“Finding the Patterns in the Big Data From Human Microbiome Ecology”

Invited Talk

Exponential Medicine

November 10, 2014

Dr. Larry Smarr

Director, California Institute for Telecommunications and Information Technology

Harry E. Gruber Professor,

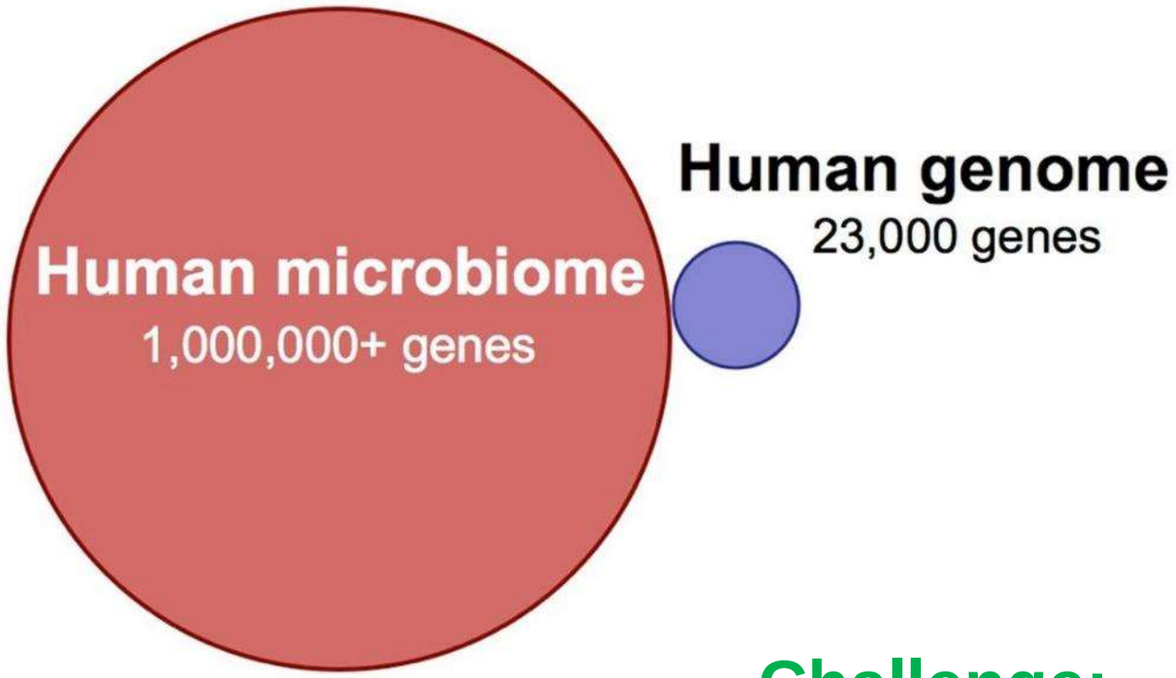
Dept. of Computer Science and Engineering

Jacobs School of Engineering, UCSD

http://lsmarr.calit2.net

How Will Detailed Knowledge of Microbiome Ecology 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

Challenge:
Map Out Microbial Ecology and Function in Health and Disease States

To Map Out the Dynamics of Autoimmune Microbiome Ecology Couples Next Generation Genome Sequencers to *Big Data* Supercomputers

Example: Inflammatory Bowel Disease (IBD)

- **Metagenomic Sequencing**
  - **JCVI Produced**
    - **~150 Billion DNA Bases From Seven of LS Stool Samples Over 1.5 Years**
  - **We Downloaded ~3 Trillion DNA Bases From NIH Human Microbiome Program Data Base**
    - **255 Healthy People, 21 with IBD**
- **Supercomputing (**Weizhong Li, JCVI/HLI/UCSD):
  - **~20 CPU-Years on SDSC's Gordon**
  - **~4 CPU-Years on Dell's HPC Cloud**
- **Produced Relative Abundance of**
  - **~10,000 Bacteria, Archaea, Viruses in ~300 People**
  - **~3Million Filled Spreadsheet Cells**

Illumina HiSeq 2000 at JCVI



SDSC Gordon Data Supercomputer



Dell Solutions Center
Industry Solutions Lab
SANGER DSU

J. Craig Venter INSTITUTE SDSC DELL NIH HMP CRBS it²

How Best to Analyze The Microbiome Datasets to Discover Patterns in Health and Disease?

Can We Find New Noninvasive Diagnostics In Microbiome Ecologies?

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

But You Need to Think of All These Phyla of Animals When You Consider the Biodiversity of Microbes Inside You



