

肠道和环境微生物组学研究

王 鹏

2020.1

目录

CONTENTS

04

数据挖掘方法总结

03

研究案例思路分享

02

宏(元)基因组学研究策略

01

微生物群落分析概述

微生物群落分析概述

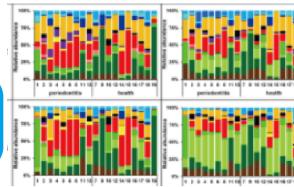
微生物存在历史

35亿年



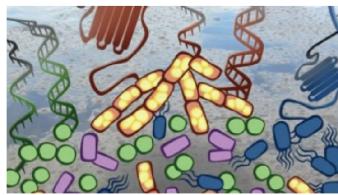
Diversity & Dynamics

Types | Distribution | Changes



Community structure

Who | Abundances | Composition



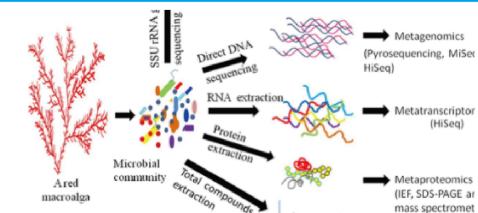
Organisms

Microbial Community

Genes

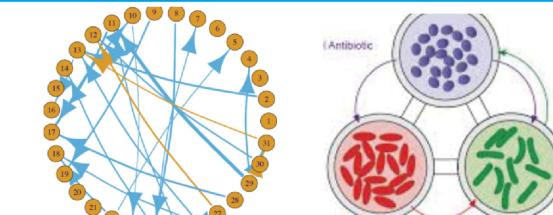
Ecosystem function

Who | Abundances | Composition



Interaction & Communication

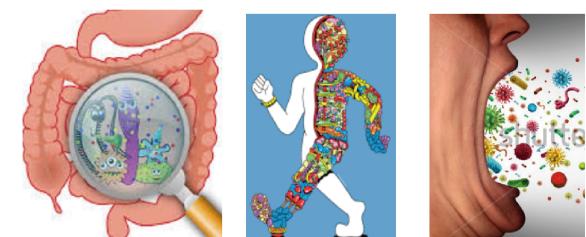
Competition | Partners | Signaling



Soil Water Food



Gut Skin Oral



Novogene
诺禾致源

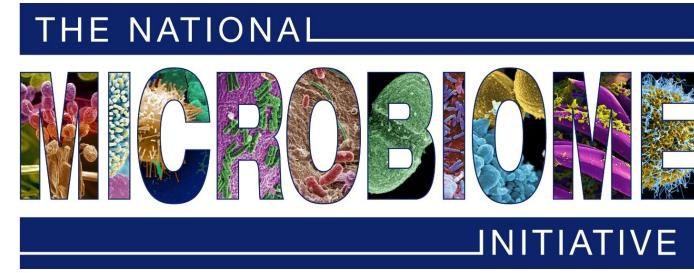
OTUs、metaOTUs、CAG、MLG.....

Providing advanced genomic solutions!

Metagenome-wide association studies: fine-mining the microbiome



<https://hmpdacc.org/>



<http://themicrobiome.com/en/>



<http://www.earthmicrobiome.org>



<http://americangut.org>



<http://www.metahit.eu>

Nevogene 諾禾致源

百家争鸣



中科院微生物组计划

基于微生物组学策略干预代谢性疾病及并发症的机制

家养动物肠道微生物组功能解析与调控

活性污泥微生物组功能网络解析与调节机制

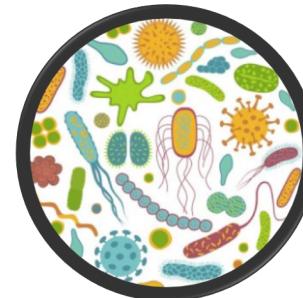
微生物组功能解析技术与计算方法学

中国微生物组数据库与资源库建设

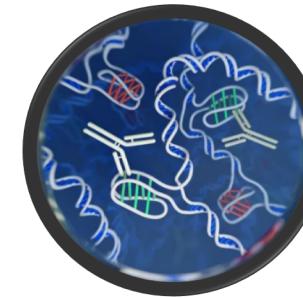
Providing advanced genomic solutions!

Microorganisms

Micro**biota**



Micro**biome**



Genomes and products

Biota: The animal and plant **life** of a particular region, habitat, or geological period.

Biome: A large naturally occurring **community** of flora and fauna occupying a major habitat

Cell



Design. Experiment. Repeat.



Design



Novogene
诺禾致源



Sampling

nature
International weekly journal of science



Get Data

Science
AAAS



Data Analysis



Conclusion



Providing advanced genomic solutions!



Temporal development of the gut microbiome in early childhood from the TEDDY study (2018)



Drosophila Histone Demethylase KDM5 Regulates Social Behavior
through Immune Control and Gut Microbiota Maintenance

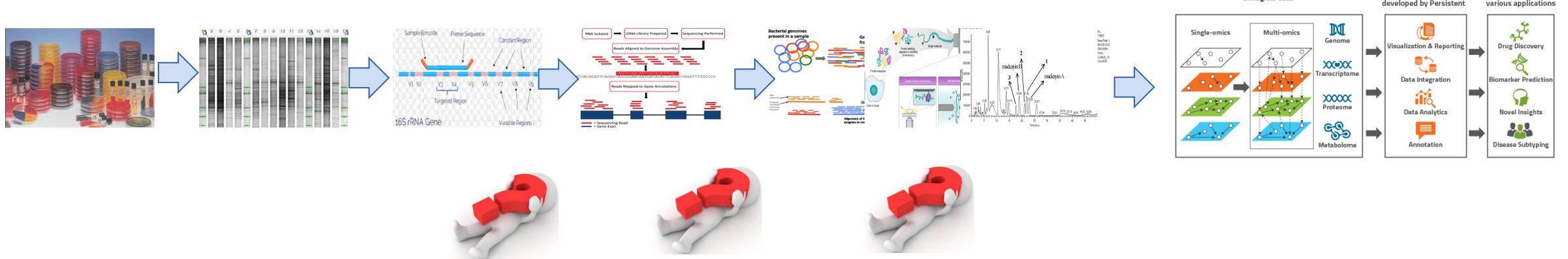


Enterotypes of the human gut microbiome VS Rethinking “Enterotypes”

Nevogene
诺禾致源

Providing advanced genomic solutions!

文章亮点总结



思路新颖

Novel insights from uncultivated genomes of the global human gut microbiome (2019)

样本充足

Temporal development of the gut microbiome in early childhood from the TEDDY study

分析透彻

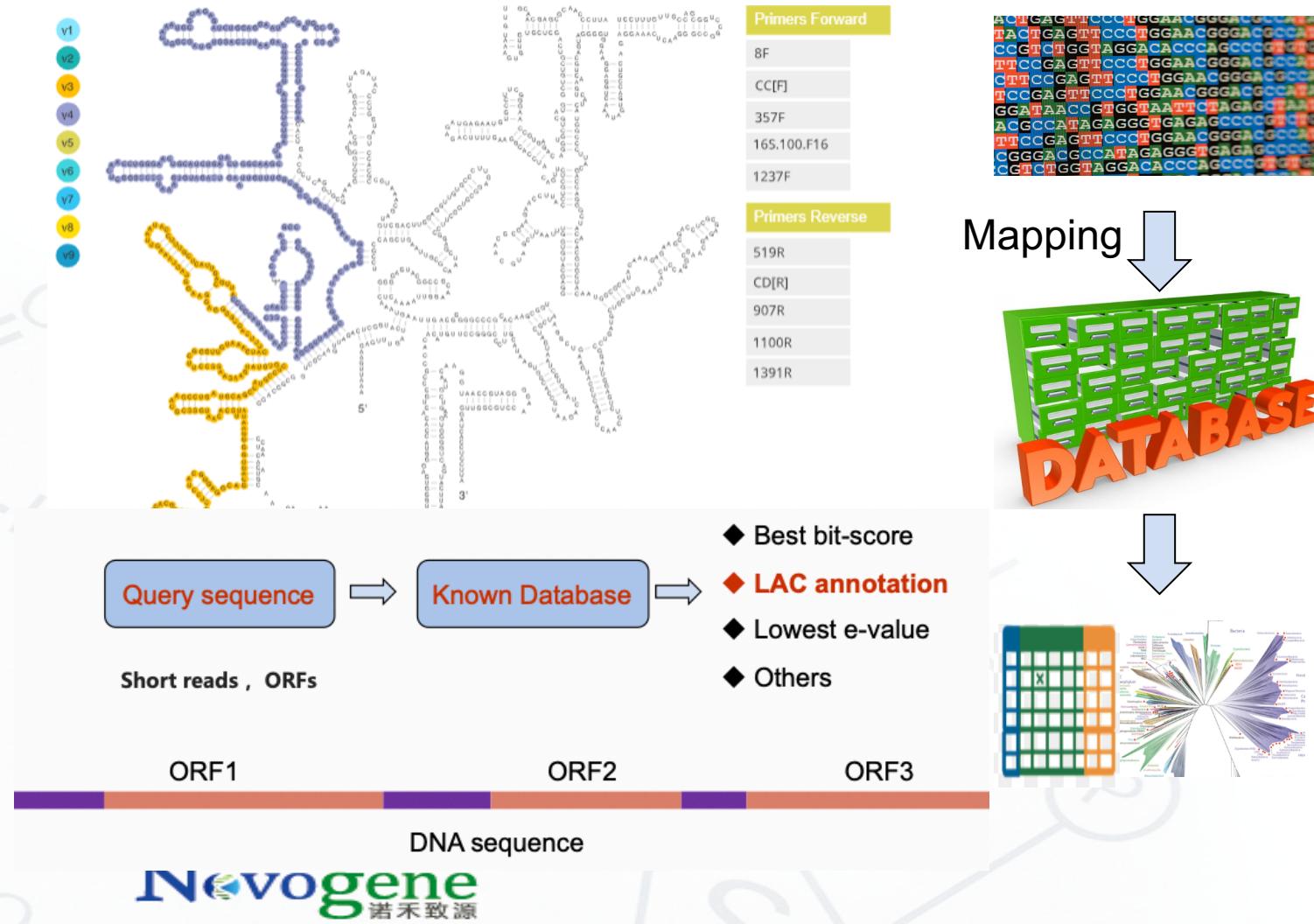
Cheese Rind Communities Provide Tractable Systems for In Situ and In Vitro Studies of Microbial Diversity

