Meta-Analysis of Colorectal Cancer Metagenomics Studies

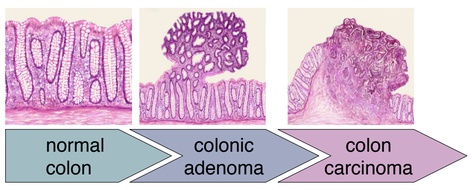
Jakob Wirbel[1], Paul Pyl[2], Ece Cevirgen[1], Alessio Milanese[1], Jonas S Fleck[1], Konrad Zych[1], Manimozhiyan Arumugam[2], Peer Bork[1], Georg Zeller[1]

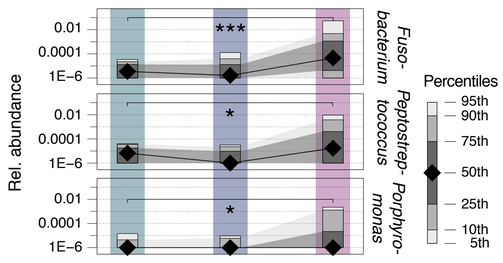
(1) Structural and Computational Biology Unit, EMBL Heidelberg
(2) Novo Nordisk Center for Basic Metabolic Research, Copenhagen

EMBL

Introduction

Colorectal Cancer and the Gut Microbiome





- Several microbes have been associated with colorectal cancer (CRC), e.g. *Fusobacteria* [Kostic et al., *Genome Res*, 2011]
- Metagenomics analysis of faecal samples can distinguish CRC cases and controls
- But it is not clear, how consistent these associations are

Faecal Metagenomics Studies Included in the Meta-Analysis

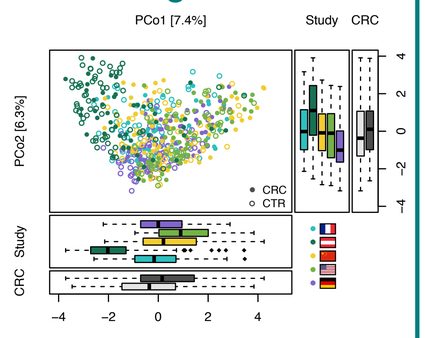
In this meta-analysis, we included four published and one additional unpublished feacal shotgun metagenomics datasets.

Country	Publication	Cases	Controls
	Zeller et al., *Molecular Systems Biology,* 2014	53	61
	Feng et al., *Nature Communications,* 2015	46	63
	Yu et al., *Gut,* 2015	74	54
	Vogtmann et al., *PLoS ONE,* 2016	52	52
	unpublished data from DKFZ, Heidelberg	60	60
	Total	285	290

Objectives

- How **reproducible** are the CRC-microbiome associations in the face of **technical variation**?
- Can we get closer towards a **„common truth“** by pooling data across several studies?
- How well do the statistical models trained on one study **generalize** across studies?

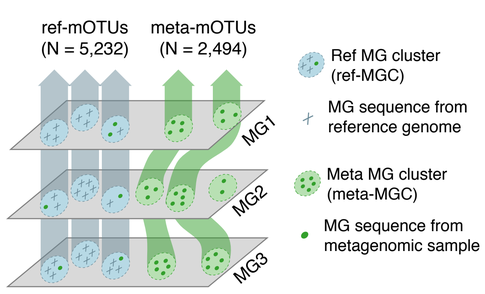
Challenges



- No good effect size measure for microbiome data
- Batch effects are predominant (see above)
- High variability in healthy gut microbiome abundance profiles

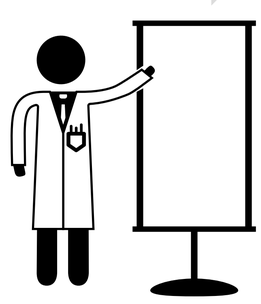
Methods

mOTU Profiler



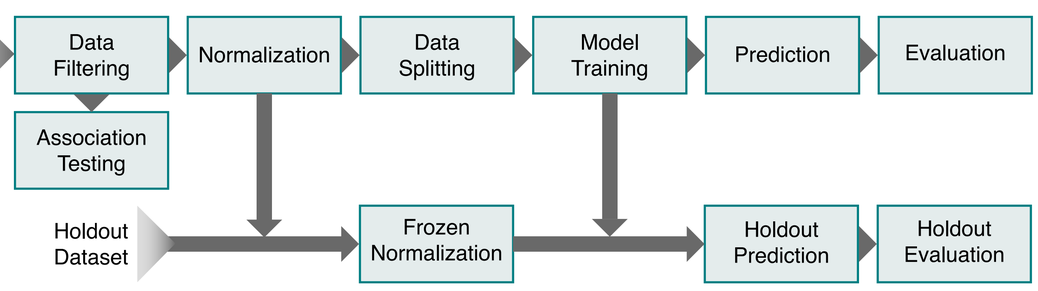
- Estimation of the **species abundance** in the metagenomic samples was performed using the **mOTU profiler**
- mOTUs enable quantification of microbial species **with (ref-mOTU) and without (meta-mOTU)** reference genomes using shotgun sequencing data

