# Metagenomics

02-715 Advanced Topics in Computational Genomics

# Metagenomics

- Investigation of the microbes that inhabit oceans, soils, and the human body, etc. with sequencing technologies

- Cooperative interactions between microbes and their hosts
  - microbial participation in host functions such as defence, metabolism and reproduction

# Metagenomics

- Human microbiom
  - Humans have more bacterial cells ($10^{14}$) in habiting our body than our own cells ($10^{13}$)
  - Consists of archaea, bacteria, and viruses
  - What are the composition and gene content of human microbiome?
  - What are the differences of microbiome composition across individuals?
  - What are the differences of microbiome composition across body parts?

# Tools for Studying Human Microbiomes

Wooley et al., PLoS Comp Bio, 2010

# Challenges

- Single species genomics
  - The full genome is sequenced
  - We know which species this genome comes from
  - We work towards the full annotation of the full genome of the given organism
- Metagenomics
  - Fragments of genomes from different species
  - Reference genome may not be available
  - Often impossible to determine the species of origin
  - Often impossible to reconstruct the genome of individual species
  - Often low coverage for some species
  - Danger of assembling a genome with mixed species

# Sampling in Metagenomics

- Number of samples and number of species captured
  - Rarefaction curve: the number of species as a function of the number of individuals sampled
  - The number of operational taxonomic units are chatarcterized by 16S (prokaryotic) or 18S (eukaryotic) rDNA
  - A rough estimate of the species diversity in the sample
- Filtering
  - Experimental (before sequencing) or computational (after sequencing) filtering of species that are relevant to the given study
- Meta data
  - Physical, chemical, and environmental characteristics of the microbe's location
  - Sampling date, time, depth, light intensity, pH, pathology, medical history etc.

# Tools for Studying Human Microbiomes

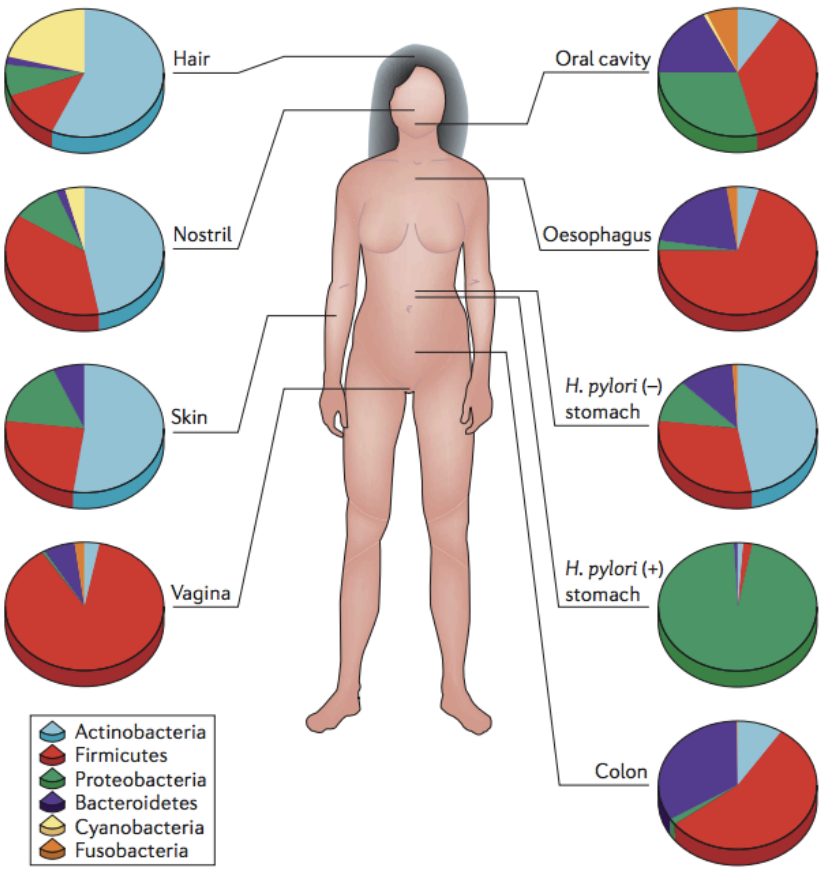
- Assigning unassembled sequences generated by shotgun high-throughput sequencing to the known gene sequences.
  - the assessment of interactions that occur
    - within the microbiome
    - between a microbiome and its host

- However, a substantial fraction of the metagenome (~33%) is not well-represented by reference genomes