Phylum Chordata



Phylum Cnidaria



Phylum Echinodermata



Phylum Annelida



Phylum Mollusca

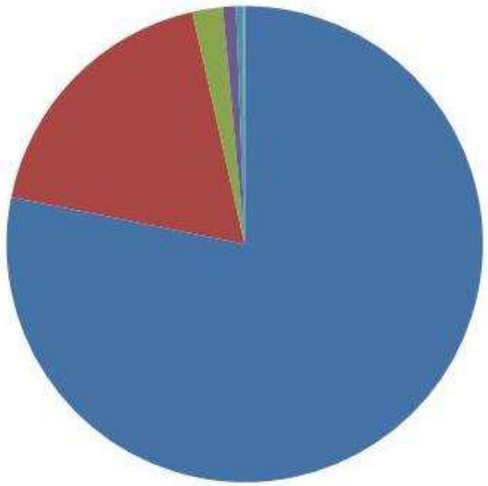


Phylum Arthropoda

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

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

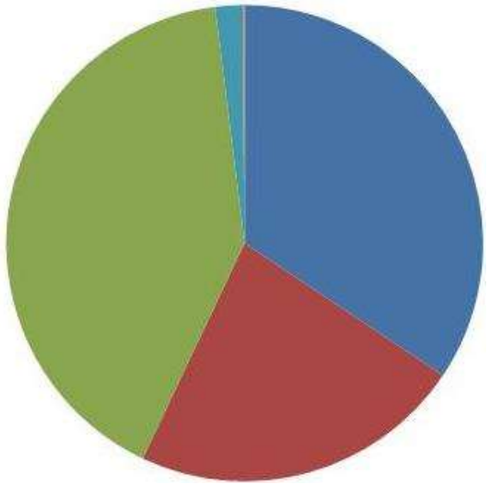
Average HE



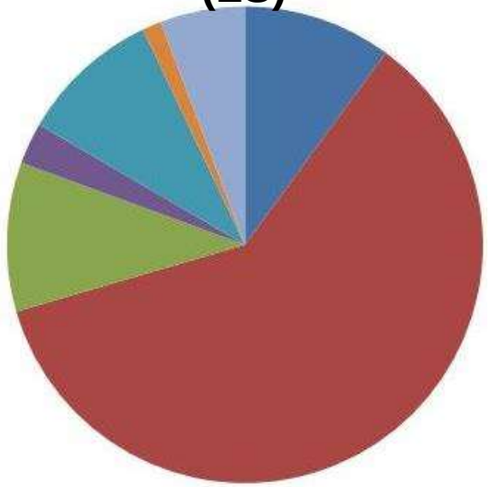
Most Common Microbial Phyla

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

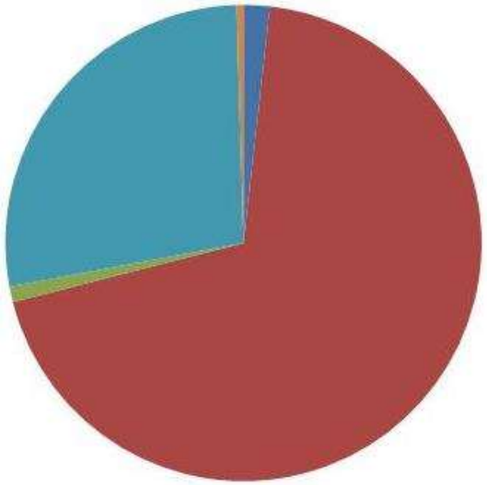
Average Ulcerative Colitis



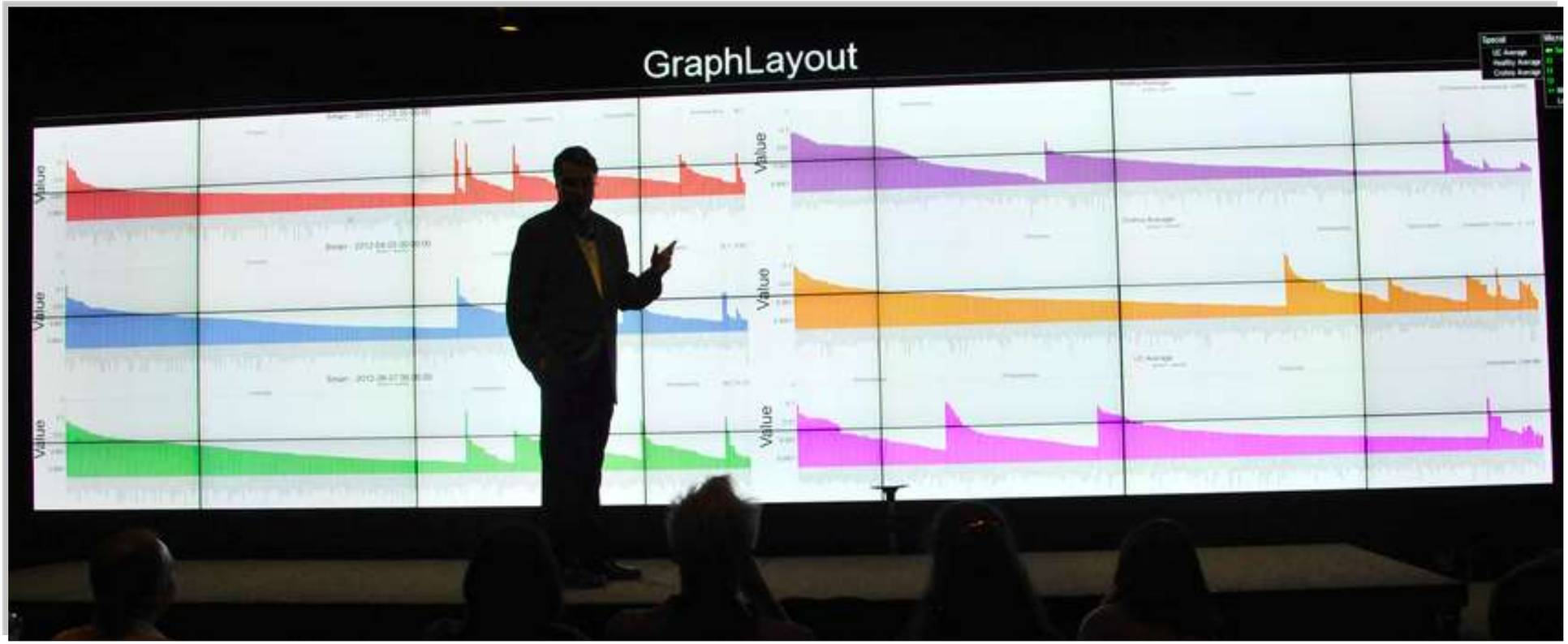
Average Colonic Crohn's Disease (LS)



Average Ileal Crohn's Disease



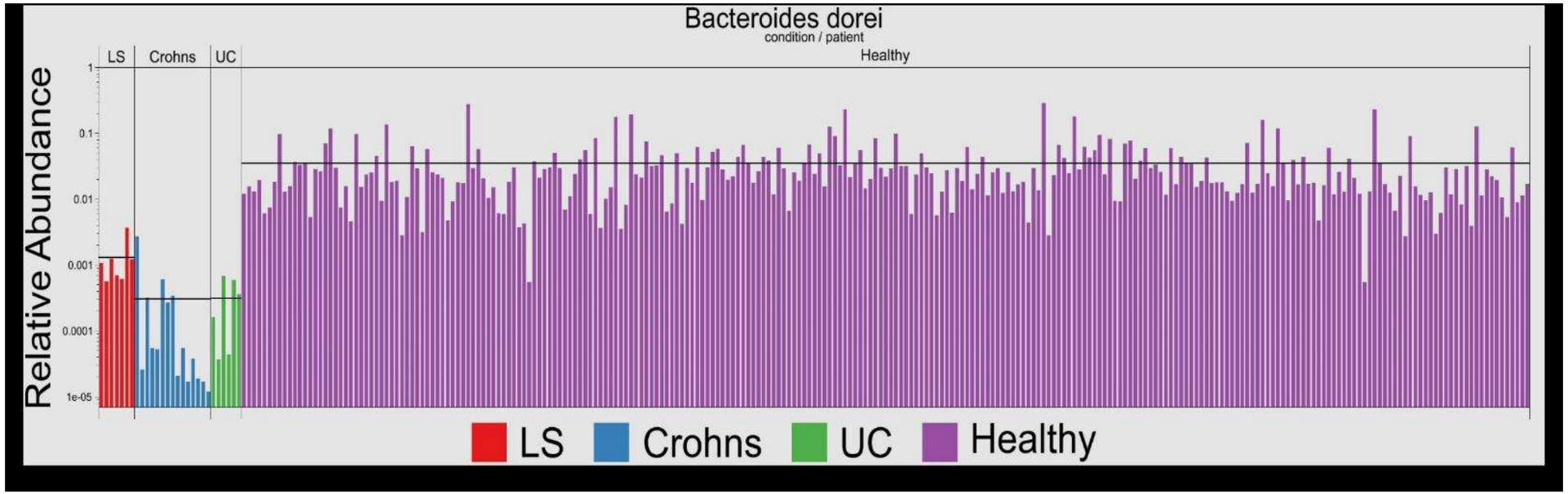
Using Scalable Visualization Allows Comparison of the Relative Abundance of 200 Microbe Species

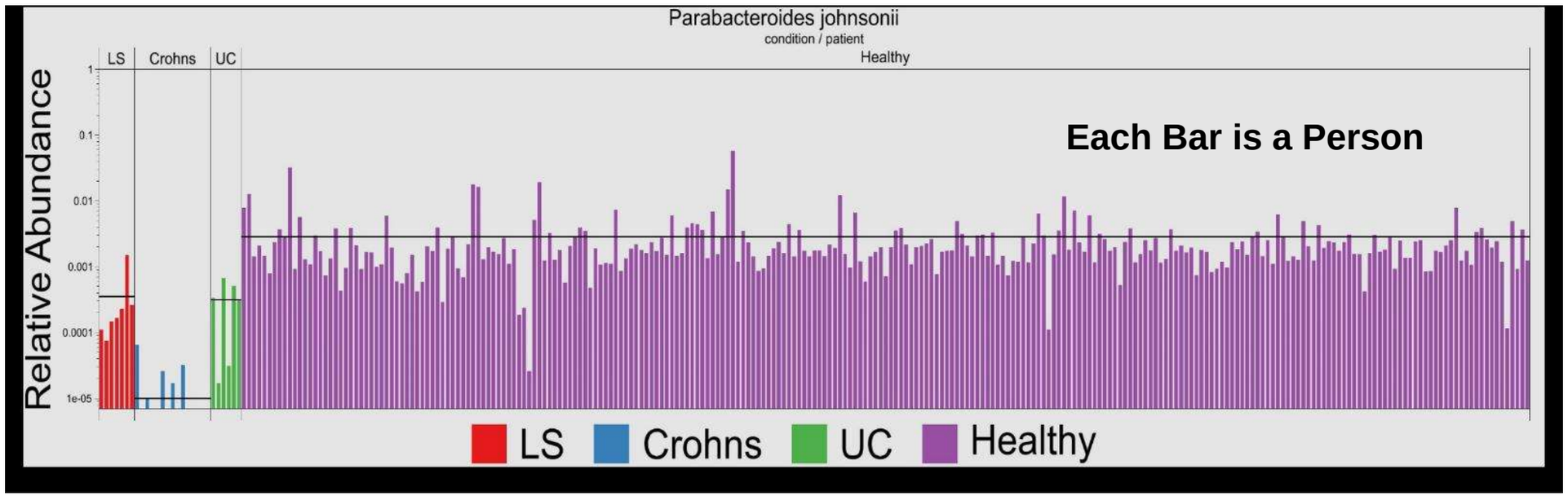


Comparing 3 LS Time Snapshots (Left)
with Healthy, Crohn's, Ulcerative Colitis (Right Top to Bottom)

Calit2 VROOM-FuturePatient Expedition

Our Scalable Visualization Analysis Found That Some Species Can Differentiate IBD vs. Healthy Subjects





Using Ayasdi Advanced Analytics to Interactively Discover Hidden Patterns in Our Data

