“Determining the Human Gut Microbiome using Genome Sequencing and Dell’s Cloud Computing”

Dell Webinar

April 29, 2014

Dr. Larry Smarr

Director, California Institute for Telecommunications and Information Technology

Harry E. Gruber Professor,

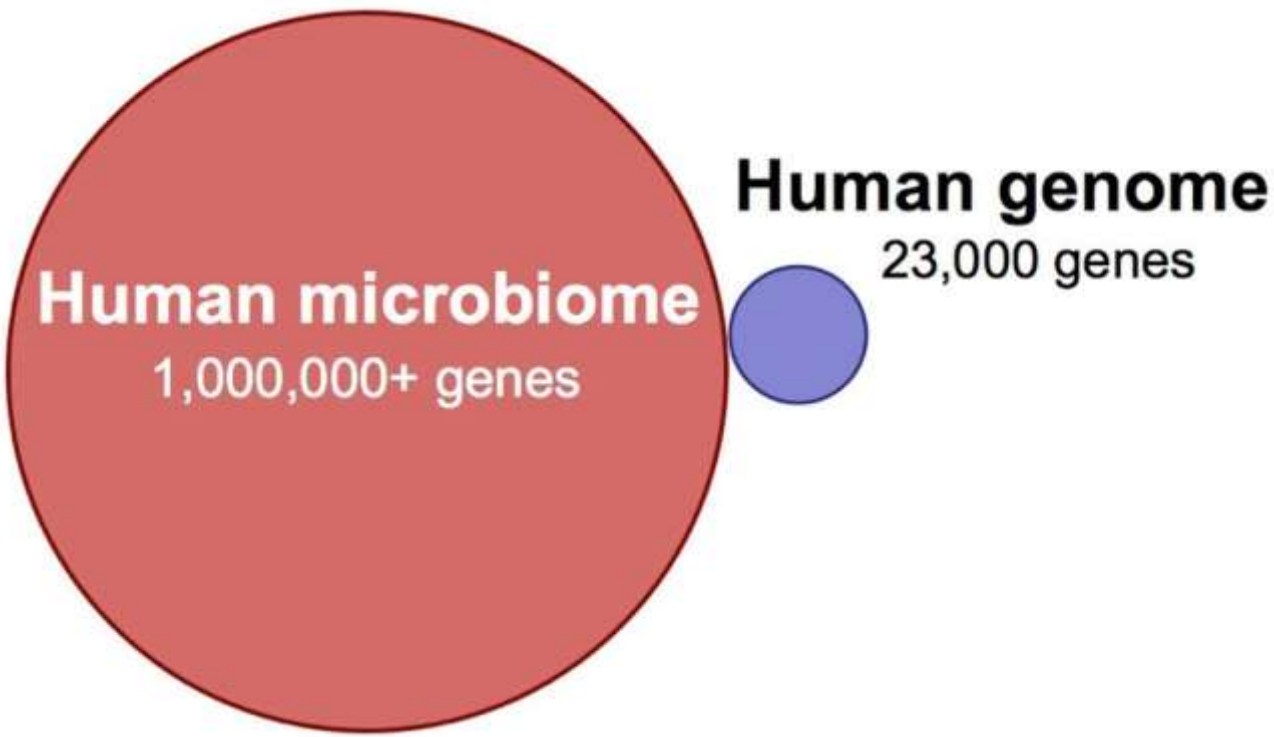
Dept. of Computer Science and Engineering