意义重大

Human gut microbes impact host serum metabolome and insulin sensitivity (2016)

组学研究通用流程

http://themicrobiome.com/media/16S_viewer.cfmv



16S: Representative OTUs

Meta: Assembly based unique genes

Meta: Read-based sequence

SILVA, GreenGenes, Customized

NT/NR/eggNOG/CAZy/KEGG/CARD

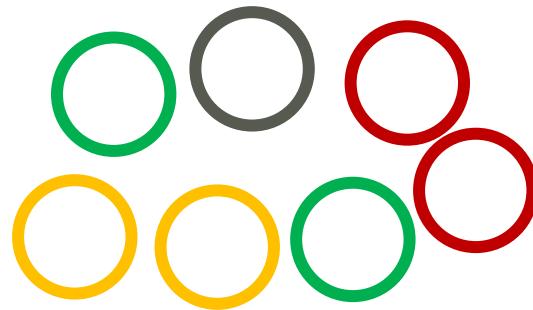
Targeted species, Known reference

Taxonomic matrix and function table

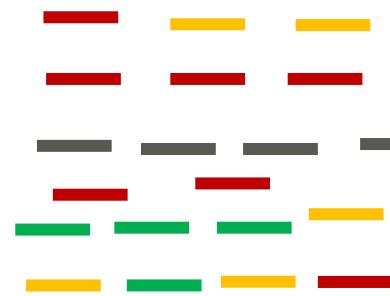
Abundance and phylogenetic tree

Providing advanced genomic solutions!

宏基因组分析流程



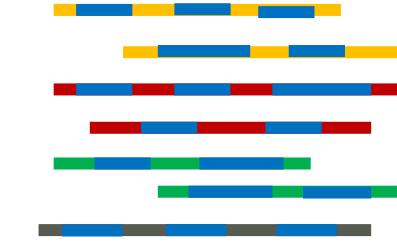
Microbial community



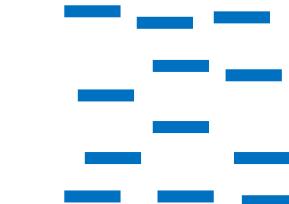
Shotgun sequencing



Assembled contigs



Gene prediction

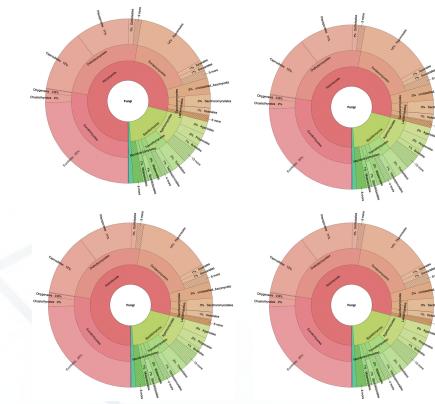


Gene catalogue

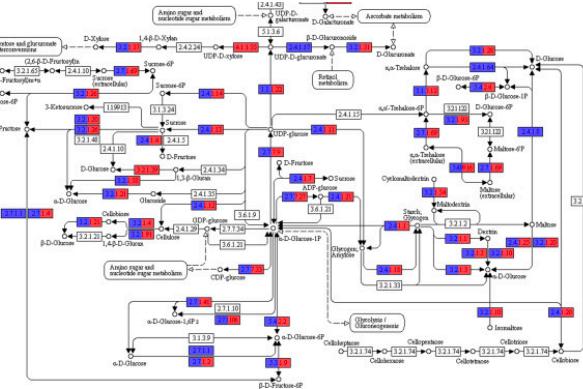
Non-redundant gene catalogue

	Sam1	Sam2	Sam3	Sam4
gene1	1	3	4	2
gene2	5	3	4	5
gene3	1	3	4	4
...	4	6	8	9
gene..	4	4	9	7

Gene abundance matrix



Taxonomic annotation



Function analysis

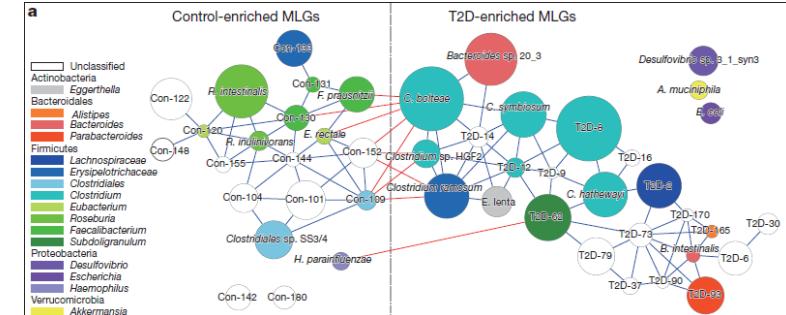
Brief introduction of assembly-based metagenomics

宏基因组纵向深入 (Binning)

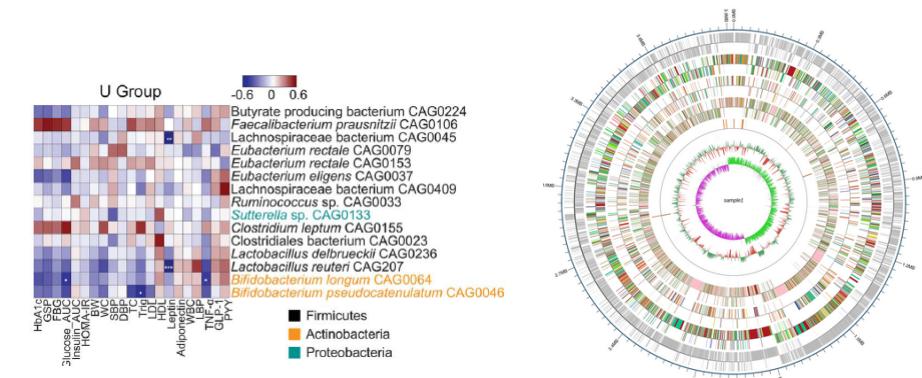
Metagenome-wide association study (MGWAS)

- Co-abundance gene groups (CAG)
- Metagenomic linkage groups (MLG)
- Metagenomic species (MGS)
- Metagenome-assembled genomes (MAG)

