Integrative 'omics of the infant stool microbiota during normal development

Anne G. Hoen[1,2], Erika Dade[2], Wimal Pathmasiri[3], Susan McRitchie[3], Juliette C. Madan[1,2,4], Susan Sumner[3], Margaret R. Karagas[1,2]

[1]Center for Molecular Epidemiology at Dartmouth, Lebanon, New Hampshire, USA; [2]Department of Epidemiology, Geisel School of Medicine at Dartmouth, Hanover, New Hampshire, USA; [3]NIH Eastern Regional Metabolomics Research Center at the University of North Carolina Nutrition Research Institute, Kannapolis, North Carolina, USA; [4]Division of Neonatology, Department of Pediatrics, Children's Hospital at Dartmouth, Lebanon, New Hampshire, USA

Introduction

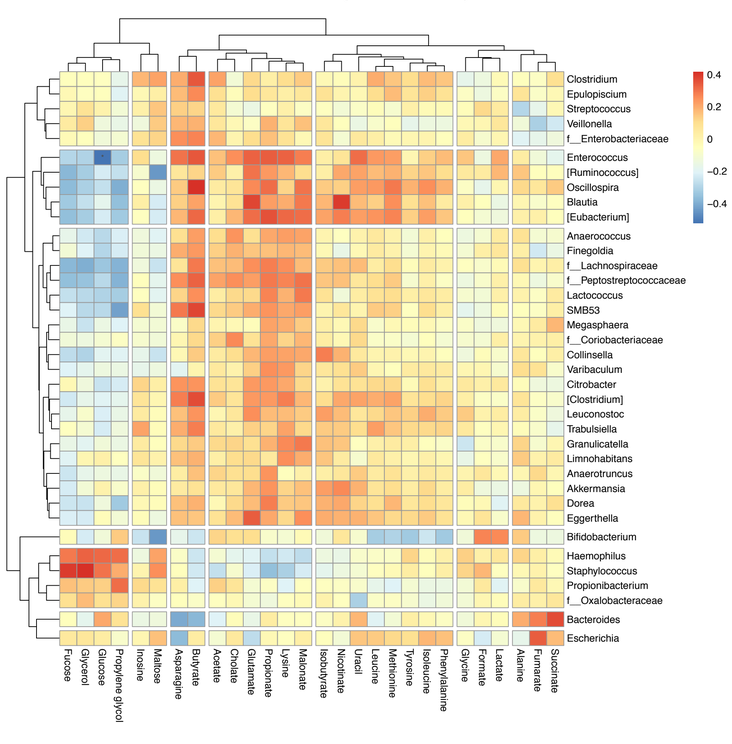
The intestinal microbiota plays a critical role in infant development, with important functions for nutrient metabolism and immune maturation. We performed both 16S rRNA gene sequence-based taxonomic profiling and ^{1}H NMR-based metabolomic profiling on 409 infant stool samples in order to identify relationships between taxonomic and functional patterns of infant gut microbiota development in the first year of life. In a subset of 132 samples, we also performed shotgun metagenomic sequencing to measure the concordance between metabolic pathways identified genomically and microbial community phenotype in the gut of normally developing infants.

Methods

Stool samples were collected from infants followed in the New Hampshire Birth Cohort Study, a large cohort of mother-infant dyads in New Hampshire, USA. We performed amplicon sequencing of the bacterial 16S v4v5 hypervariable region, ^{1}H NMR spectroscopy, and shotgun metagenomic sequencing on common sets of infant stool samples collected at either 6 weeks or 12 months of life. DADA2 v1.4 was used to infer the relative abundance of amplicon sequence variants (ASVs) in each sample and taxonomic identities were assigned using the Greengenes v13_8 reference database. Chenomx NMR Suite software was used semi-quantify selected microbial metabolite concentrations in ^{1}H NMR spectra. HUMAnN2 v0.11.1 was used to map metagenomic sequences to metabolic pathways and compute relative pathway abundances. Spearman correlations were computed for each ASV-metabolite pair and for each pathway-metabolite pair; rows or columns were omitted if all $q > 0.1$. GUniFrac was used to compute pairwise distance (alpha=0.5) matrices for pairs of samples within the same collection time point according to ASV profiles. Euclidian distances were computed for pairs of samples within the same collection time point according to the principal components of metabolomic and pathway abundance profiles where eigenvalues > 1 (capped at 30). Associations between pairwise distance matrices were evaluated using the Mantel test based on Pearson's product-moment correlation.

