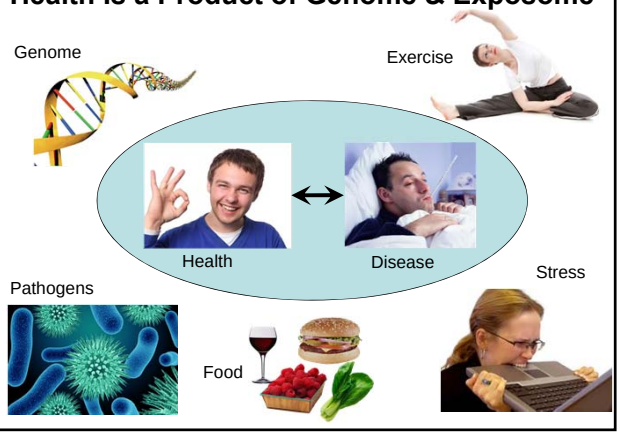
Using Deep Omics Profiling and Big Data for Monitoring Individual Physiology, Health and Disease

Michael Snyder
Stanford University

August 1, 2016

Conflicts: Personalis, Genapsys, Abcam, SensOmics

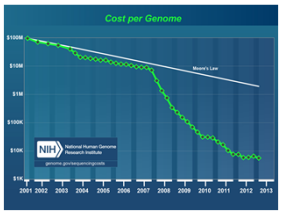
Health Is a Product of Genome & Exposome



Drivers of Big Data

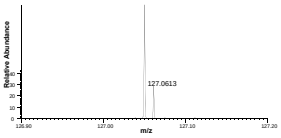
DNA Sequencing

Human Genome Cost <$2K



http://www.genome.gov/

Mass Spectrometry



Genetic Variation Among People: Three Types

1) Single nucleotide variants (SNVs)

GATTTAGATCGCGATAGAG
GATTTAGATCTCGATAGAG

4 Million/person

2) Short Indels (Insertions/Deletions 1-100 bp)

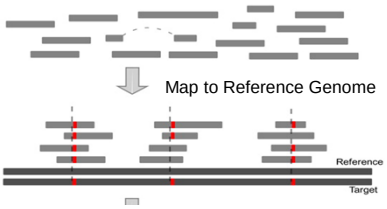
GATTTAGATCGCGATAGAG
GATTTAGA------TAGAG

500-800K/person

A Personal Genome Sequence is Determined by Comparing to a Reference Genome Sequence

Step 1
Generate Reads
30X: 150 b
Paired Ends

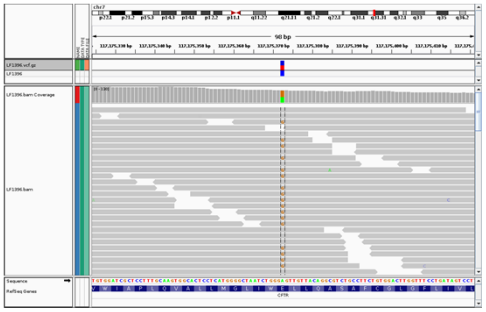
Step 2
Call SNPS & Indels
using correctly mapped reads



Reveals ~4M SNVs

Snyder et al. Genes Dev 2010;24:423-431

Plot showing aligned sequence reads



IGV plot: http://www.broadinstitute.org/software/igv/home

This individual is heterozygous for a variant in the *CFTR* gene http://jchoigt.wordpress.com/

Sequencing Genomes of Healthy People: Incorporate into Medicine

Genomic

```
GGTTCCAAAAGTTTATTGGATGCCGTTTCA
GTACATTTATCGTTTGCTTTGGATGCCCTA
ATTAAAAGTGACCCTTTCAAACTGAAATTC
ATGATACACCAATGGATATCCTTAGTCGAT
AAAATTTGCGAGTACTTTCAAAGCCAAATG
AAATTATCTATGGTAGACAAAACATTGACC
AATTTCATATCGATCCTCCTGAATTTATTG
GCGTTAGACACAGTTGGTATATTTCAAGTG
ACAAGGACAATTACTTGGACCGTAATAGAT
TTTTTGAGGCTCAGCAAAAAAGAAAATGGA
AATTAATTTTGAAGTGCCATTGA….
```

