"Large Memory High Performance Computing Enables Comparison Across Human Gut Microbiome of Patients with Autoimmune Diseases and Healthy Subjects"

XSEDE 2013 – Gateway to Discovery

San Diego, CA

July 24, 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

http://lsmarr.calit2.net

J. Craig Venter INSTITUTE

This Talk Based on XSEDE Selected Paper

Large Memory High Performance Computing Enables Comparison Across Human Gut Microbiome of Patients with Autoimmune Diseases and Healthy Subjects

Sitao Wu, Weizhong Li, Larry Smarr,
UC San Diego (CRBS, Calit2)

Karen Nelson, Shibu Yooseph, Manolito Torralba
J. Craig Venter Institute, Rockville, MD

J. Craig Venter INSTITUTE

SDSC

CRBS

it²

I Arrived in La Jolla in 2000 After 20 Years in the Midwest and Decided to Move Against the Obesity Trend

Age 41



Age 51



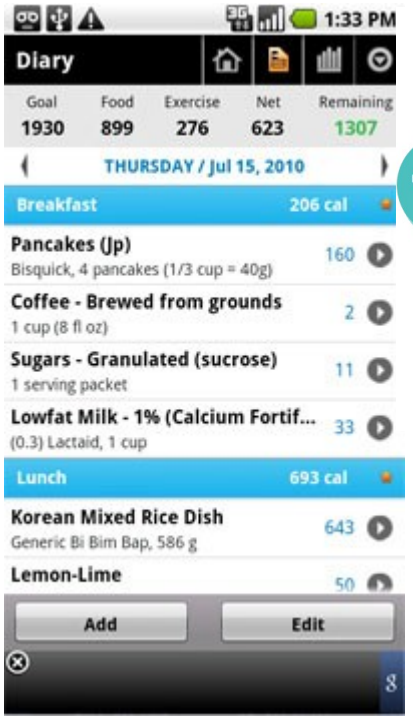
Age 61



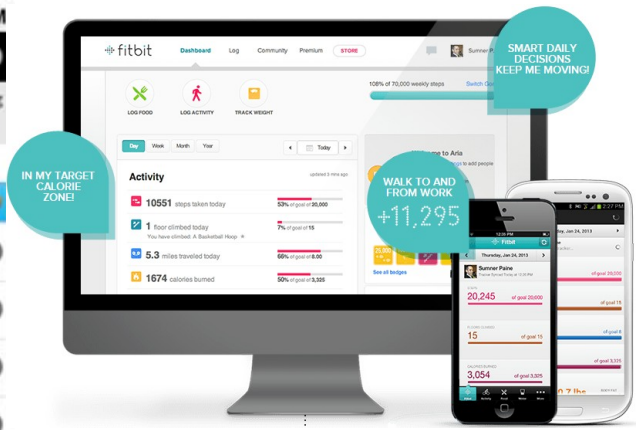
I Reversed My Body's Decline By Quantifying and Altering Nutrition and Exercise

http://lsmarr.calit2.net/repository/LS_reading_recommendations_FiRe_2011.pdf

Challenge-Develop Standards to Enable MashUps of Personal Sensor Data Across Private Clouds