- ① Recovering draft genomes 序列再利用
- ② Correlated with parameters 提高种株识别率



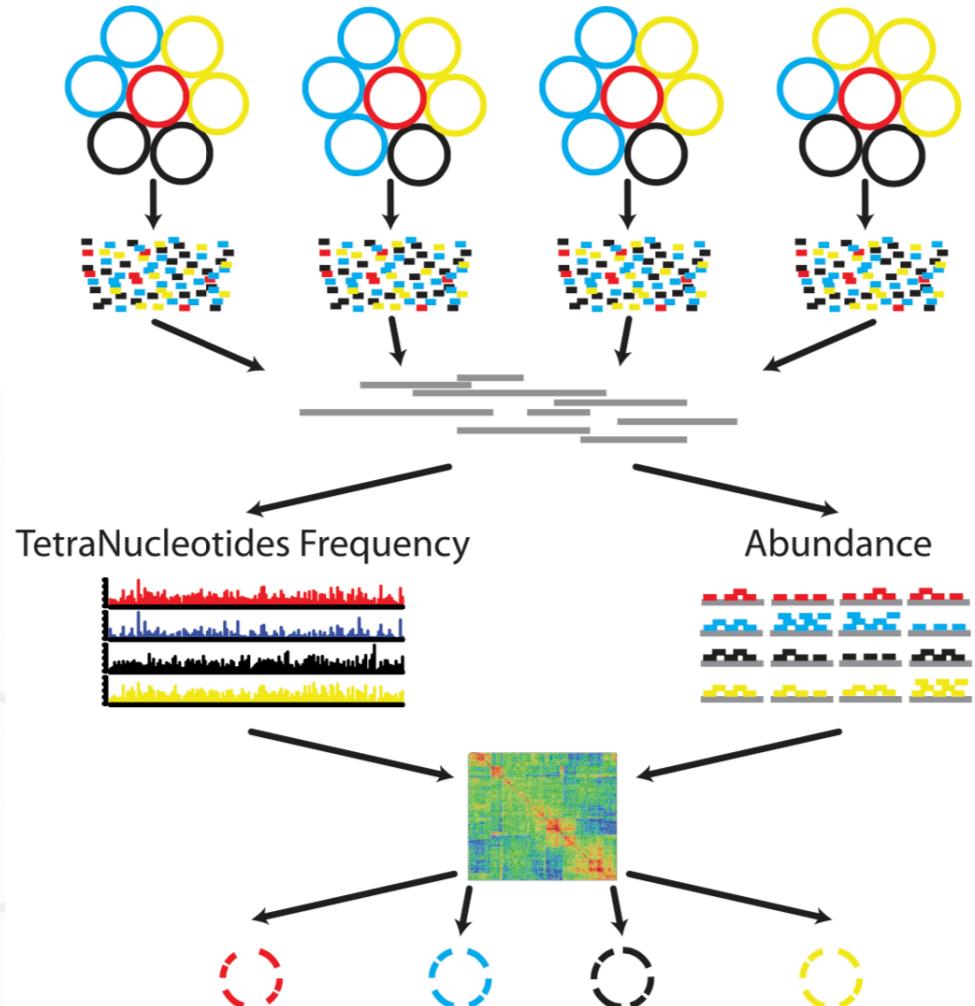
doi:10.1038/nature11450



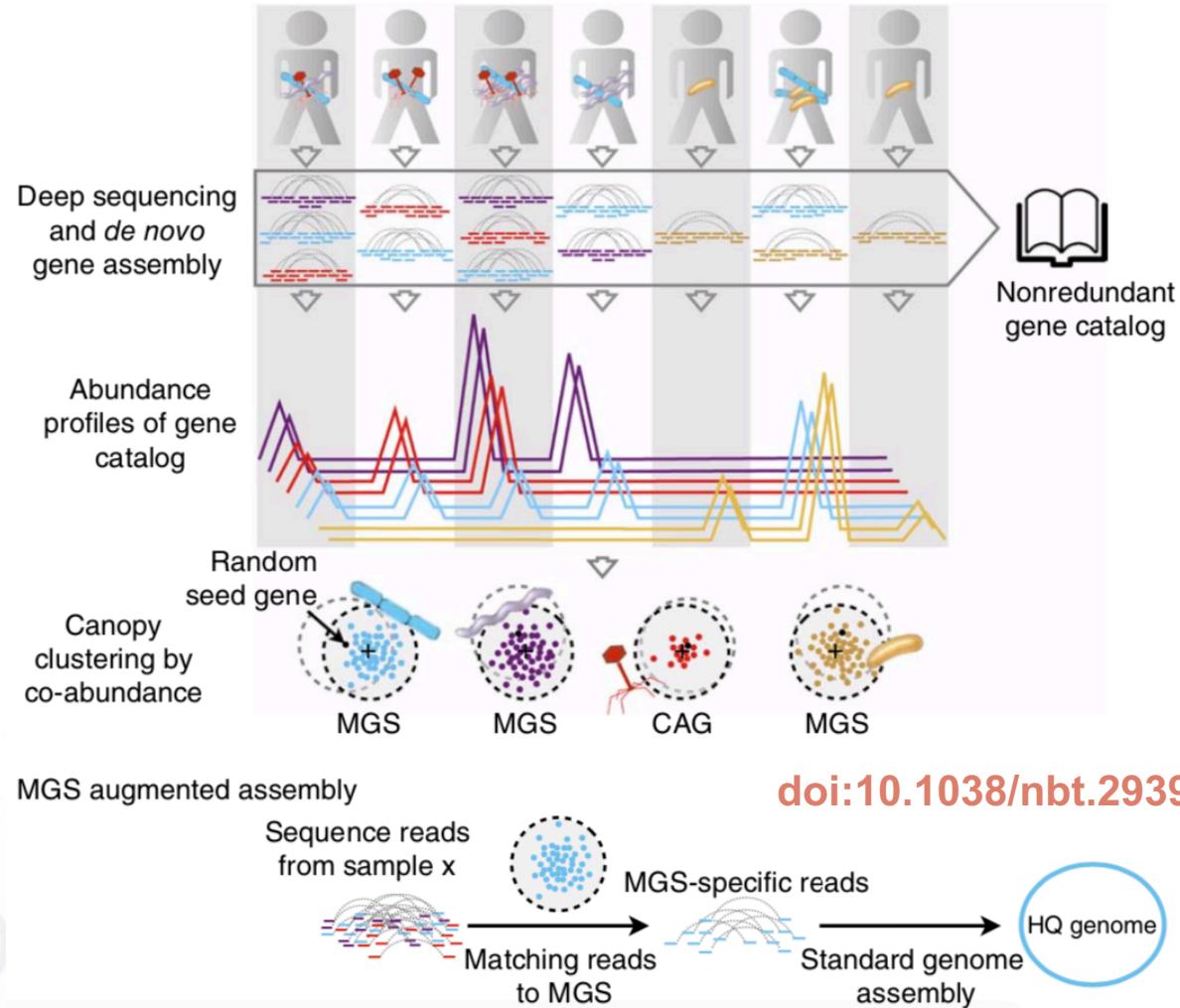
DOI: 10.1126/science.ao5774

Binning的方法

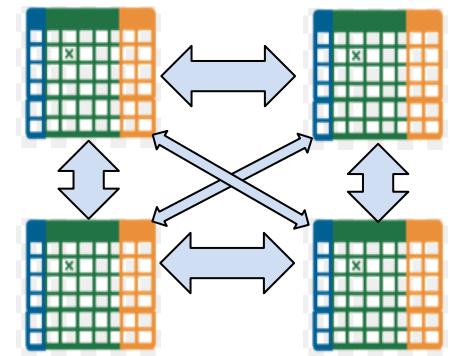
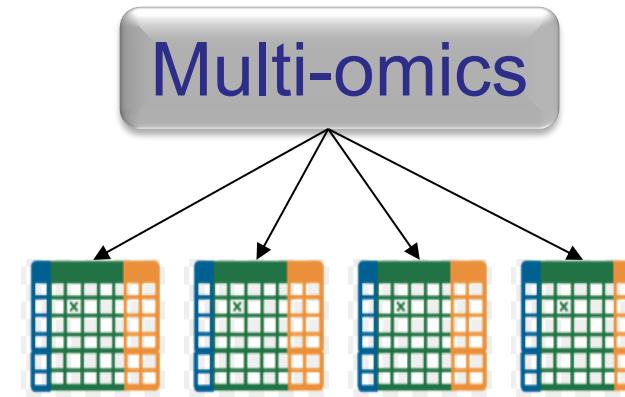
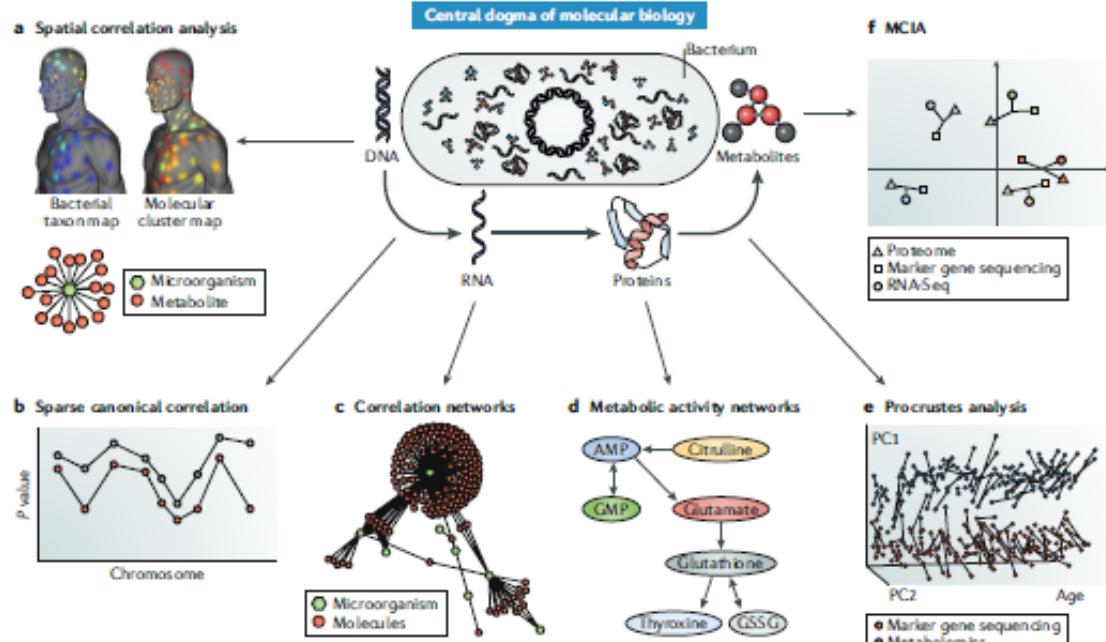
Based on contigs