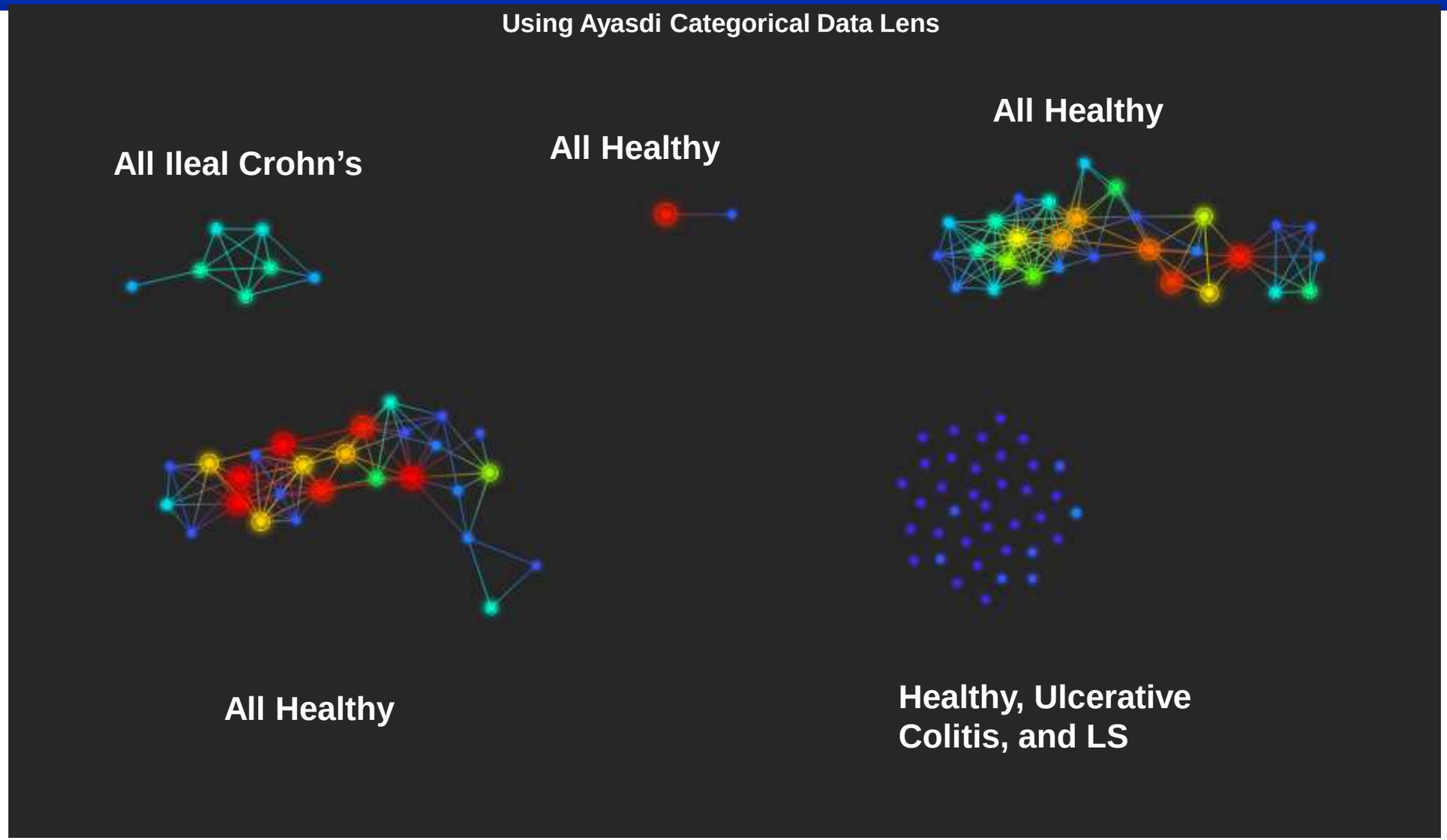
Ayasdi Advanced Analytics

topological data analysis



Visit Ayasdi in the Exponential Medicine Healthcare Innovation Lab

Using Ayasdi's Topological Data Analysis to Separate Healthy from Disease States



Ayasdi Cure™
Turn Data into Therapies

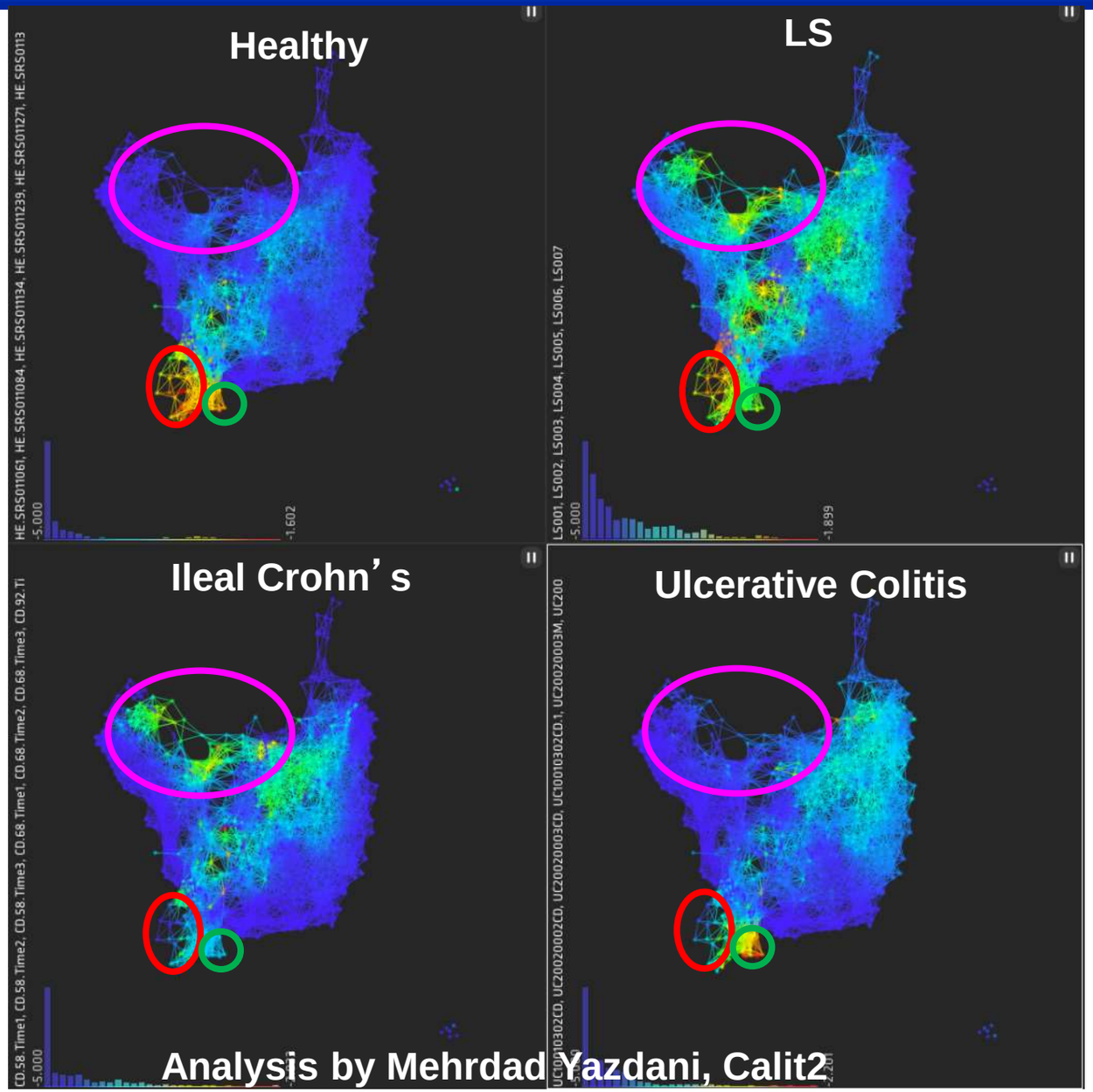
Analysis by Mehrdad Yazdani, Calit2

Ayasdi Interactively Identifies Microbial Species That Statistically Best Separates Health and Disease States

Ayasdi Confirms Our Two Species and Provides Many Others

Column Name	KS Statistic ▼1	KS p-value	T-test p-value
Actinomyces.sp..oral.taxon.448	1.000	<.001	<.001
Bacteroides.fragilis	1.000	<.001	<.001
Bacteroides.dorei	1.000	<.001	<.001
Tannerella.sp..6_1_58FAA_CT1	1.000	<.001	<.001
Bacteroides.coprophilus	1.000	<.001	<.001
Bacteroides.faecis	1.000	<.001	<.001
Parabacteroides.johnsonii	1.000	<.001	<.001
Paraprevotella.xylaniphila	1.000	<.001	<.001
Prevotella.disiens	1.000	<.001	<.001
Alloprevotella.tannerae	1.000	<.001	<.001
Prevotella.buccalis	1.000	<.001	<.001
Bacteroides.salyersiae	1.000	<.001	<.001
Prevotella.oris	1.000	<.001	<.001
Bacteroides.oleiciplenus	1.000	<.001	<.001
Prevotella.timonensis	1.000	<.001	<.001
Actinomyces.odontolyticus	0.9938	<.001	<.001
Actinomyces.sp..oral.taxon.180	0.9938	<.001	<.001
Alistipes.finegoldii	0.9938	<.001	<.001
Bacteroides.sp..2_1_7	0.9938	<.001	<.001
Bacteroides.salanitronis	0.9938	<.001	<.001
Prevotella.bivia	0.9938	<.001	<.001

