

“Using Dell’s HPC Cloud & Advanced Analytic Software to Discover Radical Changes in the Human Microbiome in Health and Disease”

**Dell Booth Talk
Supercomputing 2014
New Orleans, LA
November 18, 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

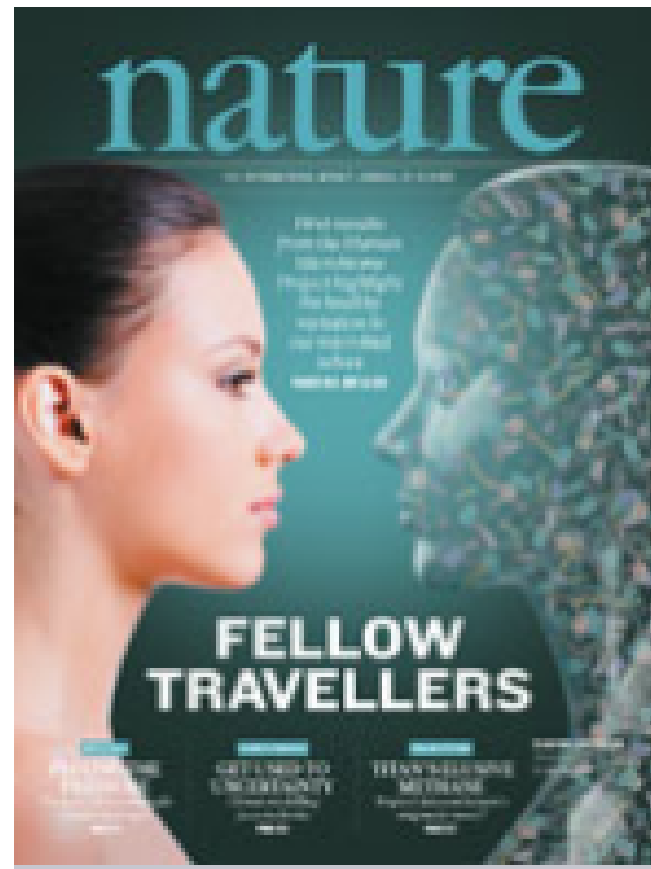
In my SC14 Invited Talk November 19th at 3:30-5pm I will describe how the human body contains ten microbial cells for every human cell, and that these microbes contain 100 times the number of DNA genes that our human DNA does. In this talk, I will discuss the technical details of how we mapped our complex software pipeline onto the Dell HPC Cloud to convert ~3 Trillion DNA bases into a high resolution views of the human gut microbiome ecology across ~300 subjects, some healthy and some with autoimmune disease. Dell then provided access to its analytical experts and advanced analytical software to enable detailed analysis of the dramatic changes in these ecologies. The data mining across of 3/4 million data points led to discoveries of distinct microbial ecology signatures in states of human health and disease.



Intense Scientific Research is Underway on Understanding the Human Microbiome



June 8, 2012



June 14, 2012



August 18, 2012

You Are a SuperOrganism: The Human Genome Contains <1% of the Bodies Genes

Human Microbiome Project



**There are 10 Times
More Bacterial Cells Than
of Human Cells in Your Body**

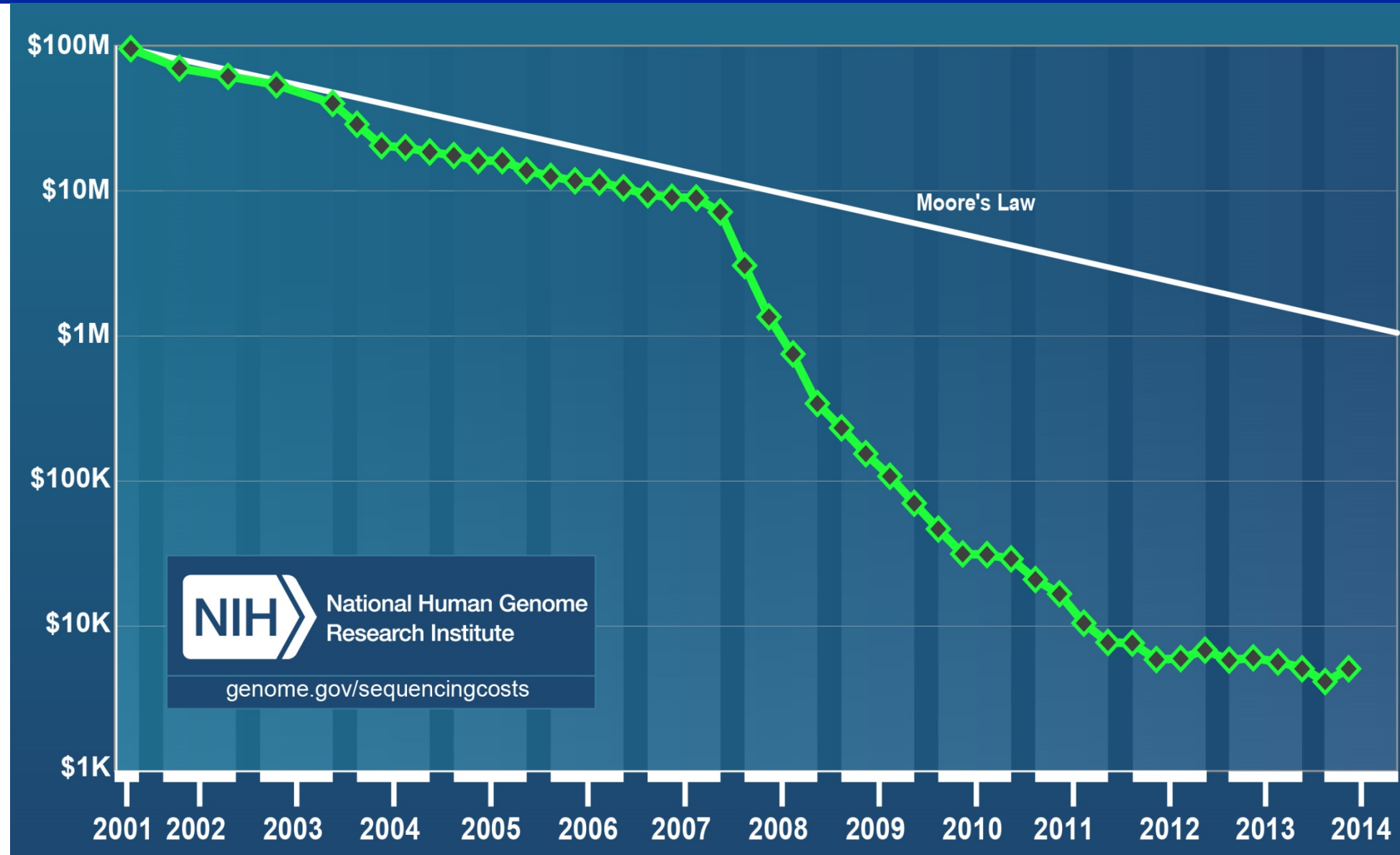
**Inclusion of the Microbiome
Will Radically Change Medicine**



