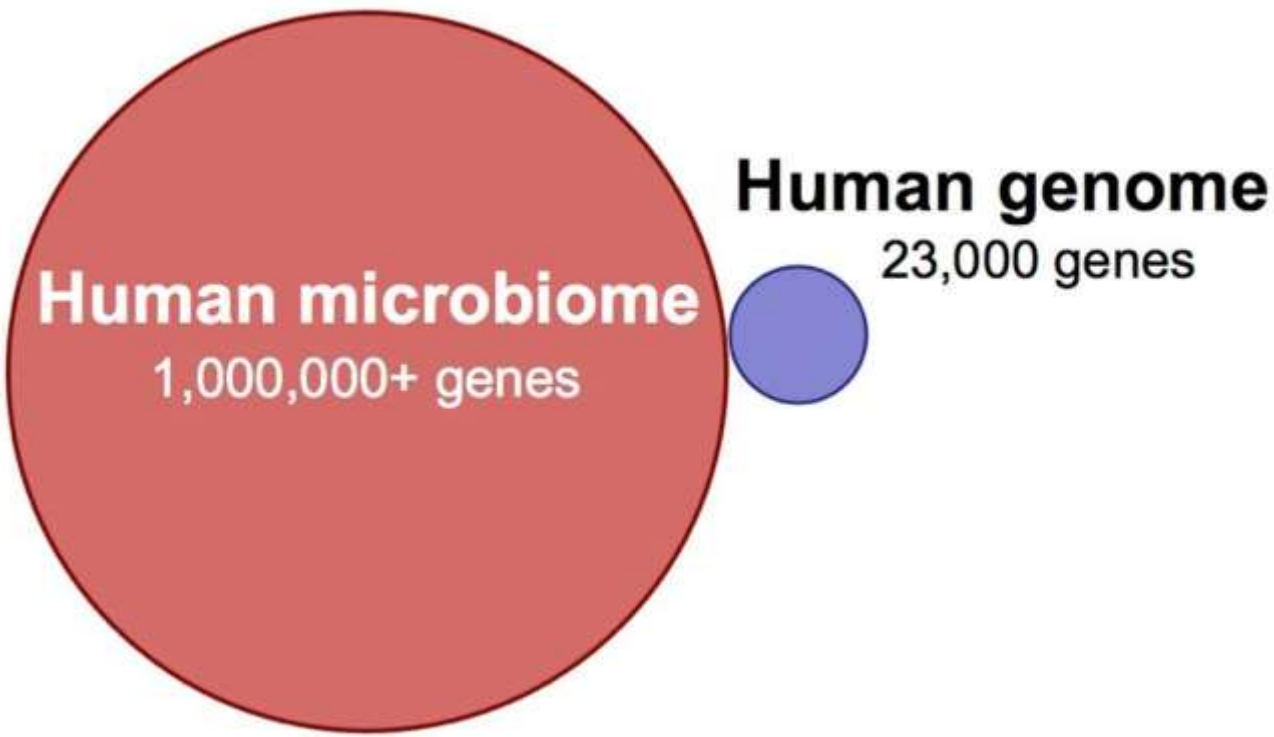
Nature Medicine **13**, 26 - 28 (2007)
doi:10.1038/nm0107-26

IL-23: a master regulator in Crohn disease

Markus F Neurath[1]

Inclusion of the Microbiome Will Radically Change Medicine and Wellness

Your Body Has 10 Times As Many Microbe Cells As Human Cells



99% of Your DNA Genes Are in Microbe Cells Not Human Cells

I Will Focus on the Human Gut Microbiome, Which Contains Hundreds of Microbial Species

J. Craig Venter™ INSTITUTE

it²

Intense Scientific Research is Underway on Understanding the Human Microbiome



June 8, 2012



June 14, 2012



August 18, 2012

When We Think About Biological Diversity We Typically Think of the Wide Range of Animals






But All These Animals Are in One SubPhylum Vertebrata of the Chordata Phylum

All images from Wikimedia Commons.
Photos are public domain or by Trisha Shears & Richard Bartz

Think of These Phyla of Animals When You Consider the Biodiversity of Microbes Inside You



Phylum Chordata



Phylum Cnidaria



Phylum Echinodermata



Phylum Annelida



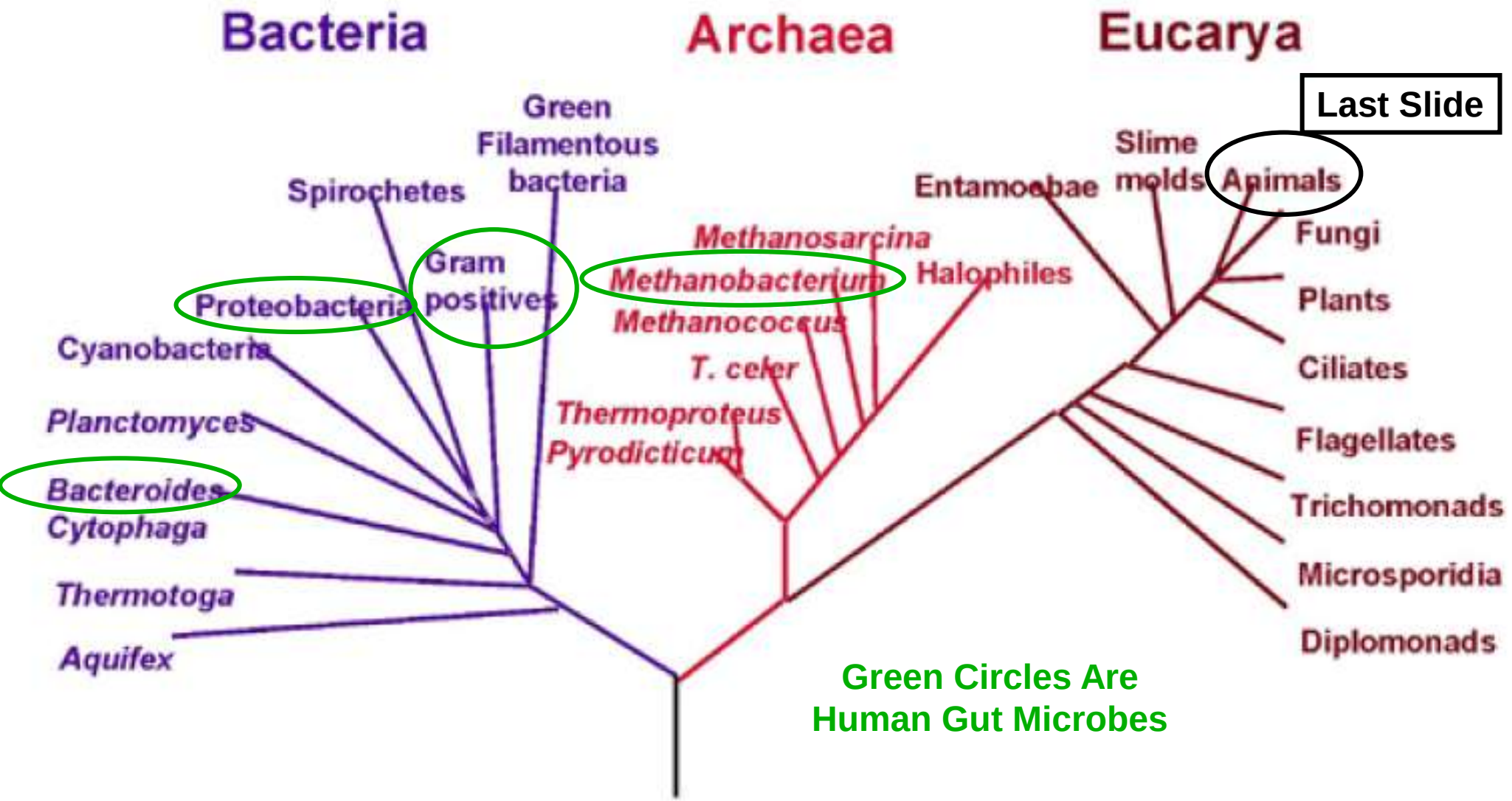
Phylum Mollusca



Phylum Arthropoda

All images from WikiMedia Commons.
Photos are public domain or by Dan Hershman, Michael Linnenbach, Manuae, B_cool