# Gene Finding in Metagenomic Data

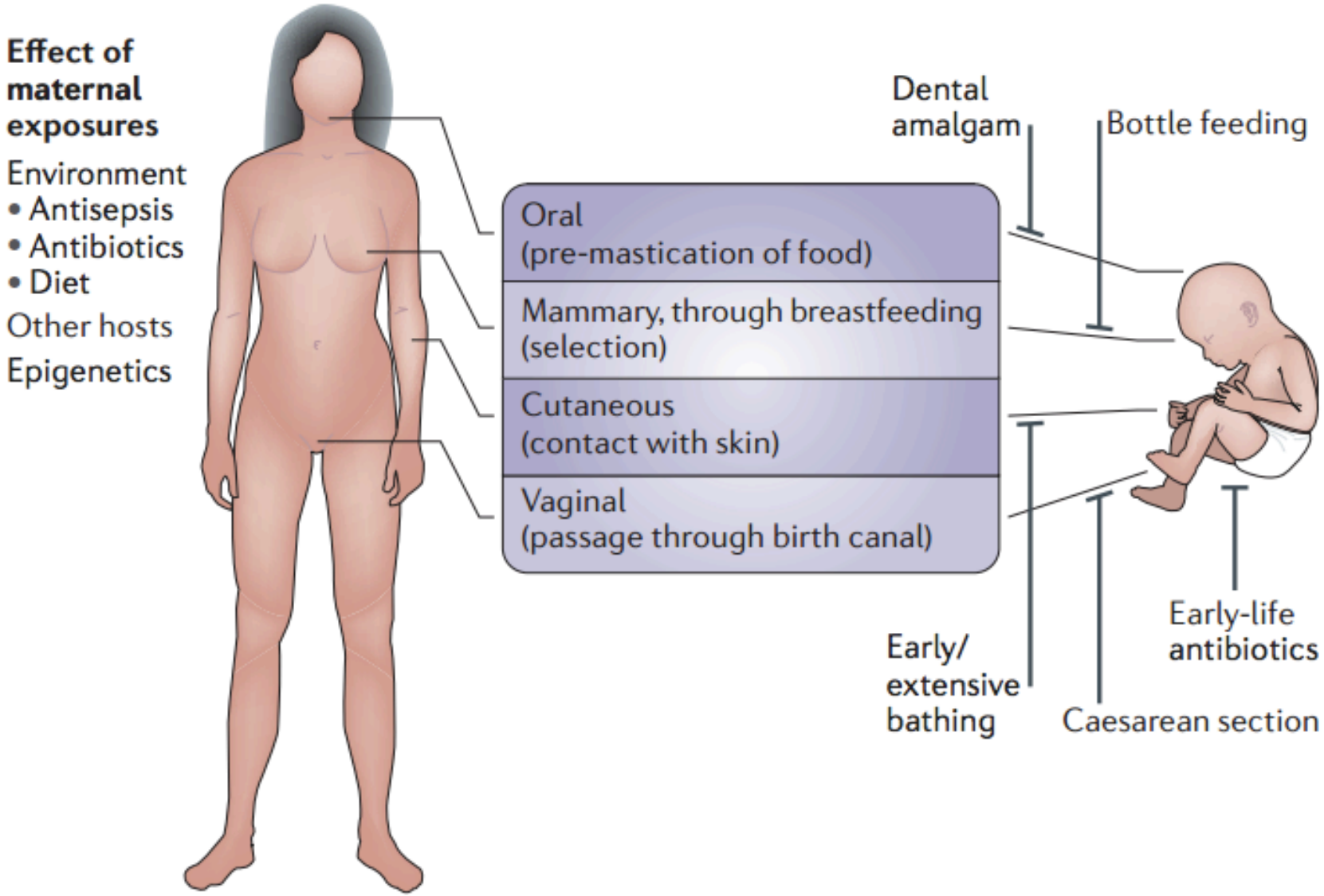
- Genes with known homologs
  - BLAST against known databases
    - We can obtain information on gene family members
- Genes without known homologs
  - Ab initio gene prediction tools based on Markov models are not effective for partial ORFs that are frequently observed in metagenome data
  - Incremental clustering methods starting from short translatable regions have been developed
- For functional annotation of genes, partial ORFs may be directly annotated without constructing the full ORFs.

