Title
Precocious Infant Gut Microbiome Accelerates Triglyceride Absorption and Promotes Childhood Obesity Risk

Author
Yong, Germaine Jia Min

Publication Date
2020

Precocious Infant Gut Microbiome Accelerates Triglyceride Absorption and Promotes Childhood Obesity Risk

by
Germaine Jia Min Yong

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Biomedical Sciences

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Jacquelyn Maher
7F58589DDF43483...

Jacquelyn Maher
Chair

DocuSigned by:
Kaveh Ashrafi

Kaveh Ashrafi

DocuSigned by:
Susan Lynch
1A38FCA893D34C5...

Susan Lynch

Committee Members

Dedication and Acknowledgments

It takes a village to get through graduate school.

From the bottom of my piggly heart,

Sue, for being a professional/personal role model and pillar of support

Kaveh and Jackie, for your thoughtful insight and advice

Grad students, for entertaining all wild ideas and scrumptious home-cooked meals

Kei, for teaching me the ropes and endless conversations on everything under the sun

Lynch lab, for being with me every step of the way

Daddy, for your unconditional love and interest in everything I do

Mummy, for *always* rooting for me even if science makes you fall asleep

Koko, for your boundless generosity since the day I insisted we share birthdays

Nugget, for your unconditional love and living in a converted garage with me

Glen, for being the inspiration and *source* (read poop) of my daily work

XH, for supporting me in all that I do and being my NB

Thanks ah.

Precocious Infant Gut Microbiome Accelerates Triglyceride Absorption and Promotes Childhood Obesity Risk

Germaine Jia Min Yong

Abstract

The composition and metabolic activity of the gut microbiome are implicated in childhood obesity, though etiology of the disease remains unclear. Using 16S bacterial and ITS2 fungal biomarker sequencing, we identify a precocious gut microbiota in more frequently formula fed 1-month-old infants that relates with higher relative risk for overweight or obesity (OW/OB) phenotypes at two years of age. Integrated shotgun metagenomic and untargeted metabolomic profiling revealed that higher-risk infant microbiomes exhibited accelerated functional maturation and broad-ranging metabolic reprogramming, including evidence for increased fermentation and amino acid catabolism. *In vitro*, exposure of enterocytes to cell-free fecal extracts of higher-risk infants who became OW/OB at 2-years reprogrammed transcription and cellular function, promoting obesity-associated gene expression and increasing triglyceride absorption. Of several microbial species screened, *Candida albicans* recapitulated accelerated triglyceride trafficking by enterocytes, via a small, secreted polar molecule in an infant formula-dependent manner. Thus, appropriately paced early life development of microbiome function and metabolism appear crucial determinants of childhood obesity.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

BH: Benjamini-Hochberg

BMI: Body Mass Index

DMM: Dirichlet Multinomial Mixture

FDR: False Discovery Rate

GABA: Gamma-aminobutyrate Acid

GMC: Gut Microbiota Class

MAZ: Microbiota-for-age Z-score

OW/OB: Overweight or Obese

RR: Relative risk

Introduction

One in three children in the United States experience overweight or obesity [defined as body mass index (BMI) ≥85th percentile for age and gender] and as a result are at increased risk of co-morbidities including heart and fatty liver disease, diabetes and asthma[1]. In adults, gut microbiome perturbation is characteristic of patients with obesity[2–5] and its functional contribution to disease pathology has been confirmed in humanized gnotobiotic mouse models[6,7]. Notably, the gut microbiome has been demonstrated to increase energy harvest[6], alter neural[8] and metabolic signaling[9,10] and control inflammation[11].

Early life environmental exposures alter the gut microbiome and host development, and are related to subsequent adiposity phenotypes in childhood. In humans, early life microbiota disruption, either due to delivery by Caesarian[12] or antibiotics, is associated with increased risk of subsequent overweight phenotypes in childhood[13,14]. Diet plays a significant role in shaping the gut microbiome[15], which in turn influences the metabolic fate of ingested nutritional substrates and the production of bioactive molecules that influence host physiology[16–19]. Of note, formula feeding in infancy has been associated with obesity in childhood[20], though the mechanism(s) by which it contributes to disease remain unclear. Breastfeeding, which protects against the development of overweight and obesity phenotypes in childhood, is a primary factor that shapes microbiome development in infancy[21].

Early life gut microbiome development progresses along a temporal gradient[21–23] and is often influenced by priority effects, in which the order and timing of species arrival

determine inter-species relationships[24]. Alterations to the composition and rate of microbial accumulation during this critical developmental window are related to risk of childhood diseases, including impaired growth[25,26], Kwashiorkor (protein malnutrition)[27] and asthma[28,29]. Although several studies have correlated early life gut microbiota perturbation with subsequent adiposity[26,30–34], none have provided direct evidence for a functional role of the infant gut microbiome in the development of childhood obesity, or identified factors that promote obesogenic features of infant microbiomes.

In an attempt to understand the origins of childhood obesity, we focus on very-early infancy and identify a compositionally and functionally precocious gut microbiome associated with increased rates of formula feeding and increased risk for overweight or obese phenotypes (OW/OB) in early childhood. We provide *in vitro* evidence that the products of these higher-risk infant gut microbiomes transcriptionally reprogram enterocytes and enhance triglyceride uptake, offering a plausible mechanism by which interactions between early life nutrition and the infant gut microbiome promotes obesogenic phenotypes in childhood.

Results

Distinct infant gut microbiota associates with OW/OB at age two years

Fecal samples (n=349) collected in early postnatal life (median age 35 days; range 21-58 days) were subjected to parallel 16S rRNA and ITS2 sequencing. Dirichlet Multinomial Mixture (DMM) modeling, which implements an unsupervised Bayesian approach to class discovery[35], was applied to the 16S rRNA dataset to sub-group participants based on fecal bacterial community composition. Three distinct gut microbiota classes (GMC1, GMC2 and GMC3) represented the best model fit (**Fig. 1a, Extended Fig. 1a**; PERMANOVA; R^2=0.11, p=1e-4).

Distinct early life GMCs significantly related to BMI and OW/OB outcomes at age two years; BMI Z-scores of GMC3 infants were significantly greater than either GMC1 (adjusted β=0.47; 95% CI 0.15-0.80; p=0.005; **Fig. 1b** and **Supplementary Table 2B**) or GMC2 groups (adjusted β=0.38; 95% CI 0.05-0.71; p=0.023; **Fig. 1b** and **Supplementary Table 2B**). Relative risk (RR) for OW/OB at age two was also significantly greater in GMC3 compared with GMC1 subjects (unadjusted RR=2.34; 95% CI 1.3-4.21; p=0.005; **Supplementary Table 2A**), despite no significant difference in age at stool sample collection in GMC3 compared to GMC1 infants (p=0.176; **Extended Fig. 1b**). GMC3 infants were more likely to have parents with a shorter duration of formal education and to have been formula fed, while GMC1 had the greatest proportion of exclusively breastfed infants (**Supplementary Table 1**). The relationship between GMC3 and OW/OB phenotypes remained significant and was strengthened following adjustment for these and additional confounding factors

(**Supplementary Table 2A**), regardless of infant feeding status (interaction p=0.80). In a multivariable model adjusting for exact age at stool sample collection, prenatal antibiotic and antifungal use, mode of delivery, and breastfeeding status at 1-month, GMC3 subjects had 2.63 times higher risk of experiencing OW/OB than GMC1 subjects (95% CI 1.45-4.76; p=0.002; **Table 1**). DMM modeling applied to older infant samples (median age 206 days; range 174-238 days; n=287) identified four additional fecal microbiota structures (GMC4-7; Unweighted UniFrac; PERMANOVA, R^2=0.15, p=0.001; **Extended Fig. 1e-g**), however these were not significantly related to OW/OB or BMI Z-score at age two years (**Supplementary Table 3**) indicating the importance of the very-early life period for risk of disease development in childhood.

Infants at higher risk for OW/OB exhibit precocious gut microbial development

Despite exhibiting increased bacterial species richness (p<2e-16, **Extended Fig. 1c**) and phylogenetic diversity (Faith's, p<2e-16, **Extended Fig. 1d**), GMC3 microbiota were depleted of key genera (e.g. *Bifidobacterium, Clostridium* and *Malassezia*) typical of this stage of postnatal development[22,36] (**Fig. 1d**-**e**, **Extended Fig. 1h-i**, **Supplementary Table 4-5**). Instead, GMC3 microbiota exhibited relative enrichment of *Lachnospiraceae*, *Ruminococcaceae, Saccharomyces* and *Candida,* genera typically associated with later stages of infant gut microbiome development[22,36] (**Fig. 1d**-**e**, **Extended Fig. 1h-i**, **Supplementary Table 4-5**) and, in human adult populations, with greater visceral fat mass[4,37,38] and insulin resistance[4]. These microbial enrichments were consistent irrespective of the lower-risk comparison group used (**Supplementary Table 6-7**) and included top age-discriminatory bacterial taxa (**Extended Fig. 1j-k**) associated with more mature infant gut microbiomes[22,39]. Consistent with this

observation, random forest predictive modeling revealed that GMC3 infants exhibited older microbiota-for-age z-scores (MAZ) ($p<2e-16$, **Fig. 1c**), indicating precocious microbial development in these infants at higher-risk for OW/OB in early childhood.