Ayasdi Enables Discovery of Differences Between Healthy and Disease States Using Microbiome Species



Ayasdi Cure™
Turn Data into Therapies

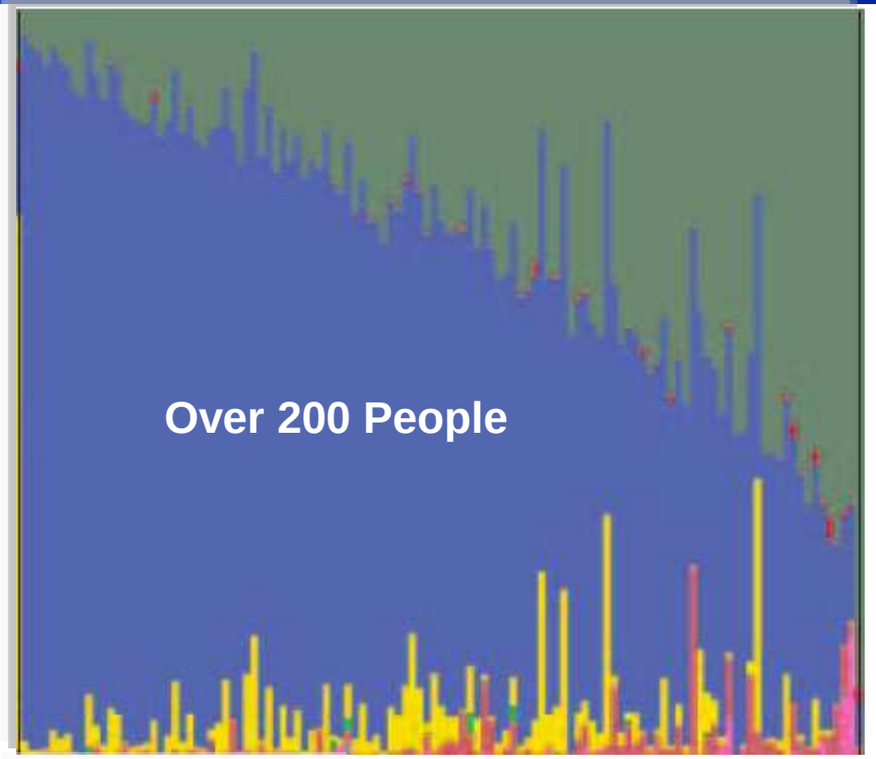
High in Healthy and LS

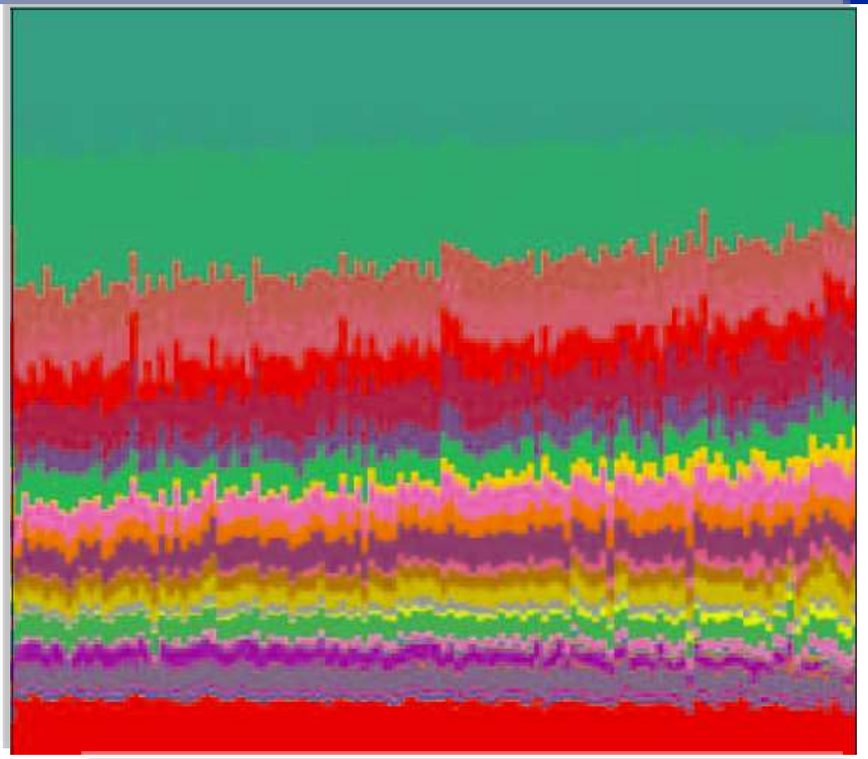
High in Healthy and Ulcerative Colitis

High in Both LS and Ileal Crohn's Disease

Using Multidimensional Scaling Lens with Correlation Metric

In a “Healthy” Gut Microbiome: Large Taxonomy Variation, Low Protein Family Variation





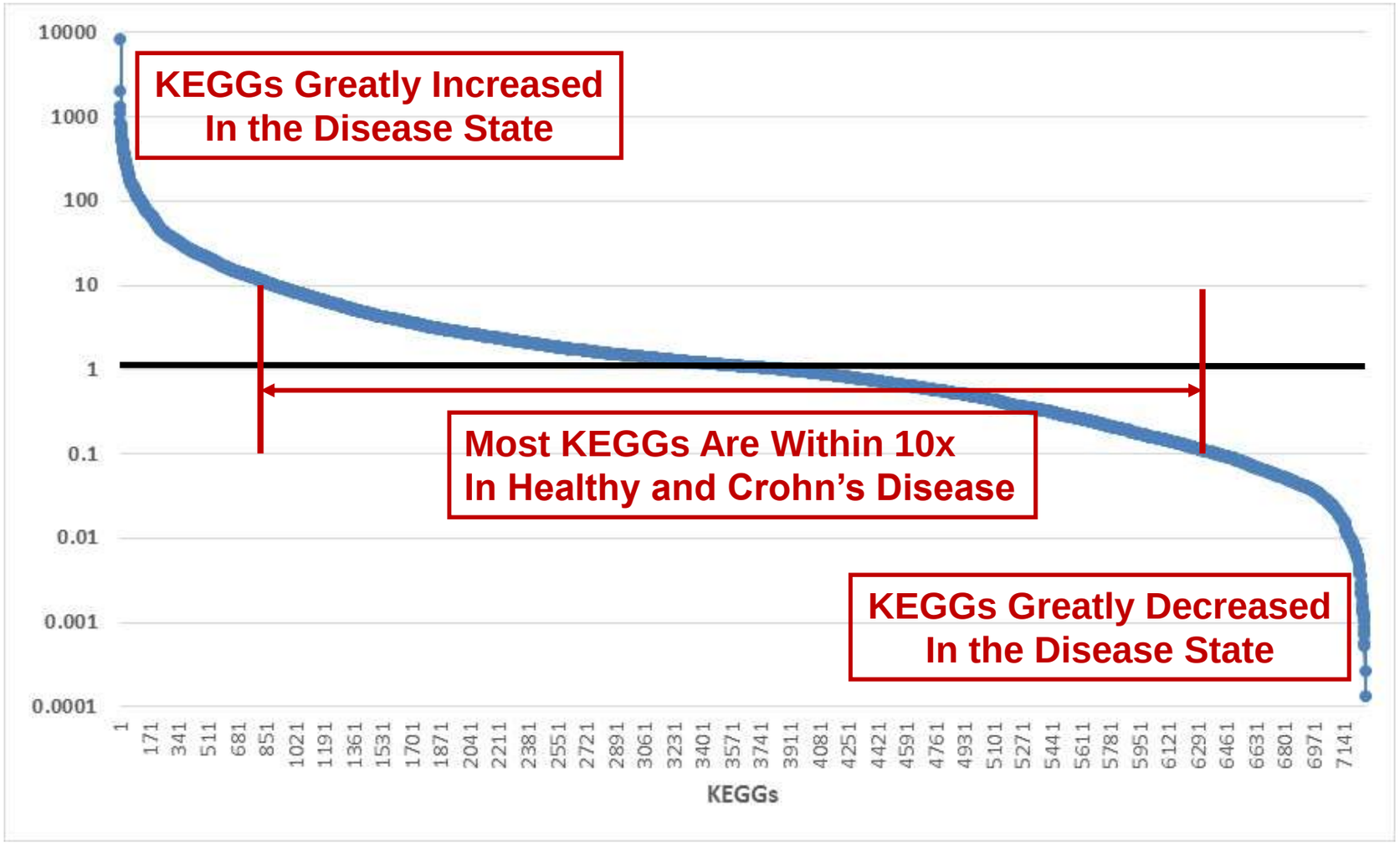
- Firmicutes
- Actinobacteria
- Bacteroidetes
- Proteobacteria
- Fusobacteria
- Tenericutes
- Spirochaetes
- Cyanobacteria
- Verrucomicrobia
- TM7