<http://commonfund.nih.gov/hmp/>



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**

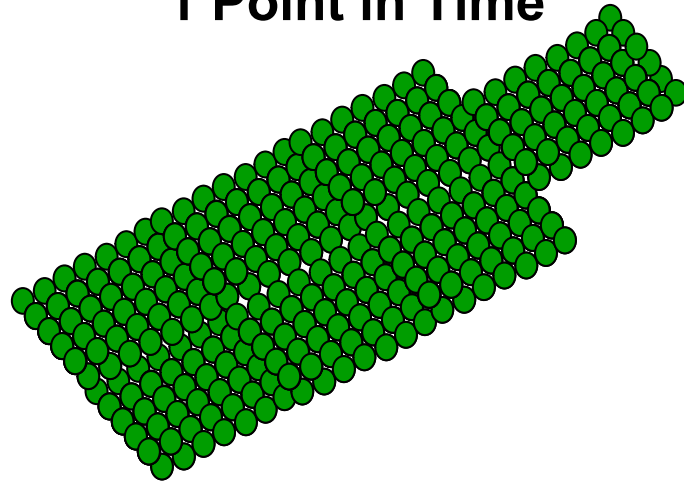


JCVI Sequenced My Gut Microbiome and We Downloaded ~270 More from the NIH Human Microbiome Project For Comparative Analysis

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

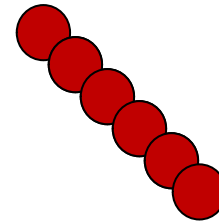
“Healthy” Individuals

250 Subjects
1 Point in Time

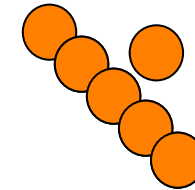


Inflammatory Bowel Disease (IBD) Patients

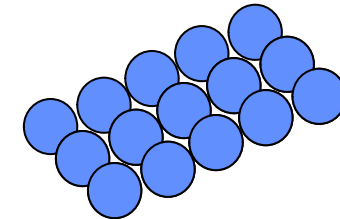
Larry Smarr
(Colonic Crohn's)
7 Points in Time



2 Ulcerative Colitis Patients,
6 Points in Time



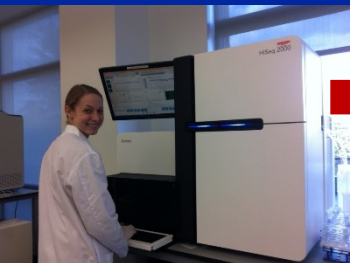
5 Ileal Crohn's Patients,
3 Points in Time



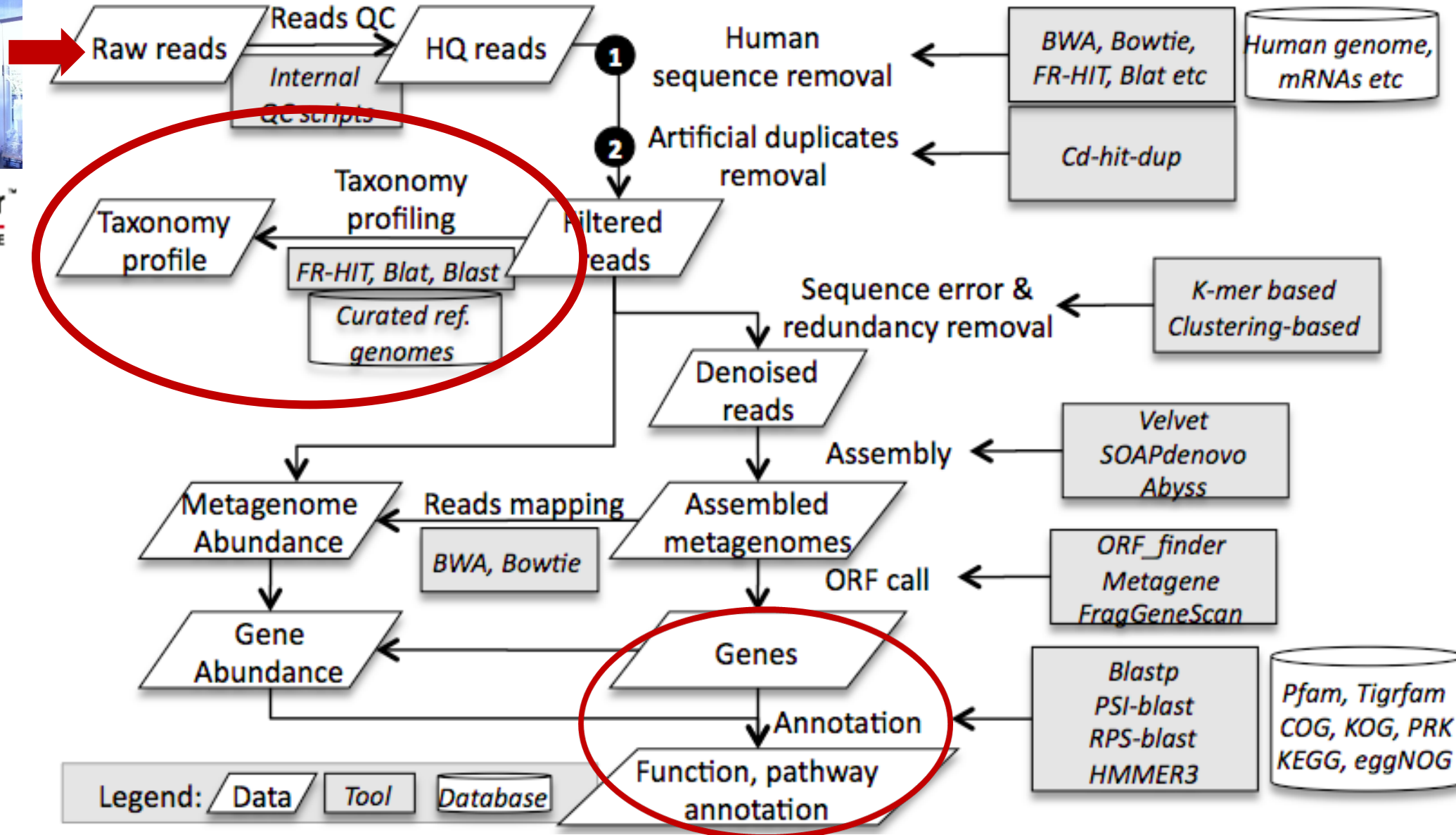
**Total of 27 Billion Reads
Or 2.7 Trillion Bases**

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

Computational NextGen Sequencing Pipeline: From Sequence to Taxonomy and Function