Family History

Medical Tests:
Few Tests (<20)

1. **Predict risk**
2. **Diagnose**
3. **Monitor**
4. **Treat**

Personal "Omics" Profiling (POP)

Genome
Epigenome
Transcriptome
Proteome
Cytokines
Metabolome, Lipidome
Autoantibody-ome
Microbiome (Gut, Urine, Nasal, Tongue, Skin)

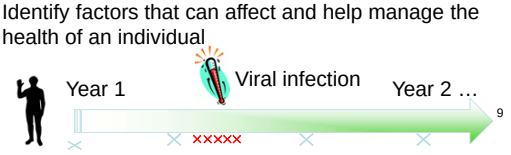
Personal Omics Profile

Initially 40K Molecules/ Measure-ments

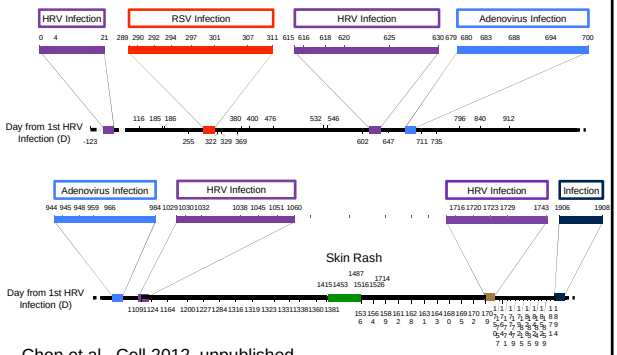
Now Billions!

General Goals

1) Understand how individuals change over time and during periods of health and disease at high resolution
2) Understand how different "omes" (microbiome, metabolome, proteome, genome) relate to one another dynamically
3) Understand how individual responses are similar and differ from one another when faced with specific perturbations
4) Identify factors that can affect and help manage the health of an individual

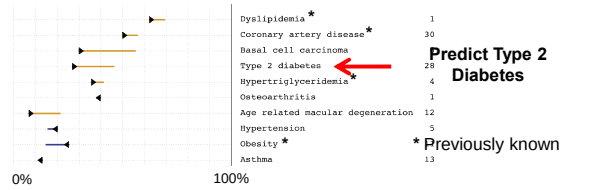


Personal Omics Profile
~76 months; >190 Timepoints; 9 Viral Infections

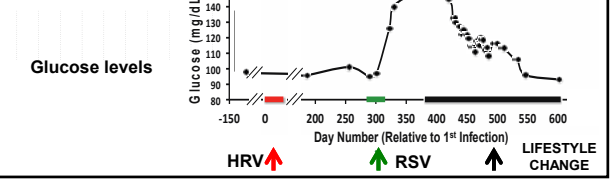


Chen et al., Cell 2012, unpublished

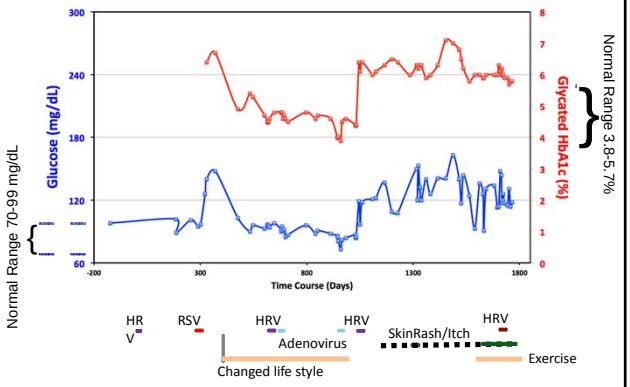
Genome Sequence (Illumina, Complete Genomics)