# Microbiomes in Human Body




- The microbiome composition varies by anatomical sites.
- Substantial interpersonal variation
- Interpersonal variation is more substantial than temporal variability

# Inheritance of the Microbiome from Mother to Baby



# Enterotypes of Human Gut Microbiome

**(Arumugan et al., Nature, 2011)**

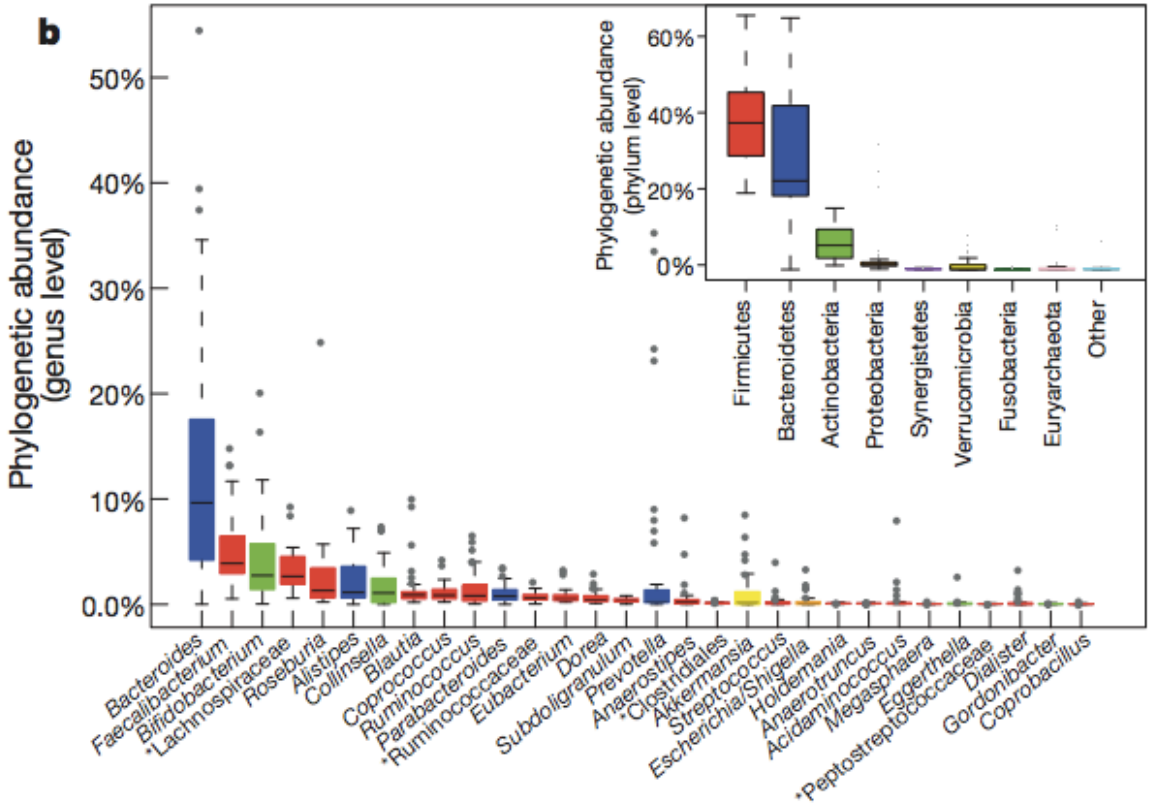
- Analysis of gut microbiome sequence data of 39 individuals including French, Spanish, Danish, Italian, American, and Japanese