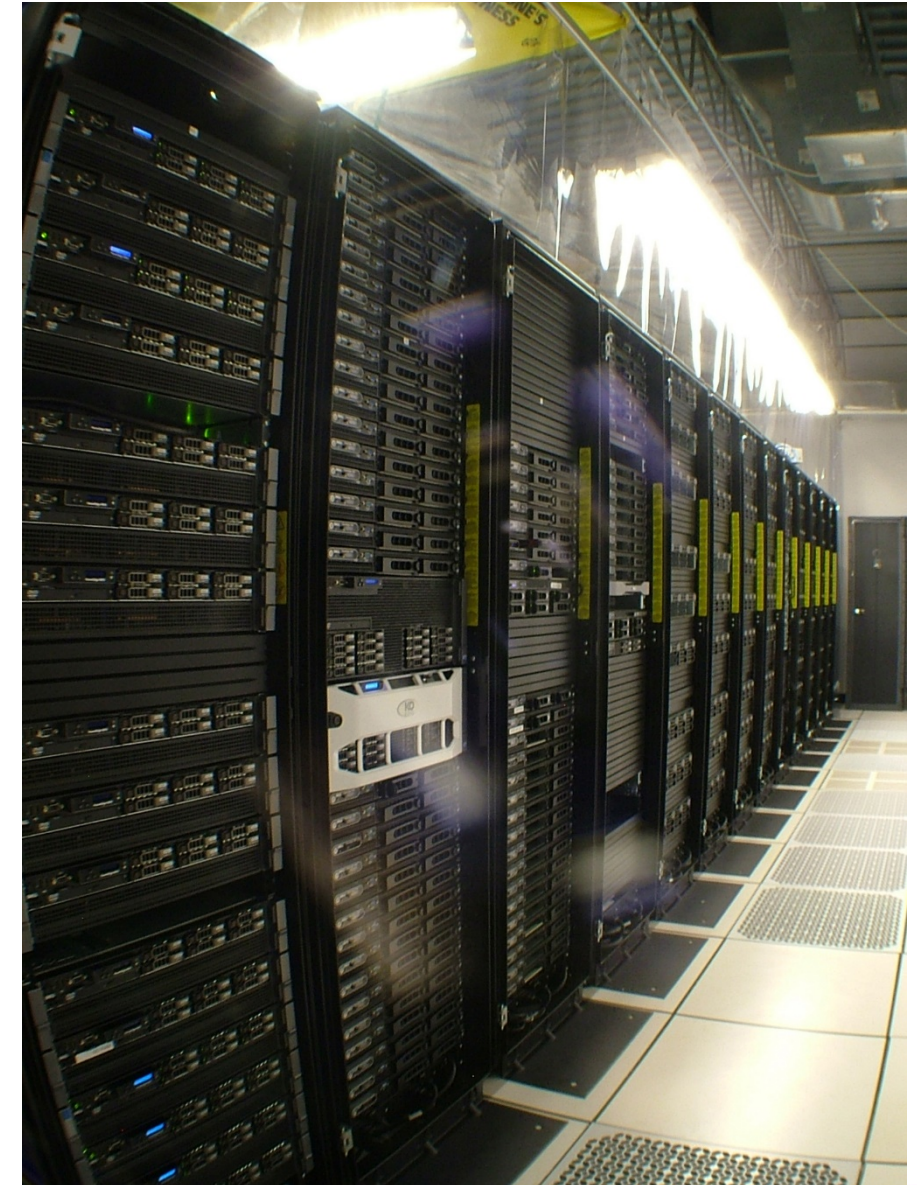
J. Craig Venter™
INSTITUTE



Dell HPC Cloud Bare Metal Solutions



- **Large-Scale Core-Count Clusters with Infiniband Interconnectivity**
- **High Memory Configurations Available for Memory Intensive Workloads**
- **User Support for the Novice or Experienced HPC Customer**
- **Remotely Accessible Over the Internet**



Dell Cloud for HPC with R-Systems



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

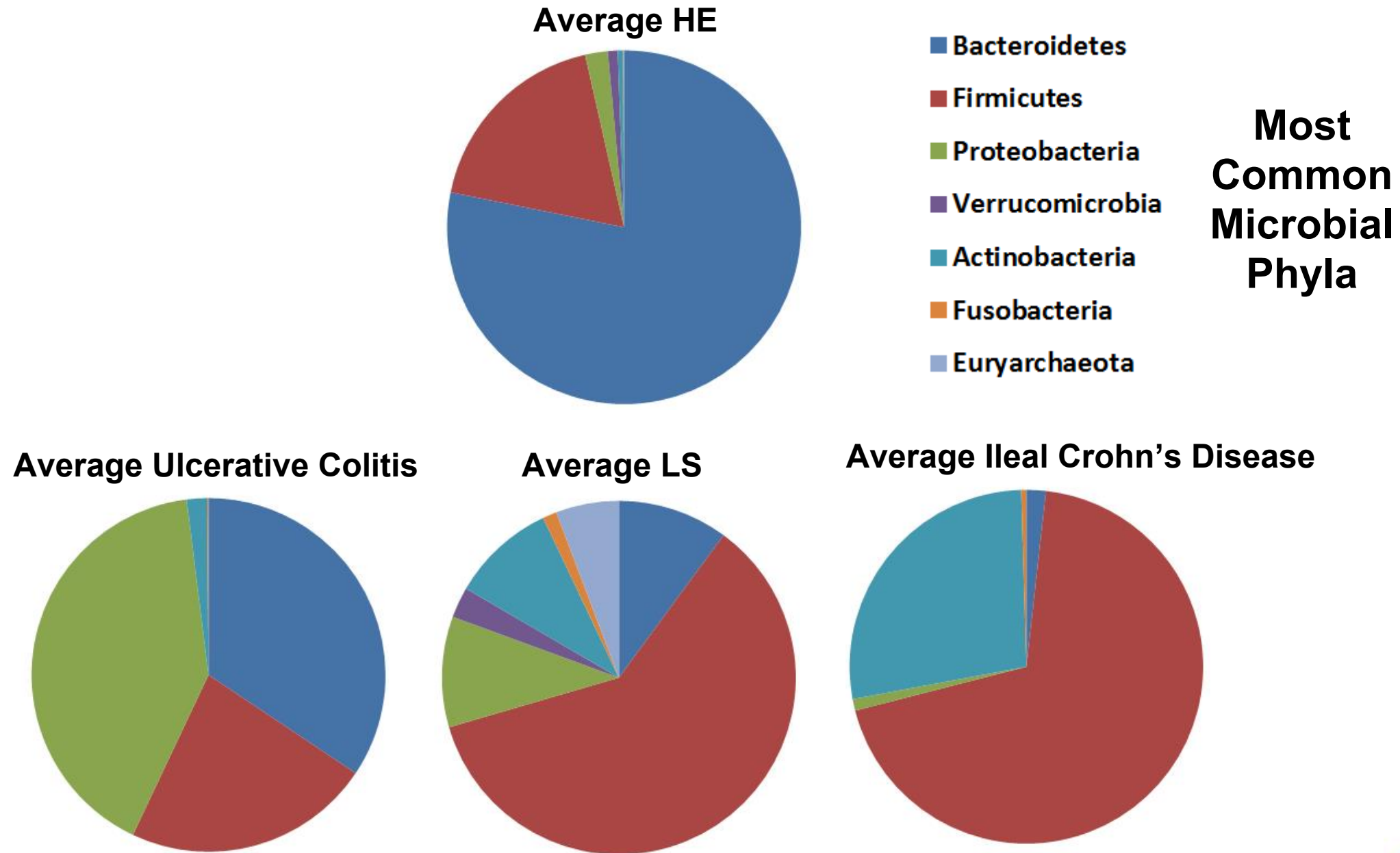
- **Dell's Sanger Cluster**
 - 32 Nodes, 512 Cores
 - 48GB RAM per Node
 - 50GB SSD Local Drive, 390TB Lustre File System
- **We Processed the Taxonomic Relative Abundance**
 - Used ~35,000 Core-Hours on Dell's Sanger
- **Produced Relative Abundance of
~10,000 Bacteria, Archaea, Viruses in ~300 People**
 - ~3Million Filled Spreadsheet Cells
- **New System: R Bio-Gen System**
 - 48 Nodes, 768 Cores
 - 128 GB RAM per Node