mOTUs
taxonomic meta-omics profiling using universal marker genes



Also check out the posters about the mOTU profiler and SIAMCAT!

SIAMCAT

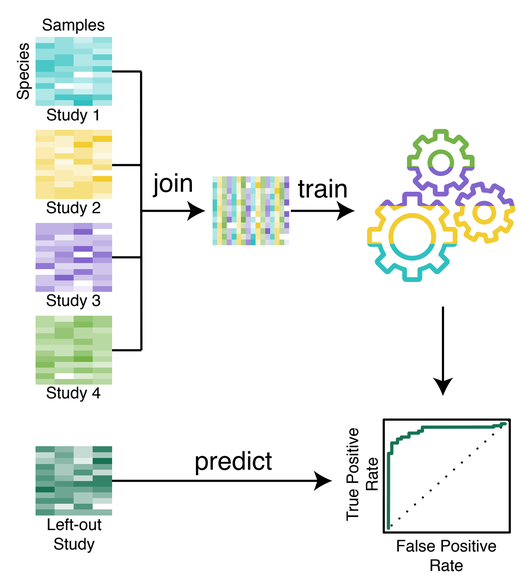


- Feature selection and model training via **LASSO** logistic regression [Tibshirani, *J R Stat Soc Series B*, 1996]
- **Avoids** a common **over-fitting** issue arising when feature selection and model training are naively combined
- **Interpretable models**, not black boxes
- Integrated into the **BioConductor environment**: Interoperability with other tools such as `mlr` and `phyloseq`

microbiome association testing and modelling
SIAMCAT

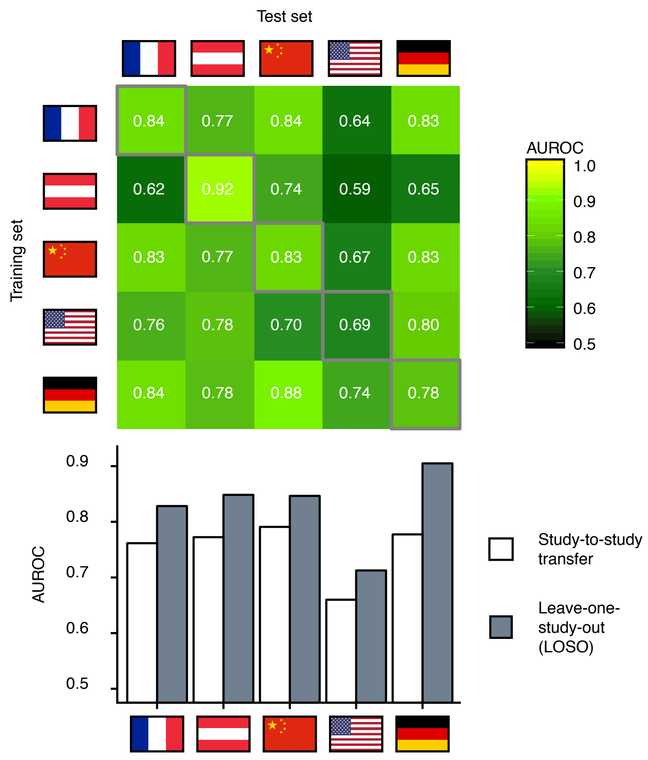
- `https://bioconductor.org/packages/SIAMCAT`

Leave-One-Study-Out (LOSO)

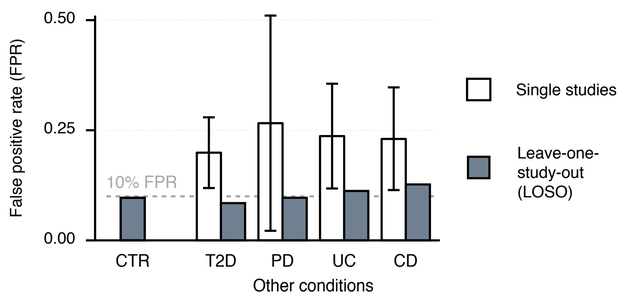


Results

Classifer Transfer



Statistical models retain high classification accuracy when applied to holdout datasets.
Pooling data in the LOSO setting improves classification results by 7.6 ± 3.0 percentage points.