- Enterotype: classification of individuals based on their microbiome ecosystem

# Enterotypes of Human Gut Microbiome

**(Arumugan et al., Nature, 2011)**

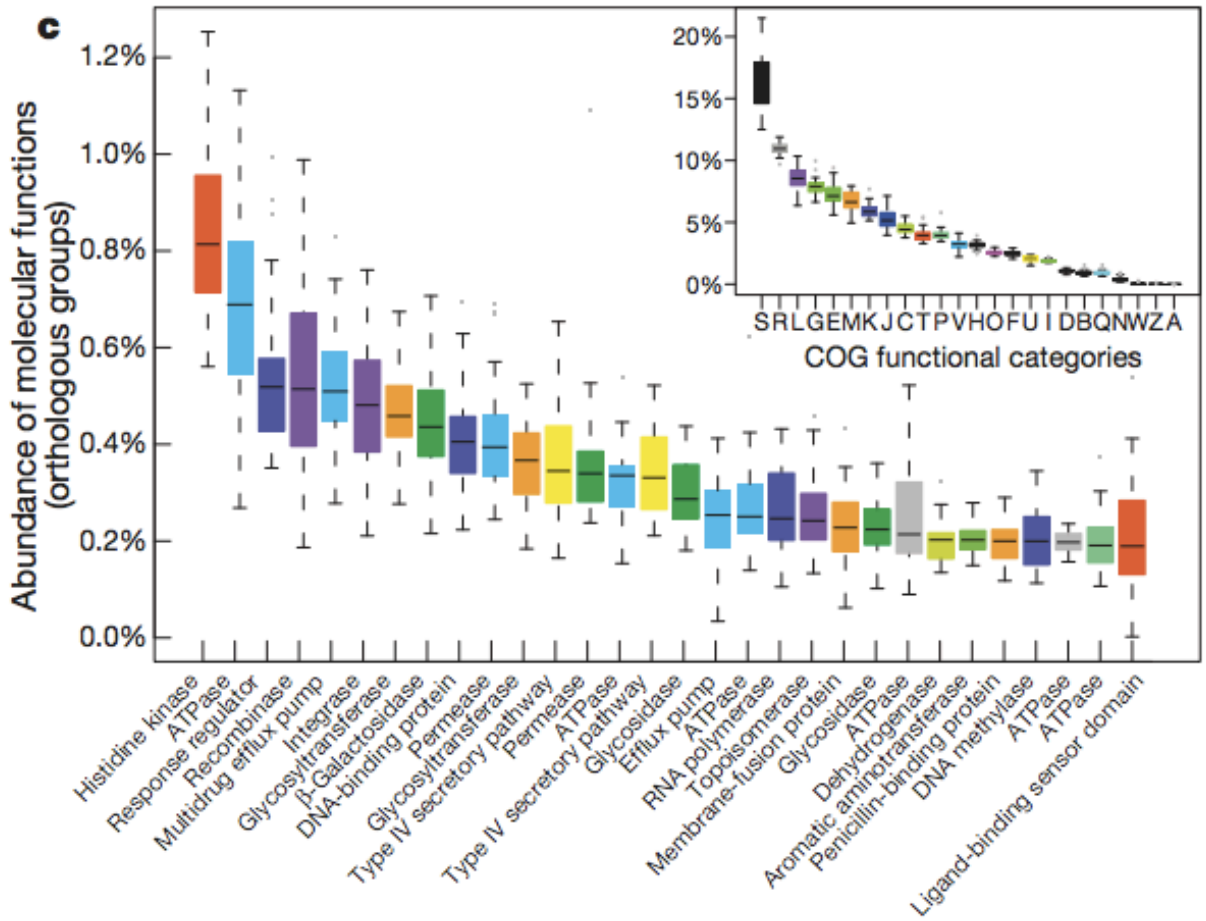
- Profiles of human gut microbiome (Boxplots represent individual variation)