Dell Solutions Center
Industry Solutions Lab
SANGER DSU

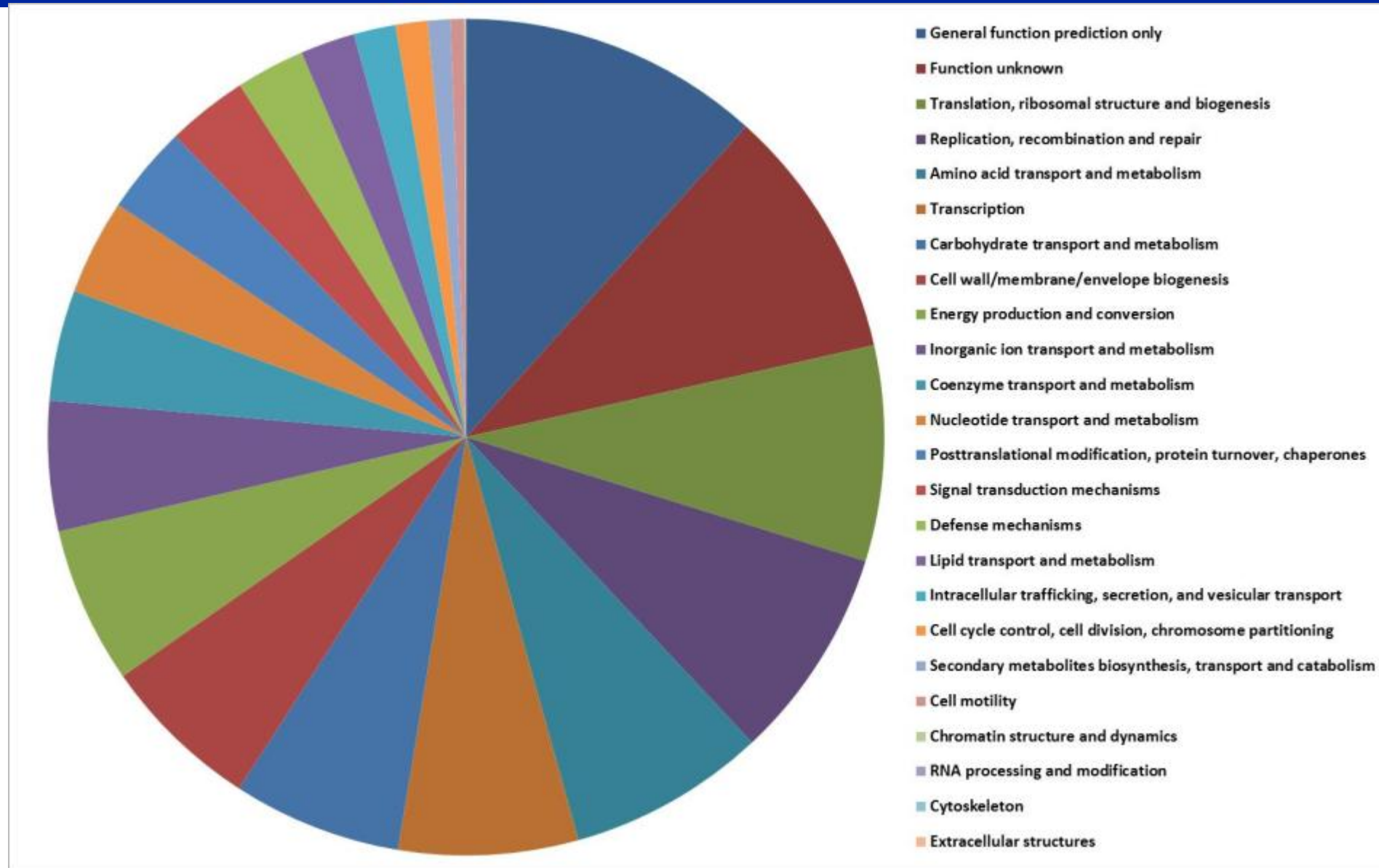


We Found Major State Shifts in Microbial Ecology Phyla Between Healthy and Two Forms of IBD



From Taxonomy to Function: Analysis of Microbiome Protein Families

For More on
Function
See My SC14
Invited Talk
Tomorrow
3:30pm
New Orleans
Theatre



Clusters of
Orthologous
Groups
(COGs)

Next Step: Compute Genes and Function For All ~300 People's Gut Microbiome

**Full Processing to Function:
Genes & Protein Families
(COGs, KEGGs)**

**Would Require
~1-2 Million
Core-Hours**

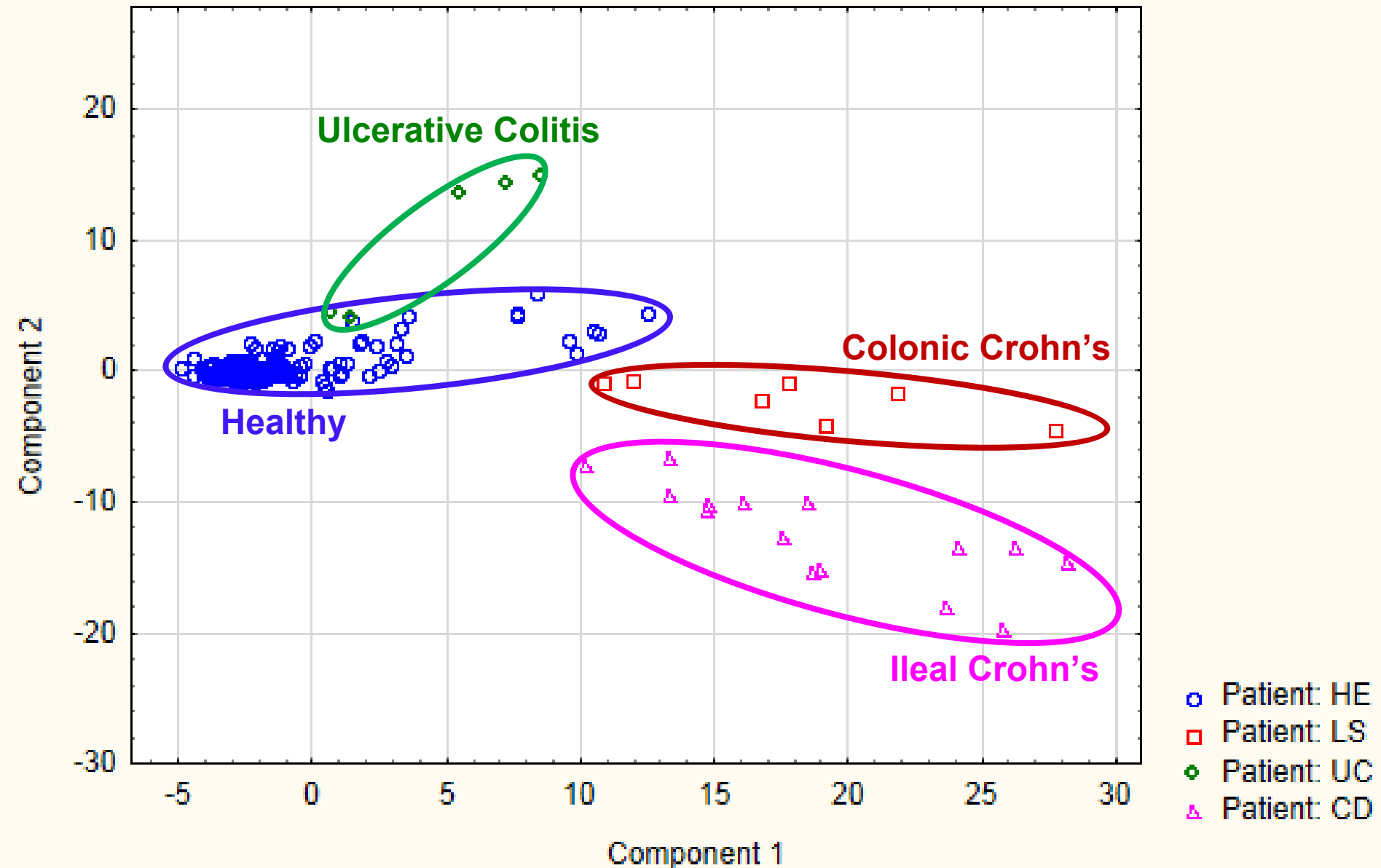
**New Internet2/CENIC 10Gbps Network to Move Data
From Dell / R Systems to Calit2@UC San Diego**