MyFitnessPal-Calories Ingested



FitBit - Daily Steps & Calories Burned



Withing/iPhone-Blood Pressure





EM Wave PC-Stress

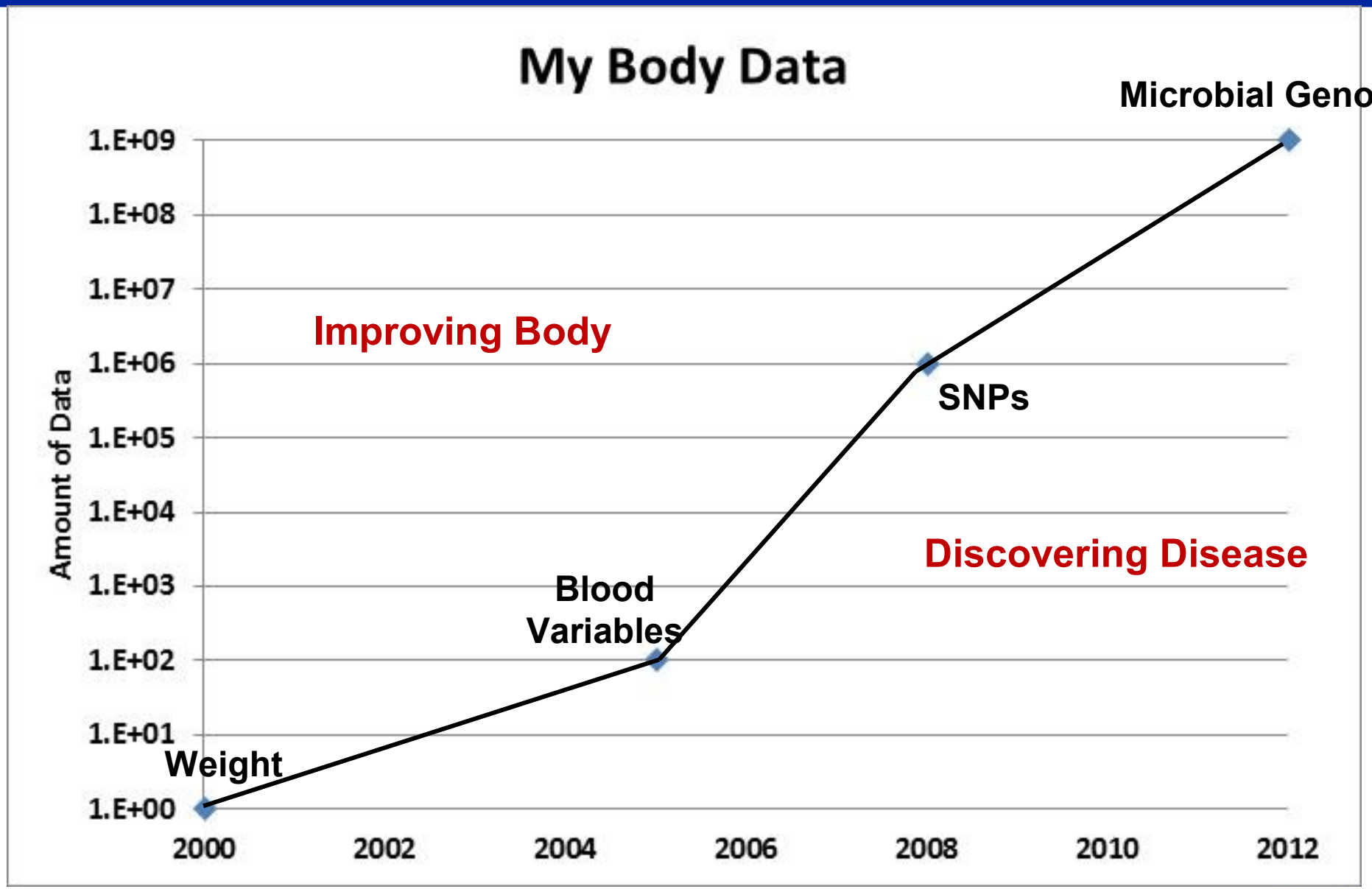


Azumio-Heart Rate



Zeo-Sleep

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



From Measuring Macro-Variables to Measuring Your Internal Variables



The Patient of the Future

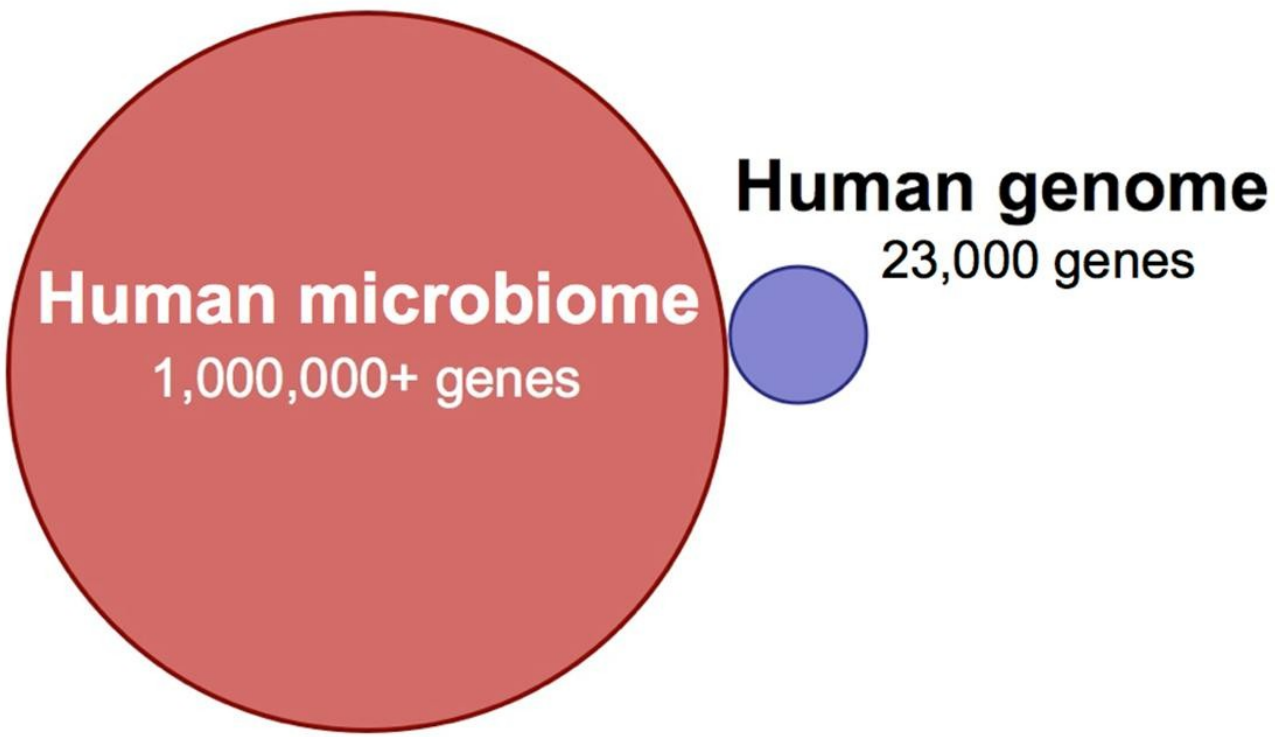
Internet pioneer Larry Smarr's quest to quantify everything about his health led him to a startling discovery, an unusual partnership with his doctor, and more control over his life.