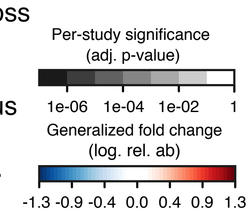


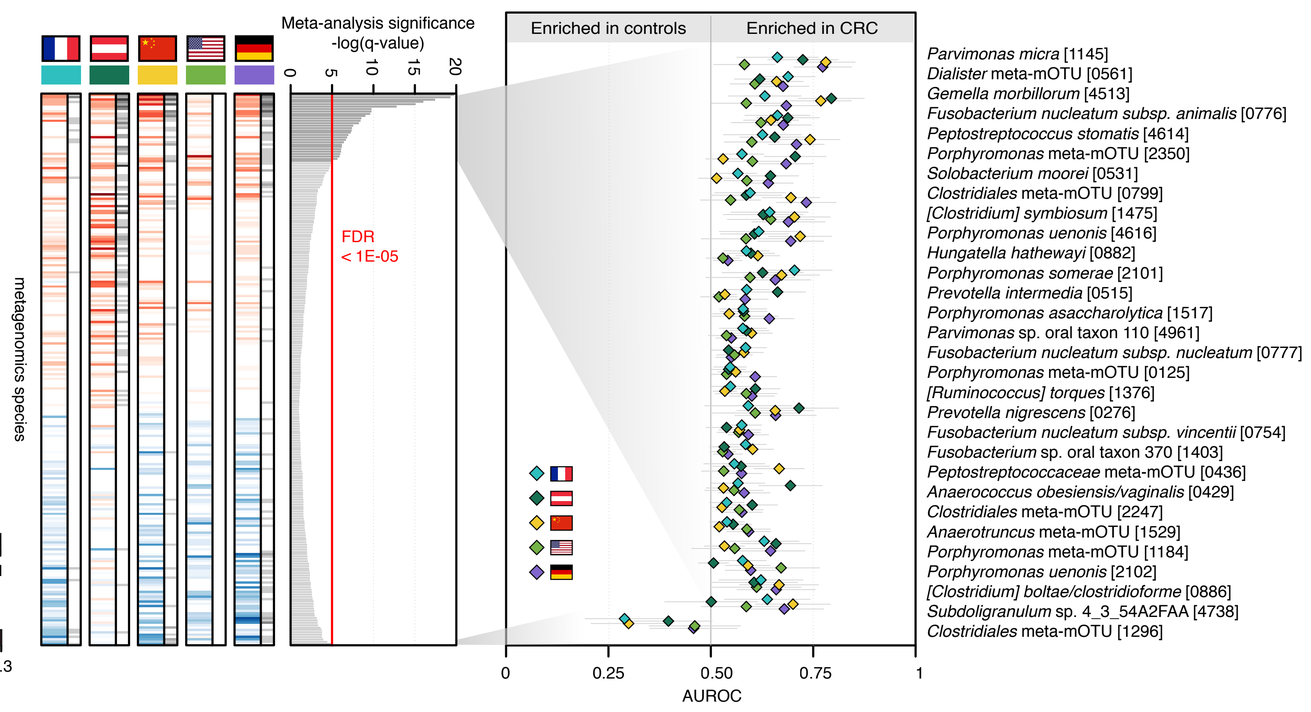
LOSO models show high disease specificity with false positive rates around 10% for patients with other diseases, significantly improving on models trained on single datasets.

Abbr.	Condition	N
CTR	Healthy controls from meta-analysis	290
T2D	Type 2 diabetes	201
PD	Parkinson's disease	31
UC	Ulcerative colitis	98
CD	Crohn's disease	63