Using Dell HPC Cloud and Dell Analytics to Discover Microbial Diagnostics for Disease Dynamics

- **Can We Distinguish Noninvasively Between Health and Disease States?**
- **Are There Subsets of Health or Disease States?**
- **Can We Track Time Development of the Disease State?**
- **Can Novel Microbial Diagnostics Differentiate Health and Disease States?**



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

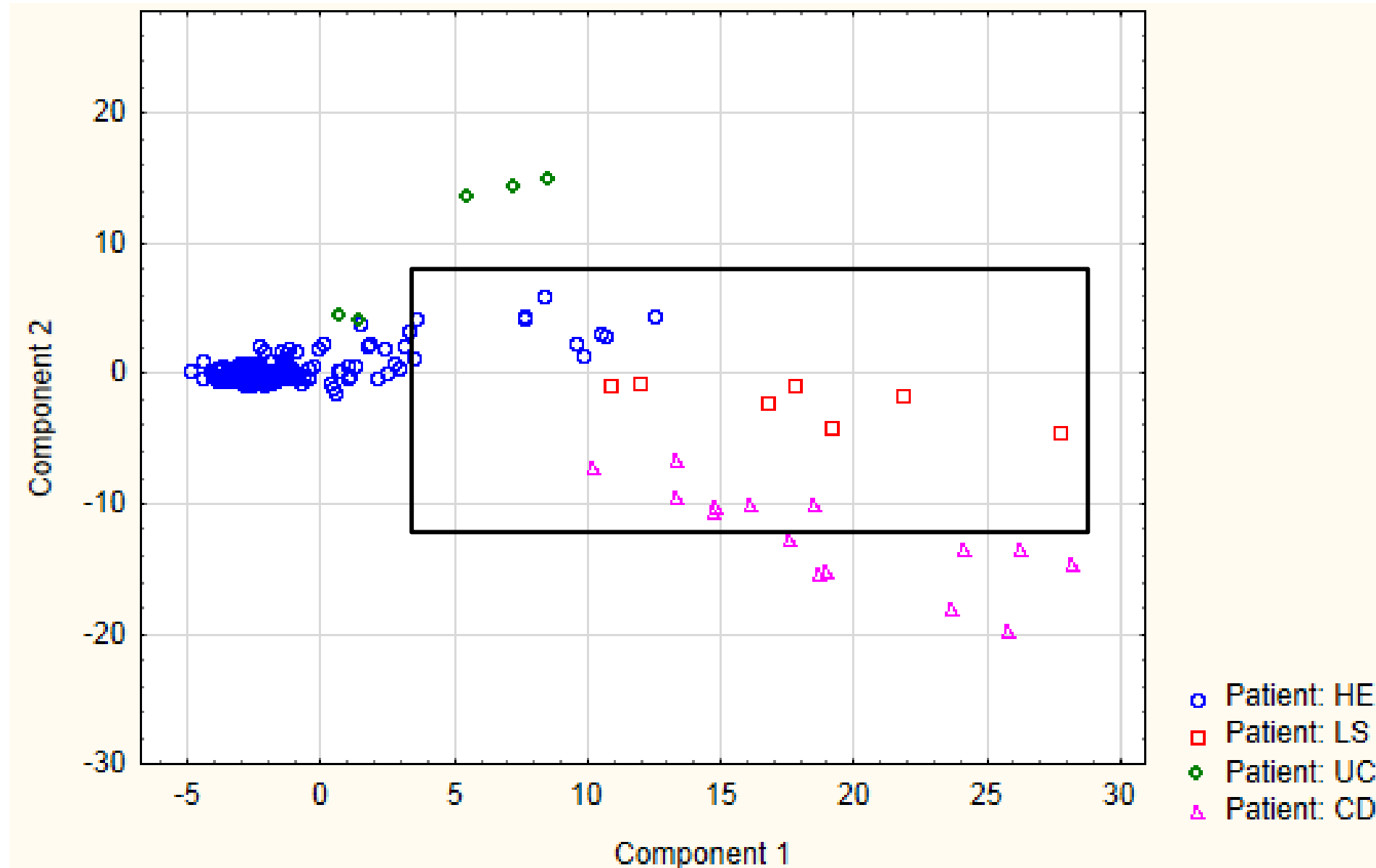


Source: Thomas Hill, Ph.D.
Executive Director Analytics

Dell | Information Management Group, Dell Software



I Built on Dell Analytics to Show Dynamic Evolution of My Microbiome Toward and Away from Healthy State – Colonic Crohn's



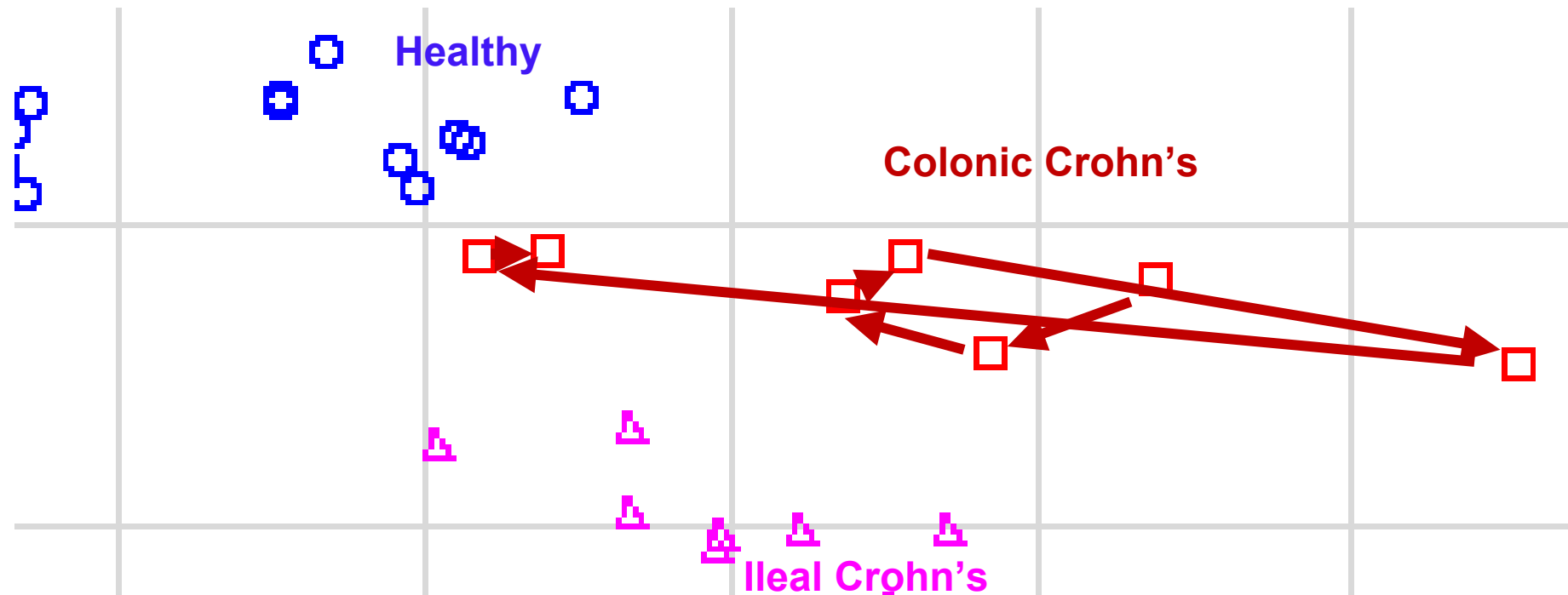
Source: Thomas Hill, Ph.D.
Executive Director Analytics
Dell | Information Management Group, Dell Software