www.technologyreview.com/biomedicine/39636

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

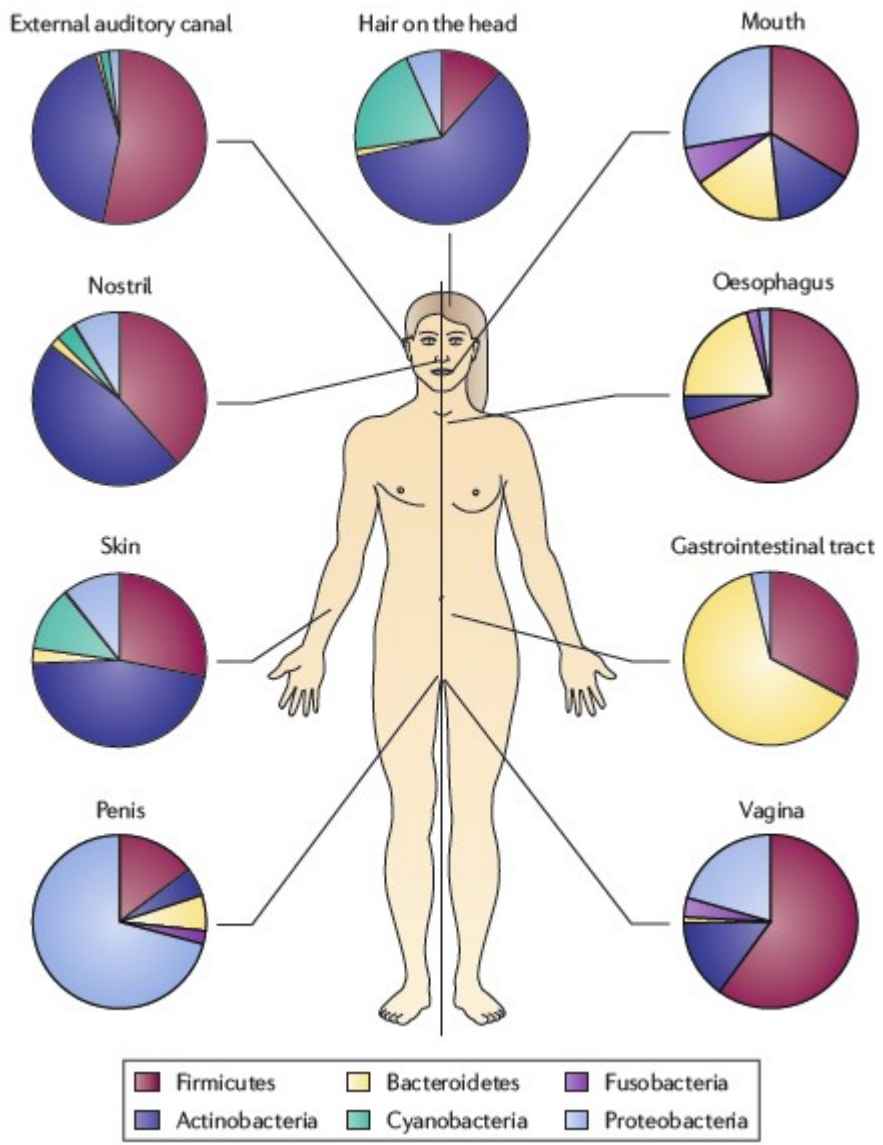
J. Craig Venter™ INSTITUTE

it²

Quantifiying the Human Superorganism: Distribution by Phyla of Microorganisms in Our Bodies

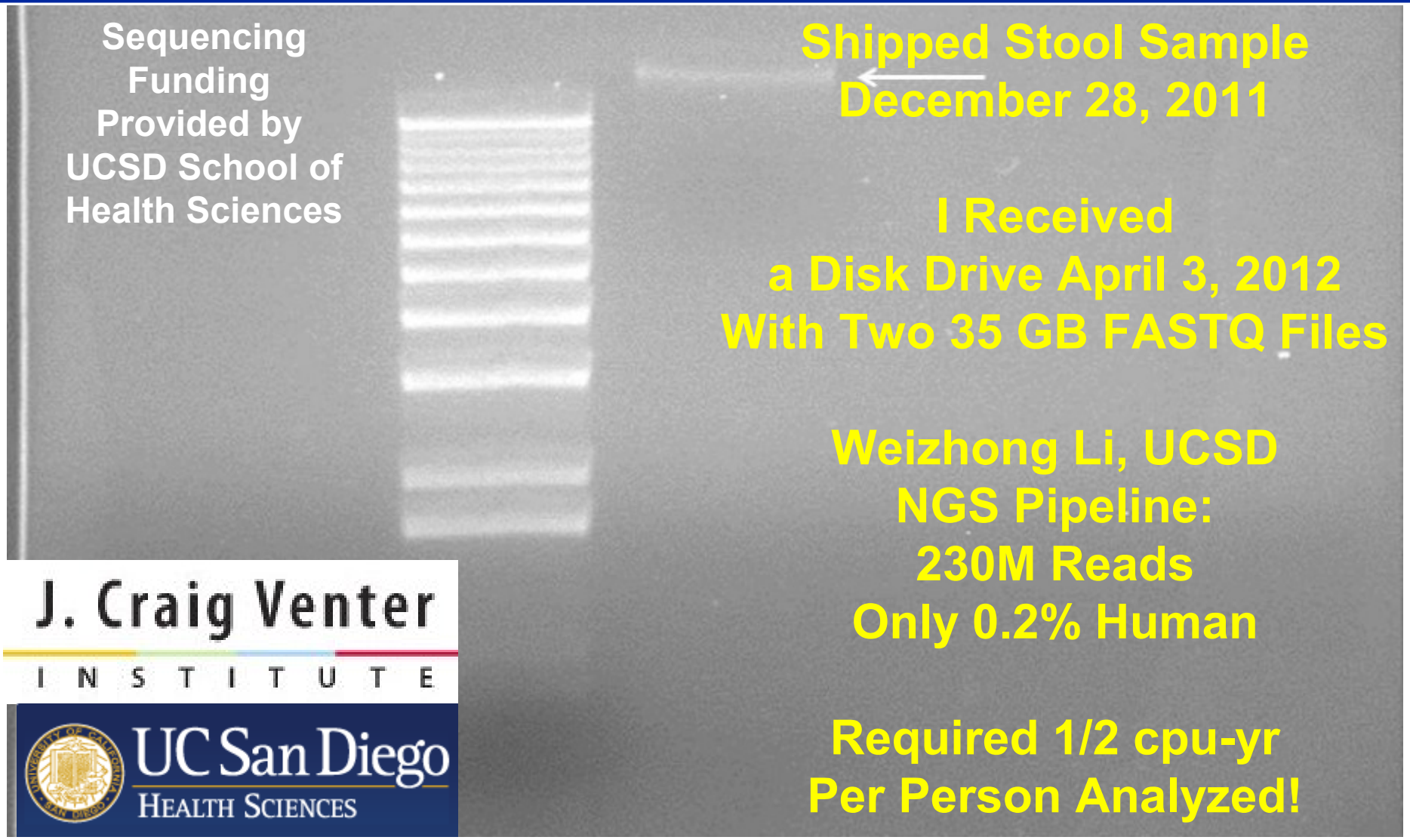
Human Microbiome Project

The NIH Common Fund
WE ACCELERATE DISCOVERY.