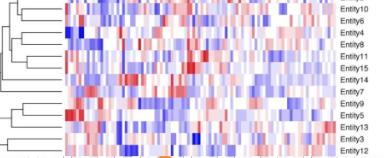
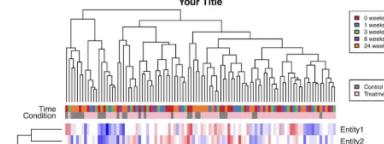
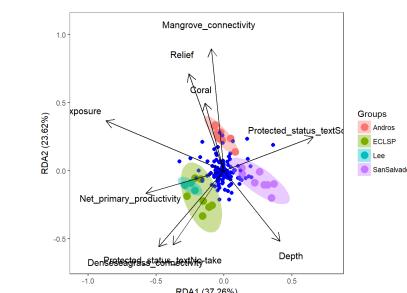
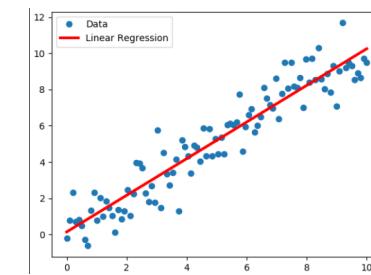
Based on genes abundance

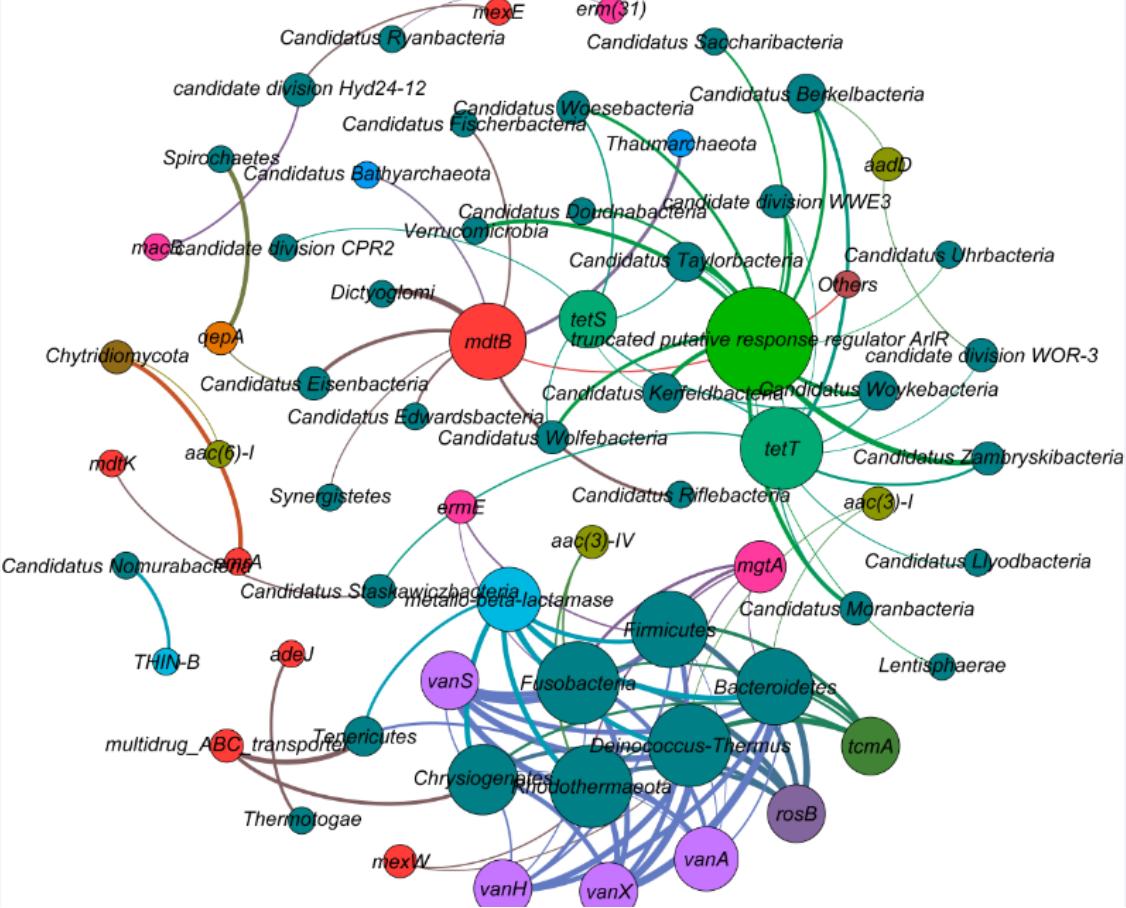
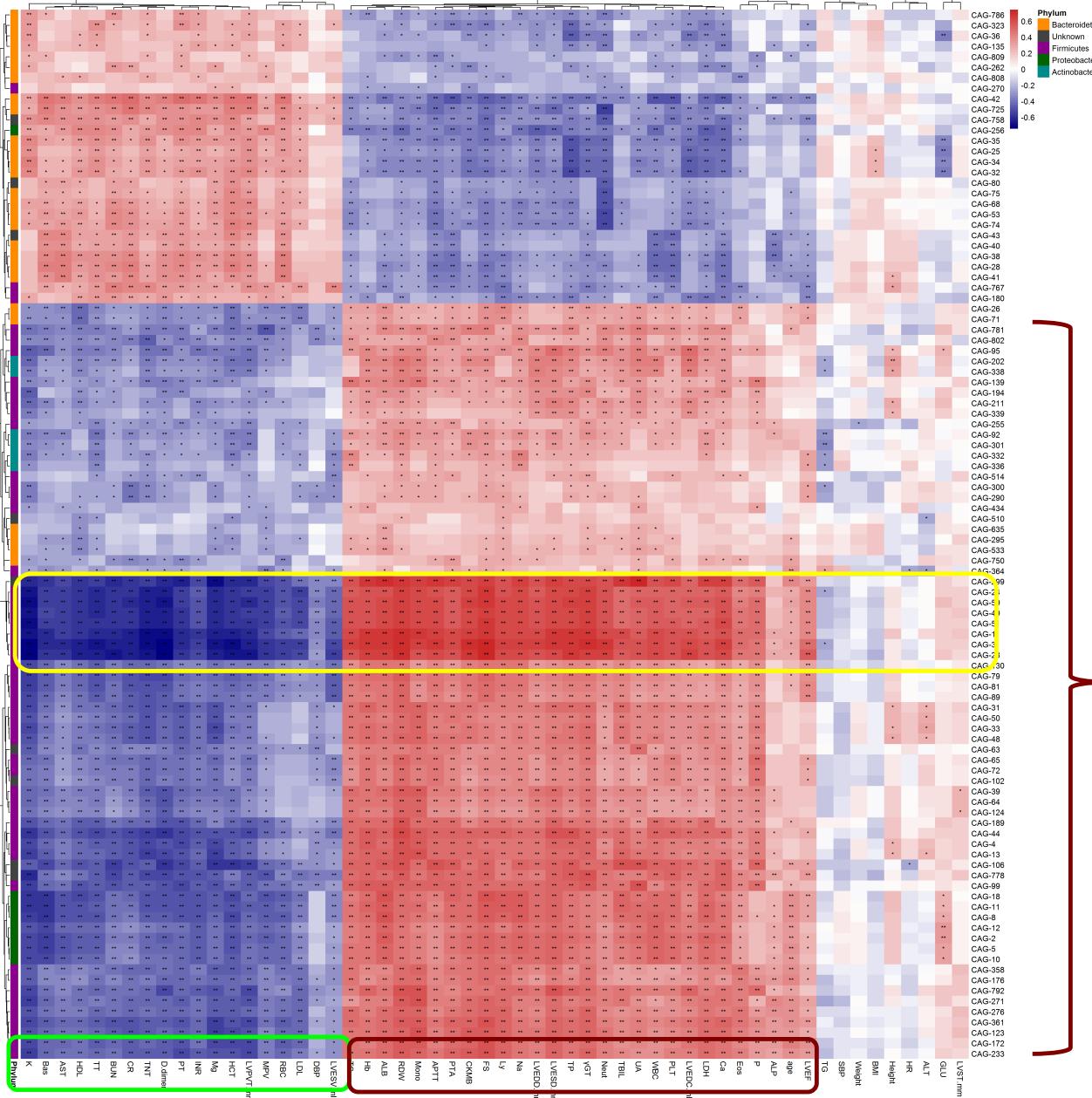


组学横向扩展（多组学关联）



- Pearson/Spearman correlation
- GLM (MaAsLin), PLS regression
- Two matrix (mantel test)
- Dimension reduction (CCA/Procrustes/envfit)