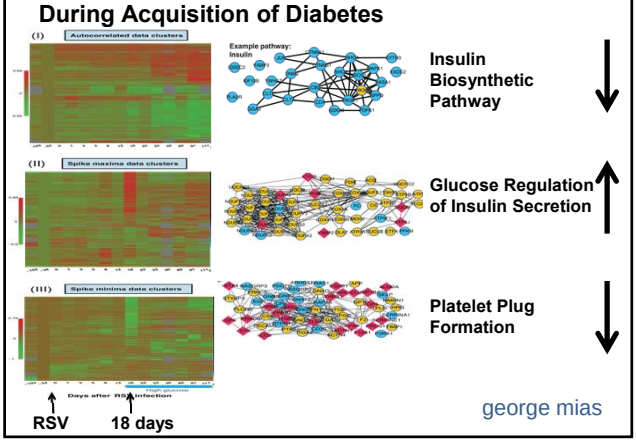
Rong Chen
Atul Butte



Extended Time Line



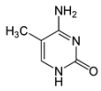
Molecules and Biochemical Pathways that Change During Acquisition of Diabetes



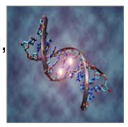
george mias

Epigenetics: DNA Methylation

- Modified bases on our DNA: Usually at CpGs
- Affects gene expression (modification in gene control regions turn genes off).
- Affected by nutrition, lifestyle factors, aging, and environment



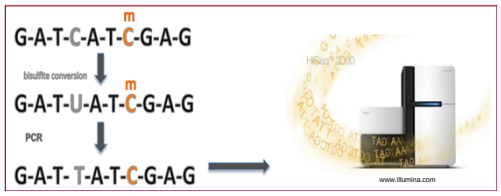
5 methylC



Map all 28 M methylated sites using whole genome bisulfite sequencing

How to detect DNA methylation: Bisulfite sequencing

- Sodium bisulfite converts **un**methylated Cytosine into Thymine Does not effect on methylated Cytosine.

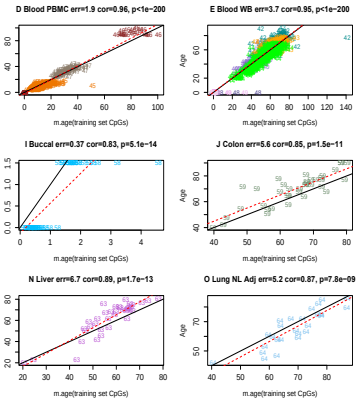


- Map all 28 M methylated bases in genome

DNA Methylation Correlates With Age

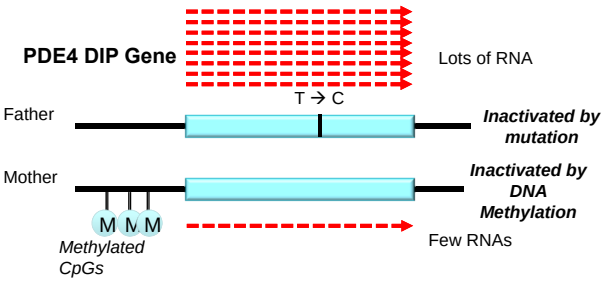
Measure Methylation at 353 CpG Sites "Clock CpGs"

Correlates Well with Age Across Many Tissues



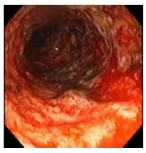
Horvath *Genome Biology* 2013 **14**:R115 doi:10.1186/gb-2013-14-10-r115

Gene Inactivation by Mutation and Methylation: PDE4 involved in eosinophilia

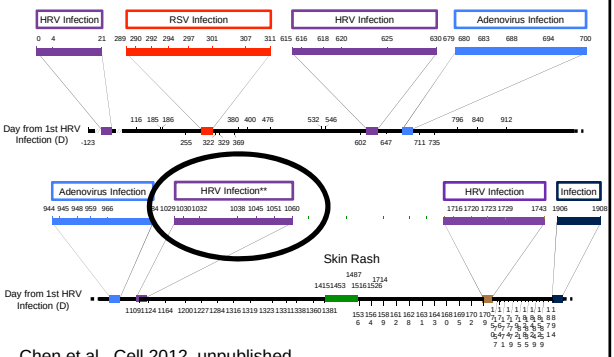


Your Microbiome is Important for Your Health

- We have 10 Trillion human cells, but 100 Trillion bacteria
- Helps digest food you eat
- Makes essential vitamins, eg, B12
- Implicated in Inflammatory Bowel Disease (Crohn's and ulcerative colitis), Diabetes, Obesity, MI