Enhanced fermentation and protein metabolism distinguish infants at higher risk of OW/OB status

Parallel shotgun metagenomic sequencing and untargeted mass spectrometry were performed on a microbiologically representative subset of infant feces based on posterior probability of GMC membership, BMI classification and feeding practices (n=60; **Supplementary Table 8**). Integrated comparative analyses indicated distinct substrate preferences for energy biogenesis between lower- and higher-risk GMCs (**Fig. 2a**). GMC3 microbiomes were enriched for gene pathways involved in degradation of glucarate, galactarate and myo-inositol (P_{FDR}=0.12 for all; **Extended Fig. 2a-b** and **Supplementary Table 9**), generating products that feed into central carbon metabolism (**Fig. 2b**), and for fermentation of pyruvate and glycerol to 1-butanol (P_{FDR}=0.12 for both; **Extended Fig. 2c-d** and **Supplementary Table 9**), the latter previously associated with soy formula feeding[16] and hepatic steatosis[40]. In contrast, GMC1 metagenomes were enriched for methylglyoxal degradation (P_{FDR}=0.12; **Extended Fig. 2e**), a detoxification pathway that produces lactate and pyruvate (**Fig. 2b**), both of which were detected in significantly elevated concentrations by parallel mass spectrometry in feces of these infants (P_{FDR}=0.047 and P_{FDR}=0.029 respectively; **Extended Fig. 3b-c**). GMC1 also exhibited enhanced energy biogenesis via glycol and ketogluconate degradation (P_{FDR}=0.13; **Extended Fig. 2f** and P_{FDR}=0.04; **Extended Fig. 2g**), indicating that lower-risk infant gut microbiomes exhibit alternative microbial strategies for energy harvest.

Consistent with more GMC1 infants receiving any breastmilk (89.3%) compared to GMC3 infants (65.6%; **Supplementary Table 8**), GMC1 fecal metabolomes exhibited significantly elevated concentrations of human milk oligosaccharides (**Extended Fig. 3a** and **Supplementary Table 10**). GMC1 was also enriched in a large range of distinct lipids, including fatty acid esters of hydroxy fatty acids (FAHFAs) and acylcarnitines (**Extended Fig. 3a** and **Supplementary Table 10**), the latter indicative of enhanced capacity for mitochondrial fatty acid β-oxidation. Metabolic profiling also indicated differences in oxidizing potential between GMC1 and GMC3 feces. While the former exhibited relatively higher concentrations of the antioxidant bilirubin (P_{FDR}=0.021, **Fig. 2c**), the latter was enriched for the microbially reduced heme catabolism product urobilinogen (P_{FDR}=0.011, **Fig. 2d**), which is associated with obesity[41] and older chronological age[42].

GMC3 microbiomes were also depleted of amino acid and vitamin production pathways (**Fig. 2a**, **Extended Fig. 2h-k**) as well as allantoin degradation (P_{FDR}=0.12; **Fig. 2e, Extended Fig. 2l**), which recycles nitrogen necessary for their synthesis. Notably, the amino acid neurotransmitter gamma-aminobutyrate acid (GABA), which regulates neuro-endocrine body weight control[43,44], obesity-induced inflammation and insulin sensitivity[45], was depleted in the feces of higher-risk GMC3 infants (P_{FDR}=0.041; **Fig. 2f**). Despite microbial deficiencies in these synthesis pathways, mass spectrometry analyses indicated significantly increased concentrations of a number of vitamins and amino acid derivatives in GMC3 infant feces (**Extended Fig. 3a** and **Supplementary Table 10**), likely attributable to a dietary surplus in these infants who were more likely formula fed. While GMC1 and GMC3 metagenomes exhibited few differences in protein

degradation capacity, mass spectrometry revealed that GMC3 feces were highly enriched for a range of catabolic amino acid by-products (**Extended Fig. 3a**), including phenylacetate (P_{FDR}=0.002, **Fig. 2g**) that triggers hepatic steatosis[10], and modulators of GABAergic signaling, such as the GABA precursor glutamine (P_{FDR}=0.006; **Extended Fig. 3d**) and weak GABA agonist 5-aminovalerate (P_{FDR}=0.024; **Extended Fig. 3e**), implicating altered enteroendocrine signaling in higher-risk infants.

Fecal extracts from higher-risk infants reprogram enterocyte physiology and accelerate triglyceride absorption

To determine the functional consequences of altered gut microbiome productivity in infants at higher- versus lower-risk for OW/OB, enterocyte transcriptional and physiological (triglyceride uptake) responses to cell-free GMC1 and GMC3 fecal products (n=17 with sufficient remaining material) were assessed in the presence of oleic acid, the most abundant unsaturated fatty acid found in human and formula milk. Variance in enterocyte transcription associated significantly with GMC (PERMANOVA R^2=0.18, p=0.005; **Extended Fig. 4a**) and OW/OB status (PERMANOVA R^2=0.12, p=0.046; **Extended Fig. 4b**), and was greatest between GMC3 OW/OB (n=7) and GMC1 Normal BMI (n=4) samples (**Extended Fig. 4c**). This latter comparison produced starkly different transcriptional programs (PERMANOVA R^2=0.33, p=0.003; **Fig. 3a**) including transcripts involved in cell proliferation (*LGALS1*, *WNT1*), organization (*BFSP1*, *LAMA1*), and barrier function (*GJB7*, *CLDN2*; **Fig. 3b-c** and **Supplementary Table 11**). GMC3 OW/OB fecal extracts increased enterocyte expression of genes regulating inflammation (*ALOX5*, *TNFSF9*, *CCL20*, *CCL22*; **Fig. 3c**), long-chain fatty acid transport (*SLC27A1*, *ACSL1*, *ACSL3*, *ACSL4*; **Fig. 3c** and **Extended Fig. 4d**) and

those regulating vesicle maturation, transport and exocytosis (*KIF5C*, *LGI3*, *TMEM59L*; **Fig. 3b**). In parallel, downregulation of genes involved in fatty acid oxidation (**Extended Fig. 4d**) and mitochondrial electron transport chain genes involved in oxidative phosphorylation (**Extended Fig. 4e**) were also observed, suggesting that products of the higher-risk for obesity infant gut microbiome induce alterations in lipid trafficking and metabolism by mammalian cells. Gene set enrichment analysis further identified genes associated with transit amplifying cells and enteroendocrine progenitors as enriched in enterocytes treated with GMC3 OW/OB extracts (**Fig. 3d**), the latter consistent with our earlier observation that GMC3 microbiomes produce metabolic products with the potential to alter enteroendocrine function. Genes associated with mucin-producing goblet cells were enriched in enterocytes treated with fecal extracts from GMC1 infants who had normal BMIs at age two (**Fig. 3d**), indicating that gastrointestinal mucus, a critical physical and biochemical barrier, may also play a protective role against childhood obesity development.

To confirm that GMC3 OW/OB-induced transcriptional reprogramming of enterocytes produced functional changes in lipid handling, cellular triglyceride accumulation was assessed. Consistent with gene expression data, fecal products from GMC3 OW/OB infants enhanced triglyceride accumulation in enterocytes compared with GMC1 infants with normal BMIs at age two (n=11; p=0.033; **Fig. 3e**).

GMC3 Candida albicans isolates accelerate triglyceride trafficking in a nutrition-dependent manner

Our data indicated that microbial members and metabolites of the higher-risk for obesity infant gut microbiome influence lipid accumulation by enterocytes. In an attempt to identify specific gut microbes capable of influencing mammalian triglyceride trafficking, members of bacterial and fungal genera enriched in GMC3 infants were isolated from fecal samples in this study or obtained from lab and commercial culture collections. Amongst all strains tested *in vitro* (22 strains from 12 genera), cell-free supernatant from multiple strains of *Candida albicans* isolated from GMC3 infants promoted the greatest egress of accumulated triglycerides from enterocytes (**Fig. 4a** and **Extended Fig. 5a**). This phenotype was consistent across *C. albicans* isolates irrespective of oxygen availability (**Fig. 4a**)**.** Fractionation of cell-free *C. albicans* supernatant identified the <3kDa polar fraction as capable of recapitulating this phenotype (**Fig. 4a** and **Extended Fig. 5b**), implicating a small polar, secreted fungal molecule. Addition of the <3kDa fraction to cell-free fecal products from low-risk GMC1 infants was sufficient to promote triglyceride egress from enterocytes (**Fig. 4b**), confirming its capacity to regulate triglyceride handling by mammalian cells.