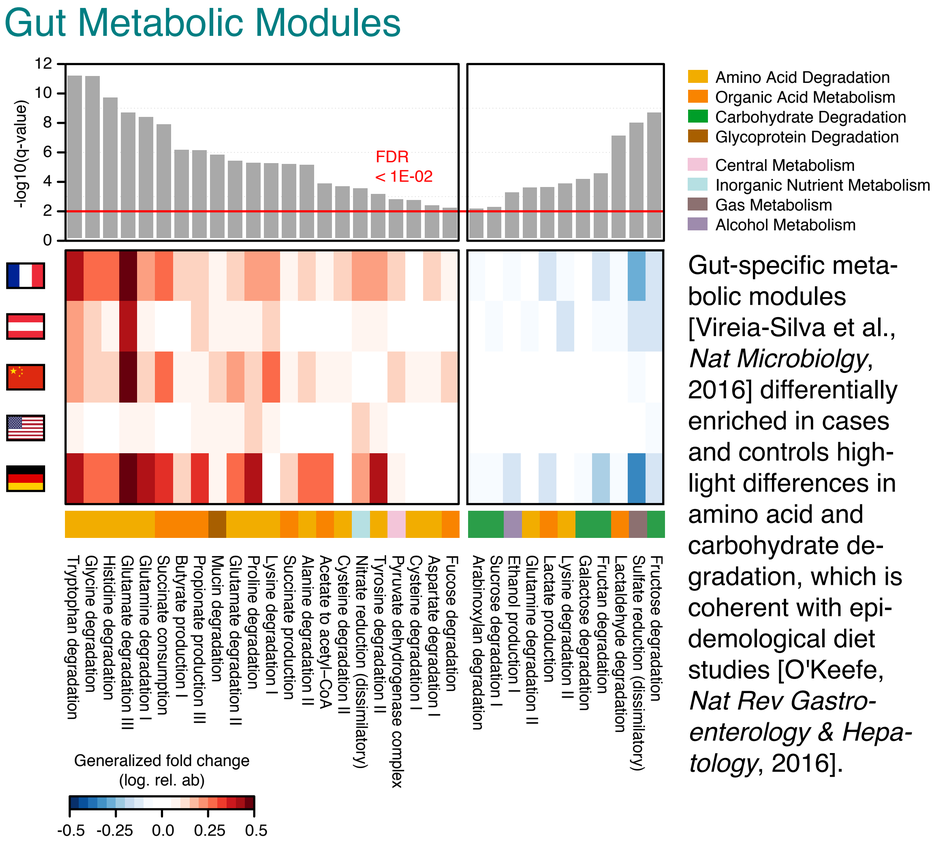
Univariate Testing

Although there is some variability in number and strength of associations across studies, there is an enriched core set of 30 species at a very stringent false discovery rate of 1E-05. Several of those are species without genomic references (meta-mOTUs).
Generally, positive associations with CRC are more consistent across studies, supporting an infectious agent hypothesis.





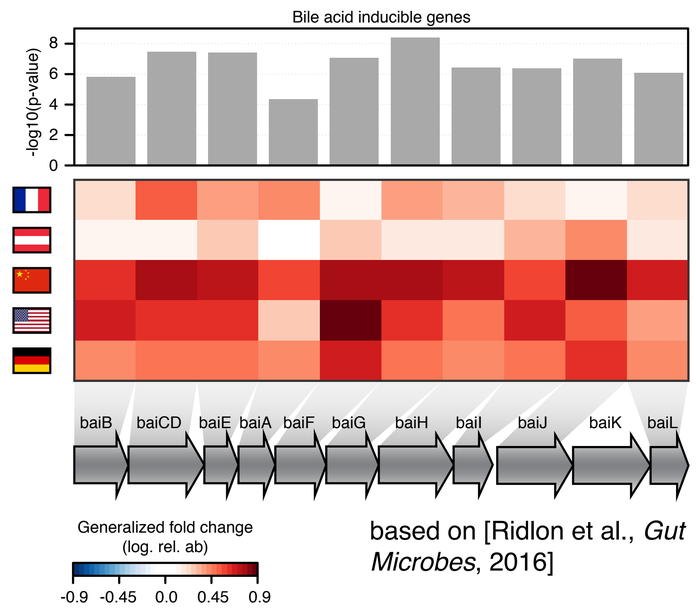
Gut Metabolic Modules



Gut-specific metabolic modules [Vireia-Silva et al., *Nat Microbiolgy*, 2016] differentially enriched in cases and controls highlight differences in amino acid and carbohydrate degradation, which is coherent with epidemological diet studies [O'Keefe, *Nat Rev Gastroenterology & Hepatology*, 2016].

Bile Acid Conversion

Conversion of primary bile acids by the gut microbiome has been hypothesized as potential contributor to CRC progression due to the DNA-damaging properties of secondary bile acids. Here, we show that bile acid converting genes are significantly enriched in CRC metagenomes.



based on [Ridlon et al., *Gut Microbes*, 2016]