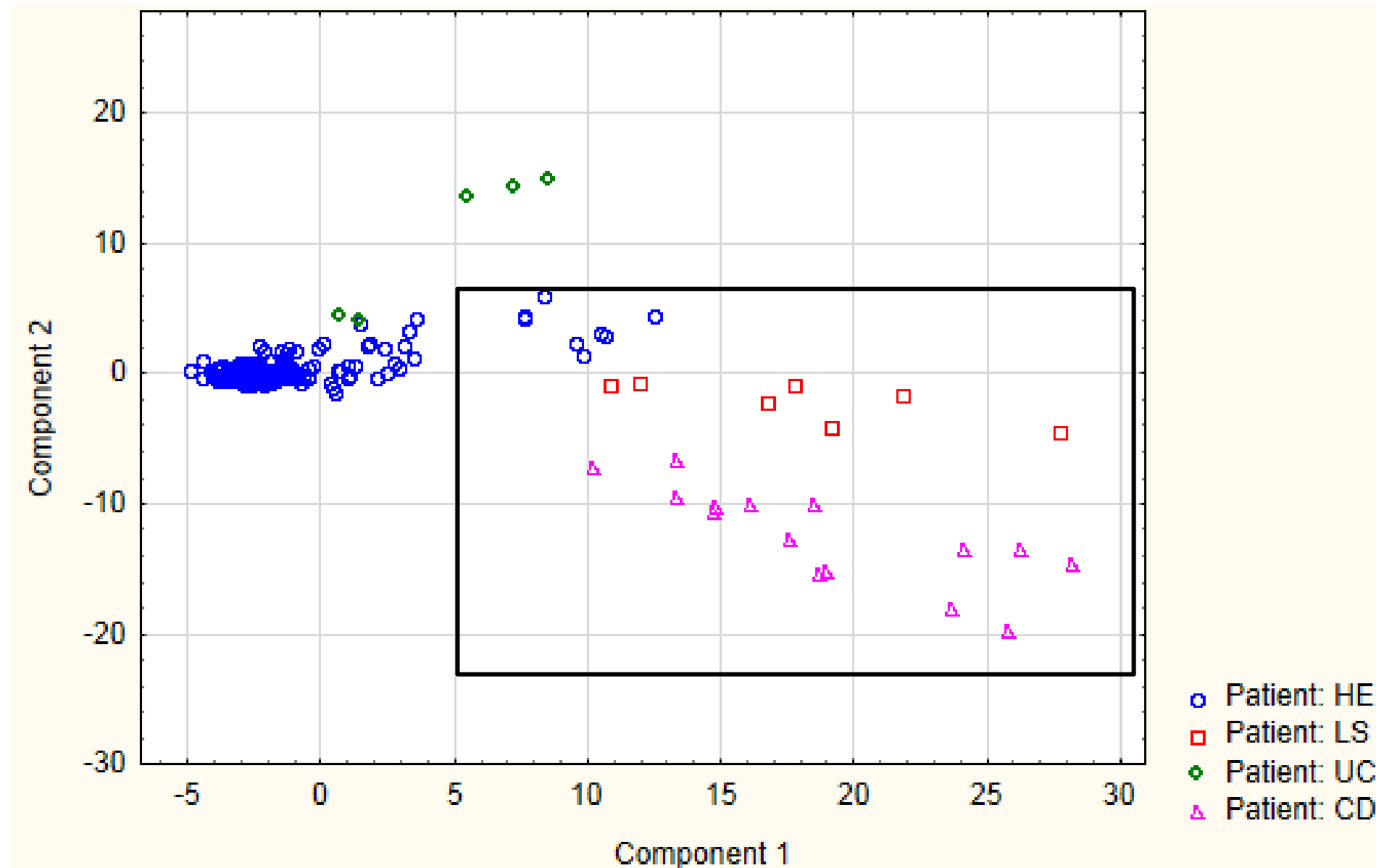
I Built on Dell Analytics to Show Dynamic Evolution of My Microbiome Toward and Away from Healthy State – Colonic Crohn's

□ Patient: LS

Seven Time Samples Over 1.5 Years



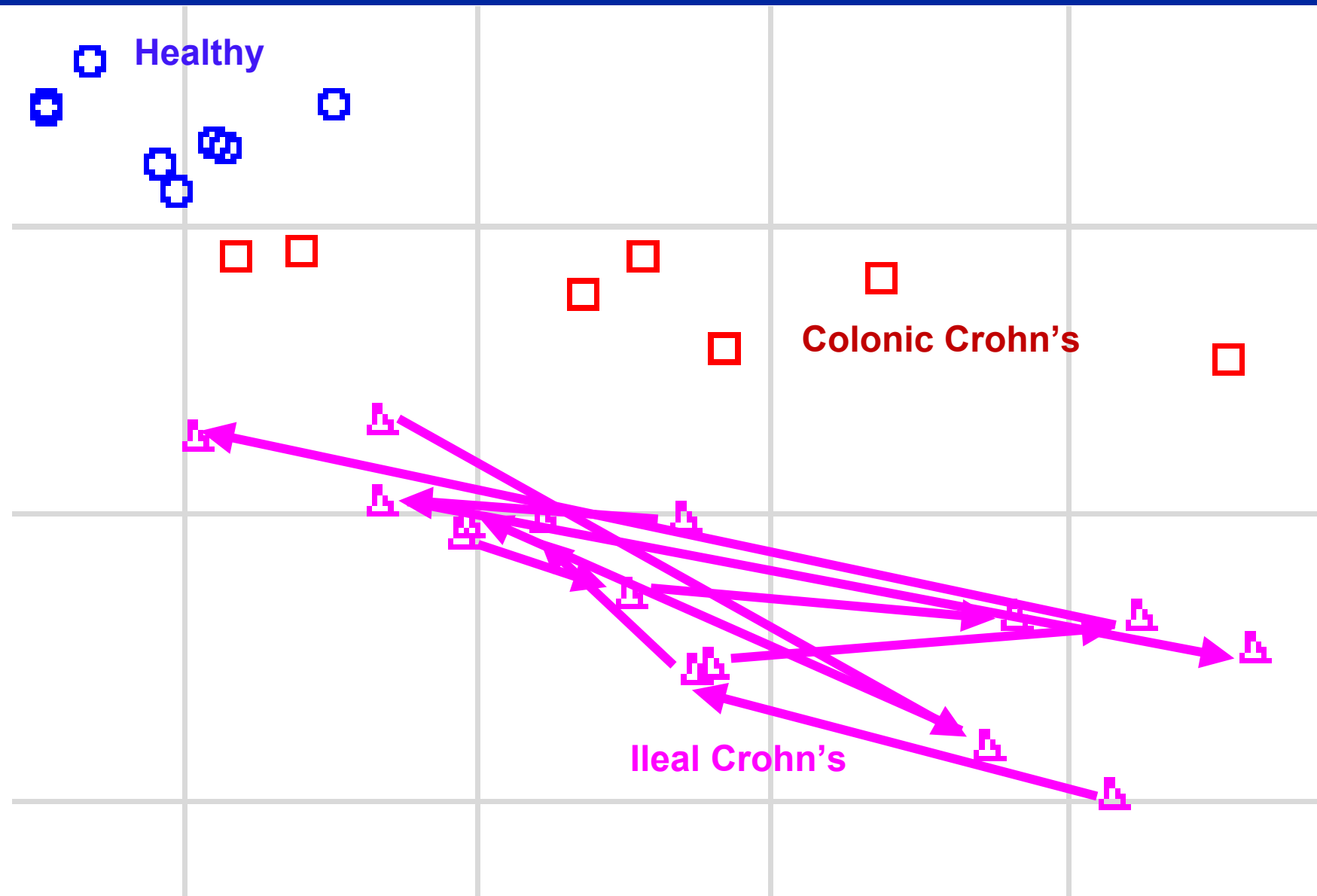
I Built on Dell Analytics to Show Dynamic Evolution of My Microbiome Toward and Away from Healthy State – Ileal Crohn's