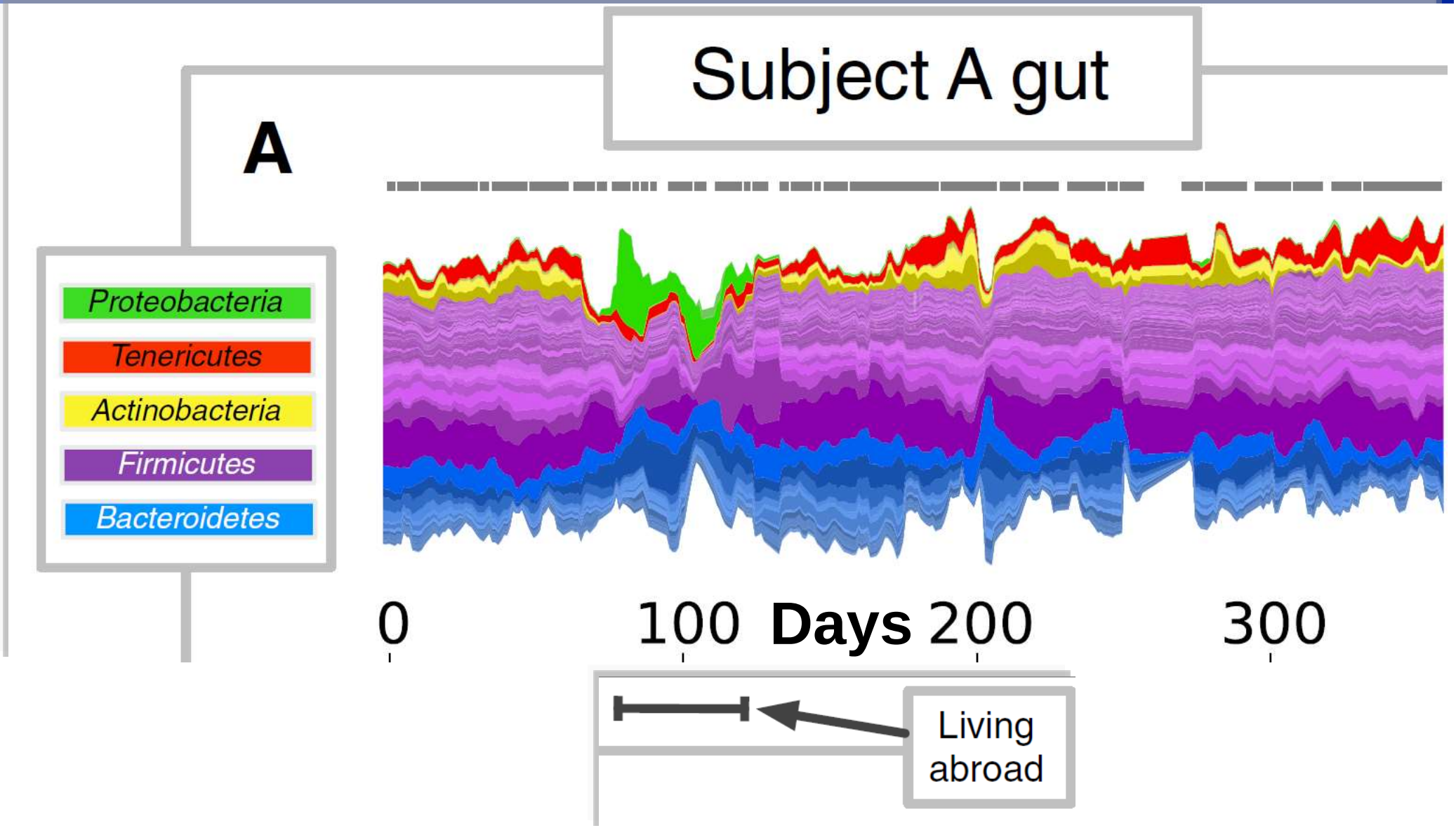
The Evolutionary Distance Between Your Gut Microbes Is Much Greater Than Between All Animals



Evolutionary Distance Derived from Comparative Sequencing of 16S or 18S Ribosomal RNA

Source: Carl Woese, et al

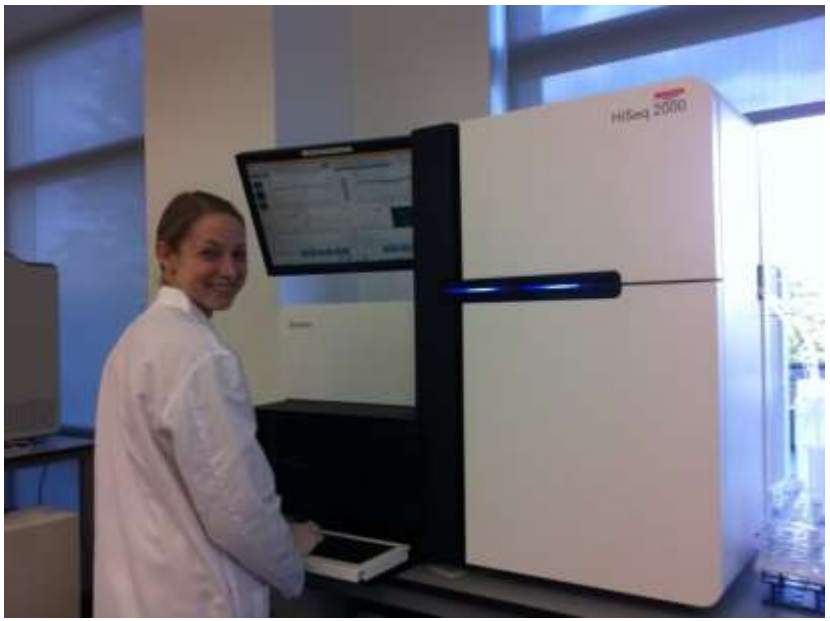
A Year of Sequencing a Healthy Gut Microbiome Daily - Remarkable Stability with Abrupt Changes



Genome Biology (2014)
David, et al.