Nature Reviews Microbiology
v.9, p. 279 (2011)

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



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

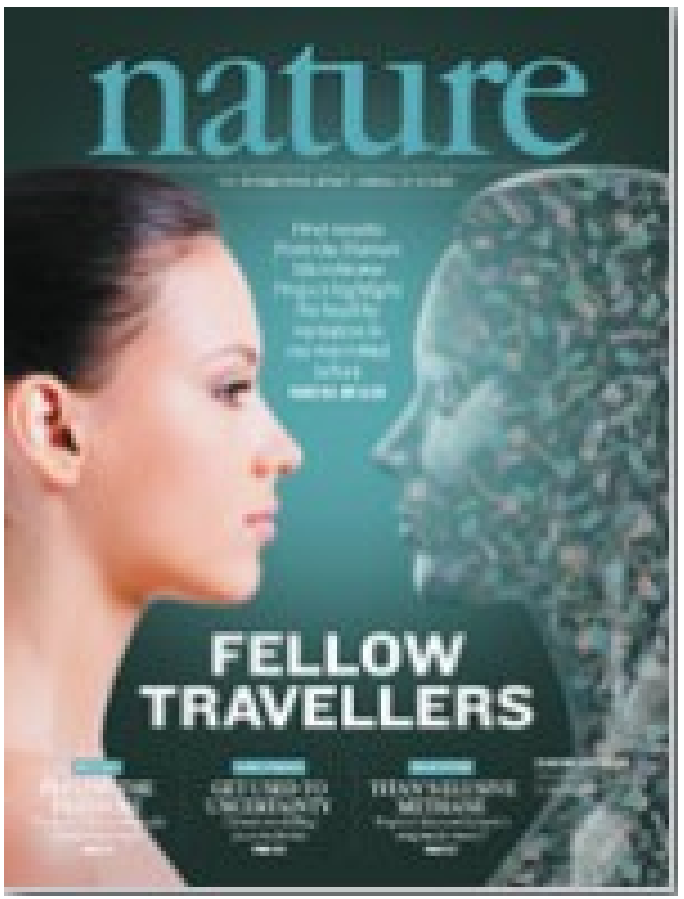
Manny Torralba, Project Lead - Human Genomic Medicine
J Craig Venter Institute

January 25, 2012

Intense Scientific Research is Underway on Understanding the Human Microbiome



June 8, 2012



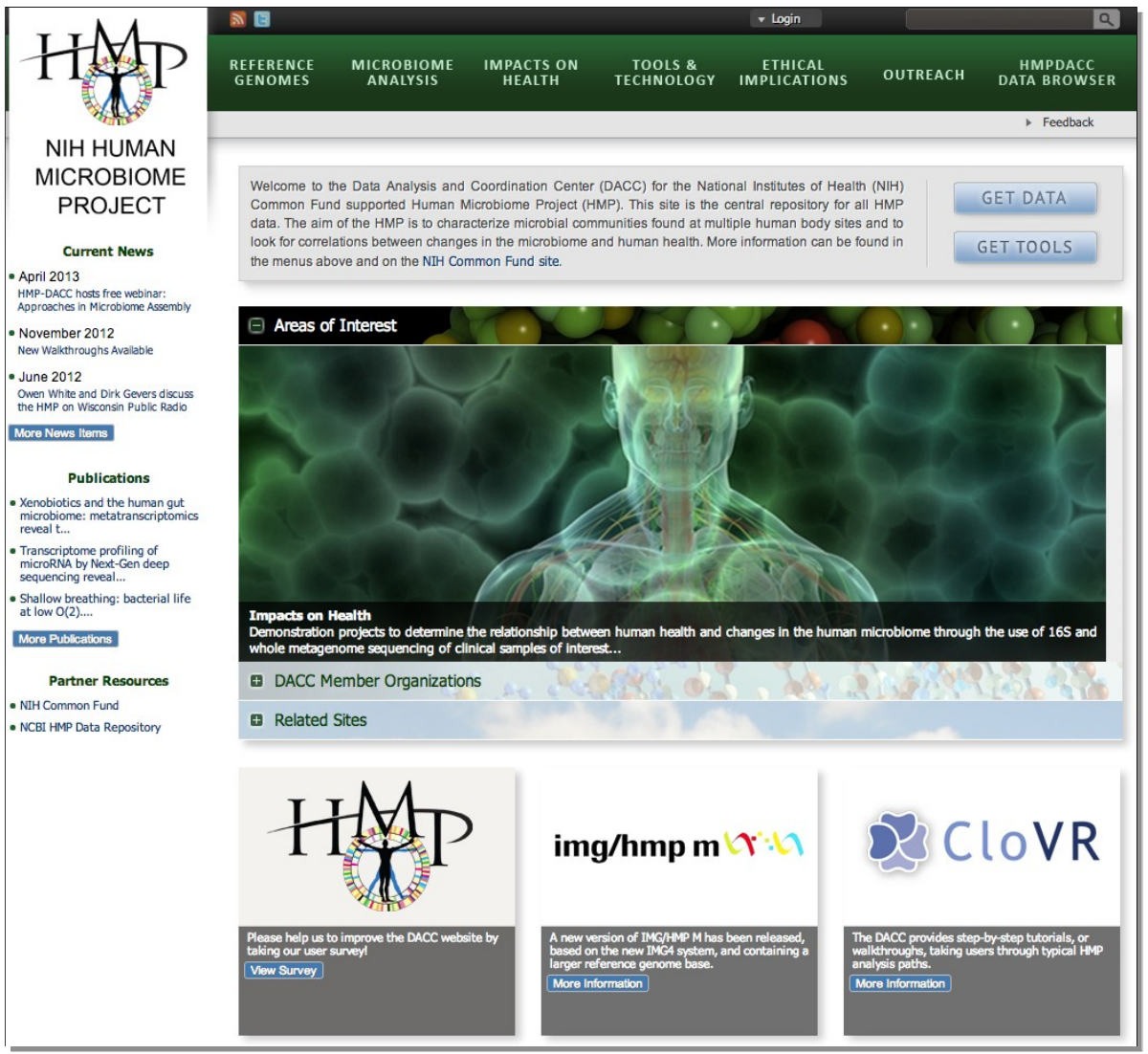
June 14, 2012

From Culturing Bacteria to Sequencing Them

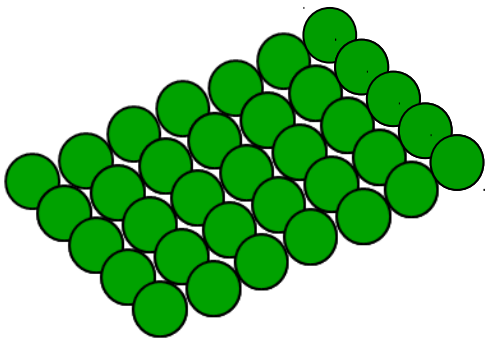
J. Craig Venter™ INSTITUTE

it²

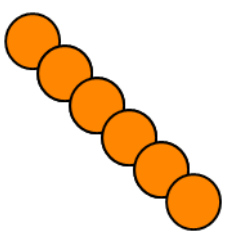
Additional Phenotypes Added from NIH HMP For Comparative Analysis