Compared with GMC1 infants who developed normal BMIs at age 2 years, *Candida* was significantly enriched in GMC3 infants who developed OW/OB status (**Extended Fig. 5c**), a significantly larger proportion of whom were formula fed. Since substrate availability is known to influence *Candida* physiology and virulence[46], we examined whether infant formula enhanced *C. albicans* growth and capacity to influence enterocyte triglyceride trafficking. *C. albicans* DN2, which promoted the greatest

triglyceride egress (**Fig. 4a**), exhibited rapid growth on either bovine- or soy-based infant formulas tested (n=3), but was inhibited by human breast milk (n=3 independent donors; **Fig. 4c** and **Extended Fig. 5d-f**). Morphologically, *C. albicans* DN2 cultured on infant formula developed floc-like structures comprised of yeast and pseudohyphal cells co-associated with extracellular DNA scaffolds (**Fig. 4d** and **Extended Fig. 5g**). In contrast, *C. albicans* DN2 cultured on breast milk exhibited a primarily yeast-like morphology, and neither formed flocs nor evidenced DNA scaffolding (**Fig. 4d** and **Extended Fig. 5g**). Cell-free supernatant from formula-grown *C. albicans* DN2 promoted triglyceride egress from enterocytes ($p=0.002$, **Fig. 4e**), while that from breastmilk cultures did not. This phenotype was cell-density dependent, since normalization of formula-grown *C. albicans* DN2 cell numbers to that observed in breast milk markedly reduced triglyceride egress by the former (**Extended Fig. 5h**).

Discussion

Early life gut microbiome development is the product of evolutionary selection and dynamic functional synergy in the context of environmental exposures. Emerging evidence indicates that alterations to the rate of gut microbiome development disrupts the chronological synchrony of host-microbe interactions, impacting microbiome assembly, productivity and early life developmental programming resulting in childhood disease[25,27–29]. Our findings indicate that accelerated gut microbiome development in early infancy increases risk of children experiencing OW/OB phenotypes at age two years. Several of the microbial pathways and metabolites enriched in higher-risk GMC3 infants have previously been correlated with formula feeding[16,47] and older chronological age[21,22]. Thus, a perturbed, functionally precocious early-life gut microbiome detected in infants more likely to be formula fed increases risk of overweight and obese phenotypes in later childhood.

Early-life nutrition plays a substantial role in shaping infant gut microbiome functional capacity and metabolic productivity[16,21,22]. Human milk is a complex, dynamic functional food, the composition of which differs across mothers and changes with length of lactation[48]. Bioactives in breast milk, such as lactoferrin and immunoglobulin-A, regulate early-life microbial colonization in the gastrointestinal tract[49,50]. Nutritional substrates, including human milk oligosaccharides, promote selection of co-evolved microbes capable of metabolizing human milk[51] and fulfilling dynamic amino acid and vitamin requirements throughout growth in infancy. In contrast, infant formula primarily provides nutritional substrates, thus failing to exert the same chronological selective pressures as breast milk on the developing microbiome.

Broad functional changes both in gene content and metabolic productivity of higher- and lower-risk for obesity infant microbiomes were evident. Of note were the differences in microbial strategies for nutrient utilization and energy biogenesis. Nutrient excess is known to promote mitochondrial dysfunction, leading to obesity-related pathologies[52]. Thus, the observed alternative microbial capacity for nutrient utilization, associated with lower-risk for obesity gut microbiomes and increased breastfeeding, may serve to reduce such nutrient excess to confer protection against childhood obesity development.

Cell-free products of higher-risk for obesity infant microbiomes induced a distinct program of enterocyte gene expression *in vitro* that was graduated by OW/OB outcomes, raising the possibility that the infant gut microbiome calibrates metabolic health in infants who subsequently experience OW/OB phenotypes. Notably, enterocytes exposed to cell-free products of higher-risk infants with subsequent OW/OB status were characterized by reduced barrier function and fatty acid oxidation, in parallel with increased expression of genes involved in inflammation and lipid handling – all previously described features of obesity[53–56]. At the cellular level, obesity is characterized by cytosolic lipid overload across multiple cell types[54,57]. Functional confirmation of accelerated triglyceride accumulation in enterocytes exposed to cell-free products of higher-risk infant microbiomes indicates that early-life microbes and their products regulate lipid loading in enterocytes. Of the strains examined, enterocyte triglyceride handling was most accelerated by *C. albicans* in an infant formula-dependent manner, underscoring the importance of early life nutrition in governing both microbial cell physiology and inter-kingdom signaling. These findings offer mechanistic

insights into the origins of childhood obesity and indicate that the infant gut microbiome plays a key role in regulating triglyceride trafficking and the risk of childhood overweight and obesity at age two years. Thus, early-life approaches to foster appropriately paced microbiome development and metabolism may help prevent childhood obesity.