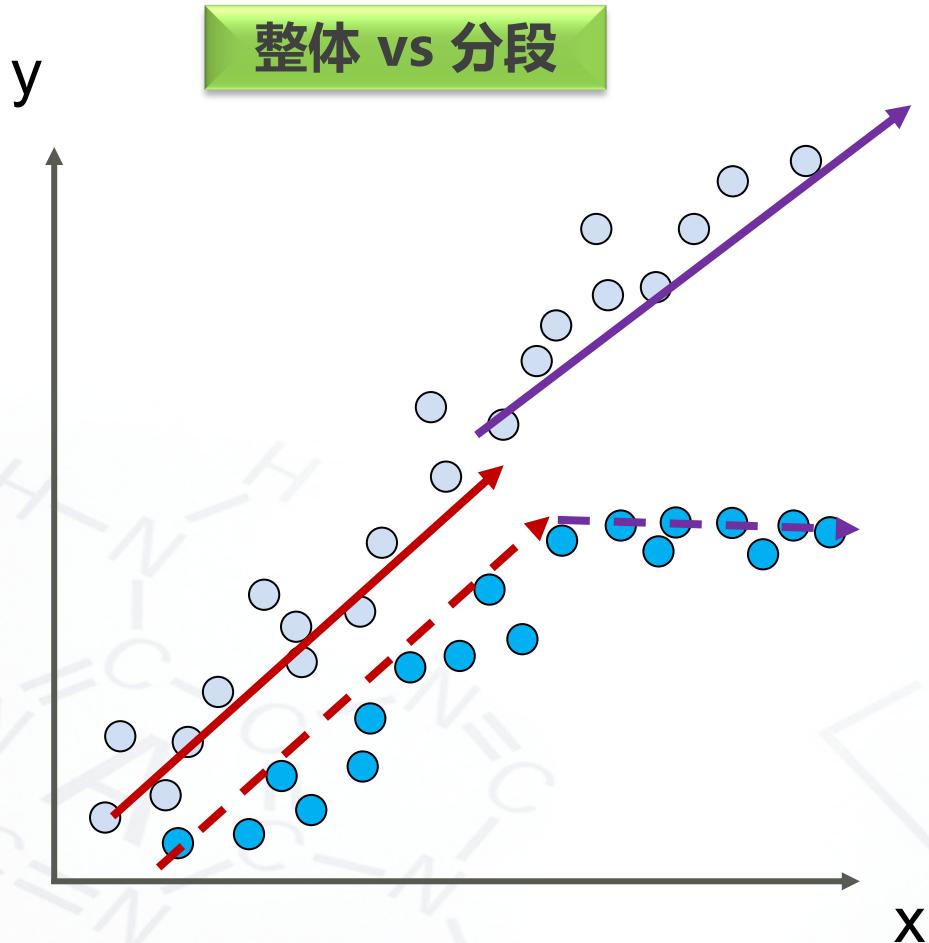


Six Degrees of Separation

(small world)

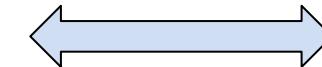
Providing advanced genomic solutions!

关联存在的问题



doi:10.1038/nrm1857
doi:10.1038/nrmicro2832

假关联



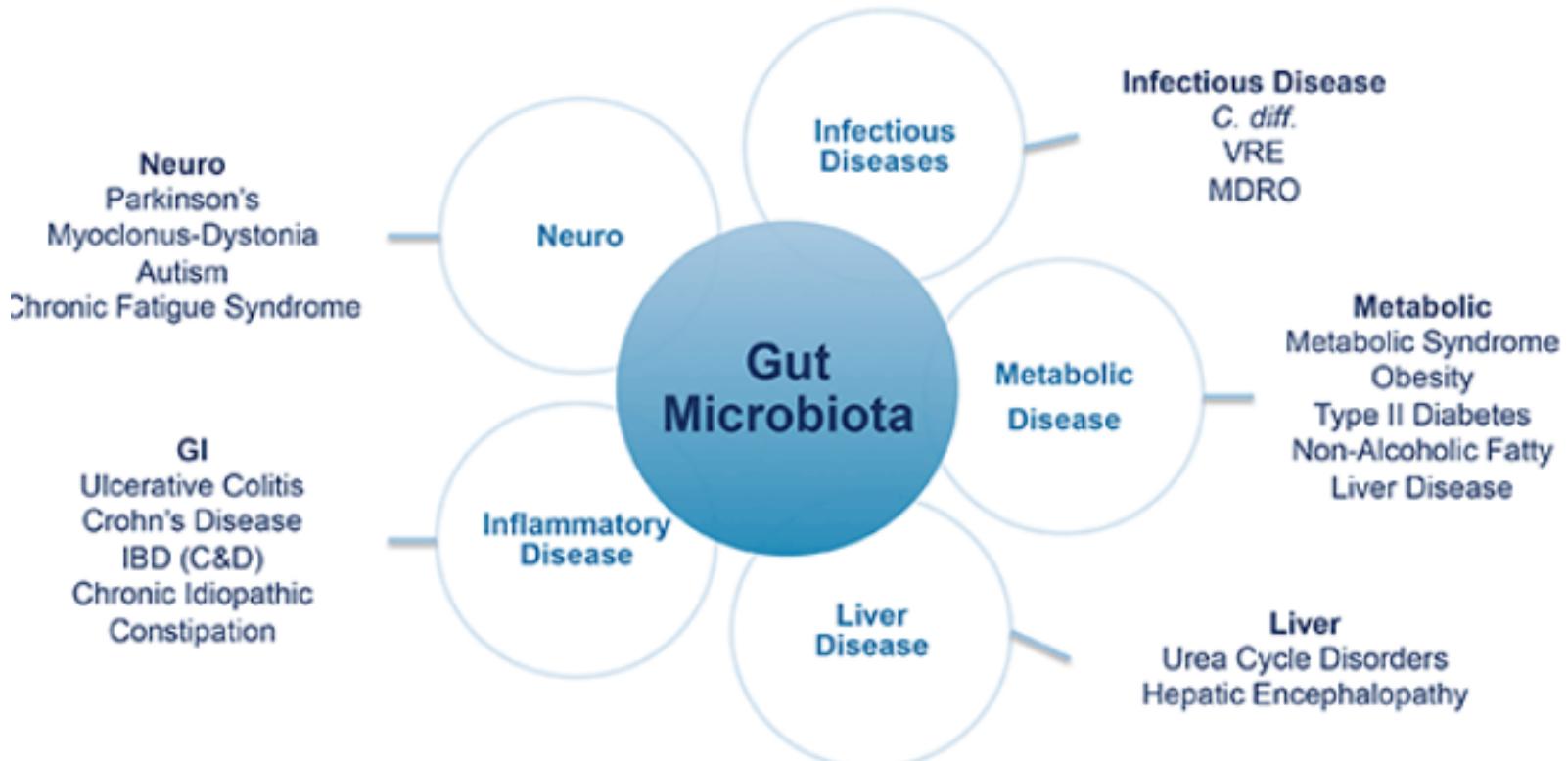
- ◆ 关联发生错乱
- ◆ 关联无指导意义



肠道微生物研究

肠道 (~10m , ~300m²)

- 最早的器官
- 最大的免疫系统
- 蕴含最多的微生物



肠道微生物研究-案例1



Step1 细化拟解决的科学问题

(关键物种和关键物质、潜在的治疗效果)

Step2 收集样本 (自闭症人群【严重程度】+ 健康人群)

Step3 16S测定肠道菌群 (是否有差异)

Step4 定位菌群功能 (代谢通路) 和代谢产物的差异

Step5 进一步筛选biomarker (具体的微生物和代谢物)

Step6 结合可培养的实验进行验证

Cell

Human Gut Microbiota from Autism Spectrum Disorder Promote Behavioral Symptoms in Mice

Part1 菌群移植，证实肠道菌群与自闭症的因果关系

Part2 差异菌群、差异菌群与行为学的关系

Part3 RNA-seq分析小鼠大脑中自闭症的相关基因

Part4 代谢组分析，寻找差异代谢物

Part5 宏基因组代谢通路与代谢物的整合分析

Part6 差异代谢物的在小鼠中的验证实验

肠道微生物研究-案例2



膳食纤维

肠道菌群

2型糖尿病



Why



How



What



临幊上认为，膳食纤维可以减缓 2 型糖尿病，但是反映形式有多种 (a, 不同菌株拥有共同的代谢纤维的基因; b, 同一菌株对不同含量的膳食纤维反应也不同)

当增加膳食纤维时，(i) 肠道生态系统中的单个菌株如何反应；(ii) 菌株其它成员互作，组成功能菌群，如何对这种新的营养环境做出反应

HOW?

- 患病样本接受不同的处理（膳食纤维+常规处理）
- 表型上是否有改善（是否接近正常指标）
- 菌群（有益和有害）是否发生改变，有哪些？
- 代谢产物（有益、有害）和代谢通路的变化
- 有益和有害代谢产物与菌群的关系

Science

DOI: 10.1126/science.aa05774

Gut bacteria selectively promoted by dietary fibers alleviate type 2 diabetes

- 样本接受膳食纤维干预，并设置对照
- 干预前后表型的变化（血糖、糖化血红蛋白等）
- 表型的变化是否与肠道微生物有关（菌群移植）
- 随着膳食干预，肠道菌群的演变规律
- 代谢通路和代谢产物（短链脂肪酸）的变化
- SCFA产生菌的鉴定 (*but*、*fhs*)
- 可培养进行分离鉴定
- *Bifidobacterium pseudocatenulatum* 的接种实验

Novogene
诺禾致源

Providing advanced genomic solutions!