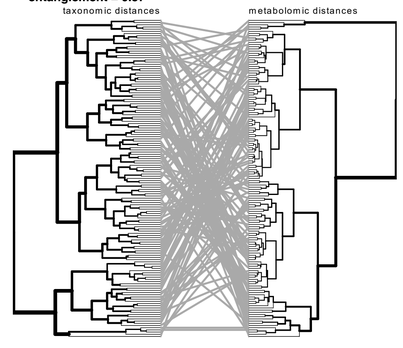
Results: Taxonomic vs. metabolomic profiles

6 weeks (N=145)

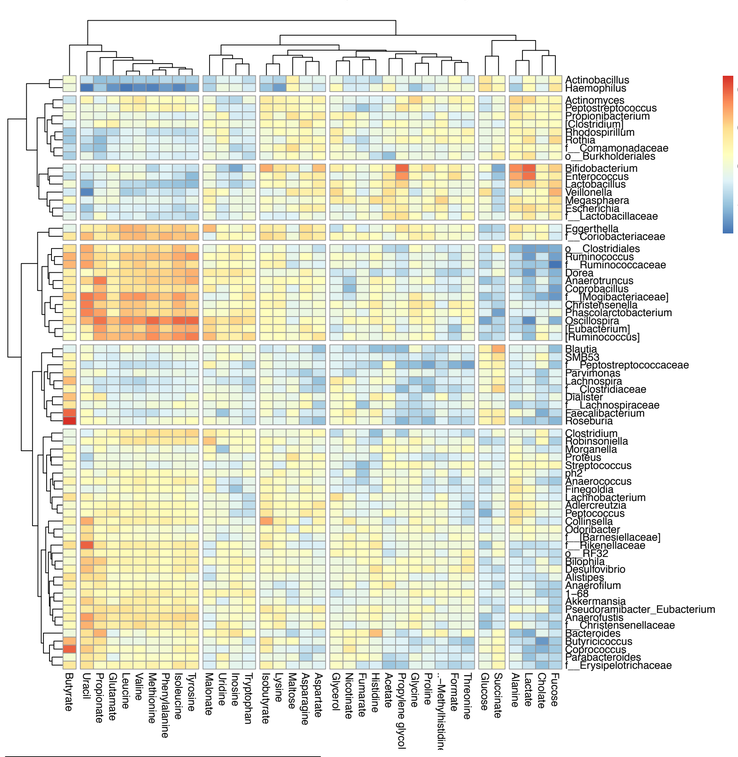


Taxon	Metabolite	R
Oscillospira	Butyrate	0.42
Staphylococcus	Glycerol	0.42
Staphylococcus	Fucose	0.39
Blautia	Nicotinate	0.39
Blautia	Glutamate	0.37
Bacteroides	Asparagine	-0.40
SMB53	Propylene glycol	-0.42
[Ruminococcus]	Maltose	-0.44
Bifidobacterium	Maltose	-0.45
Enterococcus	Glucose	-0.52

entanglement = 0.57

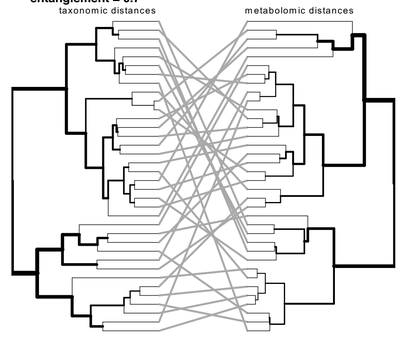


12 months (N=264)