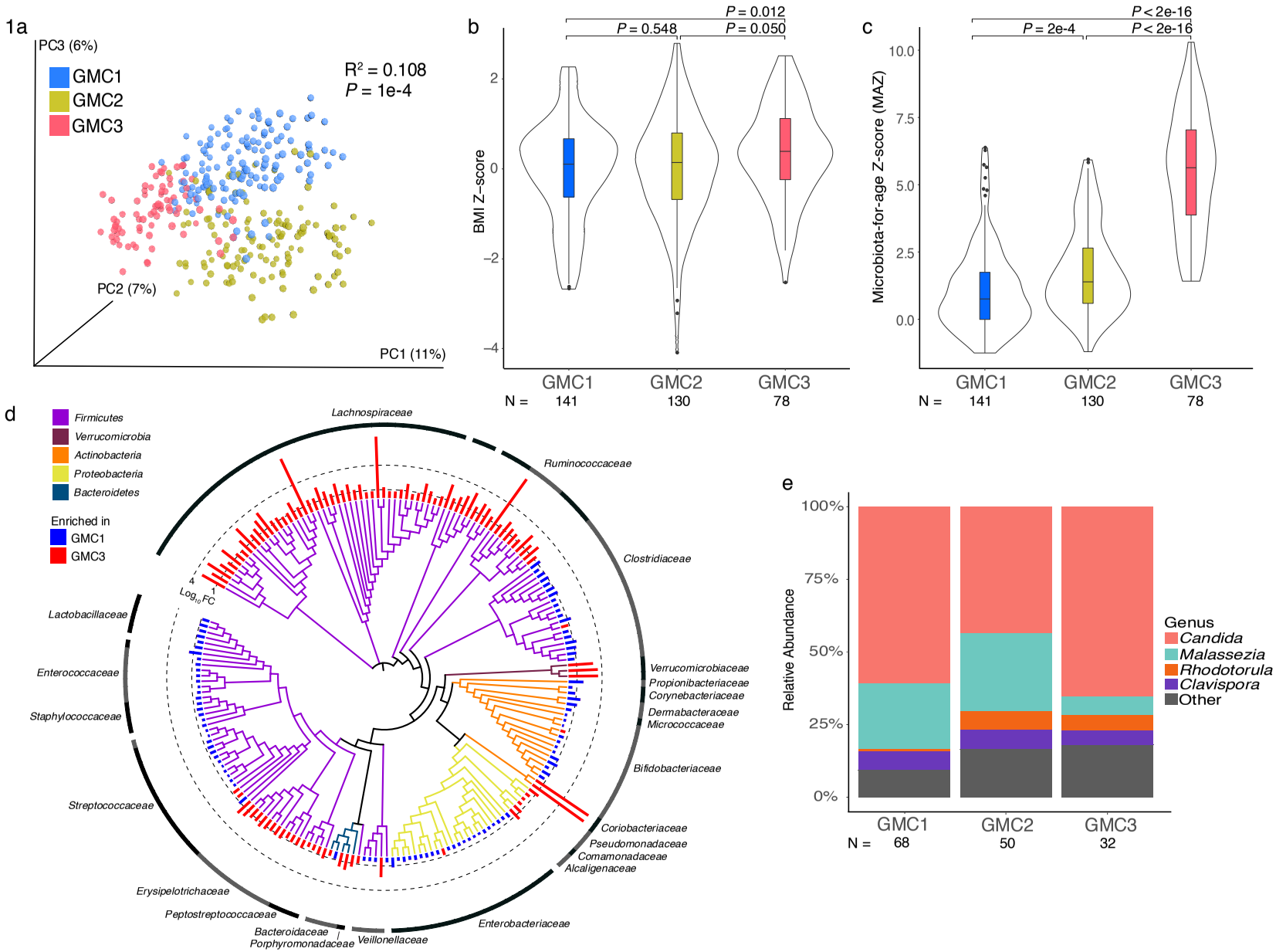

1a
PC3 (6%)
GMC1
GMC2
GMC3
R² = 0.108
P = 1e-4
PC2 (7%)
PC1 (11%)
b
P = 0.012
P = 0.548
P = 0.050
BMI Z−score
GMC1
GMC2
GMC3
N = 141 130 78
c
P < 2e-16
P = 2e-4
P < 2e-16
Microbiota-for-age Z-score (MAZ)
GMC1
GMC2
GMC3
N = 141 130 78
d
Firmicutes
Verrucomicrobia
Actinobacteria
Proteobacteria
Bacteroidetes
Enriched in
GMC1
GMC3
Log10FC
Lachnospiraceae
Ruminococcaceae
Clostridiaceae
Verrucomicrobiaceae
Propionibacteriaceae
Corynebacteriaceae
Dermabacteraceae
Micrococcaceae
Bifidobacteriaceae
Coriobacteriaceae
Pseudomonadaceae
Comamonadaceae
Alcaligenaceae
Enterobacteriaceae
Veillonellaceae
Porphyromonadaceae
Bacteroidaceae
Peptostreptococcaceae
Erysipelotrichaceae
Streptococcaceae
Staphylococcaceae
Enterococcaceae
Lactobacillaceae
e
Relative Abundance
GMC1
GMC2
GMC3
N = 68 50 32
Genus
Candida
Malassezia
Rhodotorula
Clavispora
Other

Fig. 1. Compositionally distinct gut microbiota classes (GMCs) in feces of 1-month-old infants exhibit differential microbiota maturity and relate to the relative risk (RR) of Overweight/Obesity (OW/OB) at age 2 years.

(a) Distinct GMCs identified by Dirichlet Multinomial Mixture modeling explain more than 10% of observed variation in infant fecal bacterial β-diversity (n = 349; PERMANOVA of unweighted UniFrac distances). **(b)** BMI Z-score at age 2 years is significantly different in infants with distinct GMCs (n = 349; Linear regression; P = 0.039), with GMC3 infants exhibiting significantly higher BMI Z-scores in childhood. **(c)** Microbiota-for-age Z-scores (MAZ) of GMC3 infants are significantly greater than that of lower-risk infants (n = 349; Kruskal-Wallis; P < 2e-16), indicating early microbiota maturity. MAZ scores were calculated from a random forest model trained by the 50 most age-discriminatory bacterial OTUs of normal BMI infants. **(d)** Bacterial taxonomic comparison of GMC3 and GMC1 subjects; taxa exhibiting significant differences (zero-inflated negative binomial regression (ZINB); P_{FDR} < 0.05) in mean relative abundance are shown. Values are $\log_{10}$-transformed for purposes of illustration. Bar height indicates the magnitude of between-group relative abundance difference. Bacterial phyla are color coded as indicated. **(e)** Relative abundance of fungal genera differs across GMCs. Pairwise comparisons were calculated by linear regression in **b** and two-sided Wilcoxon rank sum tests in **c**. Boxplots indicated within violin plots represent the median (center), the 25th and 75th percentiles, and the smallest and largest values within 1.5 × the interquartile range (whiskers).

Table 1. GMCs exhibit significantly different relative risks RRs of developing OW/OB phenotypes at age 2 years.

Significance of risk ratios between microbiota states was calculated on the basis of log-binomial regression.

	Microbial Community Types			RR[1] (95% CI)		
	GMC1 (N = 141)	GMC2 (N = 130)	GMC3 (N = 78)	GMC3 vs. GMC1	GMC2 vs. GMC1	GMC3 vs. GMC2
Overweight or Obese at Age 2	16	25 (19%)	21 (27%)	2.63 (1.45-4.76) P = 0.002	1.80 (0.94-3.33) P = 0.059	1.45 (0.86-2.44) P = 0.163

[1] Adjusted for age (in days) at stool sample collection, prenatal antibiotic and antifungal use, mode of delivery and breastfeeding status at 1-month

