

SELECTED ORAL PRESENTATION

Open Access

# Statistical methods for comparing the abundances of metabolic pathways in metagenomics

Bo Liu\*, Mihai Pop

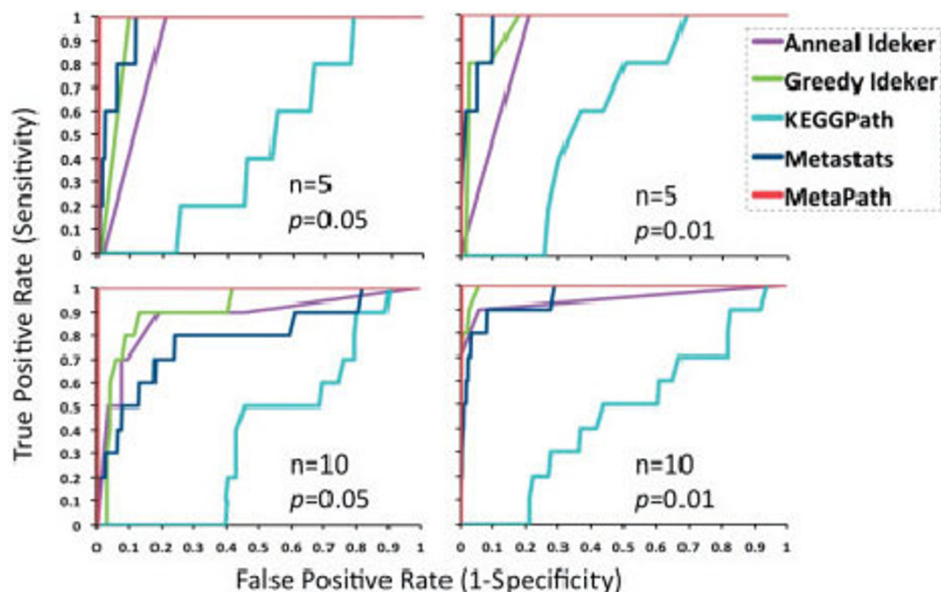
From Beyond the Genome: The true gene count, human evolution and disease genomics  
Boston, MA, USA. 11-13 October 2010

## Background

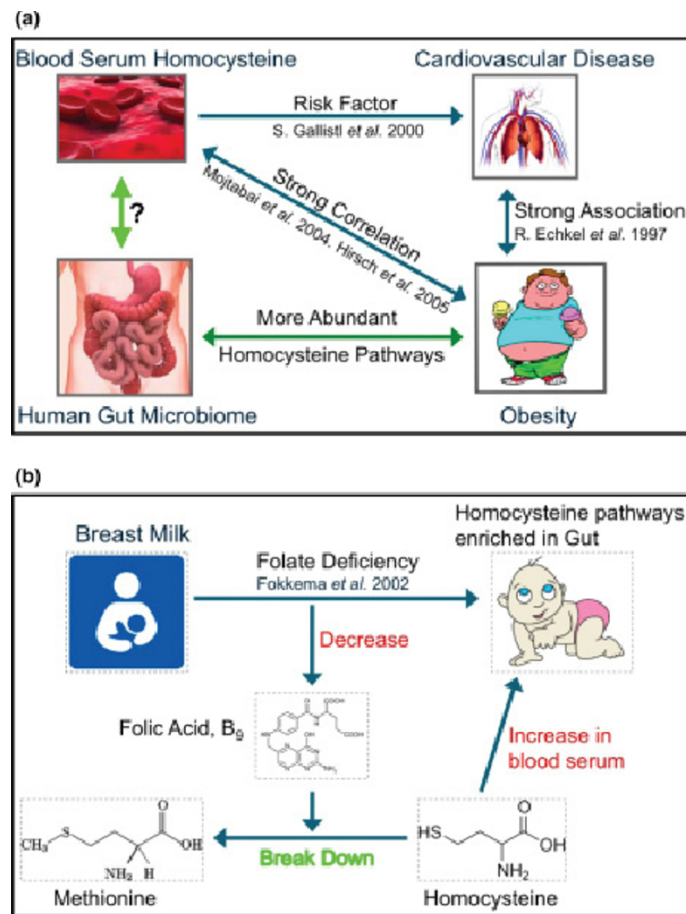
A major goal of metagenomic studies is to identify specific functional adaptations of microbial communities to their habitats. The functional profile and the abundances for a sample can be estimated by mapping metagenomic sequences to the global metabolic network consisting of thousands of molecular reactions. Here we describe our development of statistical methods that can identify differentially abundant subnetworks between metagenomic samples.

## Methods

First, we introduced a scoring function for an arbitrary subnetwork and find the max-weight subnetwork in the global network by greedy search. Then we compute  $p_{abund}$  and  $p_{struct}$  values using nonparametric approaches to answer two statistical questions: (i) Is this sub-network differentially abundant? (ii) What is the probability of finding such good subnetworks by chance? Significant metabolic subnetworks are detected on the basis of these two  $p$  values.



**Figure 1 Performance on simulated datasets.** Our methods MetaPath outperforms Anneal and Greedy (Ideker *et al*, 2002), KEGGPath and Metastats (White *et al*, 2009) on four simulated datasets.  $n$  is the number of reactions in the subnetwork and  $p$  is their significance.



**Figure 2** Homocysteine pathways are enriched in (a) obese and (b) infant subjects.

## Results

**Simulated datasets** We randomly choose a metabolic subnetwork as differentially abundant, and then simulate the abundance values from Gaussian distributions. Figure 1 shows the performance of different methods on discovering the significant subnetwork. **Real metagenomic data sets** We analyzed gut microbiome from obese or lean [1], and infant or adult subjects (Kurokawa et al, 2007), and found several interesting pathways. For example, five pathways in fatty acid biosynthesis are enriched in obese subjects, which confirm the results of a previous study that obese subjects have an increased capacity for dietary energy harvest. In addition, four and three homocysteine pathways are enriched in obese and infant subjects (Figure 2), indicating that they are highly correlated with the homocysteine levels in blood serum.

## Conclusions

We have developed statistical methods to find differentially abundant metabolic pathways in metagenomics.

The performance is better than previous approaches. Results from real metagenomic datasets confirm previous observations and also provide several new biological insights.

Published: 11 October 2010

## Reference

1. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI: An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 2006, **444**:1027-1031.

doi:10.1186/gb-2010-11-S1-O7

Cite this article as: Liu and Pop: Statistical methods for comparing the abundances of metabolic pathways in metagenomics. *Genome Biology* 2010 11(Suppl 1):O7.