Jacobs School of Engineering, UCSD

http://lsmarr.calit2.net

The Human Microbiome Ecology is Critical to Health and Disease

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



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

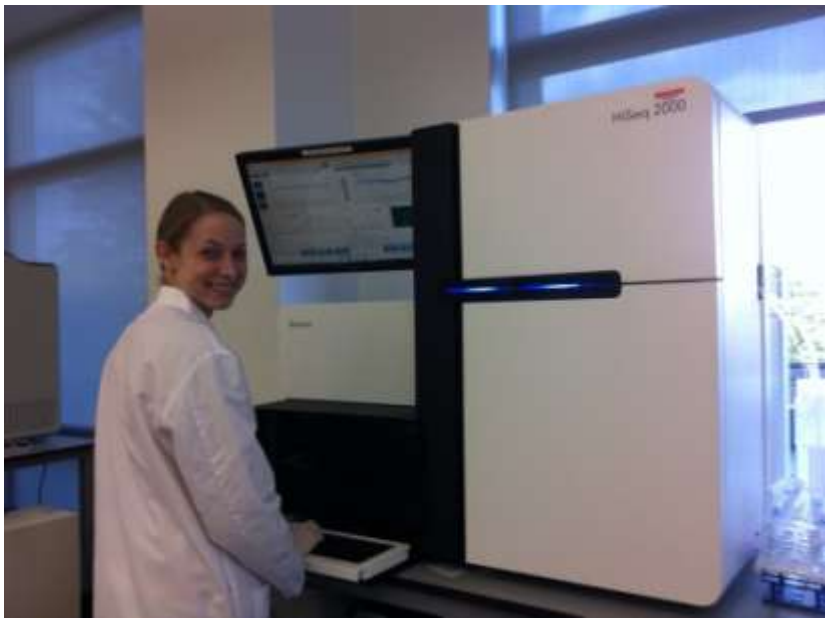
Inclusion of the Microbiome Will Radically Change Medicine

J. Craig Venter™ INSTITUTE

it²

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
- 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

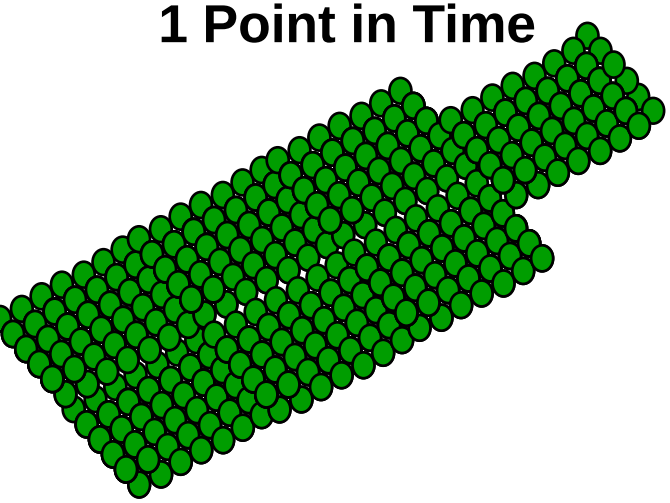
it²

We Downloaded Additional Phenotypes from NIH's Human Microbiome Program For Comparative Analysis

Download Raw Reads
~100M Per Person

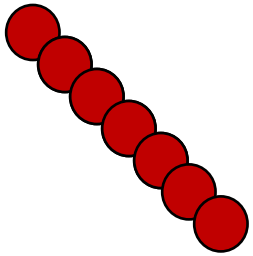
"Healthy" Individuals

250 Subjects
1 Point in Time



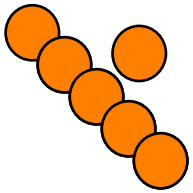
Larry Smarr

7 Points in Time
Over 1.5 Years



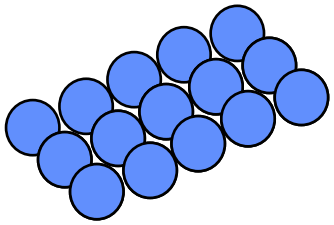
"Disease" Patients

2 Ulcerative Colitis Patients,
6 Points in Time



Inflammatory Bowel Disease

5 Ileal Crohn's Patients,
3 Points in Time



Total of ~28 Billion Reads
Or 2.8 Trillion DNA 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