Taxon	Metabolite	R
Roseburia	Butyrate	0.47
f__Rikenellaceae	Uracil	0.40
Coprococcus	Butyrate	0.40
Bifidobacterium	Lactate	0.39
Oscillospira	Tyrosine	0.39
Veillonella	Uracil	-0.32
Haemophilus	Methionine	-0.33
Haemophilus	Uracil	-0.34
Haemophilus	Leucine	-0.34
f__Ruminococcaceae	Fucose	-0.35

entanglement = 0.7



Heatmaps of correlations between hierarchically clustered taxonomic relative abundances vs. metabolite relative concentrations reveal taxa with similar metabolic profiles in the infant stool at both 6 weeks (left) and 12 months (right). The top 5 positive and negative correlations are shown in the tables below the corresponding heatmap on the left. We computed pairwise taxonomic (UniFrac) and metabolomic (Euclidean) distance matrices for each time point, displayed them as dendrograms and connected samples with lines to visualize relationships between distance matrices. Resulting tanglegrams are shown below the corresponding heatmap on the right. Finally, a Mantel test compared pairwise taxonomic (UniFrac) and metabolomic (Euclidean) distance matrices for each time point to test for a possible trend in pairs of samples with similar taxonomic profiles having similar metabolic profiles.

Results of Mantel test of correlation between taxonomic and metabolomic distance matrices

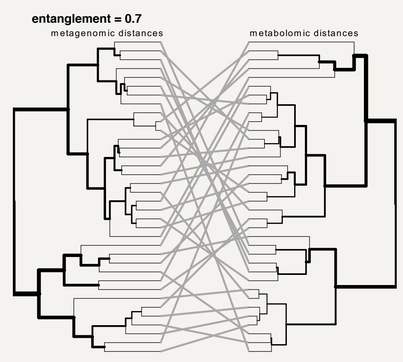
Time point	Mantel statistic *r*	*N*	p-value
6 weeks	0.15	145	0.0001
12 months	0.04	264	0.09

Results: Metagenomic vs. metabolomic profiles

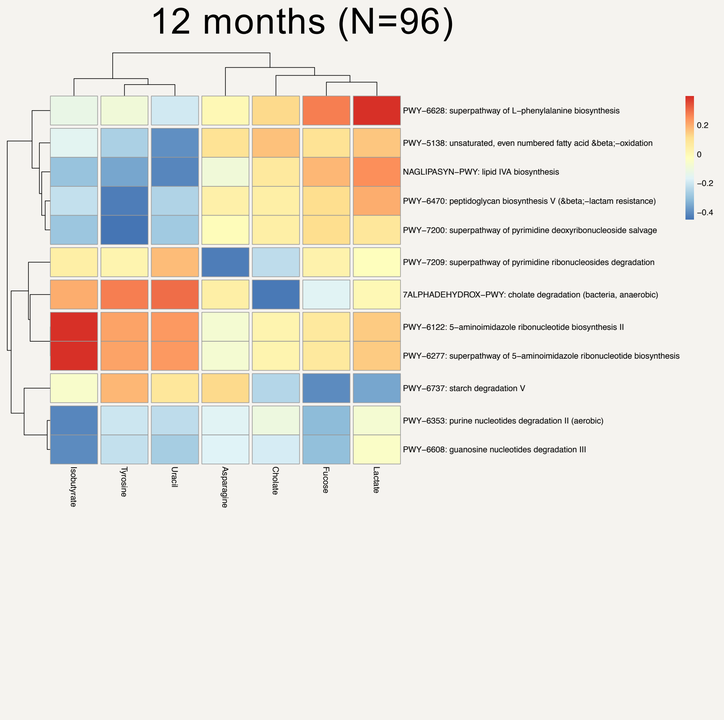
6 weeks (N=36)