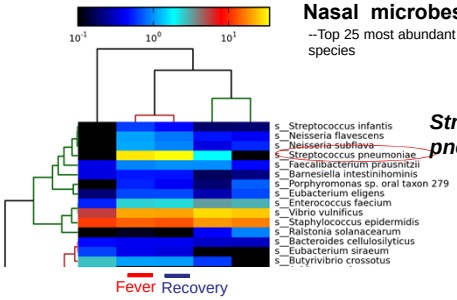


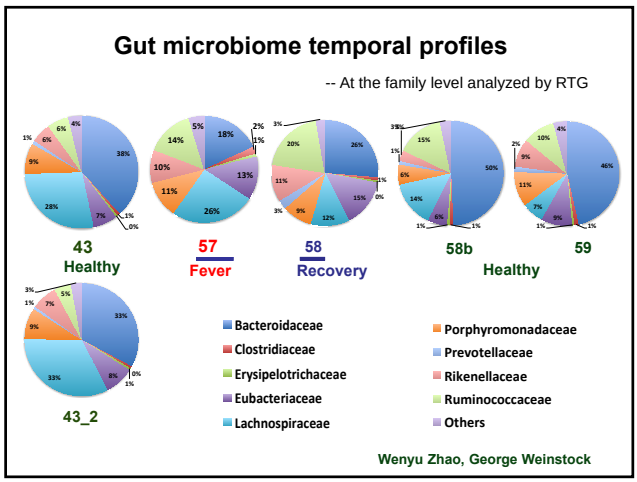
Personal Omics Profile
76 months; >190 Timepoints; 9 Viral Infections