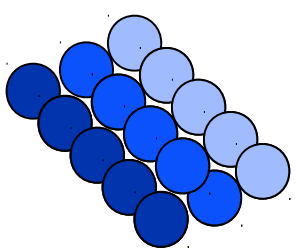
35 "Healthy" Individuals
1 Point in Time



6 Ulcerative Colitis, 1 Point in Time



5 Ileal Crohn's, 3 Points in Time



J. Craig Venter INSTITUTE

it²

Gut Microbiome Metagenomic Datasets

Sample Group	Samples	Average PE reads per sample	Total PE reads	Selected samples	Average filtered high quality PE reads per selected sample
LS	3	151,401,139	454,203,418	3	119,694,209
CD	15	91,477,052	1,372,155,785	15	57,624,823
UC	11	14,330,989	157,640,888	6	2,905,681
HE	154	68,609,164	10,565,811,331	34	39,692,865
Total	183	68,578,204	12,549,811,422	58	44,662,870

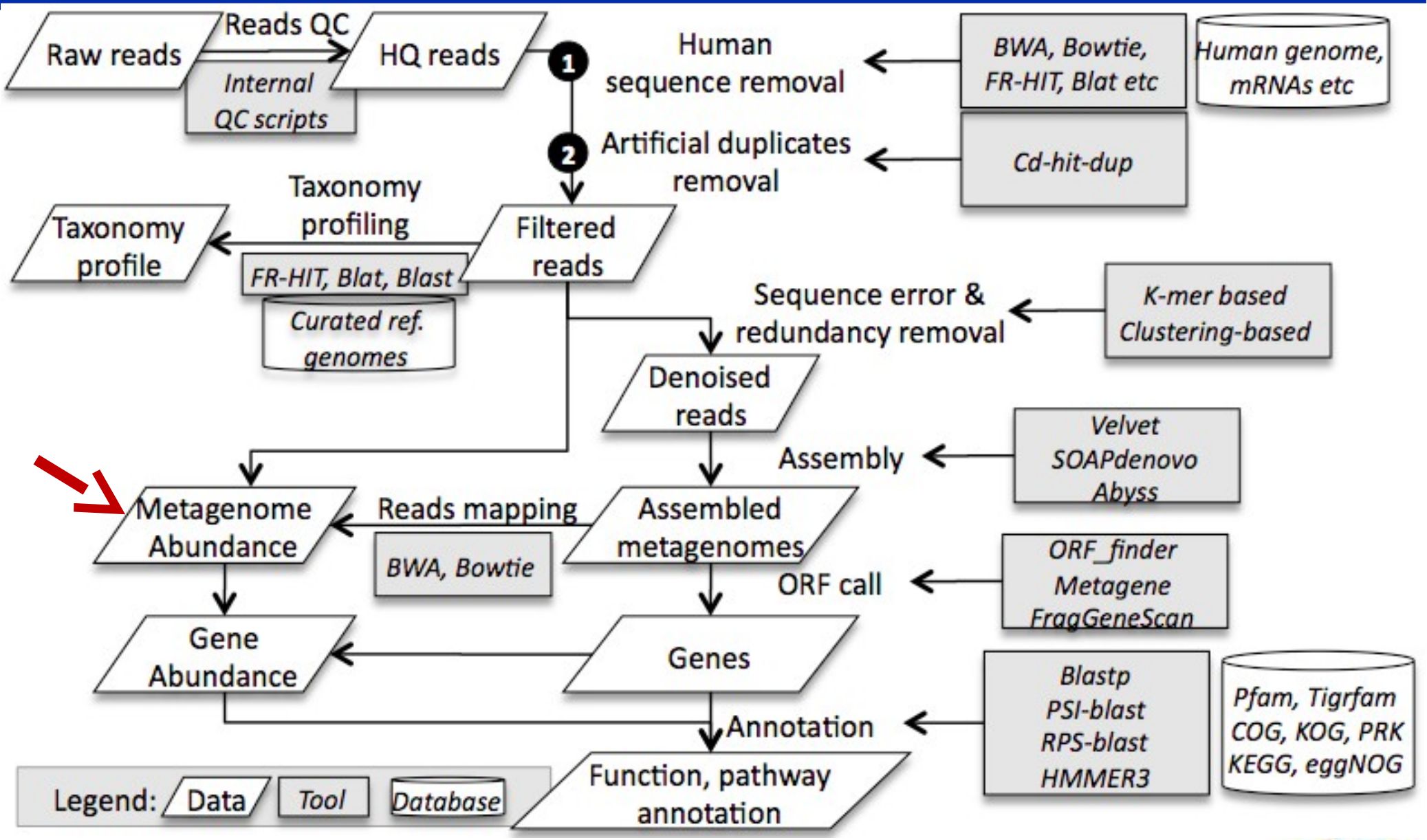
One "Read" = 100 DNA Bases
Total of 1.2 Trillion Bases!

J. Craig Venter INSTITUTE

Source: Weizhong Li, 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

Computing and Parallelization Requirements of the Computational Tools in Our Workflow

Analysis	Tool	Computing requirement			Parallelization	
		Data	CPU	RAM	Multi-threading	Map Reduce
Quality control	QC script	√				
Human DNA removal	Bowtie	√			√	
Duplicates removal	CD-HIT-DUP	√		√	√	
Mapping	FR-HIT	√	√		√	
Read clustering	CD-HIT-EST	√		√	√	
Assembly	Velvet, SOAPdenovo	√		√	√	
ORF call	Metagene, Fraggenescan	√				
ORF clustering	CD-HIT	√		√	√	
COG, KOG annotation	RPS-BLAST	√	√		√	√
KEGG annotation	BLAST	√	√		√	√
Pfam, Tigrfam annotation	HMMER3	√	√		√	√

Source: Weizhong Li, CRBS, UCSD

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

- ~180,000 Core-Hrs on Gordon
  - KEGG function annotation: 90,000 hrs
  - Mapping: 36,000 hrs
    - Used 16 Cores/Node and up to 50 nodes
  - Duplicates removal: 18,000 hrs
  - Assembly: 18,000 hrs
  - Other: 18,000 hrs
- Gordon RAM Required
  - 64GB RAM for Reference DB
  - 192GB RAM for Assembly
- Gordon Disk Required
  - Ultra-Fast Disk Holds Ref DB for All Nodes
  - 8TB for All Subjects