- Central carbohydrate metabolism
- Cofactor and vitamin biosynthesis
- Oligosaccharide and polyol transport system
- Purine metabolism
- ATP synthesis
- Phosphate and amino acid transport system
- Aminoacyl tRNA
- Pyrimidine metabolism
- Ribosome
- Aromatic amino acid metabolism

Source: Nature, 486, 207-212 (2012)

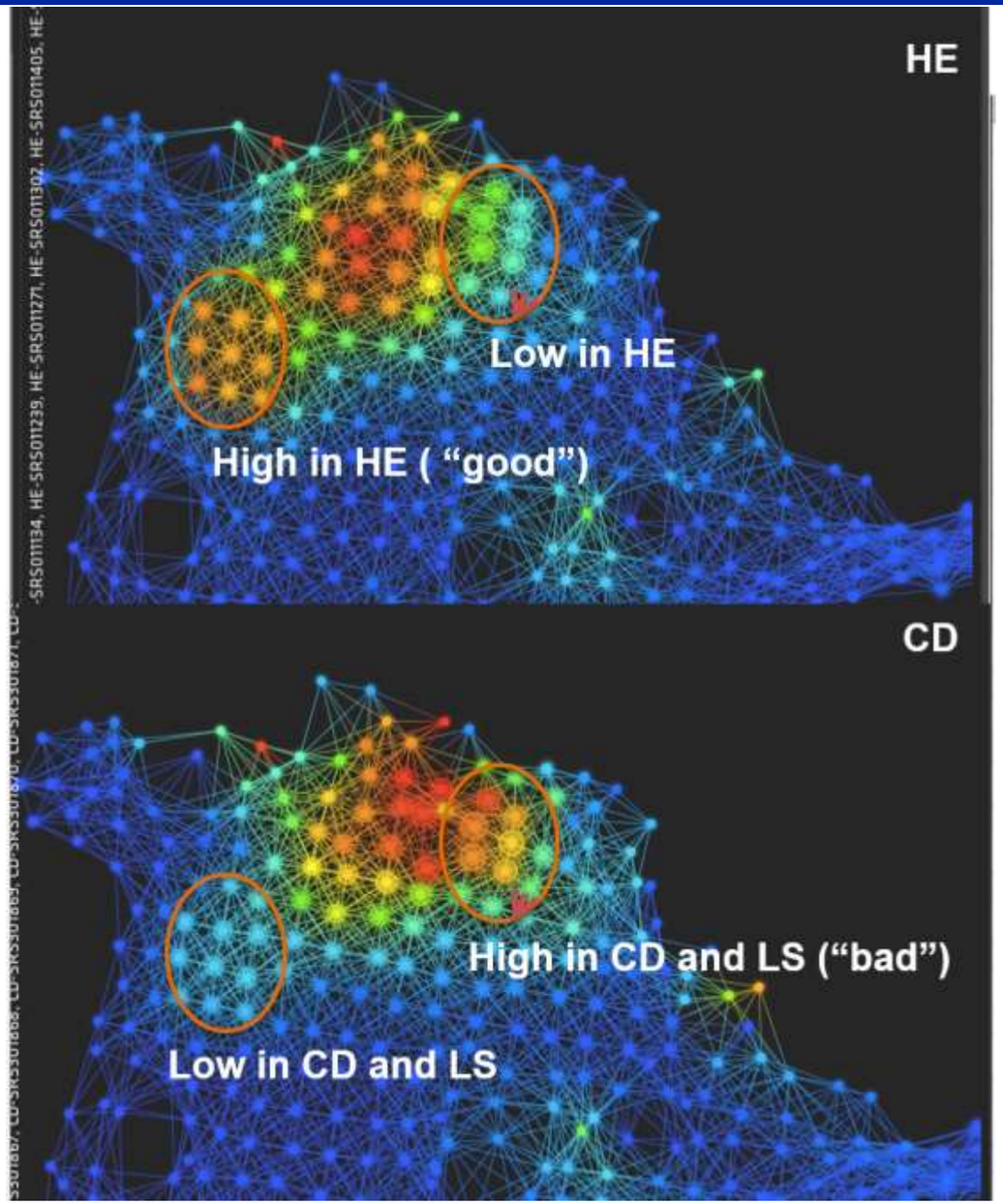
However, Our Research Shows Large Changes in Protein Families Between Health and Disease

Using KEGG Relative Abundance of Protein Families

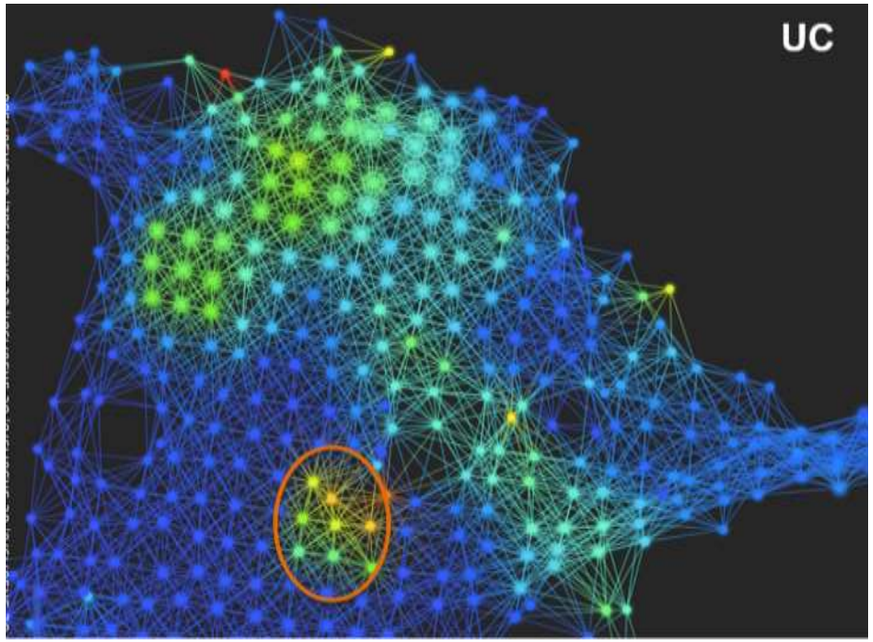


Over 7000 KEGGs Which Are Nonzero in Health and Disease States

Using Ayasdi Interactively to Explore Protein Families in Healthy and Disease States



Dataset from Larry Smarr Team
With 60 Subjects (HE, CD, UC, LS)
Each with 10,000 KEGGs -
600,000 Cells