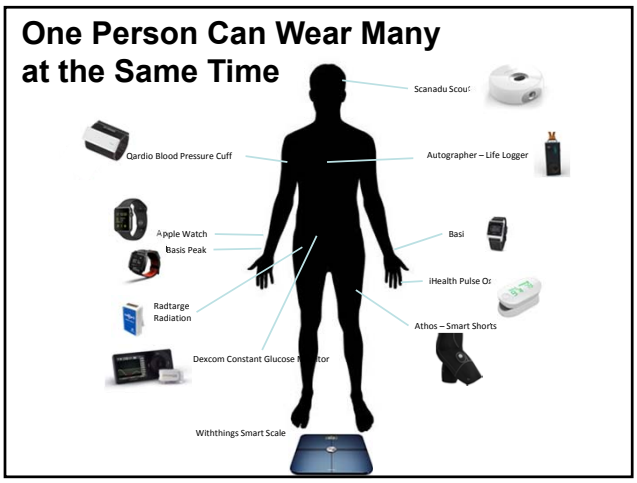


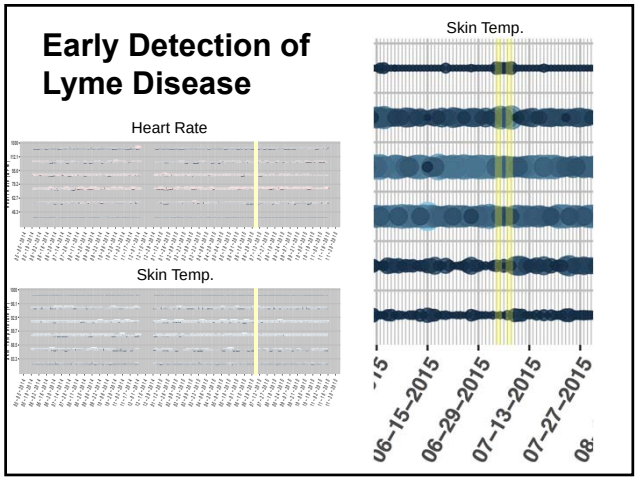
Chen et al., Cell 2012, unpublished

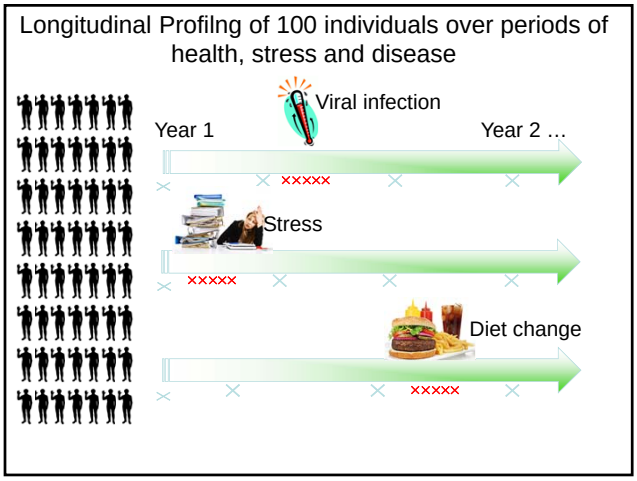
Identifying the Microbe Causing Illness



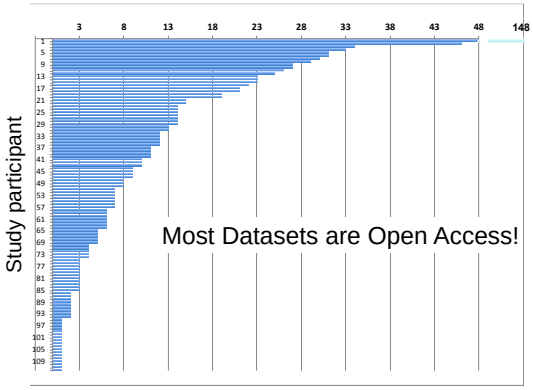

Gut microbiome temporal profiles
-- At the family level analyzed by RTG
43
Healthy
57
Fever
58
Recovery
58b
Healthy
59
43_2
Bacteroidaceae
Clostridiaceae
Erysipelotrichaceae
Eubacteriaceae
Lachnospiraceae
Porphyromonadaceae
Prevotellaceae
Rikenellaceae
Ruminococcaceae
Others
Wenyu Zhao, George Weinstock


One Person Can Wear Many at the Same Time
Scanadu Scout
Qardio Blood Pressure Cuff
Autographer – Life Logger
Apple Watch
Basis Peak
Basi
iHealth Pulse Ox
Radtarge Radiation
Athos – Smart Shorts
Dexcom Constant Glucose Monitor
Withthings Smart Scale


Early Detection of Lyme Disease
Heart Rate
Skin Temp.
Skin Temp.
06-15-2015
06-29-2015
07-13-2015
07-27-2015


Longitudinal Profilng of 100 individuals over periods of health, stress and disease
Viral infection
Year 1
Year 2 …
Stress
Diet change

>1300 collections thus far

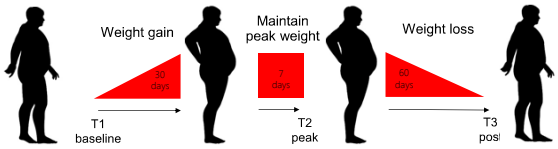


Genome Sequencing – So far

- Three have important mutations to know about
  - SHBD: high freq. of paraganglioma
  - PROC: Affects coagulation
  - RBM20: cardiomyopathy
- Two MODY variants
- All have carrier mutations and pharmacogenetic variants

Personalis, Inc

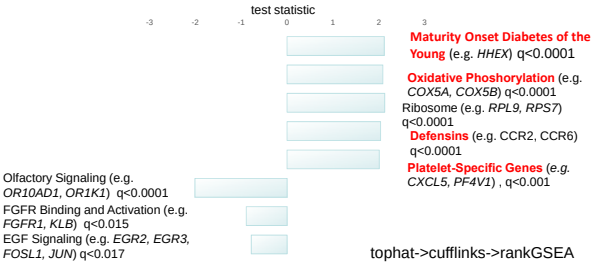
A subset of individuals undergo a dietary perturbation.



24 participants:

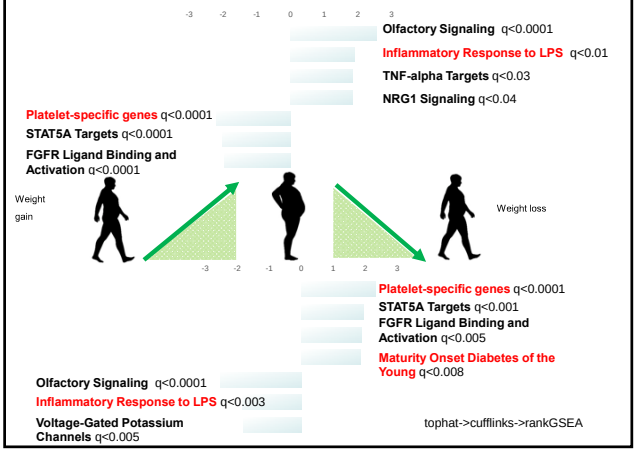
- **13 Insulin resistant**
- **11 healthy controls (BMI matched)**