SDSC

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

NSF NIH HMP

J. Craig Venter INSTITUTE

CRBS it²

We Created a Reference Database Of Known Gut Genomes

- **NCBI 2012**
  - 2036 Complete + 1826 Draft Bacteria & Archaea Genomes
  - 1397 Complete Virus Genomes
  - 39 Complete Fungi Genomes
  - 308 HMP Eukaryote Reference Genomes
- **Total 5607 genomes, ~15 GB of sequences**

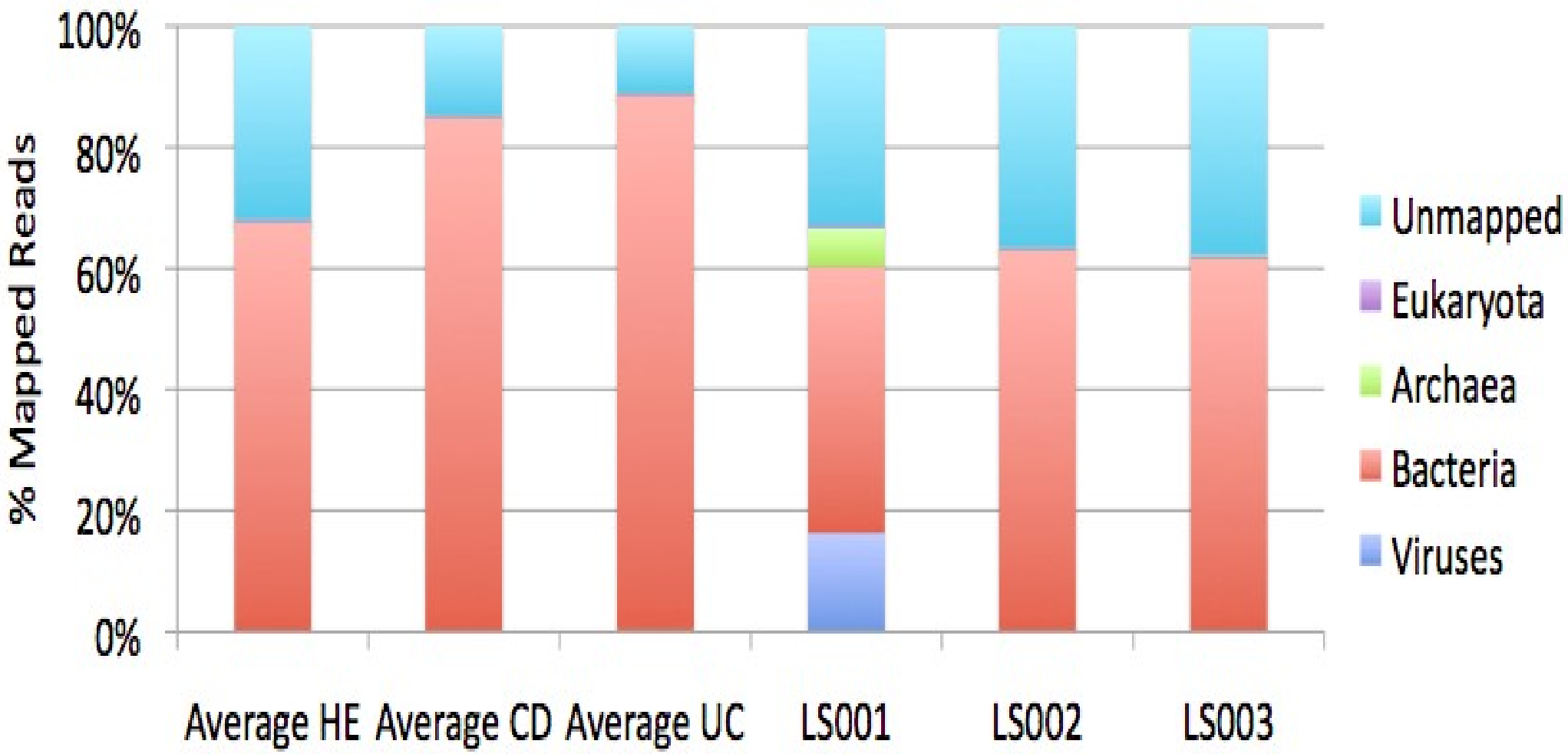
Now to Align Our 12.5 Billion Reads Against the Reference Database