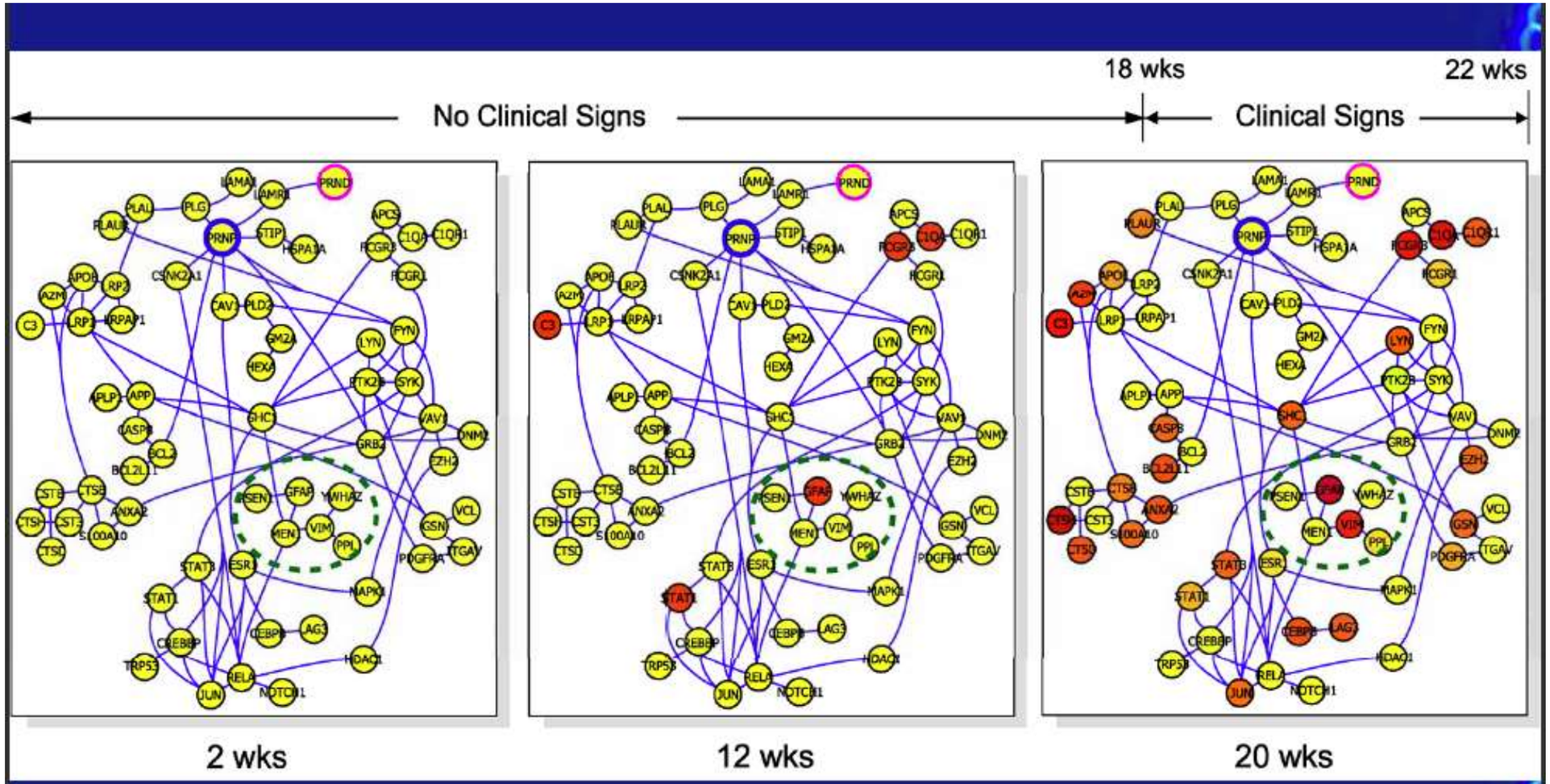


Source: Pek Lum,
Formerly Chief Data Scientist, Ayasdi



Disease Arises from Perturbed Protein Family Networks: Dynamics of a Prion Perturbed Network in Mice

Source: Lee Hood, ISB



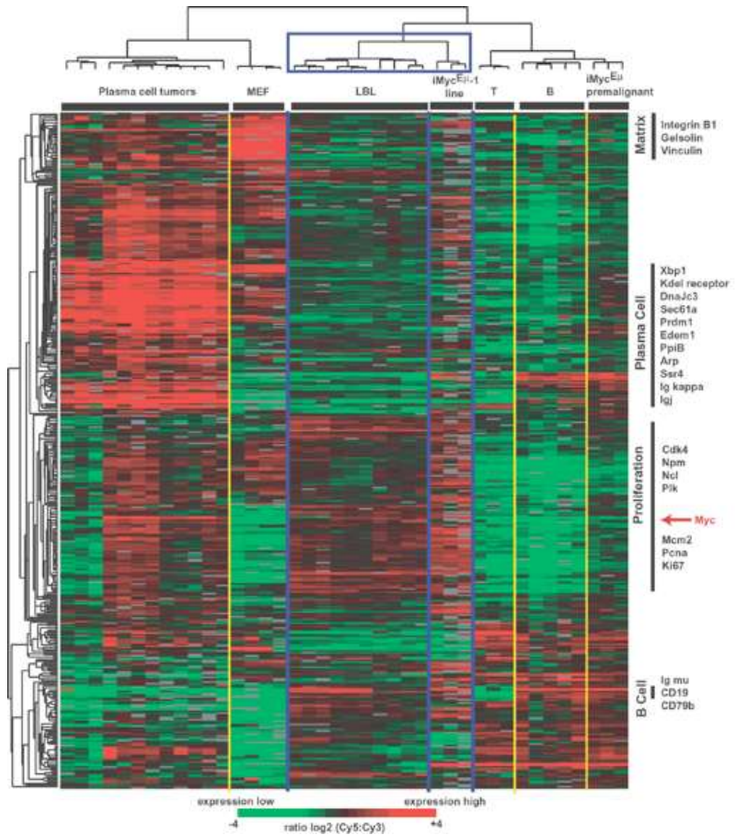
Our Next Goal is to Create Such Perturbed Networks in Humans

Institute for Systems Biology

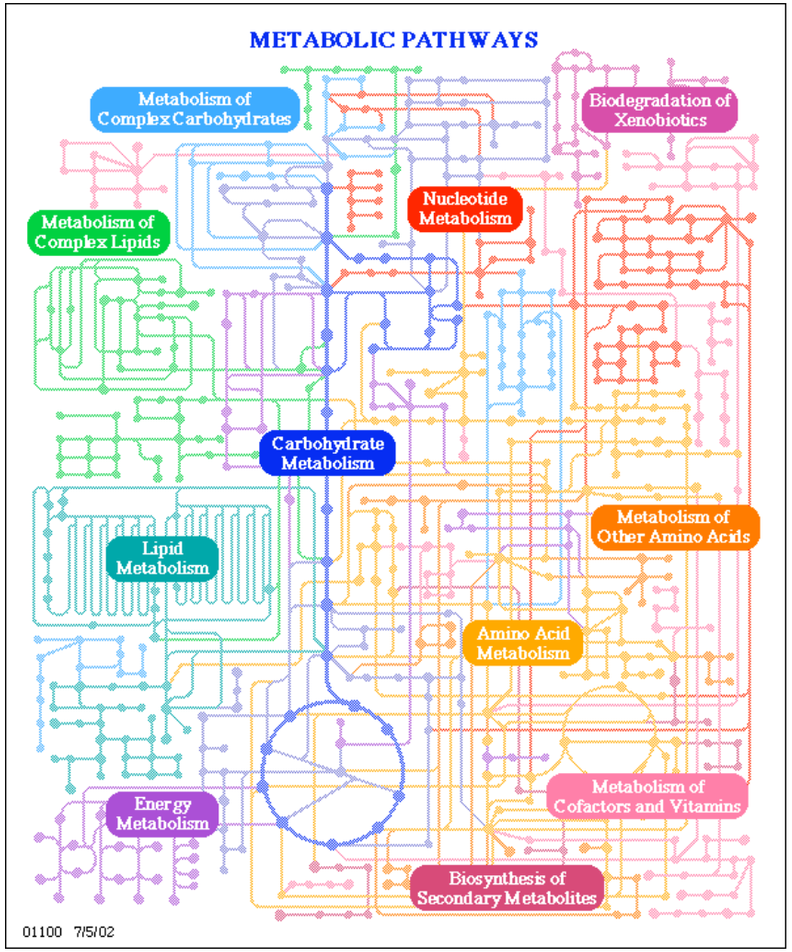
UCSD's Cytoscape Integrates and Visualizes Molecular Networks and Molecular Profiles