Baseline gene expression differences between IR and IS in blood PBMCs



tophat->cufflinks->rankGSEA

Overfeeding-induced transcriptional changes in prediabetics

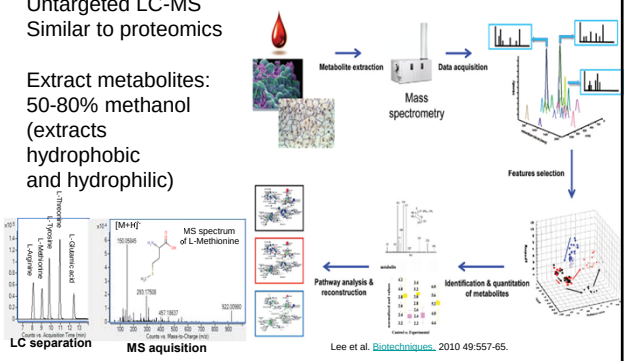


tophat->cufflinks->rankGSEA

Metabolite profiling by LC-Mass spectrometry

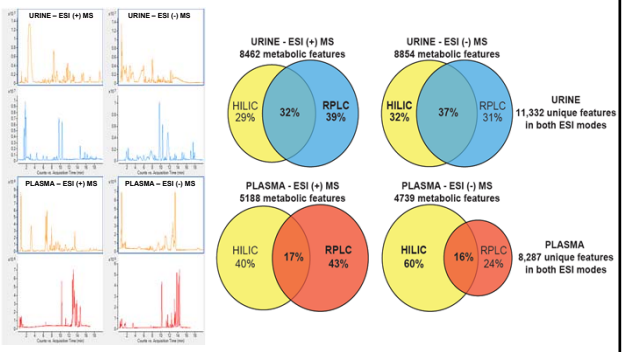
Untargeted LC-MS
Similar to proteomics

Extract metabolites:
50-80% methanol
(extracts hydrophobic and hydrophilic)



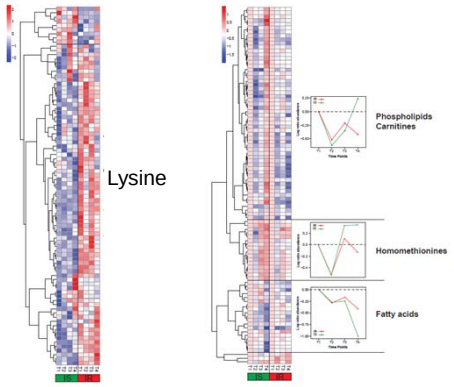
Lee et al. Biotechniques. 2010 49:557-65.

Using sophisticated LC, many thousands of MS peaks are detected in blood and urine