# Human Gut Microbiome

**(Arumugan et al., Nature, 2011)**

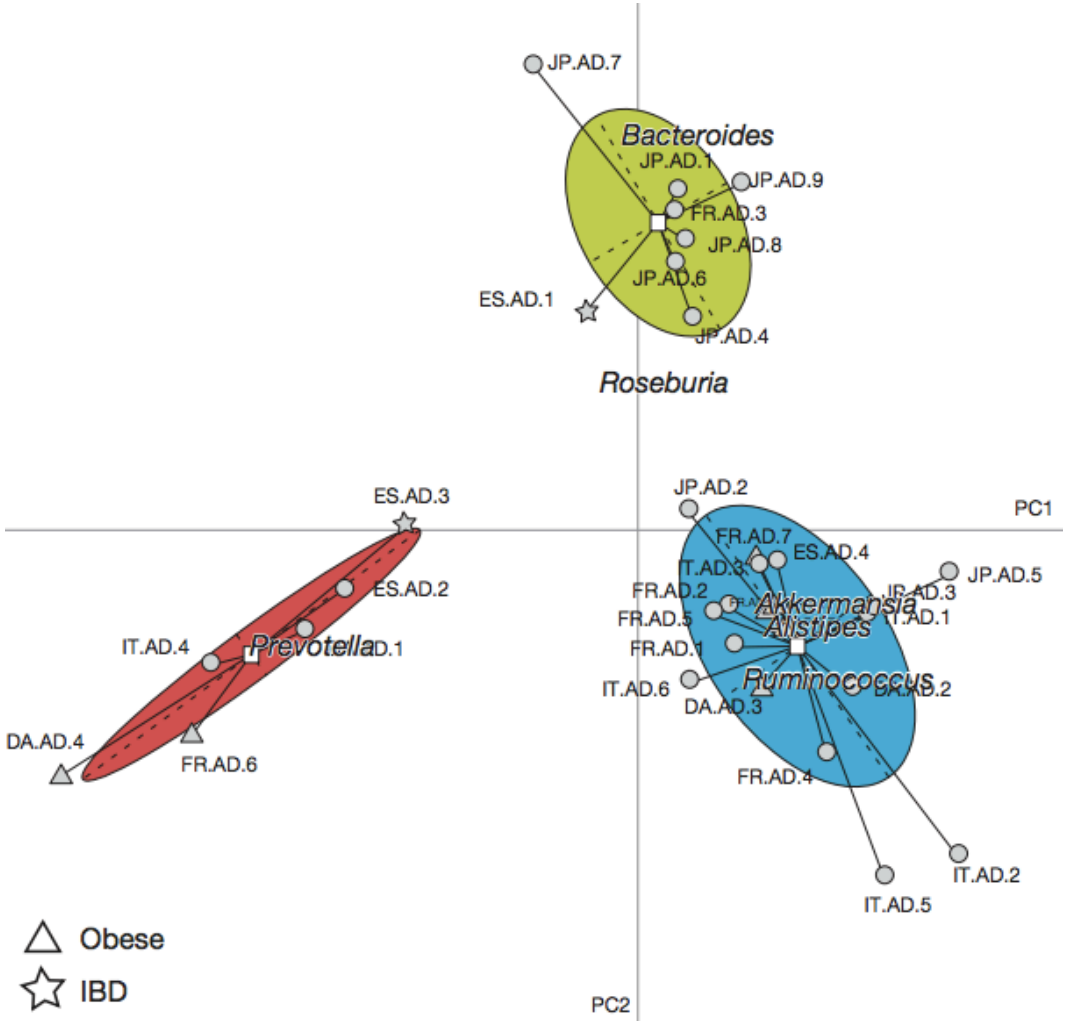
- Functional categories of the orthologous groups of genes



# Human Gut Microbiome

**(Arumugan et al., Nature, 2011)**

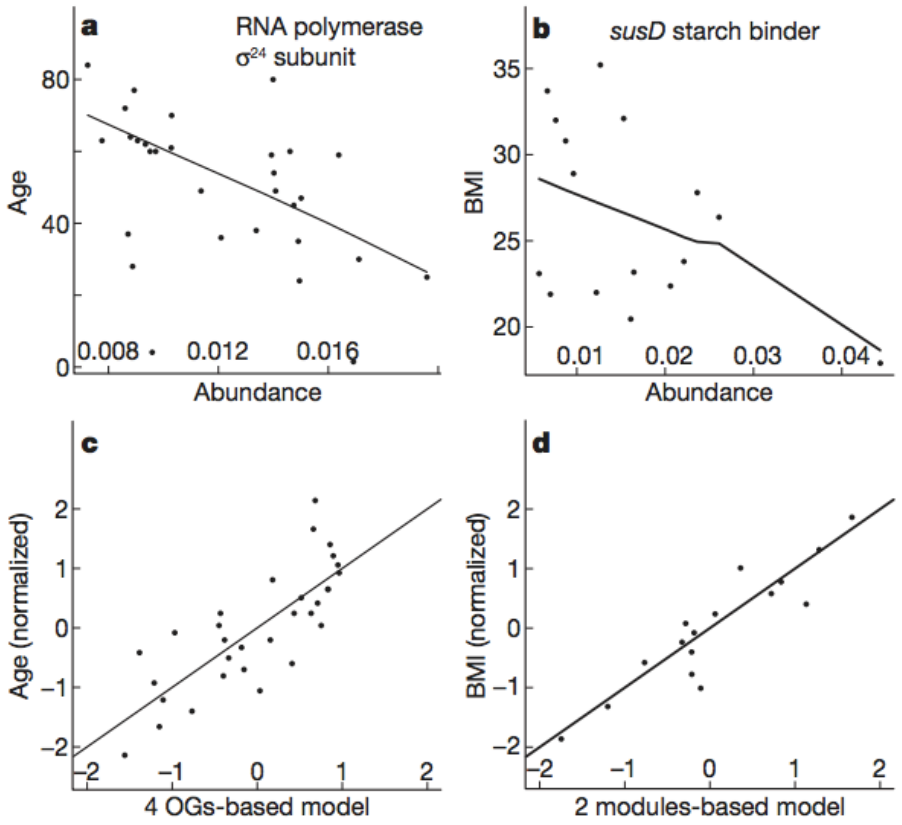
- The 39 individuals can be assigned to one of three Enterotypes
  - Bacteroides
  - Prevotella
  - Ruminococcus