Pathway	Metabolite	R
PWY-7392: taxadiene biosynthesis (engineered)	Propionate	0.71
PWY-5104: L-isoleucine biosynthesis IV	Lysine	0.71
PWY4LZ-257: superpathway of fermentation (Chlamydomonas reinhardtii)	Lysine	0.71
PWY-7392: taxadiene biosynthesis (engineered)	Acetate	0.68
PWY-6895: superpathway of thiamin diphosphate biosynthesis II	Fumarate	0.67
PWY-6121: 5-aminoimidazole ribonucleotide biosynthesis I	Fumarate	-0.59
PWY-5104: L-isoleucine biosynthesis IV	Glucose	-0.60
PENTOSE-P-PWY: pentose phosphate pathway	Threonine	-0.62
P125-PWY: superpathway of (R,R)-butanediol biosynthesis	Propylene glycol	-0.63
PWY-5384: sucrose degradation IV (sucrose phosphorylase)	Lactate	-0.65

entanglement = 0.7

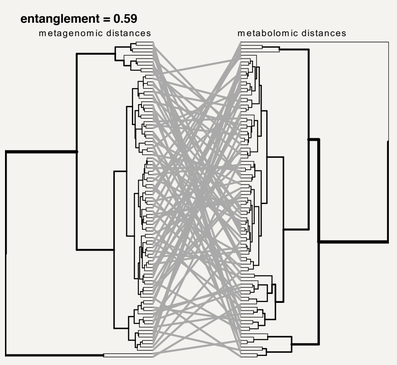


12 months (N=96)



Pathway	Metabolite	R
PWY-6122: 5-aminoimidazole ribonucleotide biosynthesis II	Isobutyrate	0.40
PWY-6277: superpathway of 5-aminoimidazole ribonucleotide biosynthesis	Isobutyrate	0.40
PWY-6628: superpathway of L-phenylalanine biosynthesis	Lactate	0.39
7ALPHADEHYDROX-PWY: cholate degradation (bacteria, anaerobic)	Uracil	0.30
PWY-6628: superpathway of L-phenylalanine biosynthesis	Fucose	0.28
NAGLIPASYN-PWY: lipid IVA biosynthesis	Uracil	-0.41
PWY-6470: peptidoglycan biosynthesis V (β-lactam resistance)	Tyrosine	-0.43
PWY-7209: superpathway of pyrimidine ribonucleosides degradation	Asparagine	-0.43
7ALPHADEHYDROX-PWY: cholate degradation (bacteria, anaerobic)	Cholate	-0.44
PWY-7200: superpathway of pyrimidine deoxyribonucleoside salvage	Tyrosine	-0.45

entanglement = 0.59



Associations between profiles of pathway abundance derived from metagenomic sequencing and metabolomic profiles were examined using an approach similar to that described above. Euclidean distances were used in the calculation of distance matrices for both metagenomic distance and metabolomic distance.

Results of Mantel test of correlation between metagenomic pathway abundance distance and metabolomic distance matrices

Time point	Mantel statistic *r*	*N*	p-value
6 weeks	0.05	36	0.31
12 months	0.06	96	0.22

Conclusions

The results of our integrative 'omics analysis are a descriptive visualization and quantification of relationships between 16S-based taxonomic relative abundance, metagenomic pathway abundance and microbial metabolite concentrations. Hierarchical clustering of ASVs and pathways according to their profiles of correlations with each microbial metabolite resulted in the empirical identification of groups of bacteria with metabolic similarities in the normally developing infant gut. We also report on associations between metabolomics profile-based pairwise "metabolic distance" and taxonomic and genomic distances. Our findings shed light on links between gut bacterial community taxonomic composition, collective functional potential and phenotype in an infant cohort.

Acknowledgements

We gratefully acknowledge the children and families who made this study possible and the staff of the New Hampshire Birth Cohort Study. This work was supported by grants from the National Institutes of Health (NIEHS P01ES022832, NIEHS P20ES018175, NIGMS P20GM104416, NLM R01LM012723, NLM K01LM011985) and the US Environmental Protection Agency (RD83459901 and RD83544201).