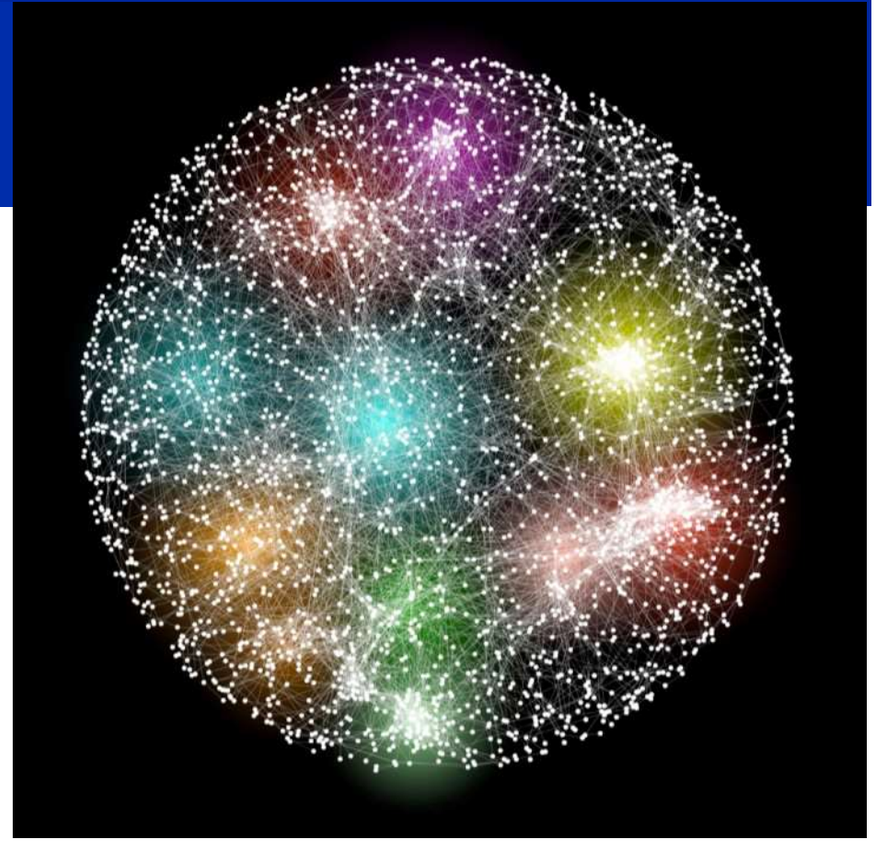
Source: Trey Ideker, UCSD

mRNA & protein expression

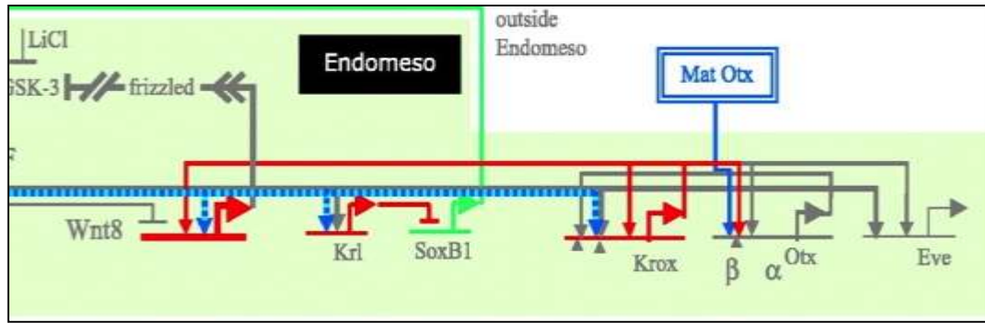


Metabolic networks



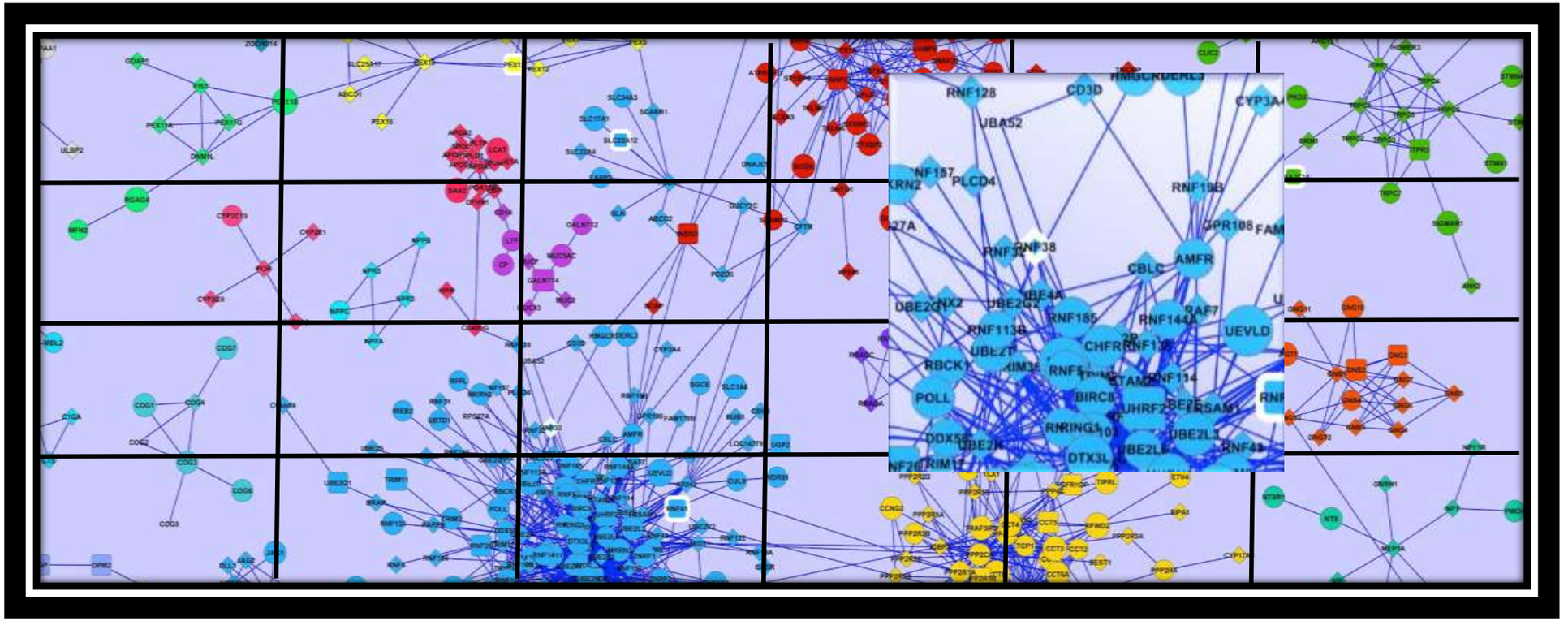


Genetic and protein interaction networks



Transcriptional networks

We Are Enabling Cytoscape to Run Natively on 64M Pixel Visualization Walls and in 3D in VR



Simulation of Cytoscape Running on VROOM

Calit2 VROOM-FuturePatient Expedition

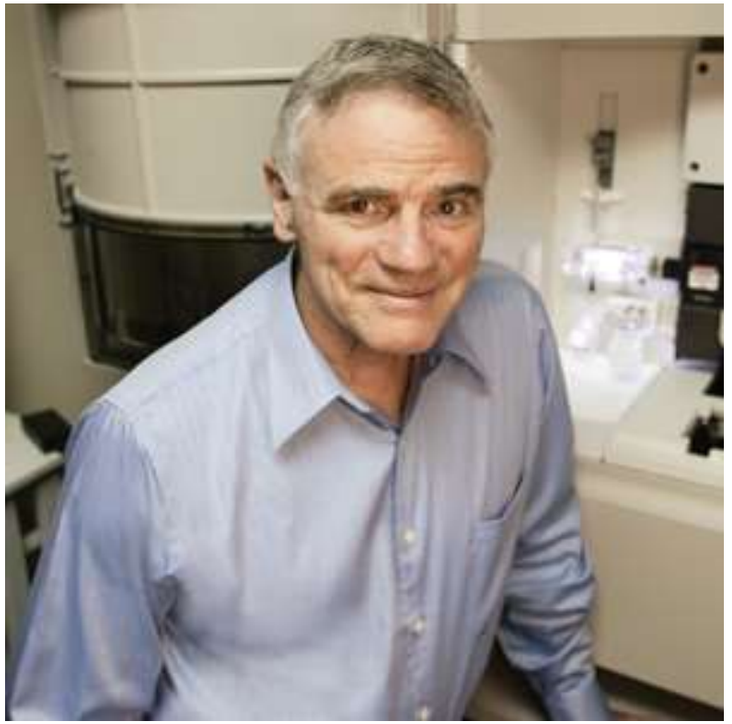
Cytoscape Example from Douglas S. Greer, J. Craig Venter Institute and Jurgen P. Schulze, Calit2's Qualcomm Institute

Next Step: Apply What We Have Learned to Larger Population Microbiome Datasets

- I am a Member of the Pioneer 100
- Our Team Now Has the Gut Microbiomes of the Pioneer 100
- We Plan to Analyze Them for Differences Using These Tools