Temporal development of the gut microbiome in early childhood from the TEDDY study

环境微生物研究



中国土壤微生物群落的分布模式



中国白酒发酵微生物菌群的研究



- 在哪里，有什么（分布模式及其规律）
- 为什么在哪里（环境选择、种群互作）

Deterministic processes



Stochastic processes

Structure and function of the global topsoil microbiome

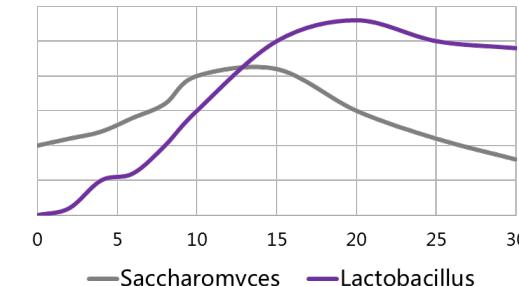
- 菌群结构和功能沿纬度的变化规律
- 菌群分布与环境、地理因子的关系
- ARGs分布规律及其与菌群的关系

气水
原料



窖泥
大曲

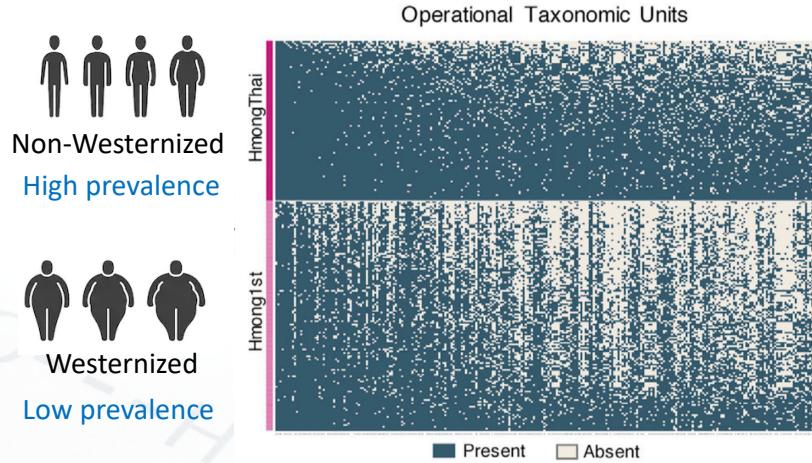
工人&器具



- 发酵过程中的微生物群落来源（有益和有害）
- 后期乳酸菌爆发的机制（发酵周期的控制）
- 酵母菌和乳酸菌相互作用的机制及其影响

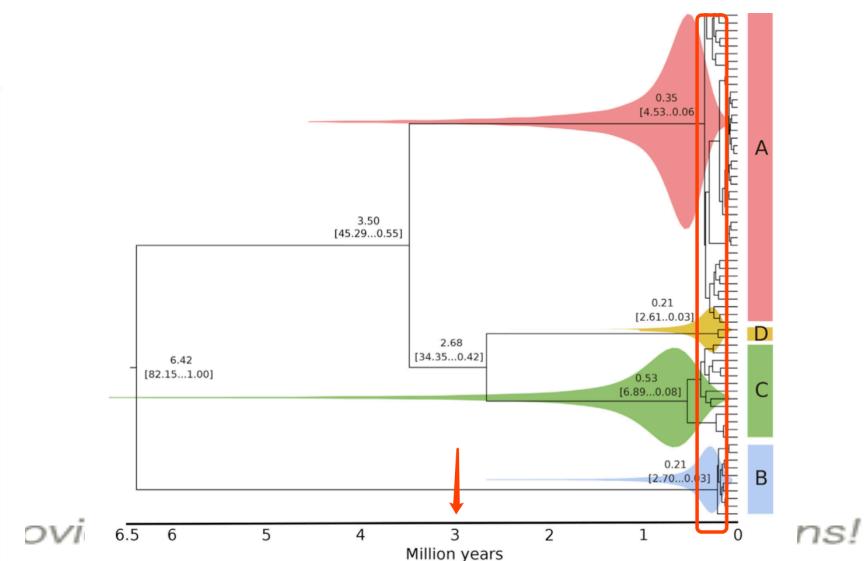
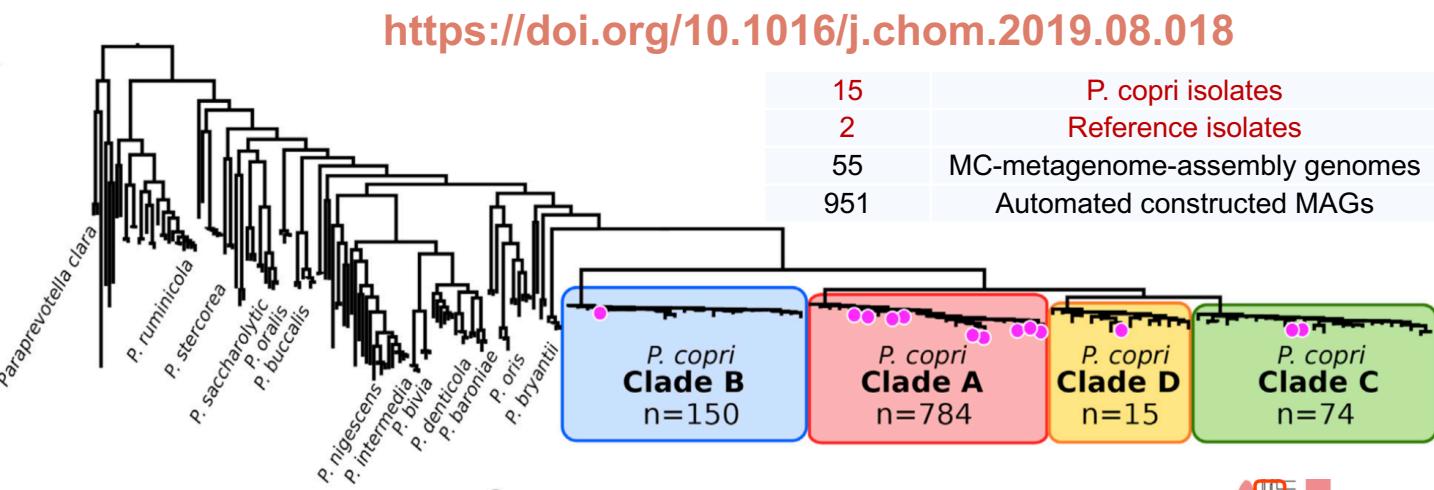
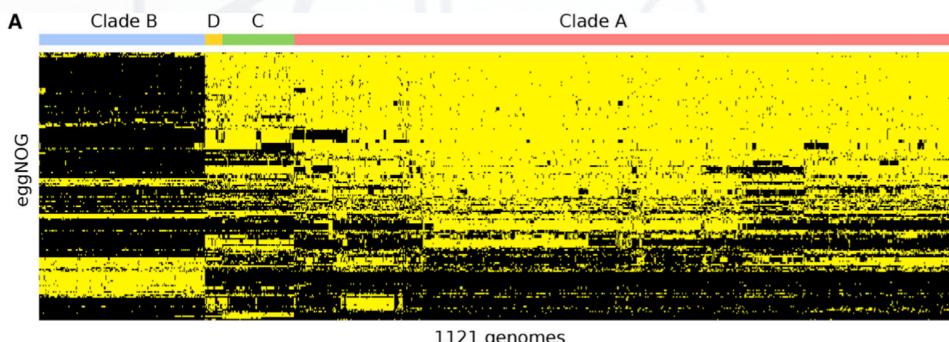
宏基因组研究新思路

Prevotella copri is 39.1% prevalent in healthy individuals from current metagenomic, it is not ubiquitous, but when present, it is often the most abundant species identified (34% of instances).



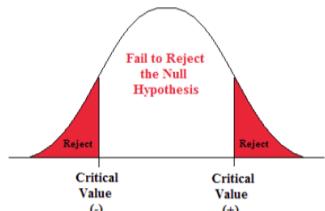
<https://doi.org/10.1016/j.cell.2018.10.029>

There is no clear evidence to suggest *P. copri* is associated with the etiology of these diseases



常用分析方法总结

How to identify biomarker



Hypothesis test

T-test

Wilcox test

ANOVA

Kruskal-Wallis

Metastats

LefSe

.....



Feature selection

Logistic regression

PCA (loading score)

PLS-DA (VIP)

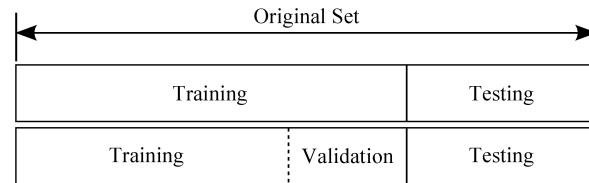
Random Forest

Lasso regression

Ridge regression

.....

Similarity analysis



Classification

Naïve Bayes classifier

Decision trees

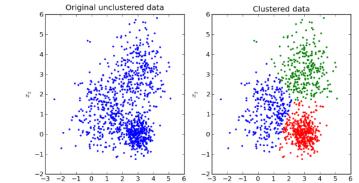
Random Forest

Support Vector Machine

Regression model

Linear discriminant

.....