To Map Out the Dynamics of My Microbiome Ecology I Partnered with the J. Craig Venter Institute

- JCVI Did Metagenomic Sequencing on Seven of My Stool Samples Over 1.5 Years
- Sequencing on Illumina HiSeq 2000
  - Generates 100bp Reads
  - Run Takes ~14 Days
  - My 7 Samples Produced
    - >200Gbp of Data
- JCVI Lab Manager, Genomic Medicine
  - Manolito Torralba
- IRB PI Karen Nelson
  - President JCVI



Illumina HiSeq 2000 at JCVI



Manolito Torralba, JCVI



Karen Nelson, JCVI

J. Craig Venter INSTITUTE

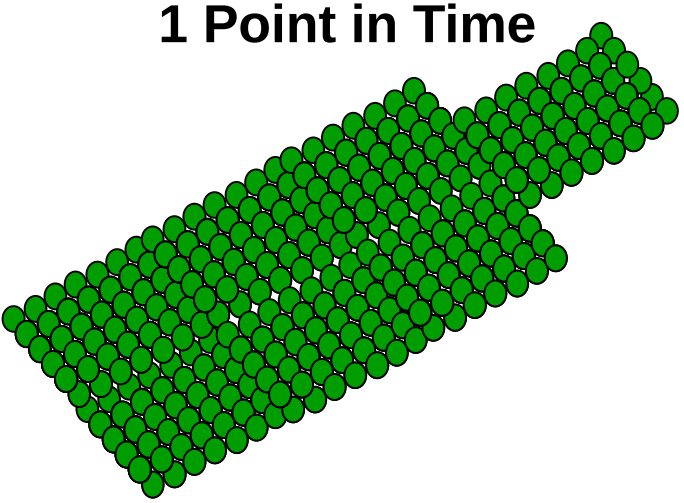
it²

We Expanded Our Healthy Cohort to All Gut Microbiomes from NIH HMP For Comparative Analysis

Each Sample Has 100-200 Million Illumina Short Reads (100 bases)

"Healthy" Individuals

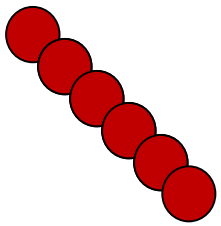
250 Subjects
1 Point in Time



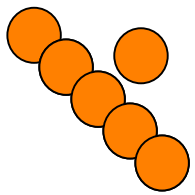
IBD Patients

Larry Smarr
(Colonic Crohn's)

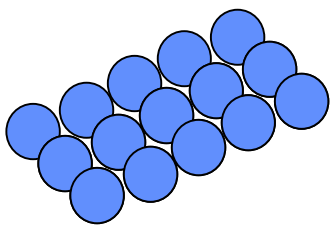
7 Points in Time



2 Ulcerative Colitis Patients,
6 Points in Time



5 Ileal Crohn's Patients,
3 Points in Time



Total of 27 Billion Reads
Or 2.7 Trillion Bases

Source: Jerry Sheehan, Calit2
Weizhong Li, Sitao Wu, CRBS, UCSD

We Created a Reference Database Of Known Gut Genomes

- **NCBI April 2013**
  - **2471 Complete + 5543 Draft Bacteria & Archaea Genomes**
  - **2399 Complete Virus Genomes**
  - **26 Complete Fungi Genomes**
  - **309 HMP Eukaryote Reference Genomes**
- **Total 10,741 genomes, ~30 GB of sequences**

NCBI National Center for Biotechnology Information

HMP

Now to Align Our 27 Billion Reads Against the Reference Database

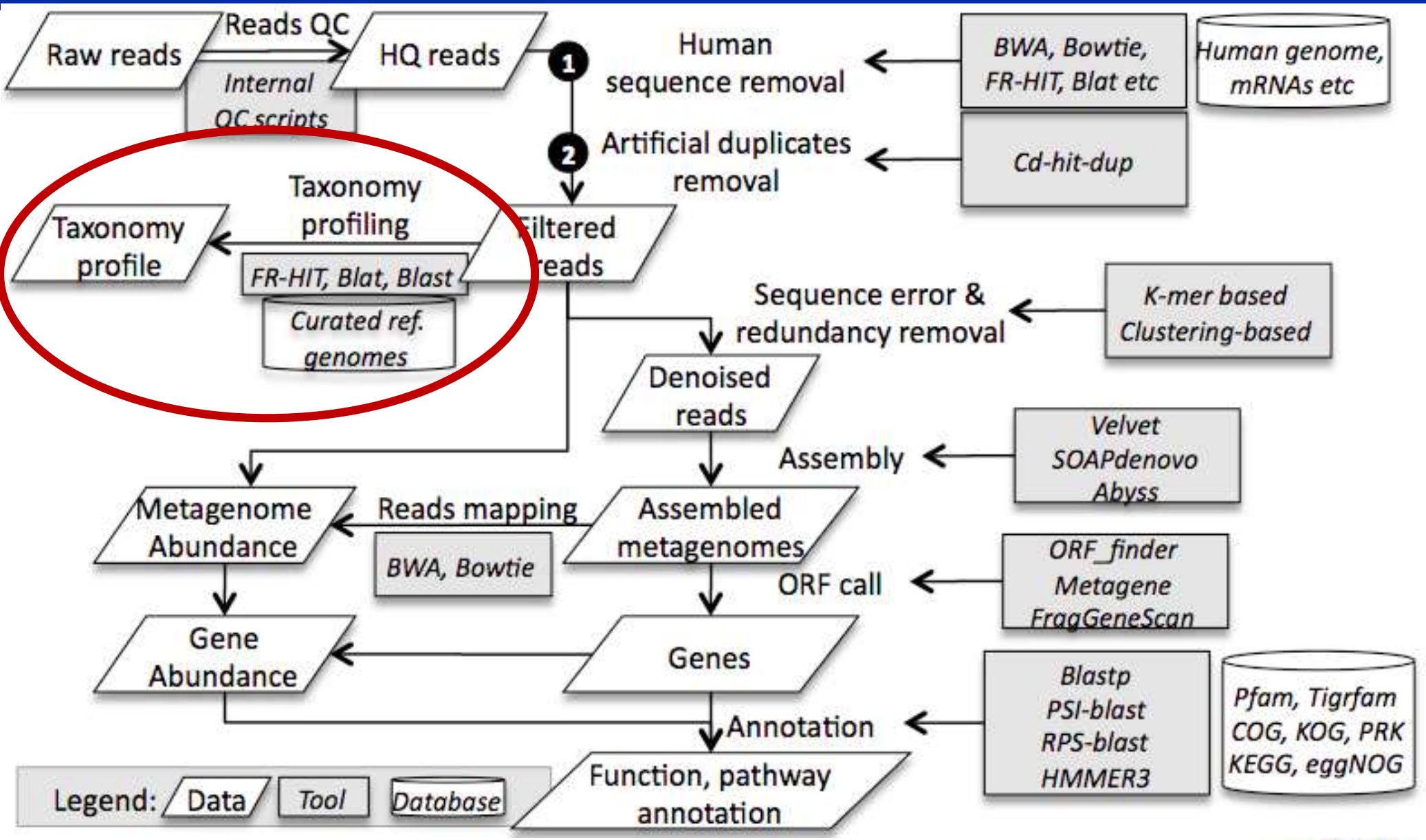
NIH

J. Craig Venter INSTITUTE

Source: Weizhong Li, Sitao Wu, CRBS, UCSD

it²

Computational NextGen Sequencing Pipeline: From "Big Equations" to "Big Data" Computing



National Human Genome Research Institute

HMP

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

CRBS

it²

We Used SDSC's Gordon Data-Intensive Supercomputer to Analyze a Wide Range of Gut Microbiomes

Source: Weizhong Li, Sitao Wu, CRBS, UCSD



Our Team Used 25 CPU-Years
To Compute
the Comparative Gut Microbiome
of My Time Samples
and Our Healthy and IBD Controls
Starting With
the 5 Billion Illumina Reads
Received from JCVI