PIONEER 100
HUNDRED PERSON WELLNESS PROJECT
Institute for Systems Biology

Will Grow to 1000
Then 10,000
Then 100,000



http://isbmolecularme.com/tag/100-pioneers/

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

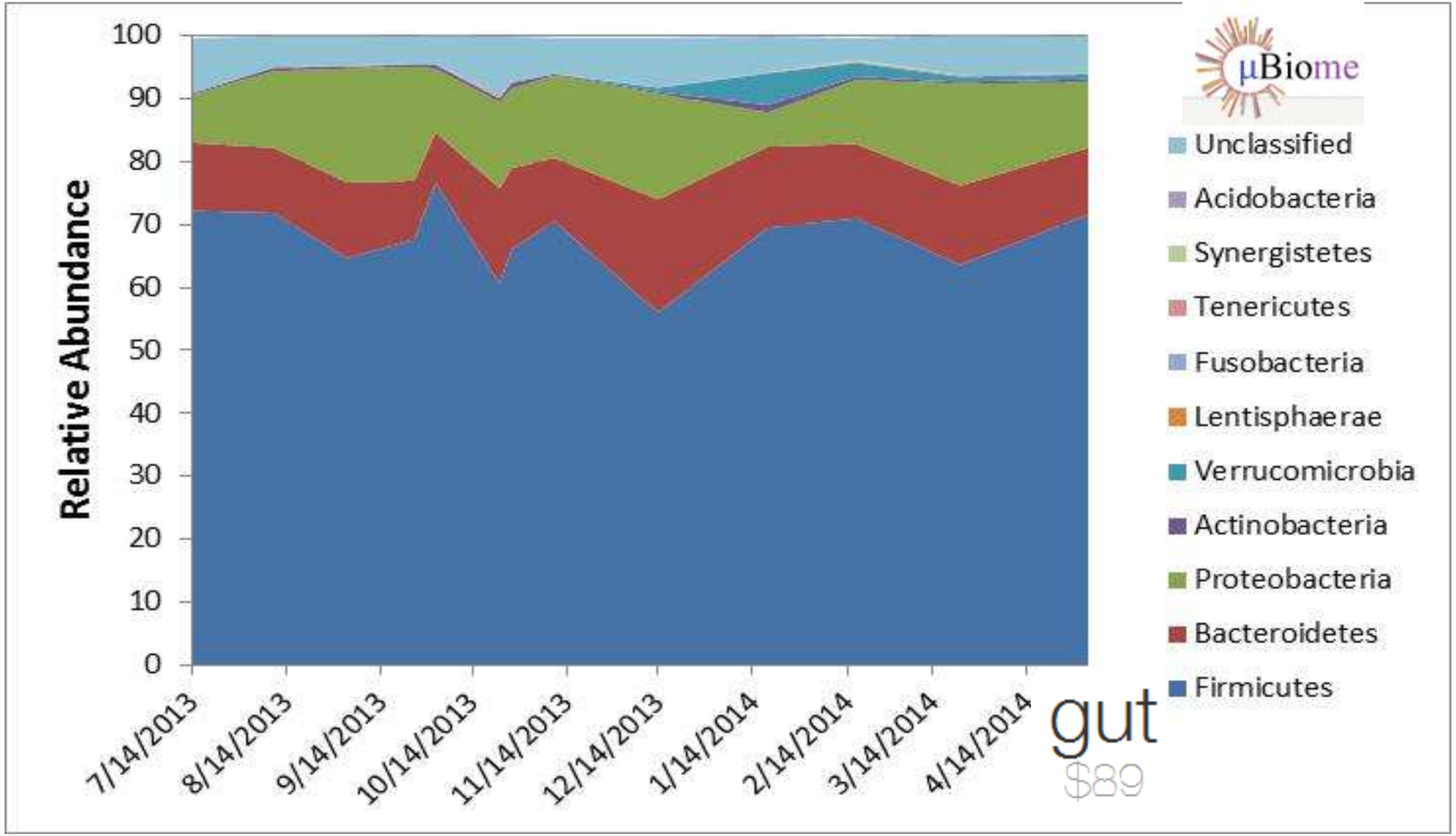
Announced Last Friday!

Inflammatory Bowel Disease Biobank For Healthy and Disease Patients

Already 120 Enrolled,
Goal is 1500

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

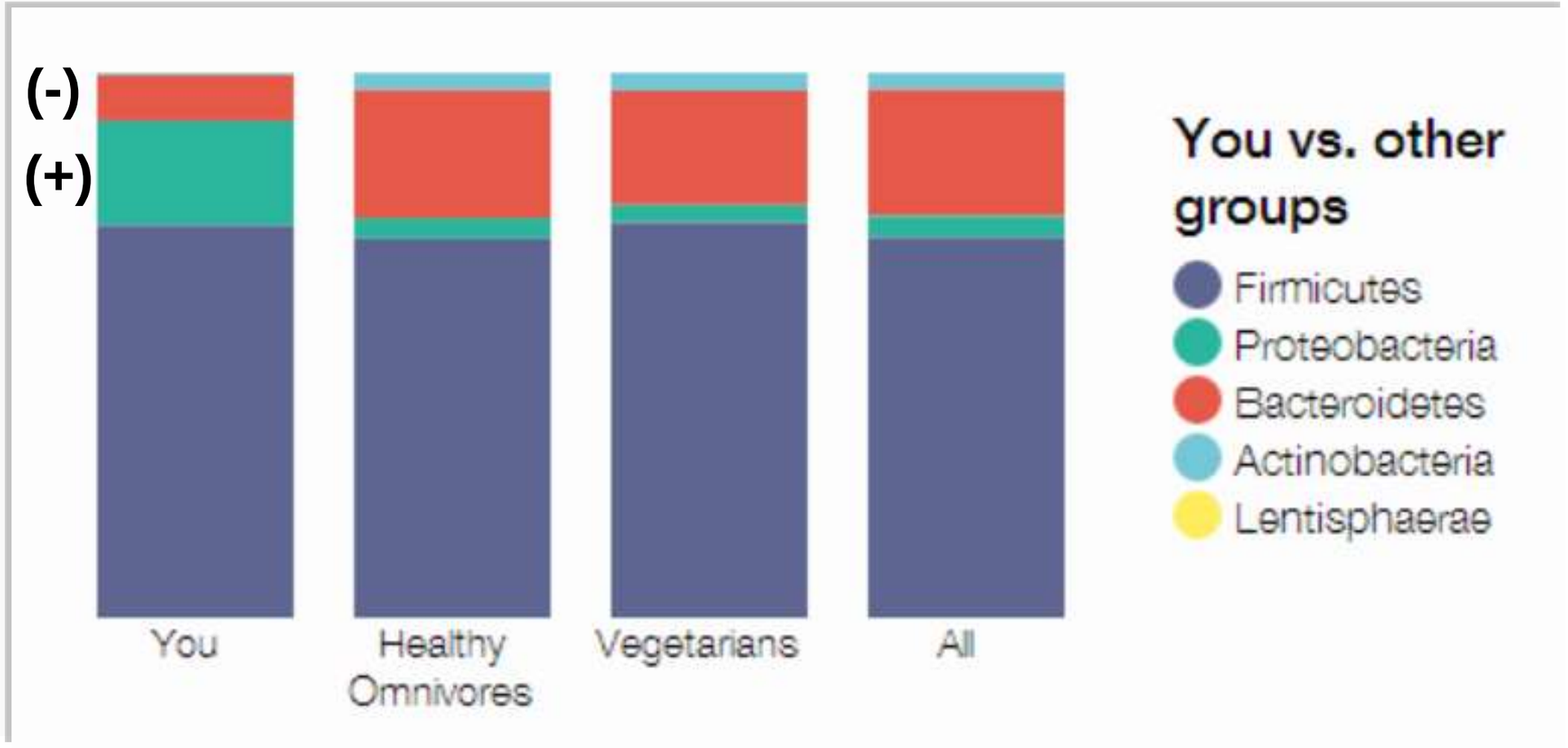
Inexpensive Consumer Time Series of Microbiome Now Possible Through Ubiome



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

By Crowdsourcing, Ubiome Can Show I Have a Major Disruption of My Gut Microbiome

LS Sample on September 24, 2014



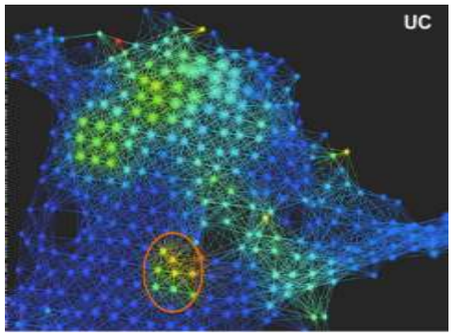
µBiome

Visit Ubiome in the Exponential Medicine Healthcare Innovation Lab

it²

Using Big Data Analytics to Move From Clinical Research to Precision Medicine

1) Identify Patient Cohorts for Treatment



2) Combine Data Types for Full View of Patient

Genetic Data

EMR Data

Financial Data

3) Precision Medicine Pathways @ Point of Care



AYASDI Cure

AYASDI Care

More data collected @ point of care

Continuous Data-Driven Improvement

Thanks to Our Great Team!

UCSD Metagenomics Team

Weizhong Li
Sitao Wu

Calit2@UCSD Future Patient Team

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



Ayasdi
Devi
Sanjnan
Pek

JCVI Team

Karen Nelson
Shibu Yooseph
Manolito Torralba

SDSC Team

Michael Norman
Mahidhar Tatineni
Robert Sinkovits

UCSD Health Sciences Team

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

This Talk Builds on My Two Prior Future Med Presentations

Download Them From:

http://lsmarr.calit2.net/presentations?slideshow=28247009

http://lsmarr.calit2.net/presentations?slideshow=16384993