Clustering

Kmeans clustering

Hierarchical cluster

Dimension reduction

PCA/PCoA/NMDS

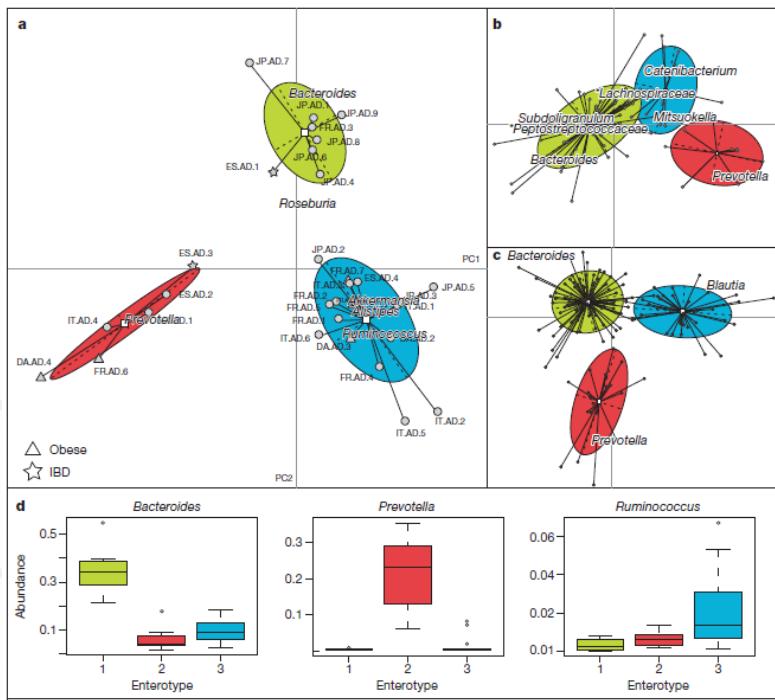
数据挖掘 Enterotype

doi:10.1038/nature09944

Enterotypes of the human gut microbiome

2011年Arumugam等人在Nature杂志

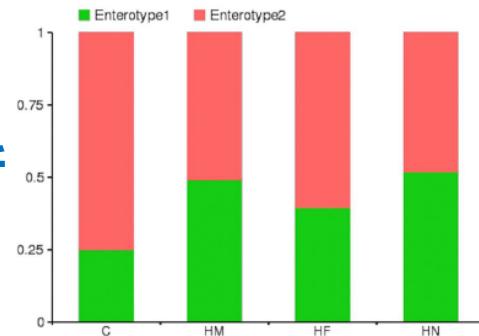
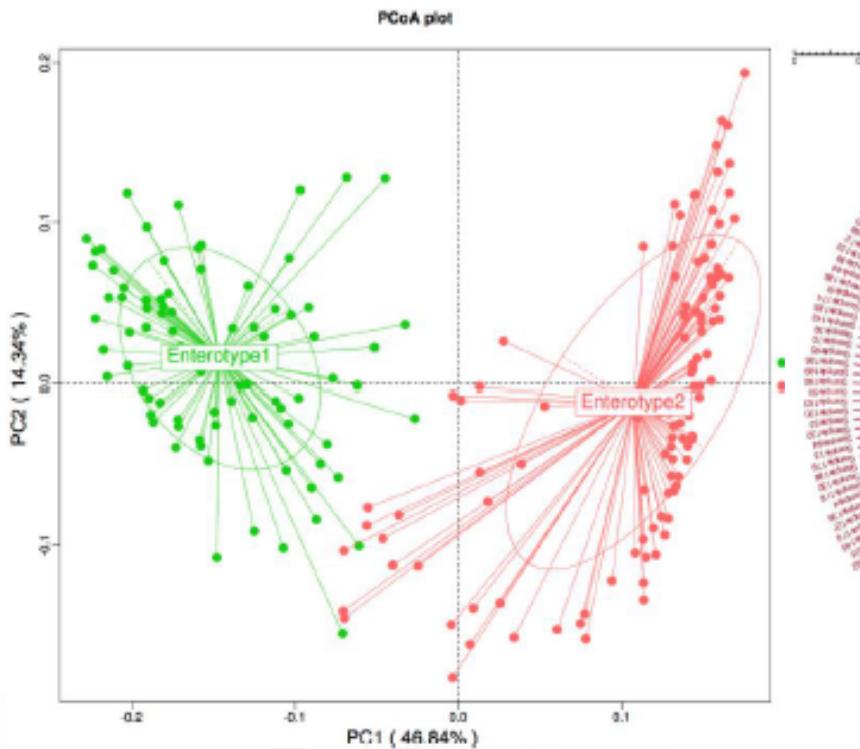
第一次提出肠型（Enterotype）概念



1、肠型的确定

2、降维分析——PCoA、PCA分析

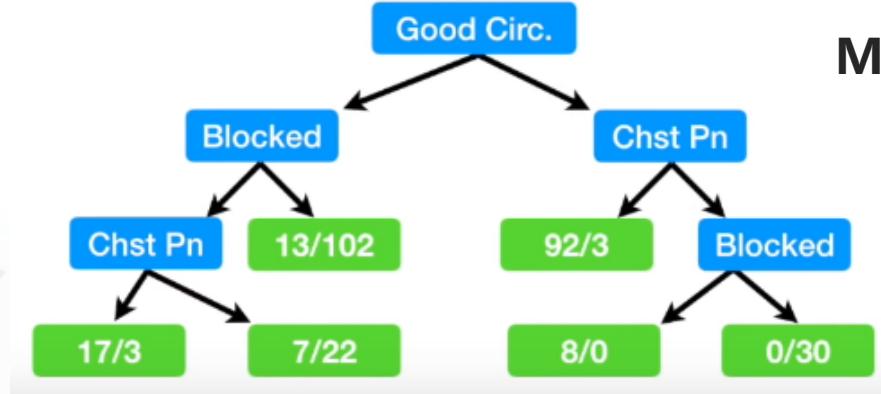
3、显著性差异物种分析



数据挖掘 RandomForest

随机森林是一种基于分类树算法的经典机器学习模型，(2001)

Decision tree

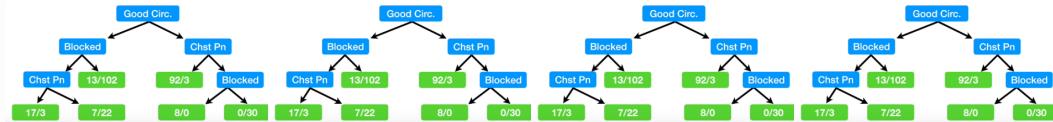


Dataset

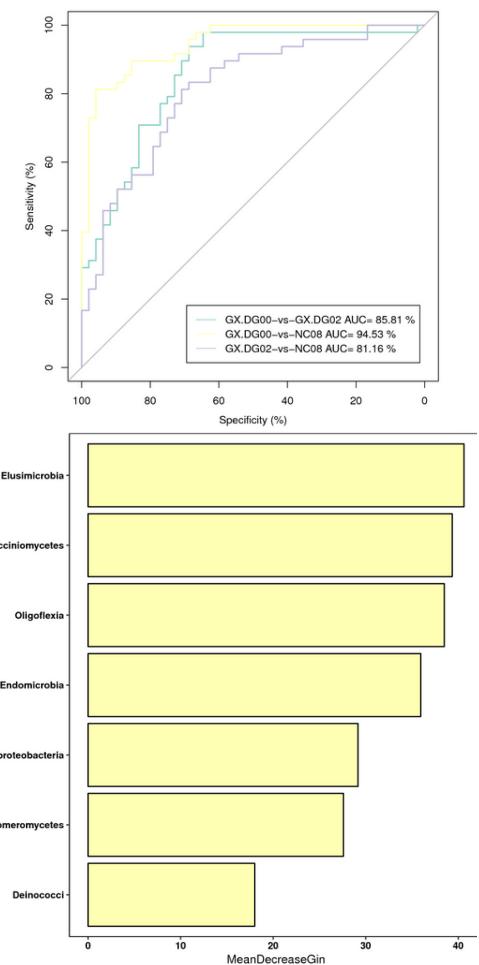
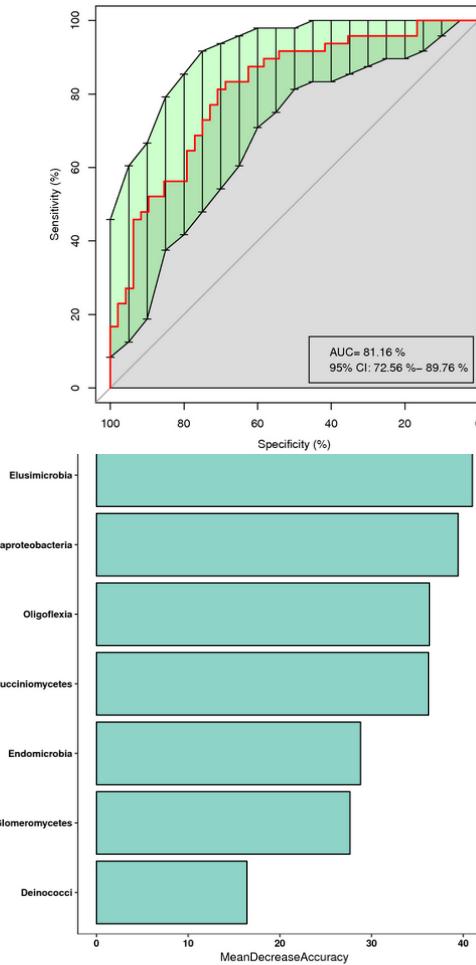
Model training

Prediction

Multi-trees



Feature selection & Classification | Result Prediction



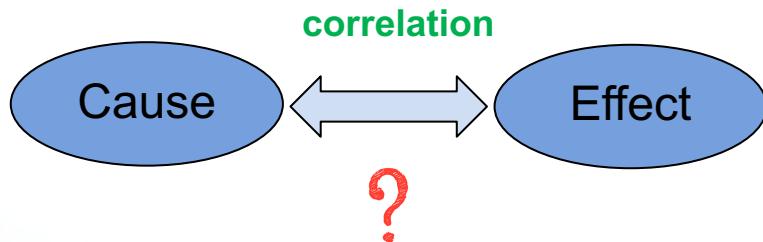
MeanDecreaseAccuracy

MeanDecreaseGin

Providing advanced genomic solutions!