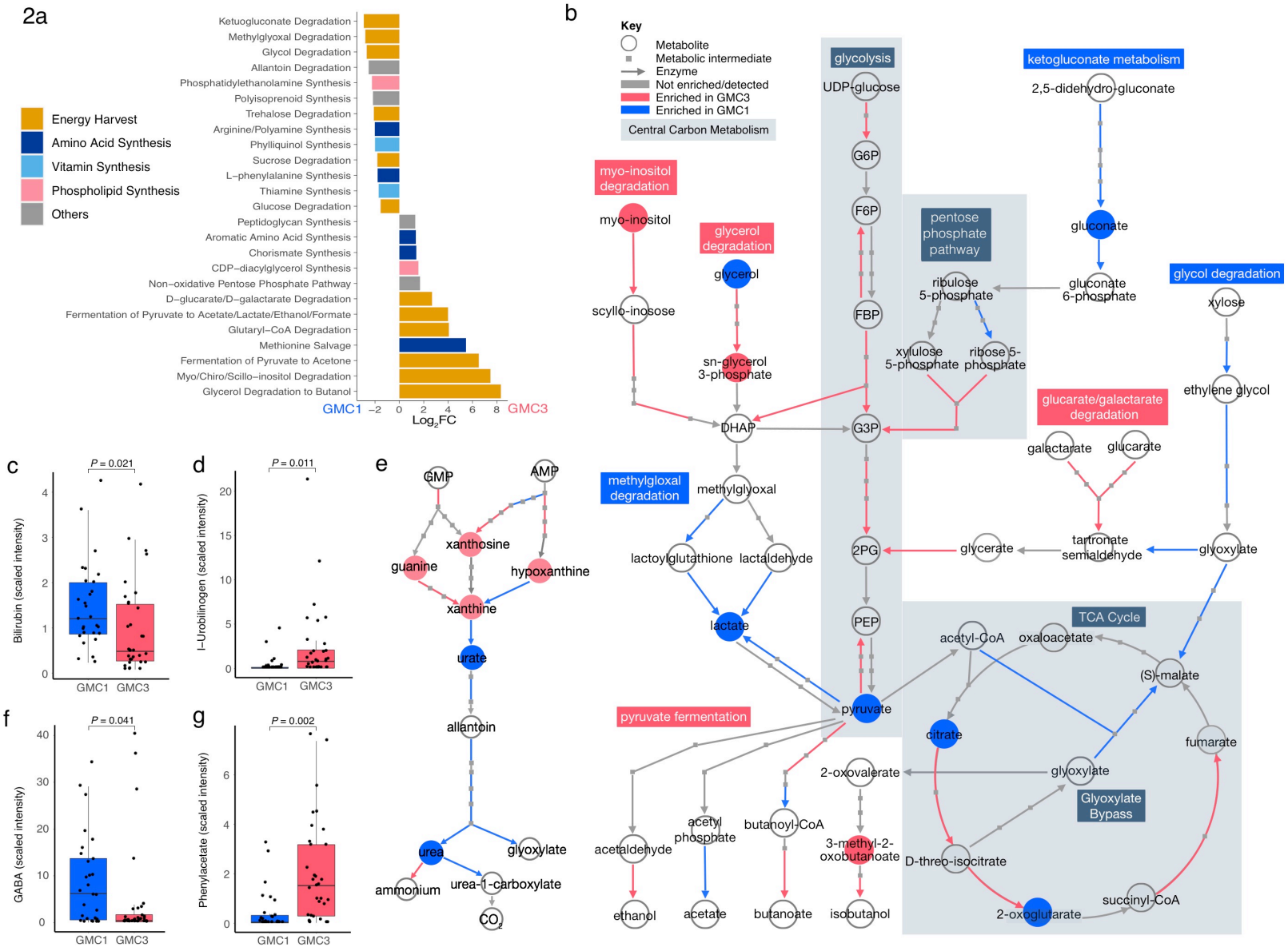

2a
Energy Harvest
Amino Acid Synthesis
Vitamin Synthesis
Phospholipid Synthesis
Others
Ketuogluconate Degradation
Methylglyoxal Degradation
Glycol Degradation
Allantoin Degradation
Phosphatidylethanolamine Synthesis
Polyisoprenoid Synthesis
Trehalose Degradation
Arginine/Polyamine Synthesis
Phylliquinol Synthesis
Sucrose Degradation
L−phenylalanine Synthesis
Thiamine Synthesis
Glucose Degradation
Peptidoglycan Synthesis
Aromatic Amino Acid Synthesis
Chorismate Synthesis
CDP−diacylglycerol Synthesis
Non−oxidative Pentose Phosphate Pathway
D−glucarate/D−galactarate Degradation
Fermentation of Pyruvate to Acetate/Lactate/Ethanol/Formate
Glutaryl−CoA Degradation
Methionine Salvage
Fermentation of Pyruvate to Acetone
Myo/Chiro/Scillo−inositol Degradation
Glycerol Degradation to Butanol
GMC1
GMC3
Log2FC
b
Key
Metabolite
Metabolic intermediate
Enzyme
Not enriched/detected
Enriched in GMC3
Enriched in GMC1
Central Carbon Metabolism
glycolysis
UDP-glucose
G6P
F6P
FBP
DHAP
G3P
2PG
PEP
pyruvate
lactate
ketogluconate metabolism
2,5-didehydro-gluconate
gluconate
gluconate 6-phosphate
pentose phosphate pathway
ribulose 5-phosphate
xylulose 5-phosphate
ribose 5-phosphate
myo-inositol degradation
myo-inositol
scyllo-inosose
glycerol degradation
glycerol
sn-glycerol 3-phosphate
glycol degradation
xylose
ethylene glycol
glucarate/galactarate degradation
galactarate
glucarate
tartronate semialdehyde
glycerate
glyoxylate
methylglyoxal degradation
methylglyoxal
lactoylglutathione
lactaldehyde
pyruvate fermentation
acetaldehyde
ethanol
acetyl phosphate
acetate
butanoyl-CoA
butanoate
2-oxovalerate
3-methyl-2-oxobutanoate
isobutanol
TCA Cycle
acetyl-CoA
oxaloacetate
(S)-malate
citrate
fumarate
D-threo-isocitrate
2-oxoglutarate
succinyl-CoA
Glyoxylate Bypass
c
P = 0.021
Bilirubin (scaled intensity)
GMC1
GMC3
d
P = 0.011
l−Urobilinogen (scaled intensity)
GMC1
GMC3
e
GMP
AMP
xanthosine
guanine
hypoxanthine
xanthine
urate
allantoin
urea
glyoxylate
ammonium
urea-1-carboxylate
CO2
f
P = 0.041
GABA (scaled intensity)
GMC1
GMC3
g
P = 0.002
Phenylacetate (scaled intensity)
GMC1
GMC3

Fig. 2. High-risk GMC3 and lower-risk GMC1 exhibit distinct functional capacities and metabolic productivity.

(a) Bar plot of gene pathways distinguishing ($\log_2$ FC > |1|; P_{FDR} < 0.25) GMC3 (n=23) and GMC1 (n=20) microbiomes. Integrated shotgun metagenomic and untargeted metabolomic analyses indicate **(b)** pathways involved in differential substrate degradation for energy biogenesis and **(e)** reduced allantoin degradation in GMC3. Enzymes (→) and metabolites (○ and ■) in **b** and **e** with $\log_2$ FC > |0.25| colored coded by GMC enrichments (red enriched in GMC3; blue enriched in GMC1). Pathways involved in central carbon metabolism are highlighted in grey shading. Legend in **b** similarly applicable to **e**. **(c)** Fecal bilirubin, **(d)** urobilinogen, **(f)** GABA and **(g)** phenylacetate levels differ between GMC3 (n=32) and GMC1 (n=28). Differences in normalized abundance between groups in **a** determined by zero-inflated compound Poisson (ZICP) regression (Supplementary Table 9A). In **c, d, f** and **g**, significance was calculated using Welch's two-sided t-test, with P_{FDR} < 0.05 considered significant, and each dot represents an independent infant stool sample. Boxplots represent the median (center), the 25th and 75th percentiles, and the smallest and largest values within 1.5 × the interquartile range (whiskers).

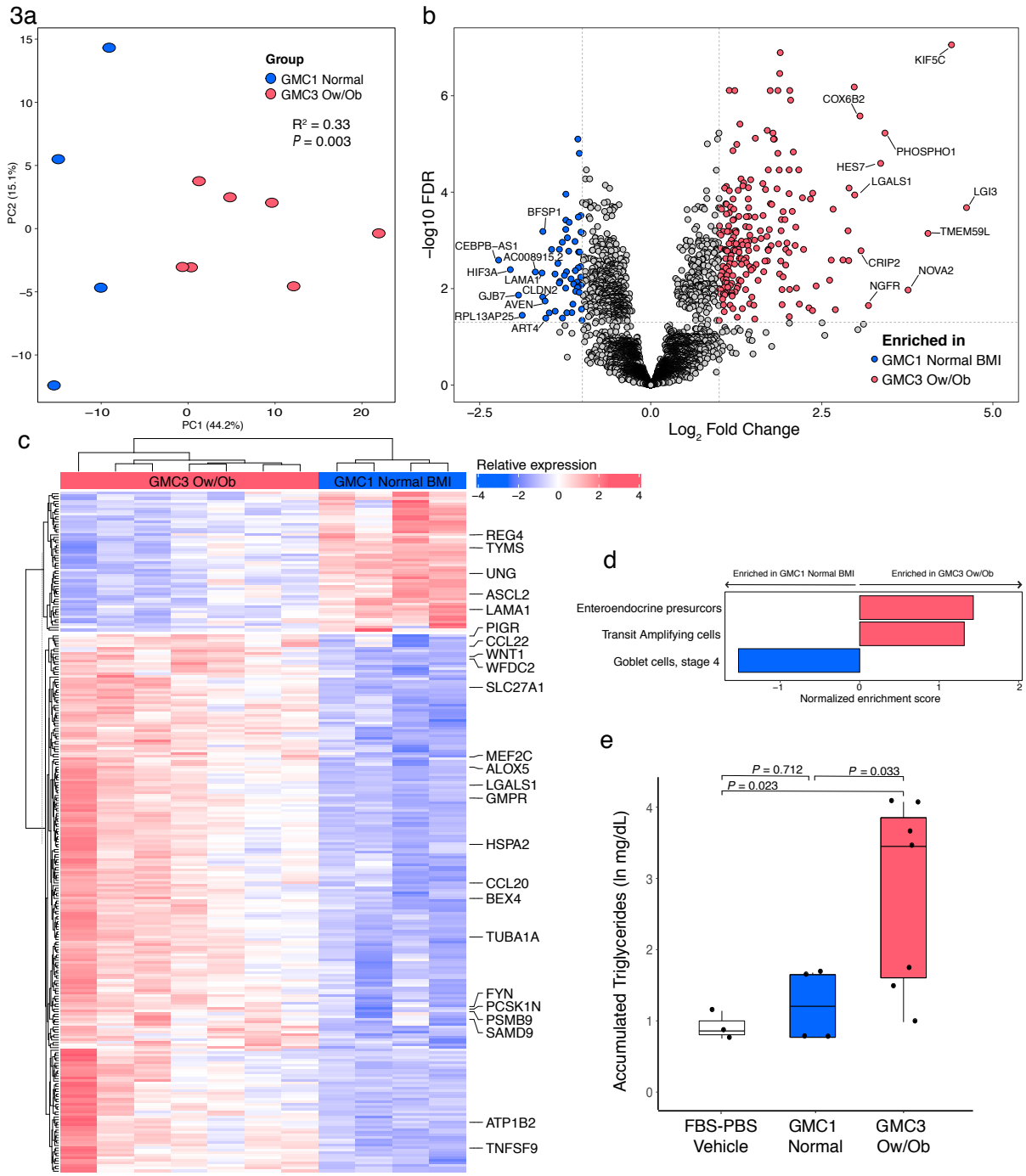
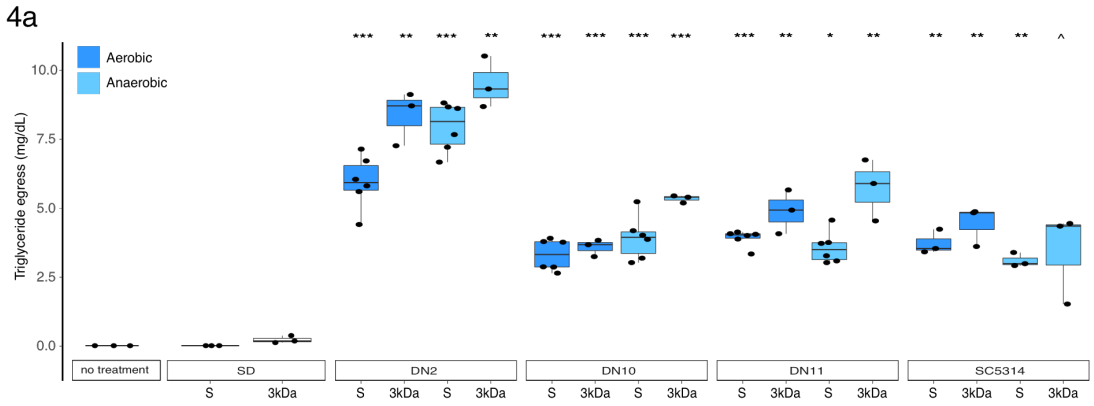
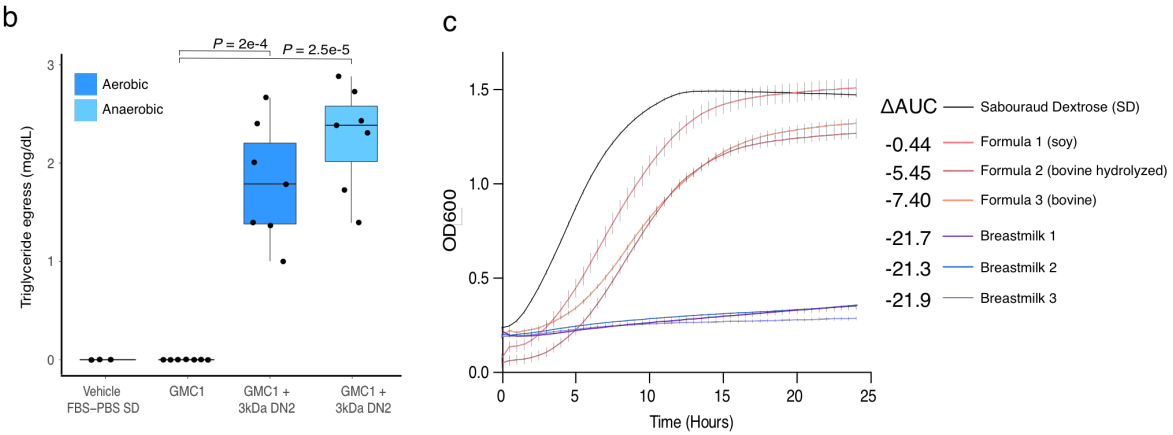