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

NSF NIH HMP J. Craig Venter INSTITUTE CRBS SDSC it²

We Used Dell's HPC Cloud to Analyze All of Our Human Gut Microbiomes



Dell Solutions Center
Industry Solutions Lab
SANGER DSU

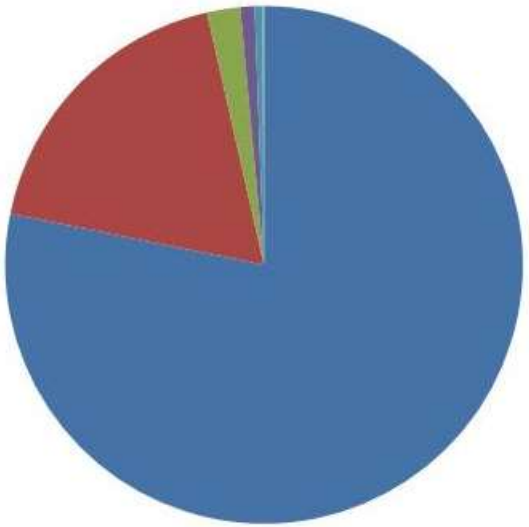
- **Dell's Sanger Cluster**
  - 32 Nodes, 512 Cores
  - 48GB RAM per Node
- **We Processed the Taxonomic Relative Abundance**
  - Used ~35,000 Core-Hours on Dell's Sanger
- **Produced Relative Abundance of ~10,000 Bacteria, Archaea, Viruses in ~300 People**
  - ~3Million Spreadsheet Cells
- **New System: R Bio-Gen System**
  - 48 Nodes, 768 Cores
  - 128 GB RAM per Node

Source: Weizhong Li, UCSD



We Found Major State Shifts in Microbial Ecology Phyla Between Healthy and Two Forms of IBD

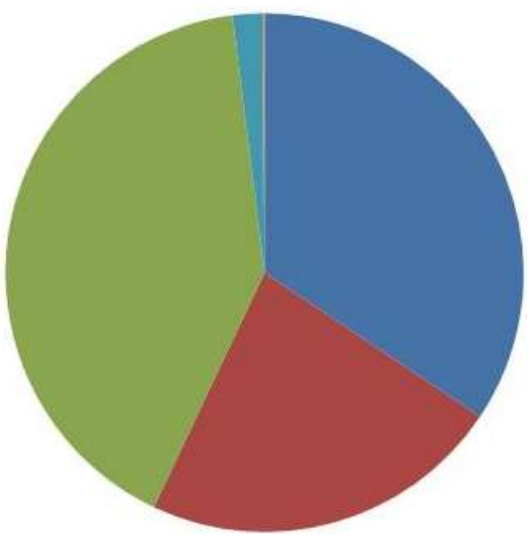
Average HE



- Bacteroidetes
- Firmicutes
- Proteobacteria
- Verrucomicrobia
- Actinobacteria
- Fusobacteria
- Euryarchaeota

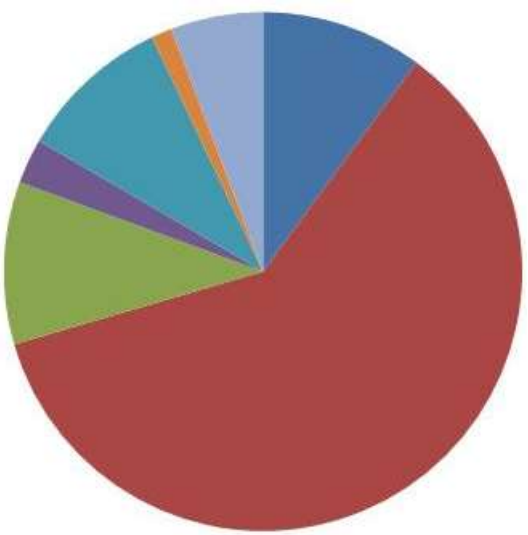
Most Common Microbial Phyla

Average Ulcerative Colitis



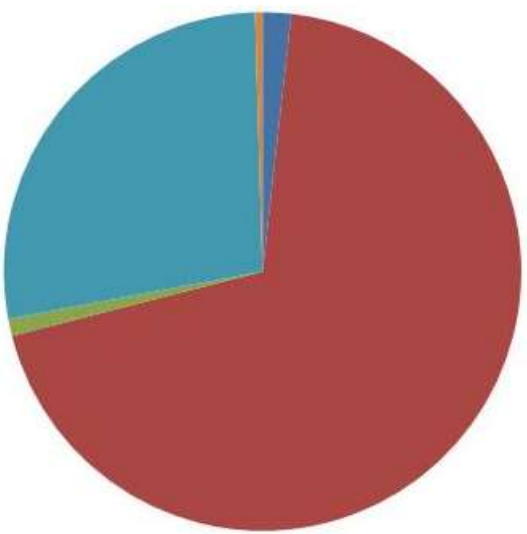
Explosion of Proteobacteria

Average LS



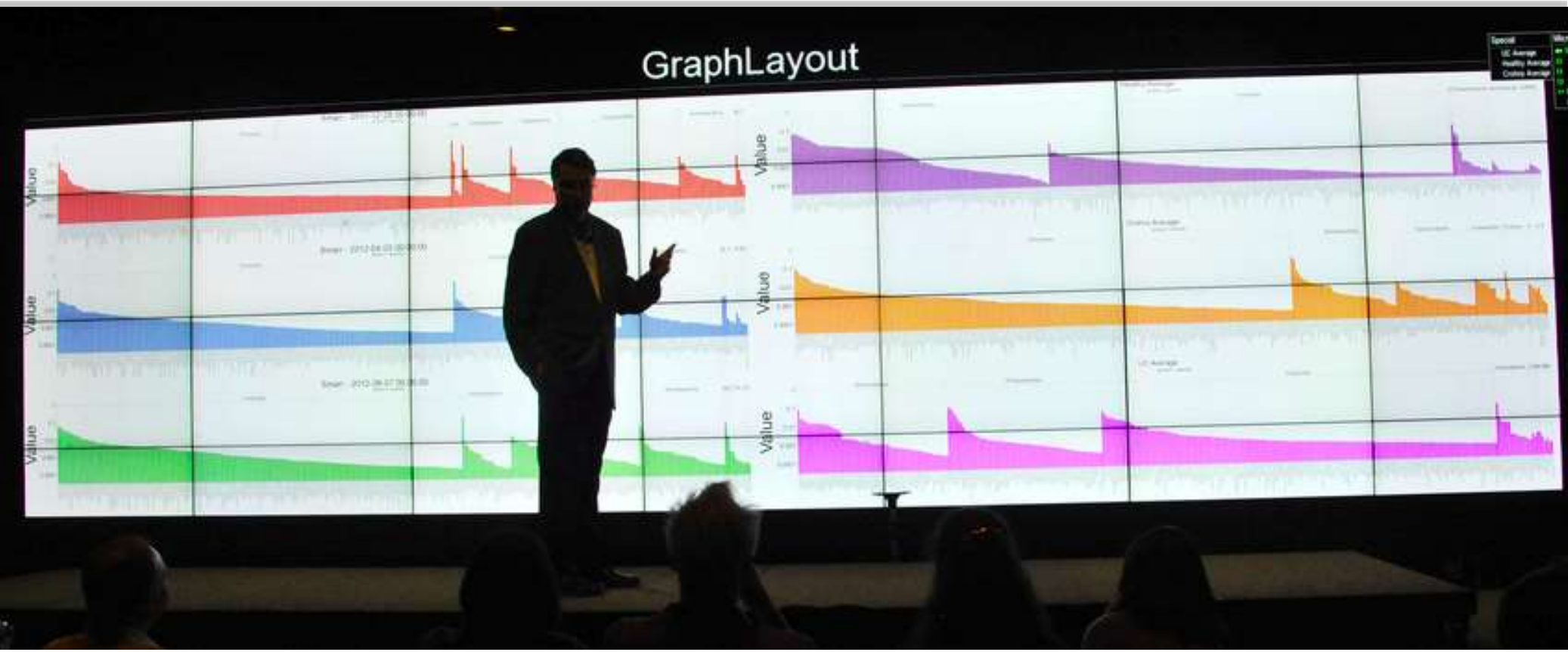
Hybrid of UC and CD
High Level of Archaea