NIH

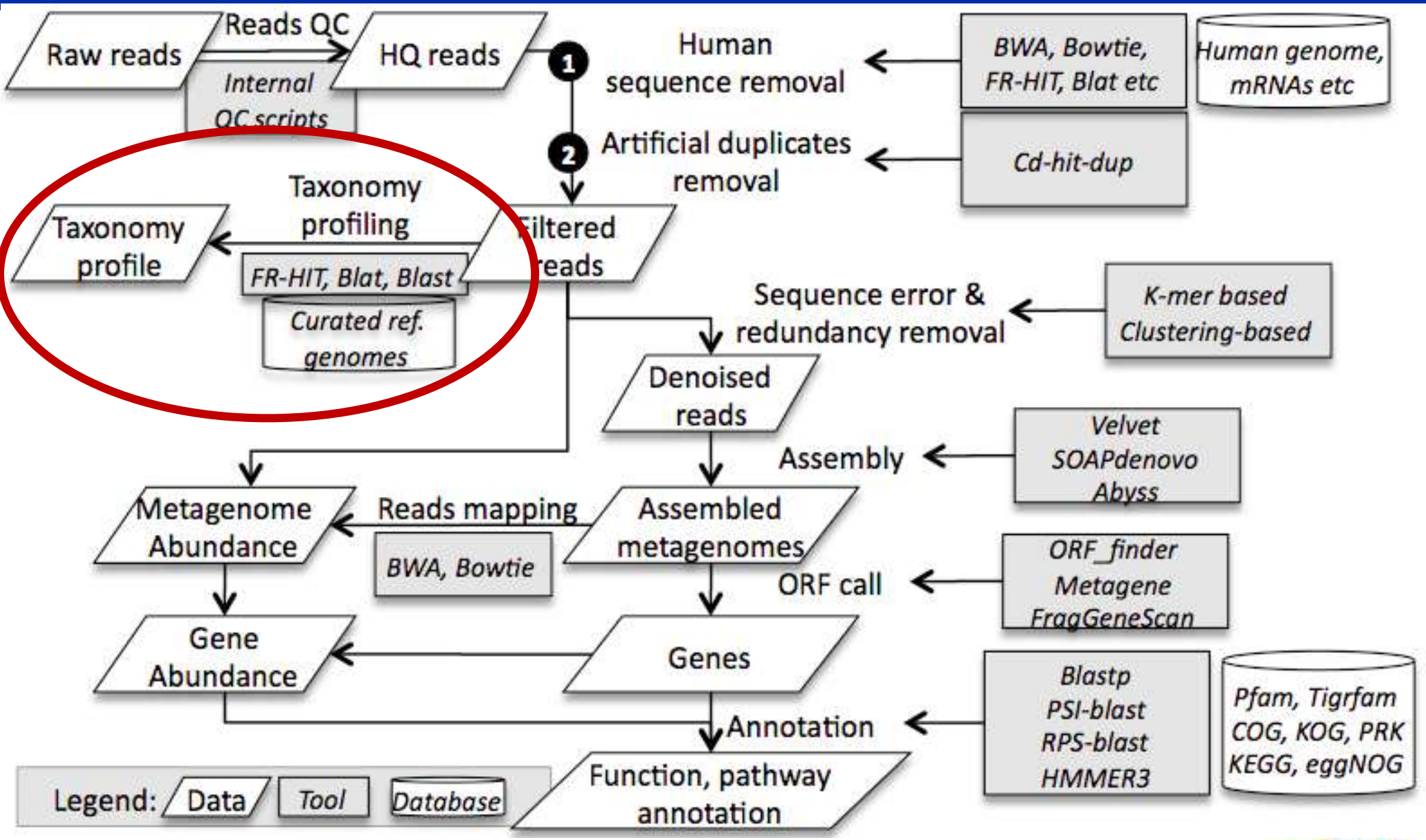
Now to Align Our 28 Billion Reads Against the Reference Database

J. Craig Venter INSTITUTE

Source: Weizhong Li, Sitao Wu, CRBS, UCSD

it²

Computational NextGen Sequencing Pipeline: From Sequence to Taxonomy and Function