3a
Group
GMC1 Normal
GMC3 Ow/Ob
R² = 0.33
P = 0.003
PC2 (15.1%)
PC1 (44.2%)
b
−log10 FDR
Log2 Fold Change
KIF5C
COX6B2
PHOSPHO1
HES7
LGALS1
LGI3
TMEM59L
CRIP2
NOVA2
NGFR
BFSP1
CEBPB−AS1
AC008915.2
HIF3A
LAMA1
CLDN2
GJB7
AVEN
RPL13AP25
ART4
Enriched in
GMC1 Normal BMI
GMC3 Ow/Ob
c
GMC3 Ow/Ob
GMC1 Normal BMI
Relative expression
REG4
TYMS
UNG
ASCL2
LAMA1
PIGR
CCL22
WNT1
WFDC2
SLC27A1
MEF2C
ALOX5
LGALS1
GMPR
HSPA2
CCL20
BEX4
TUBA1A
FYN
PCSK1N
PSMB9
SAMD9
ATP1B2
TNFSF9
d
Enriched in GMC1 Normal BMI
Enriched in GMC3 Ow/Ob
Enteroendocrine presurcors
Transit Amplifying cells
Goblet cells, stage 4
Normalized enrichment score
e
P = 0.712
P = 0.033
P = 0.023
Accumulated Triglycerides (ln mg/dL)
FBS-PBS Vehicle
GMC1 Normal
GMC3 Ow/Ob

Fig. 3. Cell-free fecal products from GMC3 infants who developed OW/OB phenotypes in childhood reprogram Caco-2 enterocyte transcription and accelerate triglyceride accumulation.

(a) Combined GMC and OW/OB groupings explain a large proportion (33%) of observed variance in Caco-2 enterocyte response to cell-free fecal products of 1 month old infants (n=11; PERMANOVA of Euclidean distances). **(b)** Volcano plot and **(c)** heat map of significantly (P_{FDR} < 0.05) differentially ($\log_2$ FC > |1|) expressed genes observed following exposure of enterocytes to cell-free fecal products of GMC3 OW/OB (n=7) or GMC1 Normal BMI (n=4) infants. Genes of interest and those associated with distinct intestinal epithelial cell types (Parikh *et al.* 2019) are labeled in **c**. **(d)** Gene-set analysis of transcripts associated with intestinal epithelial cell states (Parikh *et al.* 2019) indicate enrichment of enteroendocrine precursors and transit amplifying cell populations in response to cell-free fecal products of GMC3 infants who become OW/OB at 2 years. **(e)** Cell-free fecal products of 1-month-old GMC3 infants who develop an OW/OB phenotype at 2 years (n=7) promote increased triglyceride accumulation in enterocytes compared to cell-free products of GMC1 infants who develop normal BMIs (n=4; ANOVA; P = 0.031). For **a**-**d** and **e**, n indicates biologically independent infant samples. Each dot represents one independent infant fecal sample in **a**, one transcript in **b**, and one independent biological replicate in **e**. *DESeq2* was used to calculate significant genes using a two-sided FDR and $\log_2$ FC. Welch's two-sided t-test was used for pairwise comparisons in **e**. Boxplots indicate the median (center), the 25th and 75th percentiles and the smallest and largest values within 1.5 × interquartile range (whiskers).


4a
Aerobic
Anaerobic
Triglyceride egress (mg/dL)
10.0
7.5
5.0
2.5
0.0
no treatment
SD
DN2
DN10
DN11
SC5314
S
3kDa


b
P = 2e-4
P = 2.5e-5
Aerobic
Anaerobic
Triglyceride egress (mg/dL)
Vehicle FBS–PBS SD
GMC1
GMC1 + 3kDa DN2
GMC1 + 3kDa DN2
c
OD600
Time (Hours)
ΔAUC
Sabouraud Dextrose (SD)
-0.44 Formula 1 (soy)
-5.45 Formula 2 (bovine hydrolyzed)
-7.40 Formula 3 (bovine)
-21.7 Breastmilk 1
-21.3 Breastmilk 2
-21.9 Breastmilk 3