Average Crohn's Disease



Collapse of Bacteroidetes
Explosion of Actinobacteria

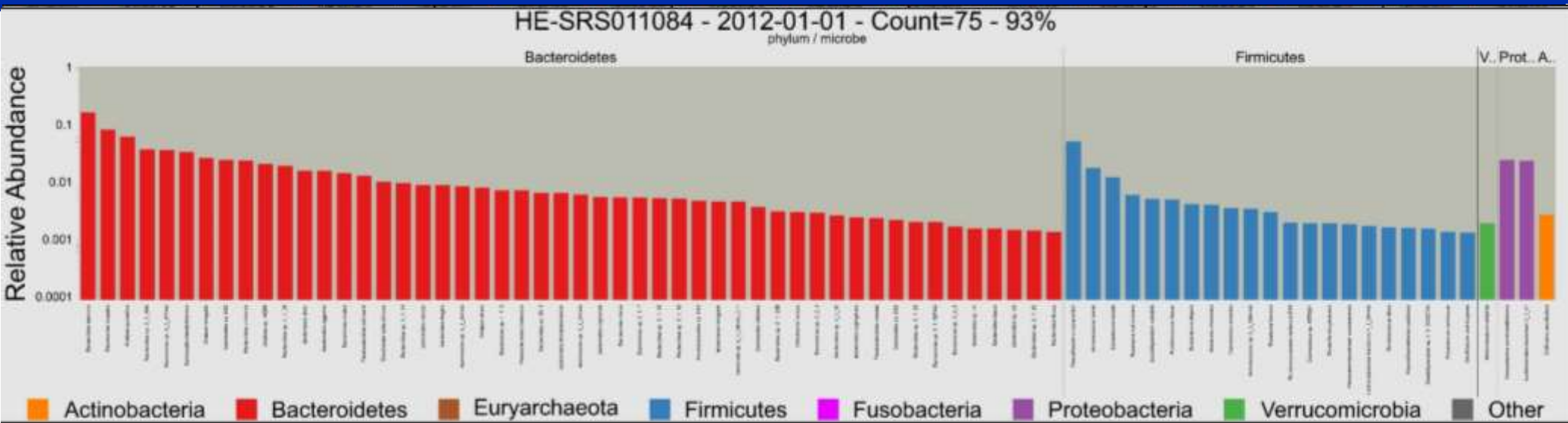
Using Scalable Visualization Allows Comparison of the Relative Abundance of 200 Microbe Species

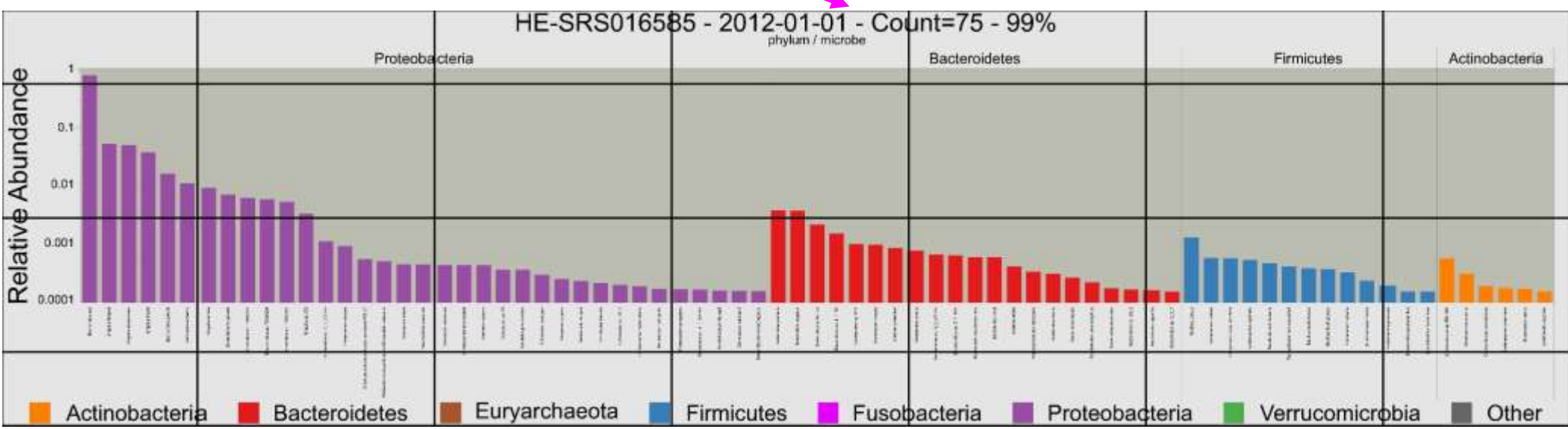


Comparing 3 LS Time Snapshots (Left)
with Healthy, Crohn's, UC (Right Top to Bottom)