National Human Genome Research Institute

HMP

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

CRBS

it²

We Used Dell's Cloud (Sanger) to Analyze All of Our Human Gut Microbiomes

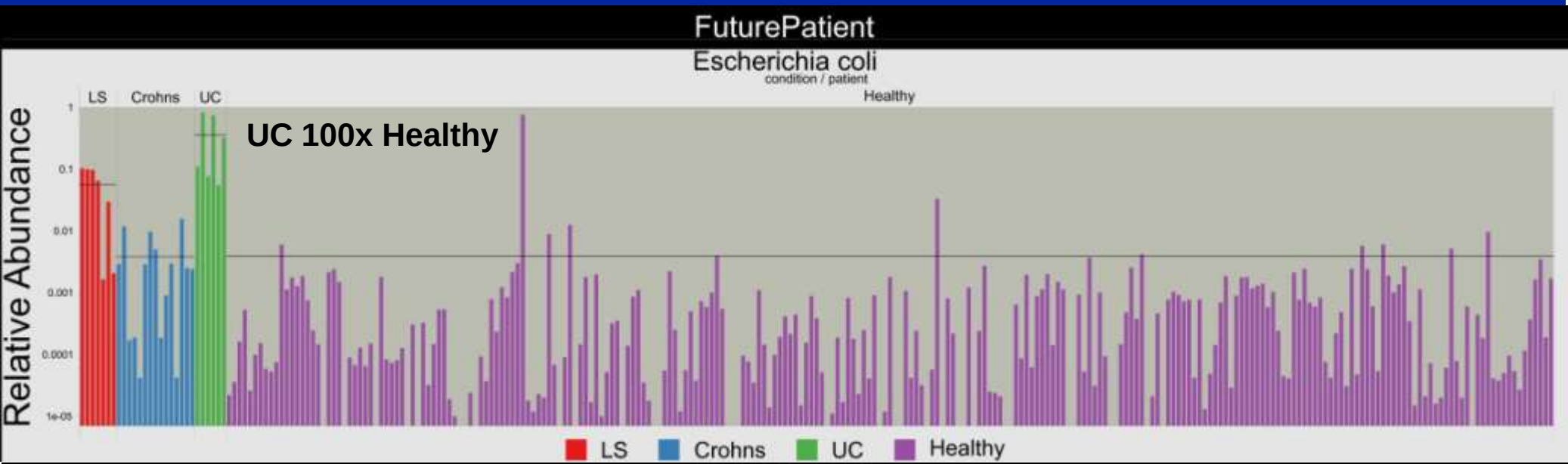
- **Dell's Sanger Cluster**
  - 32 Nodes, 512 Cores,
  - 48GB RAM per Node
  - 50GB SSD Local Drive, 390TB Lustre File System
- **We Processed the Taxonomic Relative Abundance**
  - Used ~35,000 Core-Hours on Dell's Sanger
  - With 30 TB data
- **Full Processing to Function (COGs, KEGGs)**
  - Would Require ~1-2 Million Core-Hours

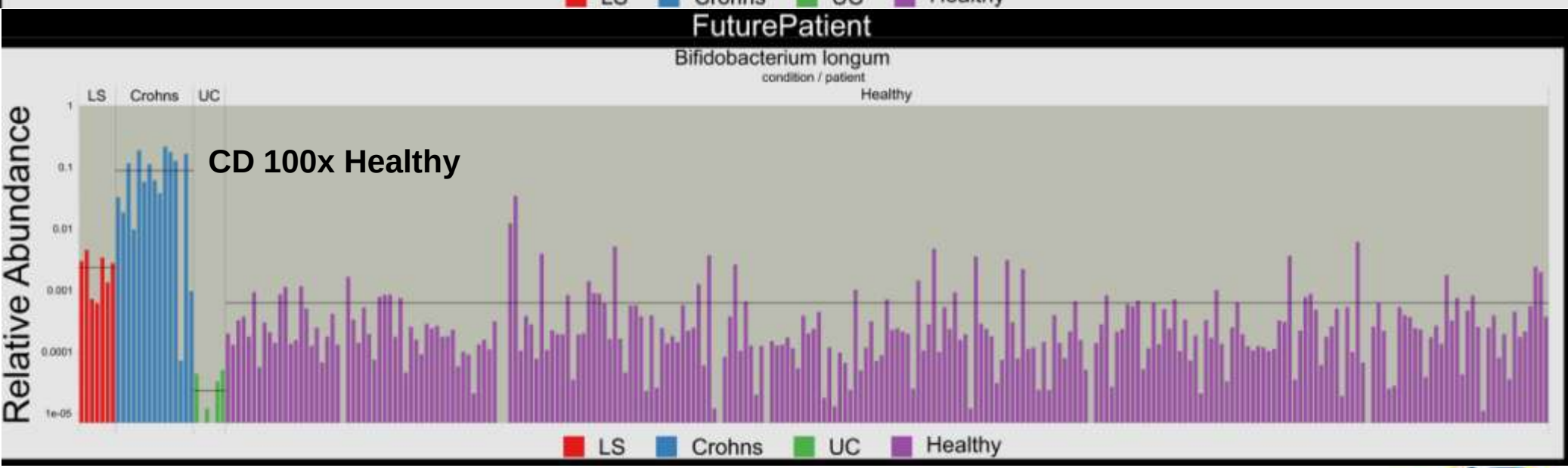


Dell Solutions Center
Industry Solutions Lab
SANGER DSU

Source: Weizhong Li, UCSD

Dell Cloud Results Are Leading Toward Microbiome Disease Diagnosis





We Produced Similar Results for ~2500 Microbial Species