# Human Gut Microbiome

**(Arumugan et al., Nature, 2011)**

- Correlations between host properties and abundant microbiome species



# Microbiome and Human Health

- Causal link between microbiome variation and disease
  - Variant microbe populations that occur in specific disease states
  - Temporal microbial changes over the course of a disease
- Associations between human conditions and microbiota characteristics

| Disease | Relevant finding |
|---|---|
| Psoriasis | Increased ratio of Firmicutes to Actinobacteria |
| Reflux oesophagitis | Oesophageal microbiota dominated by gram-negative anaerobes; gastric microbiota with low or absent *Helicobacter pylori* |
| Obesity | Reduced ratio of Bacteroidetes to Firmicutes |
| Childhood-onset asthma | Absent gastric *H. pylori* (especially the cytotoxin-associated gene A (*cagA*) genotype) |
| Inflammatory bowel disease (colitis) | Larger populations of Enterobacteriaceae |
| Functional bowel diseases | Larger populations of *Veillonella* and *Lactobacillus* |
| Colorectal carcinoma | Larger populations of *Fusobacterium spp.* |
| Cardiovascular disease | Gut-microbiota-dependent metabolism of phosphatidylcholine |

# Microbiome and Human Health

- Gut microbiota and obesity
  - In mice
    - Genetically obese mice have decreased Bacteroidetes/Firmicutes ratios compared with their lean siblings
    - Transplantation of gut microbiota from the obese (ob/ob) to germ-free mice conferred an obese phenotype – the transferred microbiomes had increased capacities for energy harvest
  - In humans
    - the relative proportions of members of the Bacteroidete phylum increase with weight loss
    - Antibiotic use in human infancy (before the age of 6 months) was significantly associated with obesity development

# Key Questions for Microbiome Research

- Understanding microbiome characteristics in relation to families
- For diseases that have changed markedly in incidence in recent decades, do changes in the microbiome have a role? (e.g., childhood-onset asthma, food allergies, type 1 diabetes, obesity, inflammatory bowel disease and autism)
- Do particular signatures of the metagenome predict risks for specific human cancers and other diseases that are associated with ageing? Can these signatures be pursued to better understand oncogenesis? (Work on Helicobacter pylori provides a clear example of this.)
- How do antibiotics perturb the microbiome, both in the short-term and long-term? Does the route of administration matter?
- How does the microbiome affect the pharmacology of medications? Can we 'micro-type' people to improve pharmacokinetics and/or reduce toxicity? Can we manipulate the microbiome to improve pharmacokinetic stability?
- Can we harness knowledge of microbiomes to improve diagnostics for disease status and susceptibility?
- Can we harness the close mechanistic interactions between the microbiome and the host to provide hints for the development of new drugs?
- can we harness the microbiome to develop new narrow-spectrum antibiotics?
- Can we use knowledge of the microbiota to develop true probiotics (and prebiotics)?

# Human Microbiome Project

| Body region | Body site | Total samples |
|---|---|---|
| Gut | Stool | 352 |
| Oral cavity | Buccal mucosa | 346 |
| | Hard palate | 325 |
| | Keratinized gingiva | 335 |
| | Palatine tonsils | 337 |
| | Saliva | 315 |
| | Subgingival plaque | 334 |
| | Supragingival plaque | 345 |
| | Throat | 331 |
| | Tongue dorsum | 348 |
| Airway | Anterior nares | 316 |
| Skin | Left antecubital fossa | 269 |
| | Left retroauricular crease | 313 |
| | Right antecubital fossa | 274 |
| | Right retroauricular crease | 319 |
| Vagina | Mid-vagina | 145 |
| | Posterior fornix | 152 |
| | Vaginal introitus | 142 |
| | Total | 5,298 |

- Sequencing microbiome obtained from 15 (males) and 18 (female) body sites from hundreds of donors

# Summary

- Human microbiomes are being investigated for gene content and function using sequencing technology
- Classifying individuals into enterotypes based on gut microbiomes
- Characteristics of human microbiomes are often related to health issues