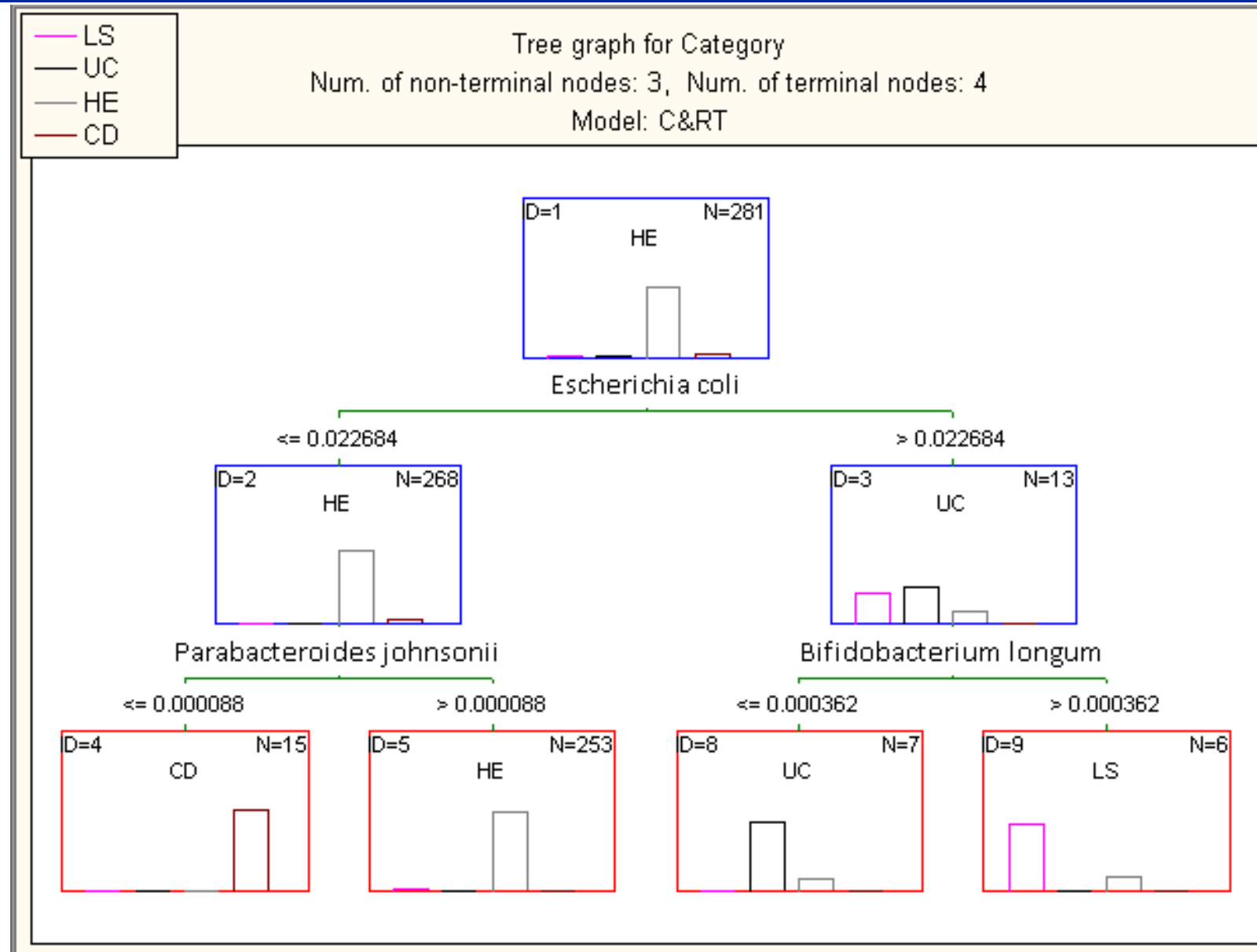
Source: Thomas Hill, Ph.D.
Executive Director Analytics
Dell | Information Management Group, Dell Software



I Built on Dell Analytics to Show Dynamic Evolution of My Microbiome Toward and Away from Healthy State – Ileal Crohn's



Dell Analytics Tree Graphs Classifies the 4 Health/Disease States With Just 3 Microbe Species



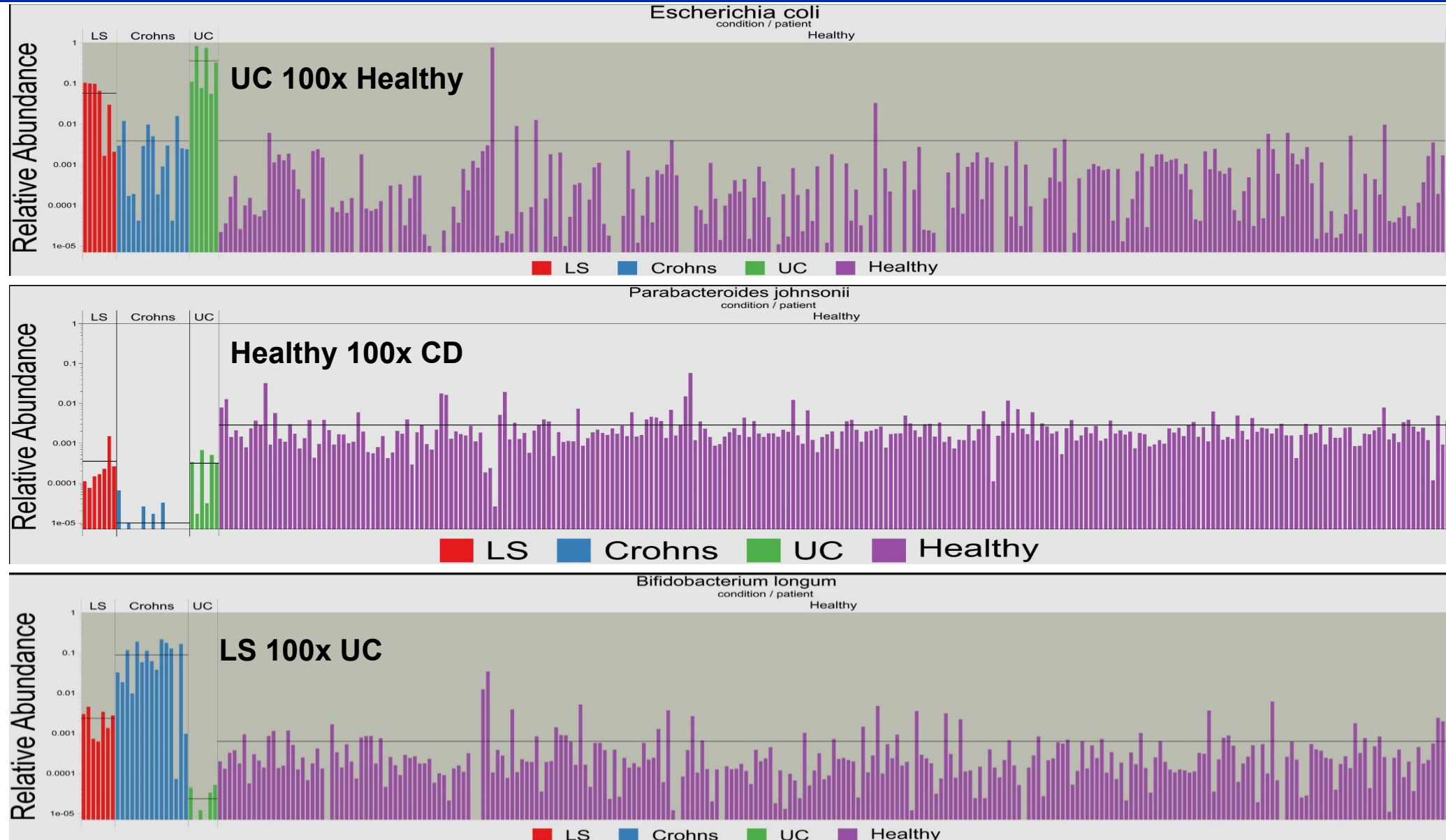
Source: Thomas Hill, Ph.D.