J. Craig Venter INSTITUTE

Source: Weizhong Li, CRBS, UCSD

it²

We Still Don't Know a Significant Fraction of the Gut Genomes



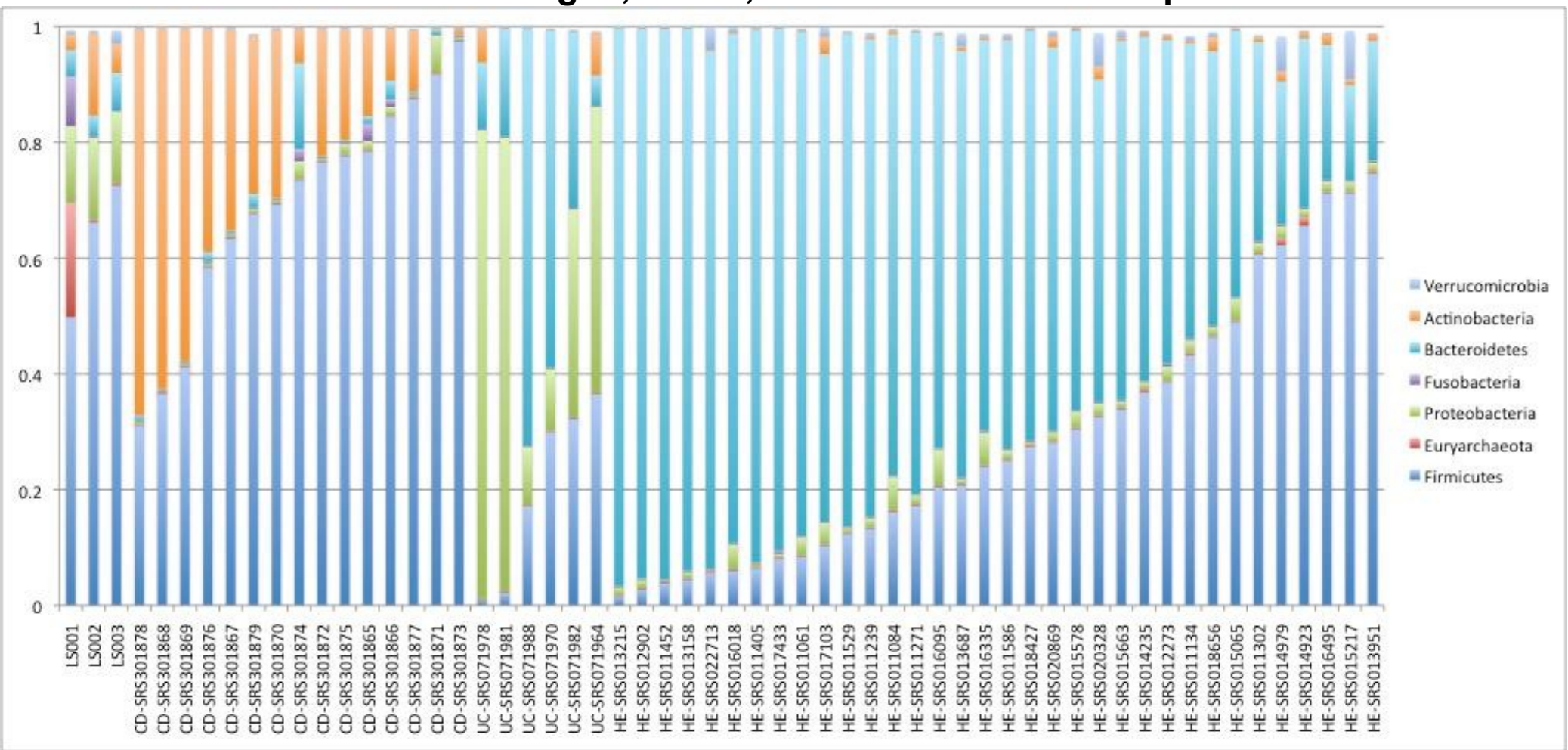
J. Craig Venter INSTITUTE

Source: Weizhong Li, CRBS, UCSD

it²

Phyla Gut Microbial Abundance Without Viruses: LS, Crohn's, UC, and Healthy Subjects

Source: Weizhong Li, UCSD; Calit2 FuturePatient Expedition



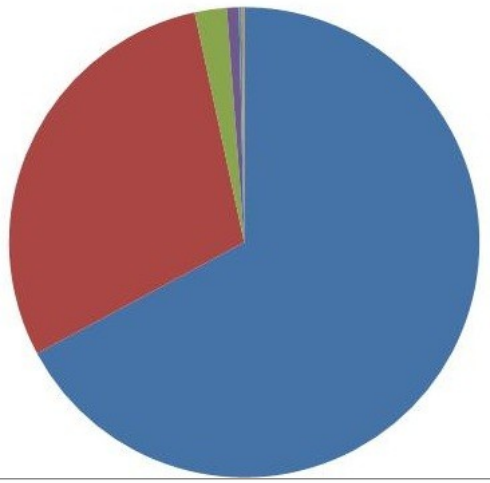
LS **Crohn's** **Ulcerative Colitis** **Healthy**

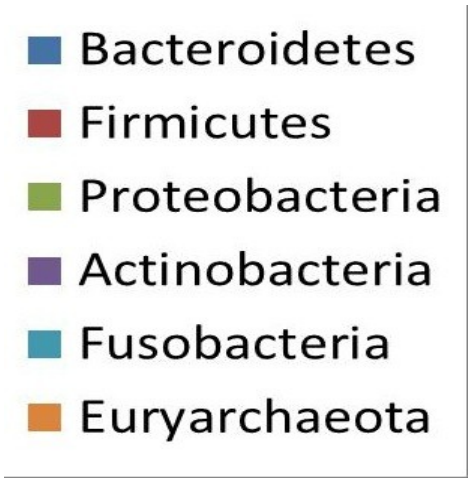
Toward Noninvasive Microbial Ecology Diagnostics

We Find Major Shifts in Microbial Ecology Between Healthy and Two Forms of IBD

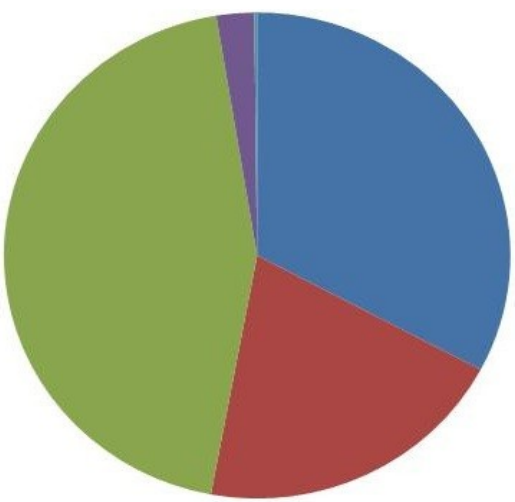
Microbiome "Dysbiosis" or "Mass Extinction"?