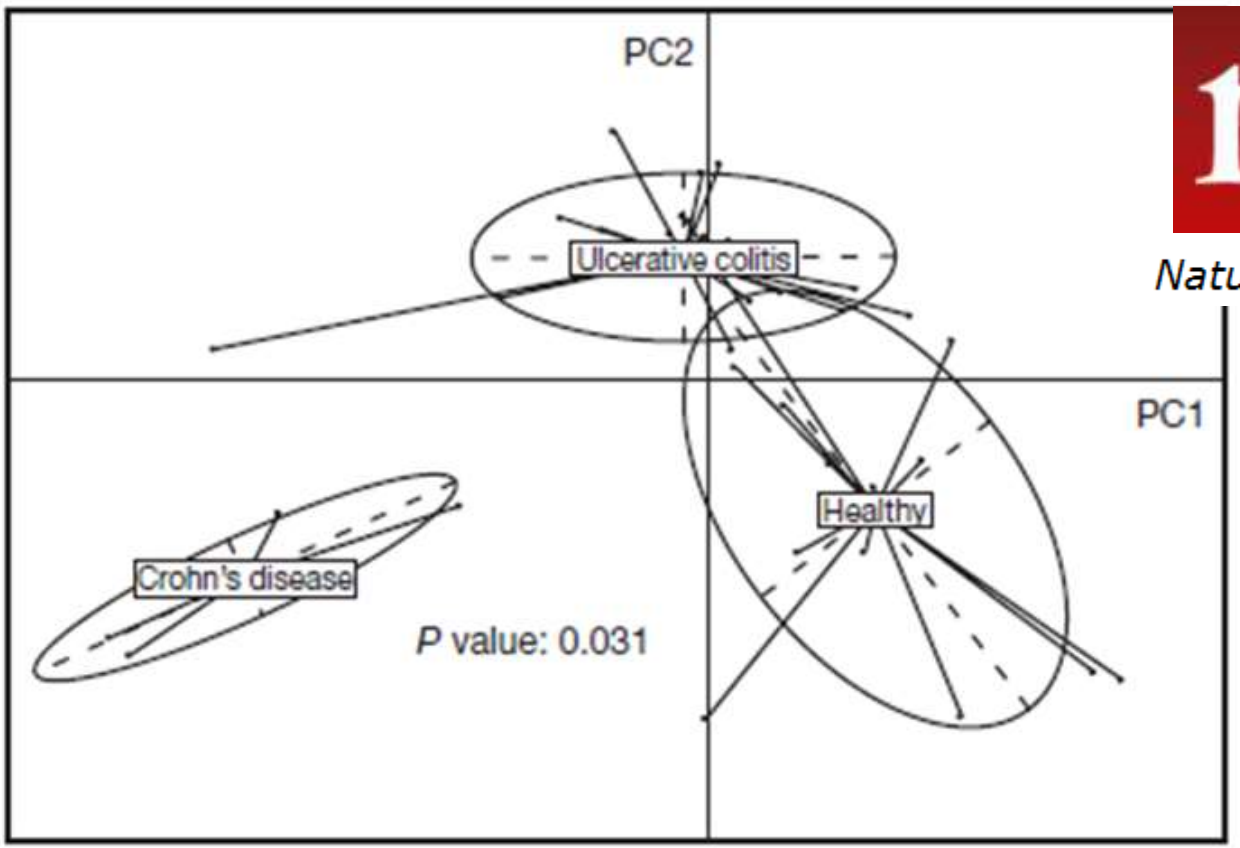
Calit2 VROOM-FuturePatient Expedition

Using Microbiome Profiles to Survey 155 Subjects for Unhealthy Candidates





Using Principal Component Analysis To Stratify Disease States From Healthy States



nature

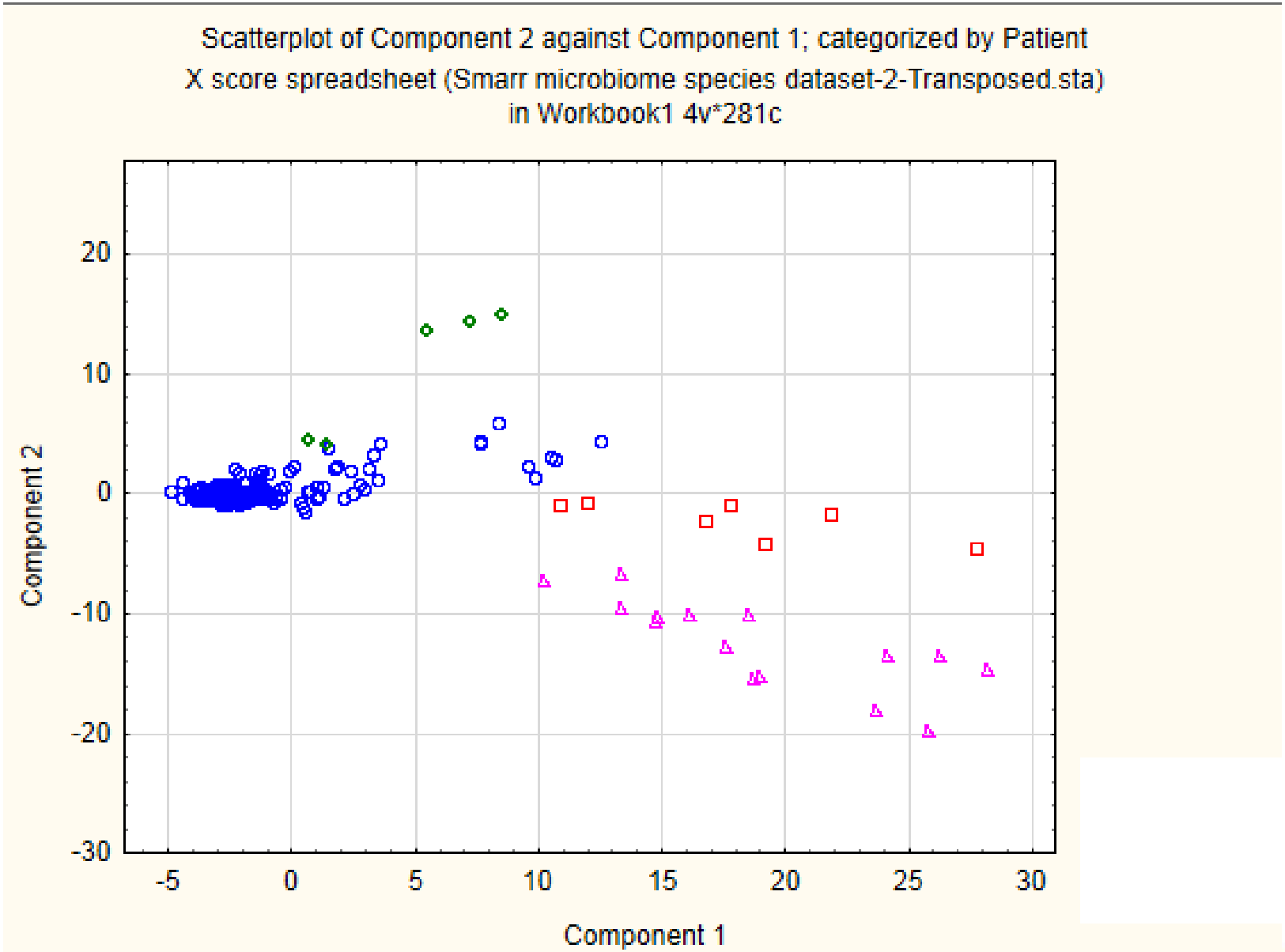
Nature **464**, 59-65 (4 March 2010)

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],

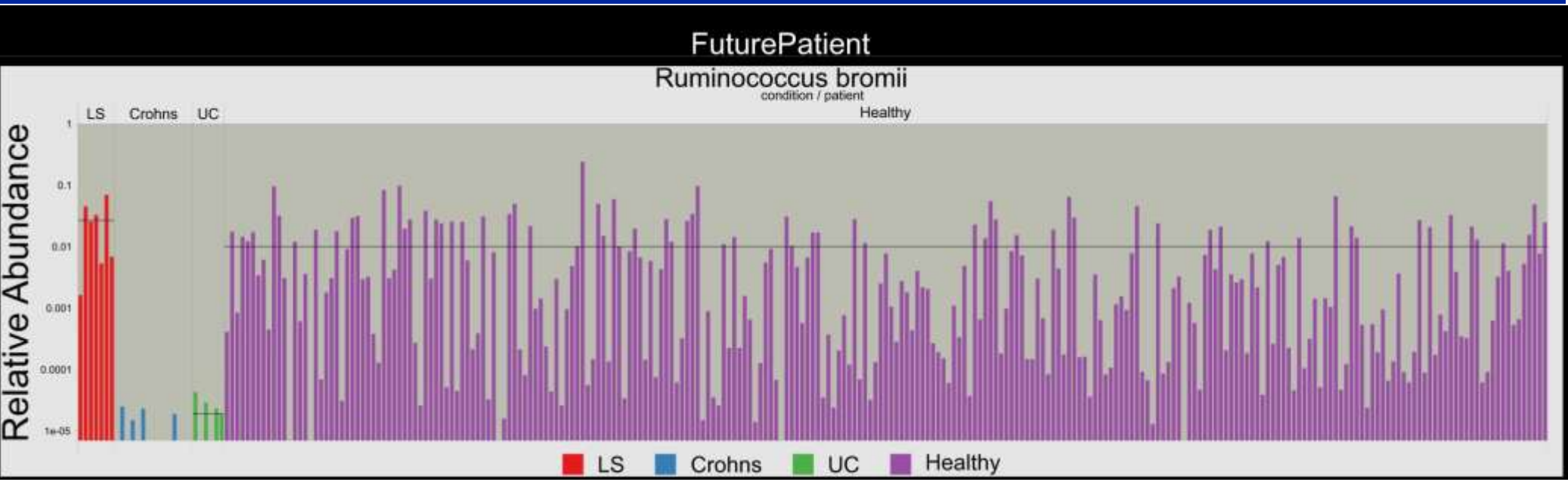
Dell Analytics Separates The 4 Patient Types in Our Data Using Microbiome Species Data