Executive Director Analytics

Dell | Information Management Group, Dell Software



Our Relative Abundance Results Across ~300 People Show Why Dell Analytics Tree Classifier Works



We Produced Similar Results for ~2500 Microbial Species



Dell Analytics Determines Best Candidates for IBD Microbial Diagnostics

P<0.001

Akkermansia muciniphila

Alistipes putredinis

Ruminococcus bromii

Parabacteroides johnsonii

Acinetobacter baumannii

Alistipes shahii

Alistipes finegoldii

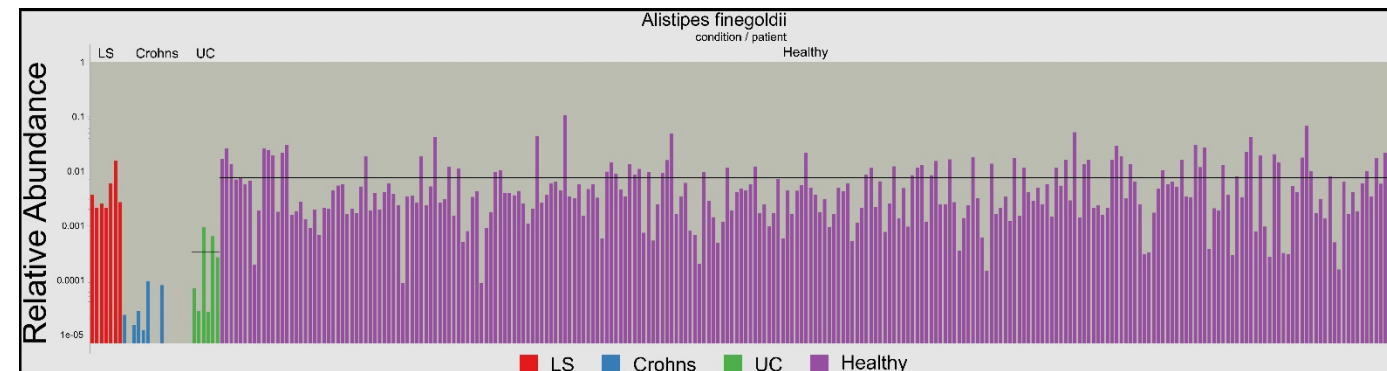
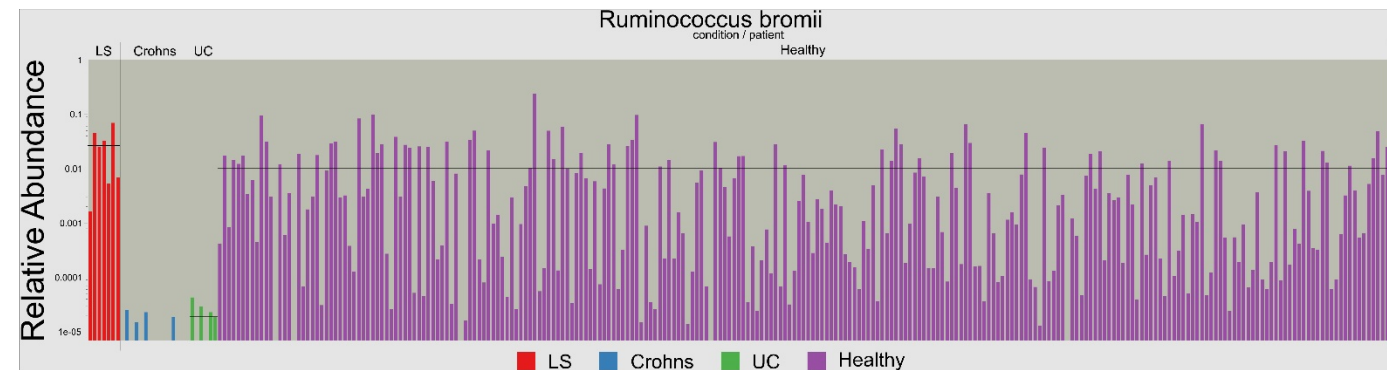
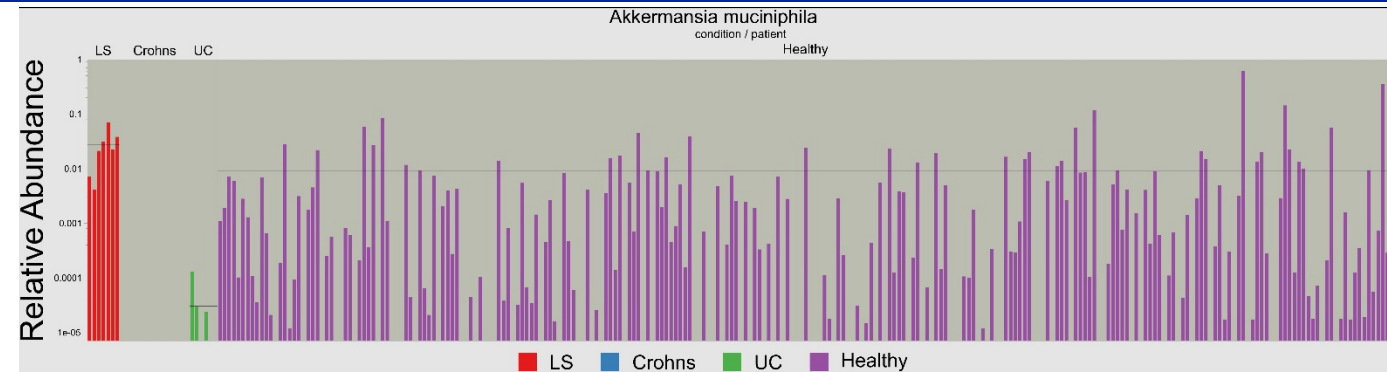
Ruminococcus albus

Methanobrevibacter smithii

Bifidobacterium animalis

Prevotella copri

Prevotella stercorea



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

Announced November 7, 2014!

**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**



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



JCVI Team

Karen Nelson
Shibu Yooseph
Manolito Torralba



SDSC Team

Michael Norman
Mahidhar Tatineni
Robert Sinkovits



Dell/R Systems and Dell Analytics

Brian Kucic
John Thompson
Tom Hill

UCSD Health Sciences Team

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