Ave. Healthy

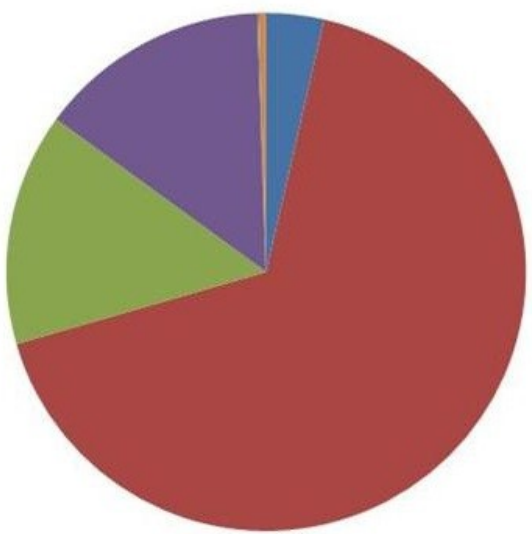




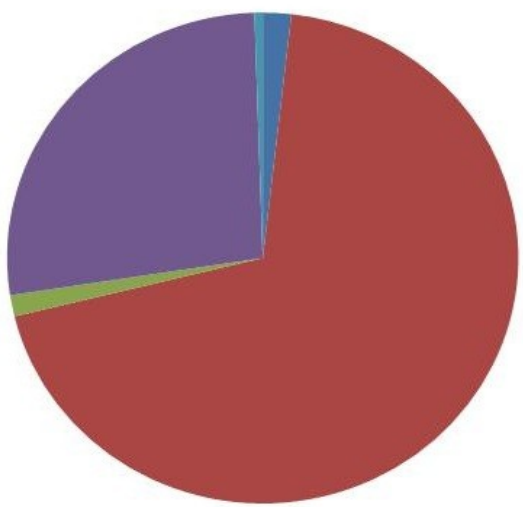
Ave. Ulcerative Colitis



LS Apr 2012



Ave. Crohn's



Explosion of Proteobacteria

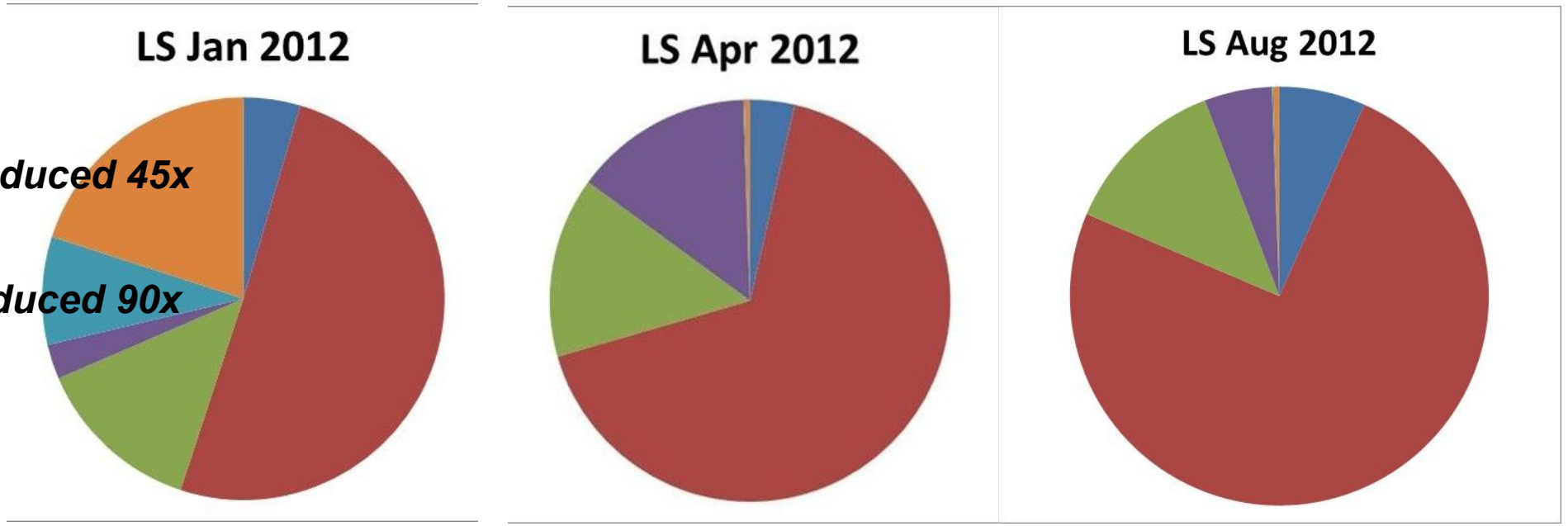
On the IBD Spectrum

Collapse of Bacteroidetes

J. Craig Venter INSTITUTE

it²

Major Changes in LS Microbiome Before and After 1 Month Antibiotic & 2 Month Prednisone Therapy



Therapy Greatly Reduced Two Phyla, But Massive Reduction in Bacteroidetes And Large % Proteobacteria Remain

Small Changes With No Therapy

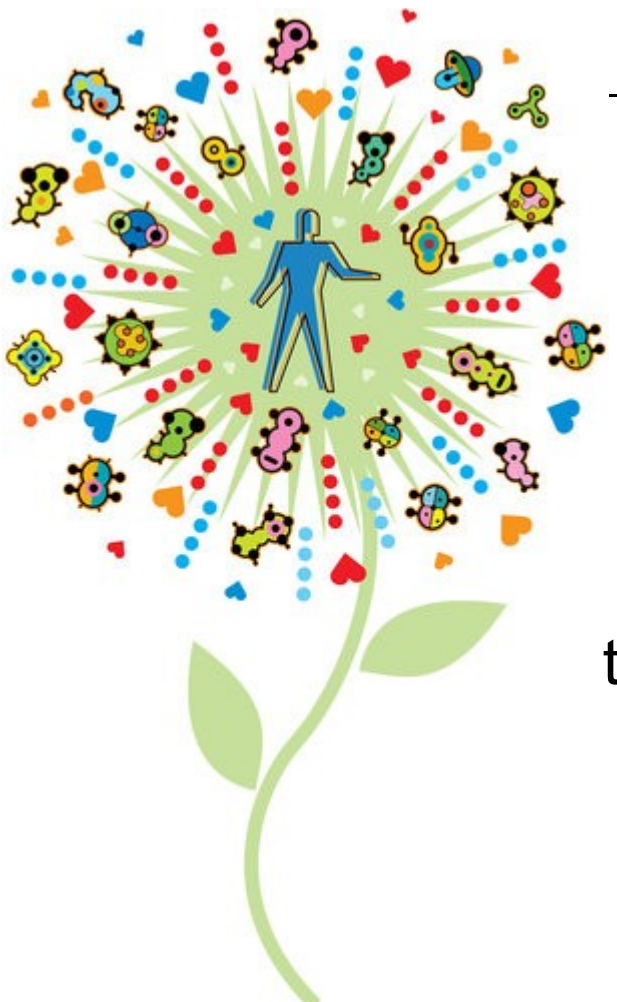
- Bacteroidetes
- Firmicutes
- Proteobacteria
- Actinobacteria
- Fusobacteria
- Euryarchaeota

How Does One Get Back to a "Healthy" Gut Microbiome?

J. Craig Venter™ INSTITUTE

it²

From War to Gardening: New Therapeutical Tools for Managing the 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."