Source: Thomas Hill, Ph.D.
Executive Director Analytics
Dell | Information Management Group, Dell Software

Connecting Diet, Gut Microbes, and Disease

Absence of Ruminococcus bromii May Impair Fermentation in IBD Patients



ISME J. 2012 Aug;6(8):1535-43. doi: 10.1038/ismej.2012.4. Epub 2012 Feb 16.

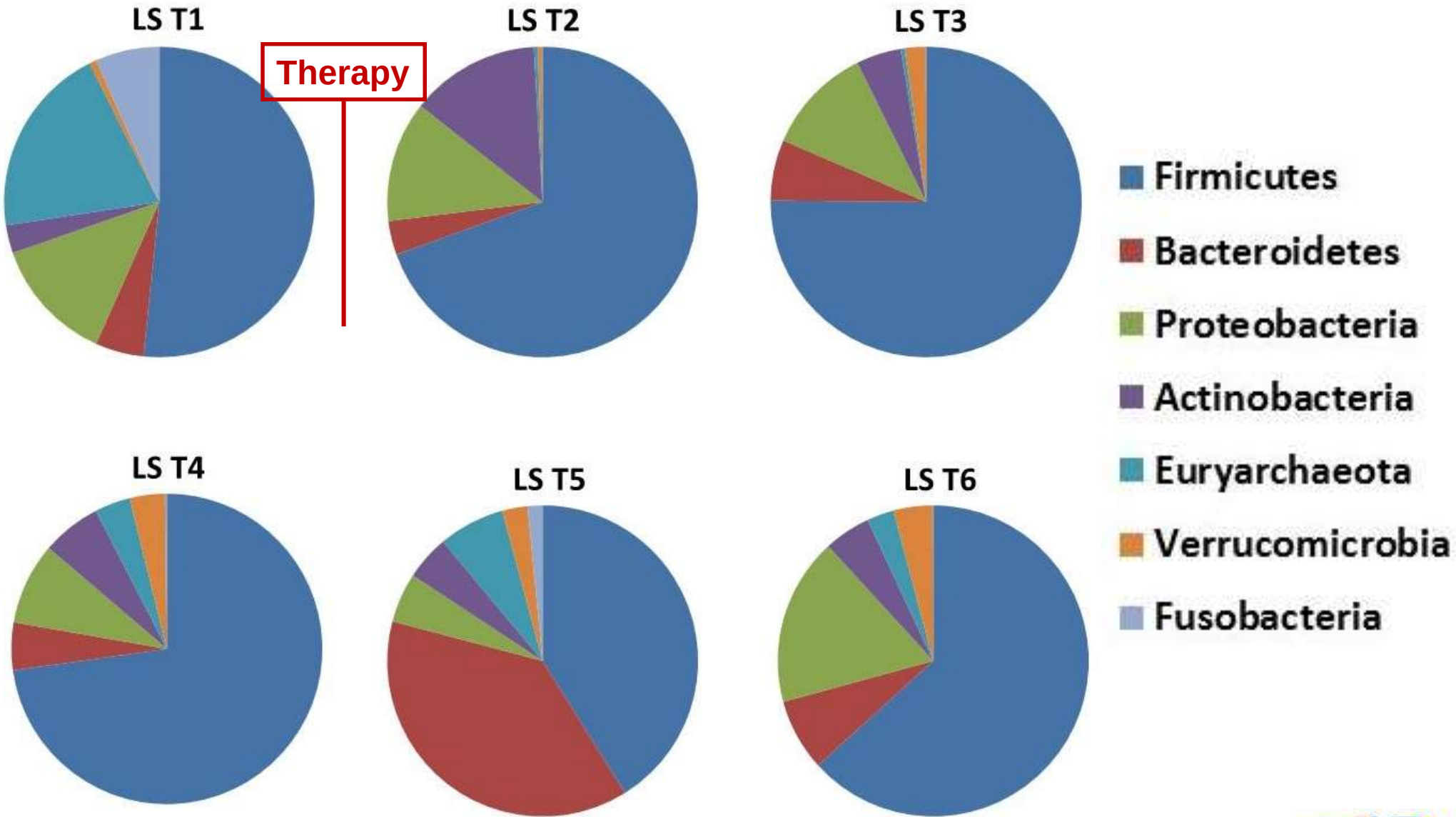
Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon.

Ze X[1], Duncan SH, Louis P, Flint HJ.

"This argues strongly that R. bromii has a pivotal role in fermentation of [resistant starch] RS3 in the human large intestine, and that **variation in the occurrence of this species and its close relatives may be a primary cause of variable energy recovery** from this important component of the diet."