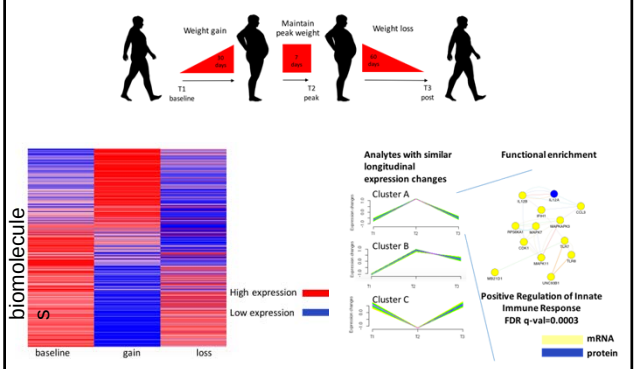


Changes in metabolites in response to overfeeding

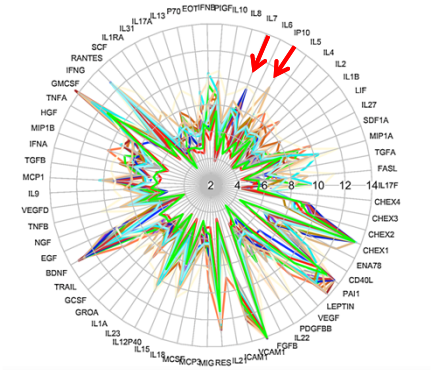
IR vs IS at T1 | T1 vs T2



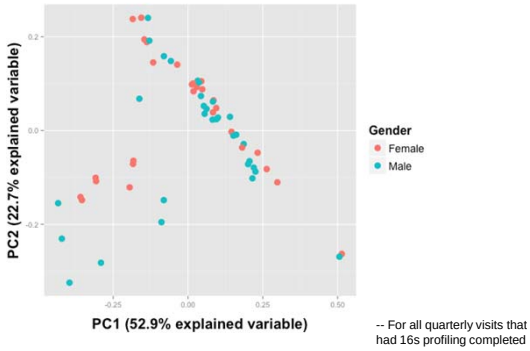
Example data: Short-term weight gain



Blood cytokine profiles: 20 subjects at baseline

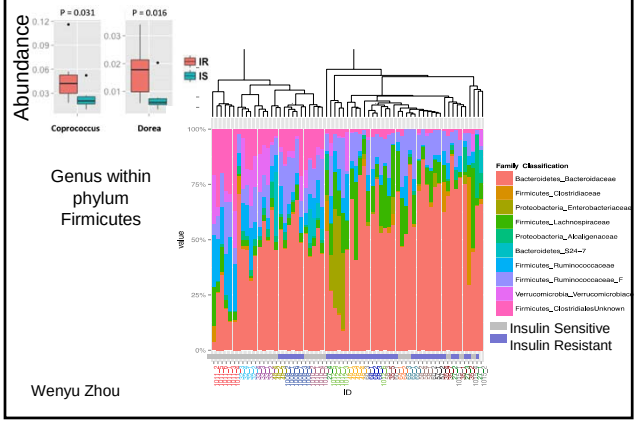


Participants' fecal 16s microbial profiles stratified by Gender



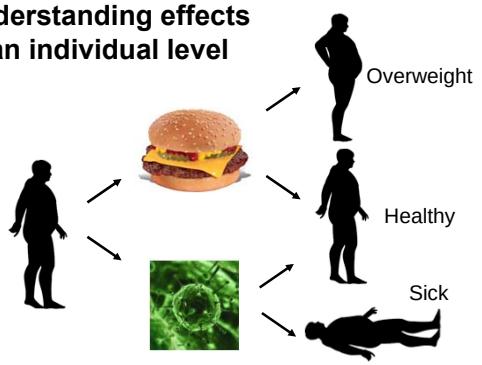
-- For all quarterly visits that had 16s profiling completed

Microbiome composition during the weight perturbation



Wenyu Zhou

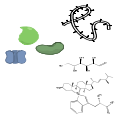
Understanding effects at an individual level



The Future?

Genomic Sequencing

Omes & Sensors: Personal Devic

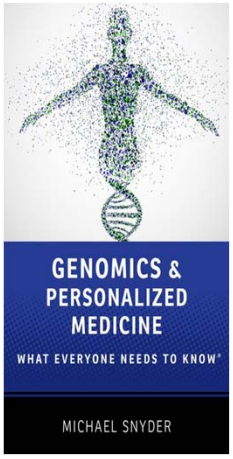


1. Predict risk
2. Early Diagnose
3. Monitor
4. Treat

Amanda Mills

Overall Summary

1) Genome sequencing is here. It can be used to manage disease and health
2) Multi-omics analyses are valuable for determining pathways and biochemical activities involved in human disease.
3) Longitudinal profiles are very valuable for understanding personal disease states
4) Everyone has a different molecular profile
5) Individuals will be responsible for their own health



Genomics and Personalized Medicine

What Everyone Needs to Know®

Michael Snyder

Available January 26, 2016 from Amazon

Stanford | Genetics and Genomics Certificate

New professional certificate - 100% online

Taught by faculty and industry experts

Certificate is comprised of 2 core courses in genetics and genomics and 4 elective courses

For more Information:

http://geneticscertificate.stanford.edu