数据挖掘 SourceTracker

SourceTracker是依据贝叶斯统计来定量化描述不同样本中微生物菌群的来源关系



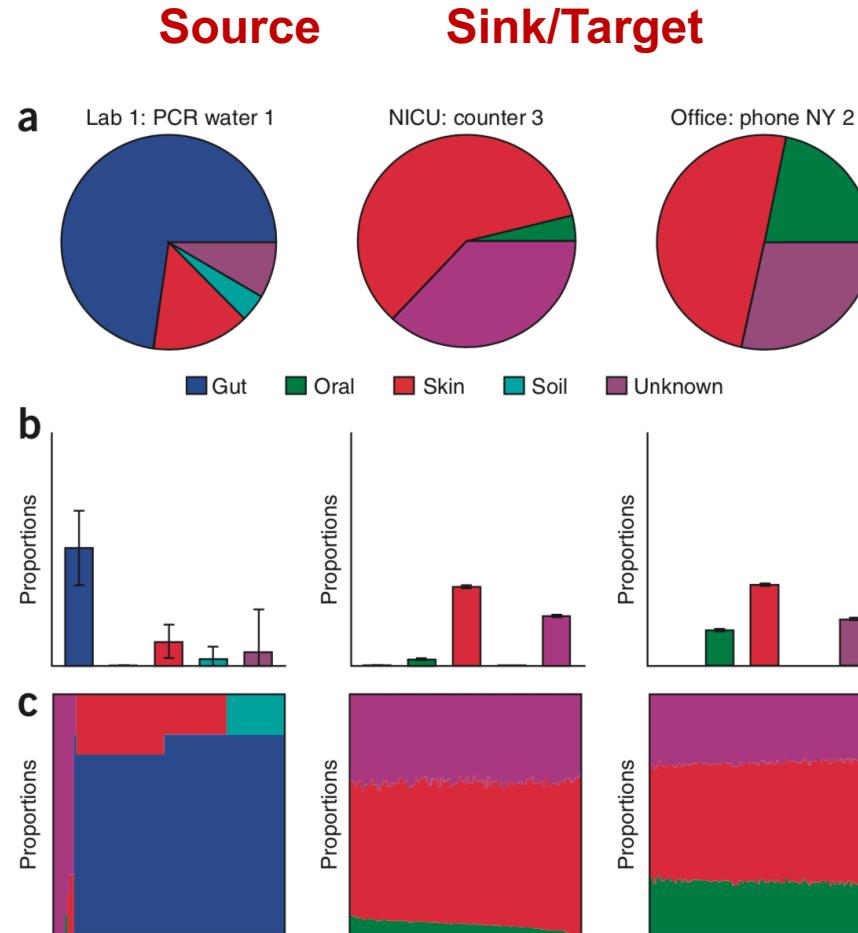
Bayesian statistics

The Posterior	The Evidence	The Prior
	The probability of getting this evidence if this hypothesis were true	The probability of H being true, before gathering evidence
$P(H E) = \frac{P(H E)P(H)}{P(E)}$		

The probability that the hypothesis (H) is true given the evidence (E)

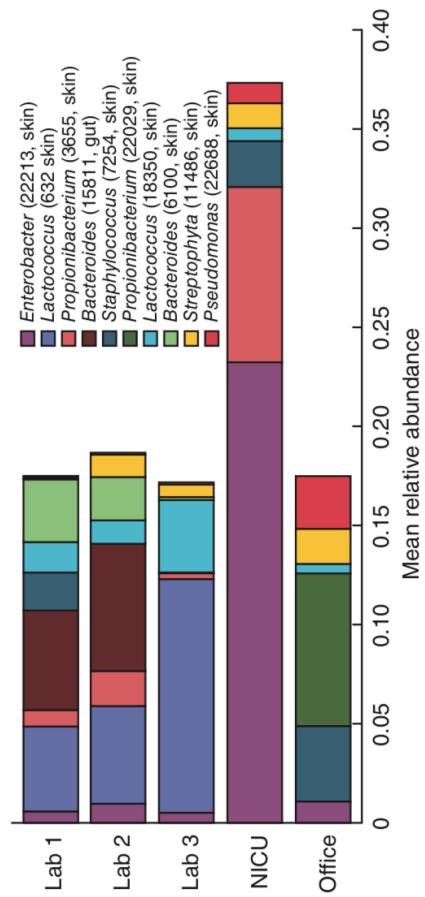
The marginal probability of the evidence (Prob of E over all possibilities)

Nevogene 諾禾致源



Providing advanced genomic solutions!

Bayesian community-wide culture-independent microbial source tracking



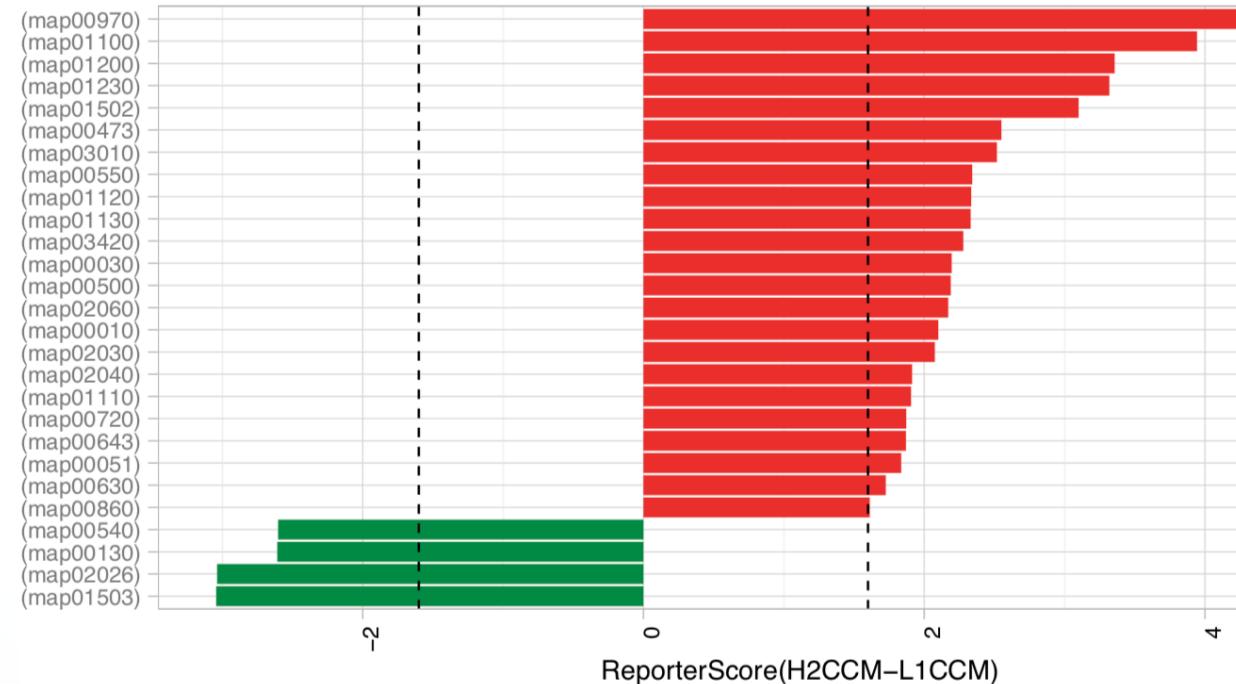
数据挖掘 Enrichment analysis

富集分析是基于KO，针对某个pathway或module计算出的Reporter score。

KO_ID	p_value	q_value	z_score
KO1	P ₁	q ₁	Z ₁
KO2	P ₂	q ₂	Z ₂
KO3	P ₃	q ₃	Z ₃
⋮	⋮	⋮	⋮
KOn	P _n	q _n	Z _n

Z pathway

Zadjust score



<https://github.com/wangpengnovos/ReporterScore>

https://github.com/wangpengnovos/diff_analysis

课程总结

- 组学分析基本流程
- 组学横纵分析(MGWAS)
- 高分组学文章的思路
- 常用数据挖掘方法

<https://github.com/wangpengnovos>



Providing advanced genomic solutions!

Thanks for your attention!

更多关注, 敬请留意: www.novogene.cn