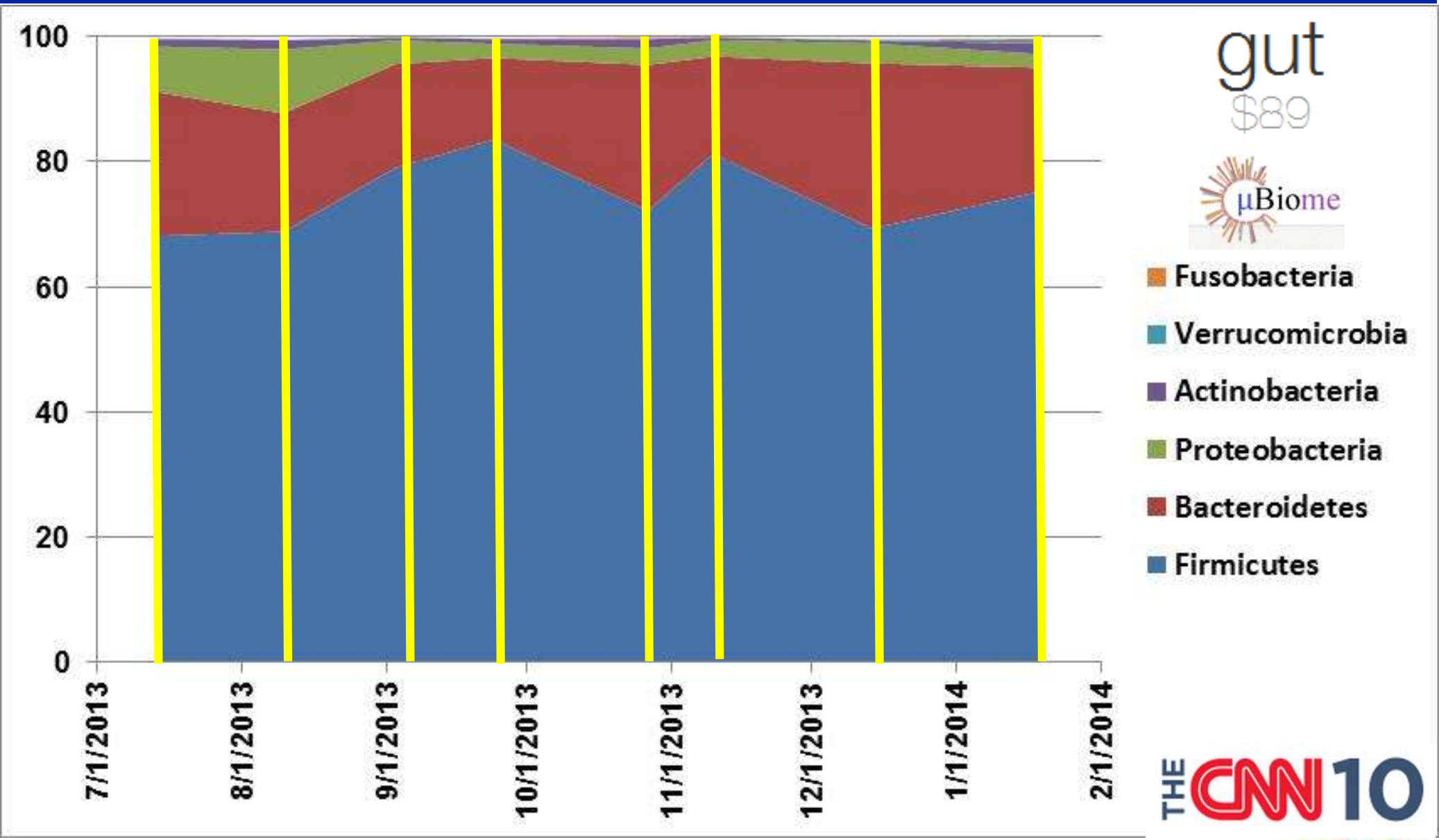
Supports Research on Importance of Resistant Starch for Gut Health by Drs. David Topping and Mark Morrison in Australia

Time Series Reveals Autoimmune Dynamics of Gut Microbiome by Phyla



Six Metagenomic Time Samples Over 16 Months

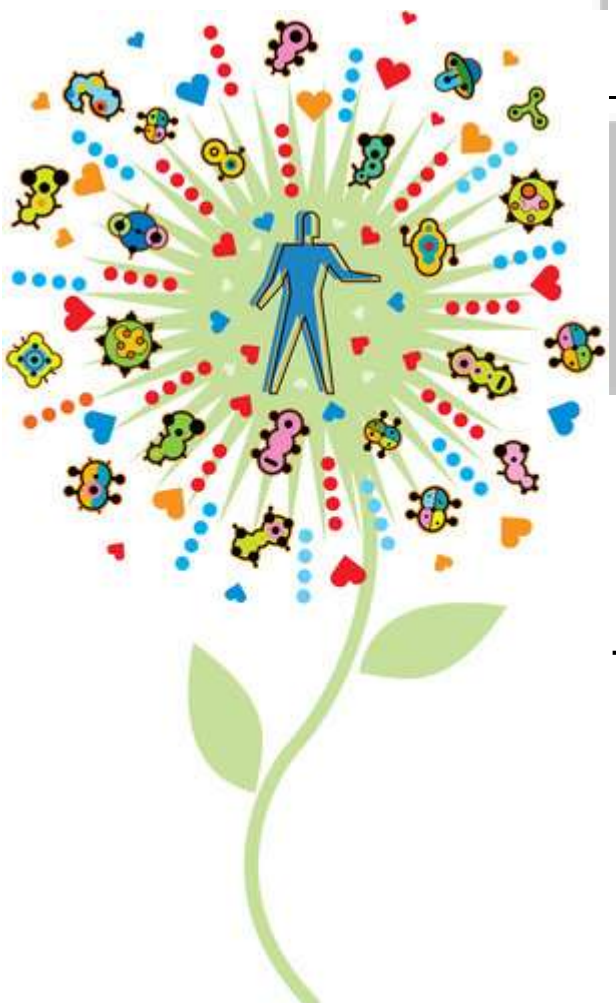
Inexpensive Consumer 16S Time Series of Microbiome Allows Similar Analysis Through Ubiome



THE CNN10

Data source: LS (Yellow Lines Stool Samples); Sequencing and Analysis Ubiome

There is a Huge New Field of Products Coming Which Enable You to "Garden" Your Microbiome



The New York Times

Science

Tending the Body's Microbial Garden

By CARL ZIMMER

Published: June 18, 2012

"I would like to lose the language of warfare,"
said Julie Segre, a senior investigator at
the National Human Genome Research Institute.
"It does a disservice to all the bacteria
that have co-evolved with us
and are maintaining the health of our bodies."

Will Medical Foods Provide New Tools for Altering Gut Microbiome?

Faecal Microbiota Transplantation (FMT) Therapy Has Been Pioneered in Australia



Professor Tom Borody,
founder and current medical director
of the Center for Digestive Diseases (CDD)
in Sydney, Australia

**Controversial, but very promising.
More experiments needed on
a variety of disease states.**

**"I think we're on the edge of something extraordinary.
The attention has switched entirely to the large bowel bacterial population
which we now know is absolutely critical to human health,"
--Dr. David Topping, Chief Research Scientist
at CSIRO Animal, Food and Health Sciences in Adelaide, South Australia**

NEWS

18 Mar 2014

Early Adopting MDs Are Creating Partnerships with Their Quantified Patients

- **"The 100 participants will be guided on this 9-month journey by a coach and when necessary, be referred to their own health care practitioners."**

- **The data sets that will be evaluated include:**
  - **Self-Tracking Devices**
  - **Medical History, Traits, Lifestyle**
  - **Blood, Urine, Saliva**
  - **Gut Microbiome**
  - **Whole Genome Sequencing**

PIONEER 100
HUNDRED PERSON WELLNESS PROJECT
Institute for Systems Biology

Will Grow to 1000, then 10,000

https://pioneer100.systemsbiology.net/

UC San Diego Is Carrying Out a Major Clinical Study of IBD Using These Techniques

Goal: Understand
The Coupled Human Immune-Microbiome Dynamics
In the Presence of Human Genetic Predispositions

Already 100 Enrolled, Goal is 1500

Drs. William J. Sandborn, John Chang, & Brigid Boland
UCSD School of Medicine, Division of Gastroenterology

Thanks to Our Great Team!

UCSD Metagenomics Team

Weizhong Li
Sitao Wu

JCVI Team

Karen Nelson
Shibu Yooseph
Manolito Torralba

Calit2@UCSD Future Patient Team

Jerry Sheehan
Tom DeFanti
Kevin Patrick
Jurgen Schulze
Andrew Prudhomme
Philip Weber
Fred Raab
Joe Keefe
Ernesto Ramirez



SDSC Team

Michael Norman
Mahidhar Tatineni
Robert Sinkovits

UCSD Health Sciences Team

William J. Sandborn
Elisabeth Evans
John Chang
Brigid Boland
David Brenner