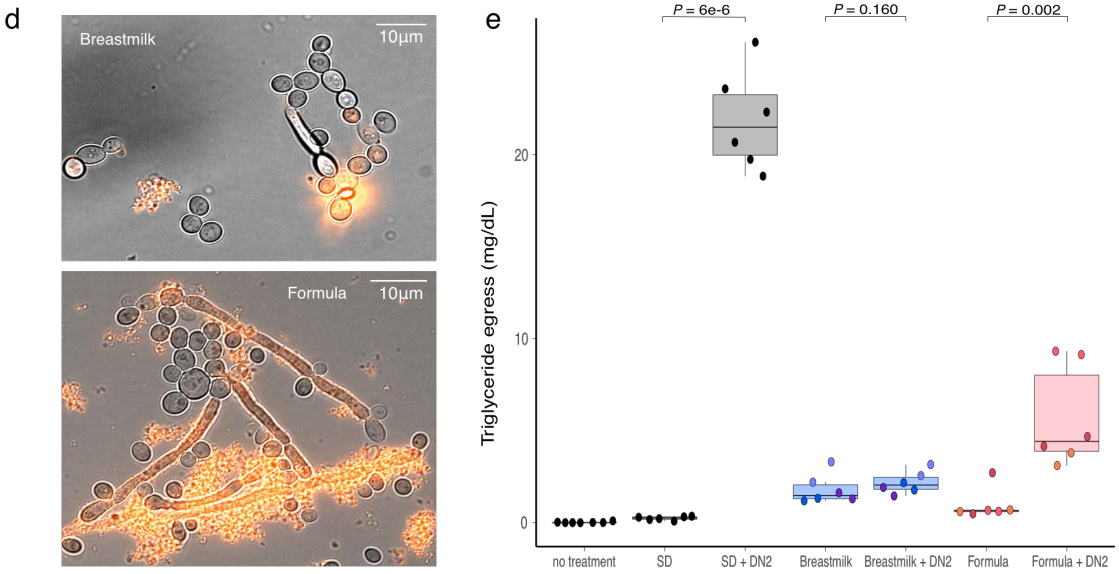
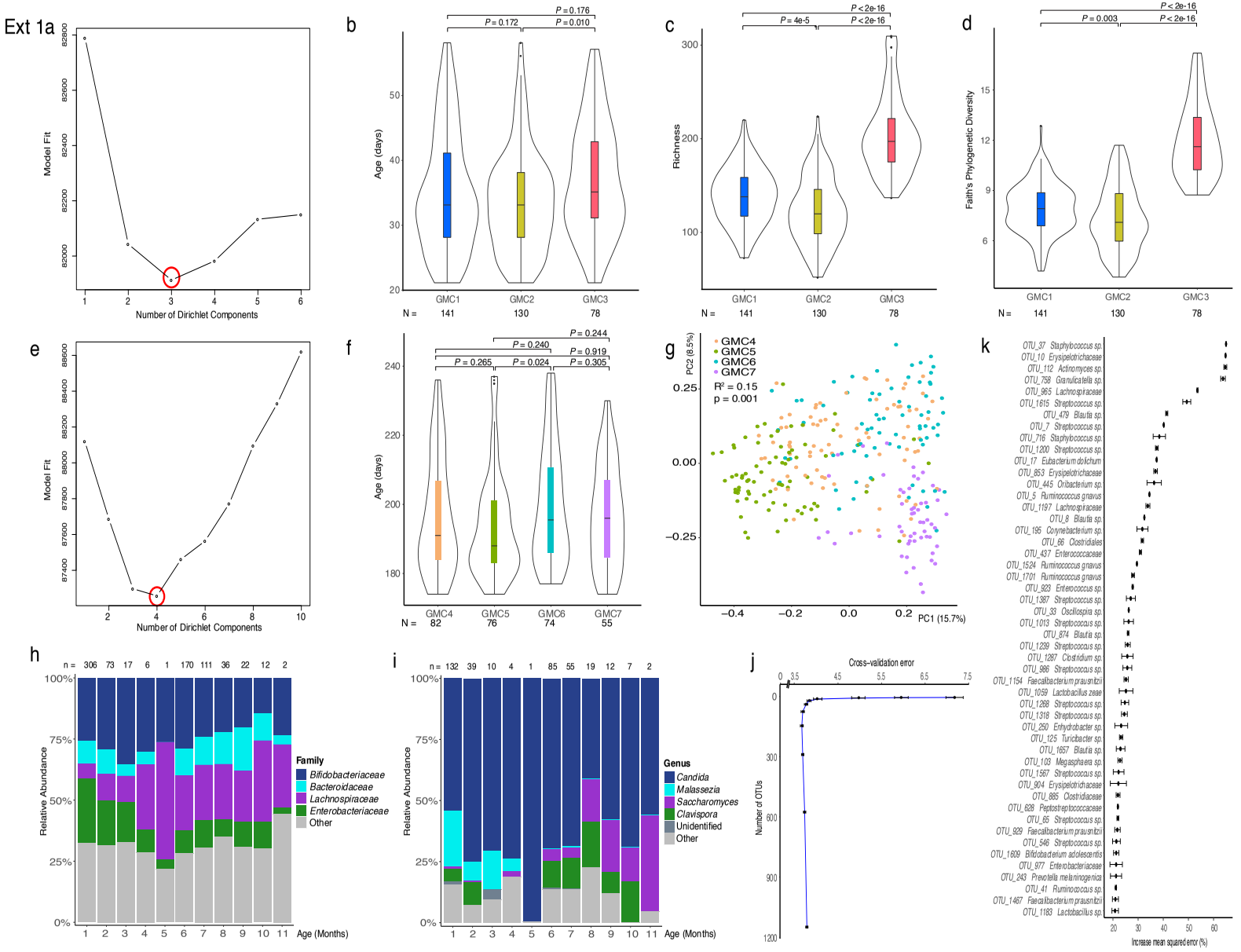

d
Breastmilk
10µm
Formula
10µm
e
P = 6e-6
P = 0.160
P = 0.002
Triglyceride egress (mg/dL)
no treatment
SD
SD + DN2
Breastmilk
Breastmilk + DN2
Formula
Formula + DN2

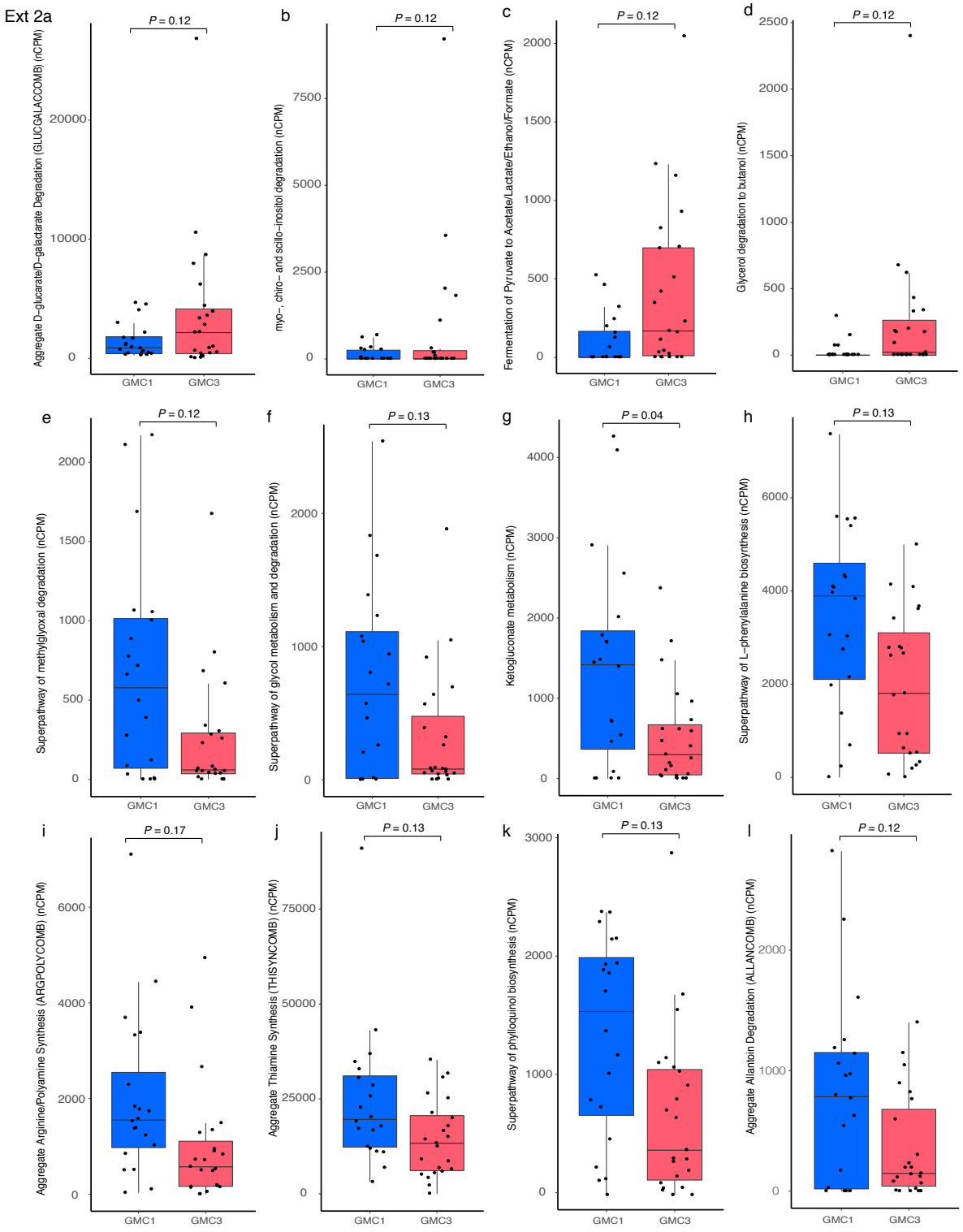
Fig. 4. *Candida albicans* isolated from GMC3 infants accelerate triglyceride trafficking in a nutrient-dependent manner.

(a) Egress of accumulated triglycerides from enterocytes is promoted by the 3kDa fraction of multiple *C. albicans* isolates (DN2, DN10, DN11 and SC5314) independent of oxygen availability. Pairwise comparisons between cell-free supernatant (S) or 3kDa fraction (3kDa) and respective Sabouraud Dextrose (SD) control were all significant (^ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). **(b)** *C. albicans* DN2 3kDa fraction is sufficient to accelerate egress of absorbed triglycerides even against a background of low-risk GMC1 cell-free fecal products. **(c)** *C. albicans* DN2 growth is inhibited in the presence of human breast milk (Breastmilk 1-3) compared with infant formula (Formula 1-3) or SD media. Representative growth curves of three independent experiments measured by optical density at 600 nm (OD600), error bars denote s.e.m. from center mean between three technical experiments. Integral of logistic regression model fitting was used to calculate area under the curve (auc) and change with respect to SD is reported as Δauc. **(d)** Representative phase contrast and fluorescence micrographs of *C. albicans* DN2 grown in breast milk (top) and infant formula (bottom) at 60x with propidium iodide stain for DNA in red. Scale bars indicate size. Images representative of three independent milk donors or formula brands per group (Extended Data 5f). **(e)** Promotion of triglyceride egress from enterocytes by *C. albicans* DN2 differs by growth media; colored dots represent independent breast milk or formula brands respectively. In **a** and **e** each dot represents a biological replicate. In **b**, each dot represents an independent fecal sample. Significance of pairwise comparisons in **a, b,** and **e** were calculated using Welch's two-sided t-test. Boxplots indicate the median (center), the 25th and 75th percentiles, and the smallest and largest values within 1.5 × the interquartile range (whiskers).


Ext 1a
Model Fit
Number of Dirichlet Components
b
Age (days)
P = 0.176
P = 0.172
P = 0.010
GMC1
GMC2
GMC3
N = 141 130 78
c
Richness
P < 2e-16
P = 4e-5
P < 2e-16
d
Faith's Phylogenetic Diversity
P < 2e-16
P = 0.003
P < 2e-16
e
f
P = 0.244
P = 0.240
P = 0.919
P = 0.265
P = 0.024
P = 0.305
GMC4
GMC5
GMC6
GMC7
N = 82 76 74 55
g
PC2 (8.5%)
PC1 (15.7%)
R² = 0.15
p = 0.001
k
Increase mean squared error (%)
h
Relative Abundance
Age (Months)
Family
Bifidobacteriaceae
Bacteroidaceae
Lachnospiraceae
Enterobacteriaceae
Other
i
Genus
Candida
Malassezia
Saccharomyces
Clavispora
Unidentified
Other
j
Cross-validation error
Number of OTUs

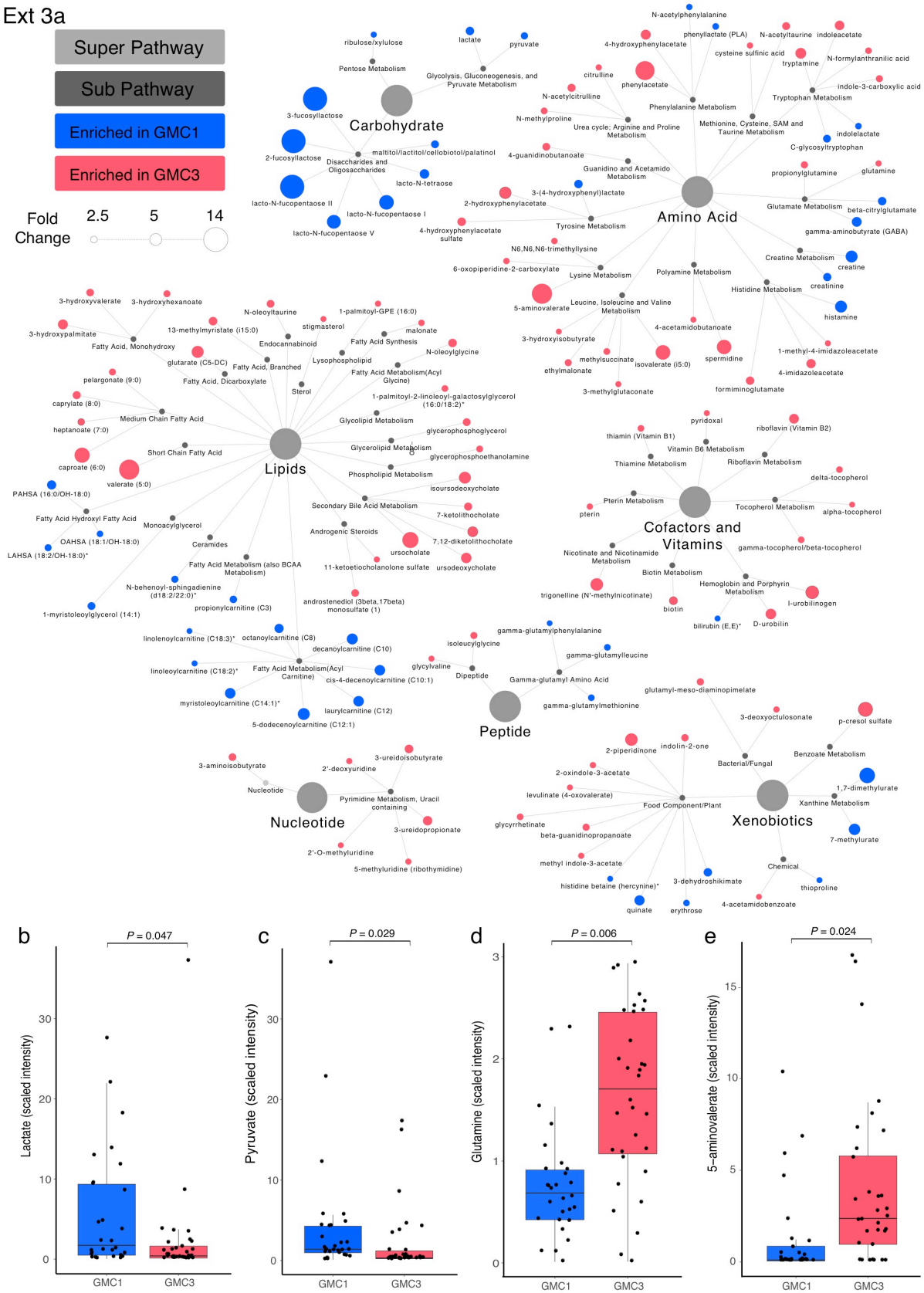
Extended Data Fig. 1. Compositionally distinct gut microbiota classes (GMCs) exist during the first year of life.

(a) Dirichlet multinomial mixture (DMM) model identifies three compositionally distinct bacterial GMCs as the best model fit in very early life (n=349, range 21 – 58 days; median = 35). Model fit was based on the Laplace approximation to the negative log model where a lower value indicates a better model fit. **(b)** GMC1 and GMC3 infants do not significantly differ in age. **(c)** Bacterial richness (n=349; Kruskal–Wallis; $P < 2e\text{-}16$) and **(d)** phylogenetic α–diversity are significantly greater in GMC3 (n=349; Kruskal–Wallis; $P < 2e\text{-}16$). **(e)** DMM model identifies four compositionally distinct bacterial GMCs as the best model fit later in infancy (n=287, range 174 – 238 days; median = 206). Model was constructed as defined in **a**. **(f)** GMC4-7 participants do not differ significantly in age (n=287; Kruskal–Wallis; $P = 0.15$), with the exception of GMC5 and GMC6 where the latter is significantly older in a pairwise comparison. **(g)** GMC designation significantly explains the observed variation in bacterial β-diversity in later life samples (n = 287; PERMANOVA of Bray Curtis distances, $R^2 = 0.15$; $P = 0.001$). Each dot represents an independent fecal sample. **(h)** Age-stratified taxa summaries (presented at the family level) of bacterial relative abundance (n = 756; number of participants per age group is provided above bars). **(i)** Age-stratified taxa summaries (presented at the genus level) of fungal relative abundance (n = 366; number of participants per age group is provided above bars). **(j)** Tenfold cross-validation indicates that 50 bacterial OTUs are sufficient for random forest predictions of the chronological age of normal BMI infants on the basis of microbiota composition. Data show mean ± s.d. computed over 100 iterations. **(k)** The 50 most informative predictors to the random forest model, ranked in descending order of their importance to model accuracy. These bacterial OTUs were included in the sparse model used to calculate MAZ. Data show mean ± s.e.m. computed over 100 iterations. Two-sided Wilcoxon rank sum test for significance of pairwise comparisons in **b**-**d** and **f**. Boxplots indicated within violin plots represent the median (center), the 25th and 75th percentiles, and the smallest and largest values within 1.5 × the interquartile range (whiskers).


Ext 2a
a
Aggregate D−glucarate/D−galactarate Degradation (GLUCGALACCOMB) (nCPM)
P = 0.12
b
myo−, chiro− and scillo−inositol degradation (nCPM)
P = 0.12
c
Fermentation of Pyruvate to Acetate/Lactate/Ethanol/Formate (nCPM)
P = 0.12
d
Glycerol degradation to butanol (nCPM)
P = 0.12
e
Superpathway of methylglyoxal degradation (nCPM)
P = 0.12
f
Superpathway of glycol metabolism and degradation (nCPM)
P = 0.13
g
Ketogluconate metabolism (nCPM)
P = 0.04
h
Superpathway of L−phenylalanine biosynthesis (nCPM)
P = 0.13
i
Aggregate Arginine/Polyamine Synthesis (ARGPOLYCOMB) (nCPM)
P = 0.17
j
Aggregate Thiamine Synthesis (THISYNCOMB) (nCPM)
P = 0.13
k
Superpathway of phylloquinol biosynthesis (nCPM)
P = 0.13
l
Aggregate Allantoin Degradation (ALLANCOMB) (nCPM)
P = 0.12
GMC1
GMC3

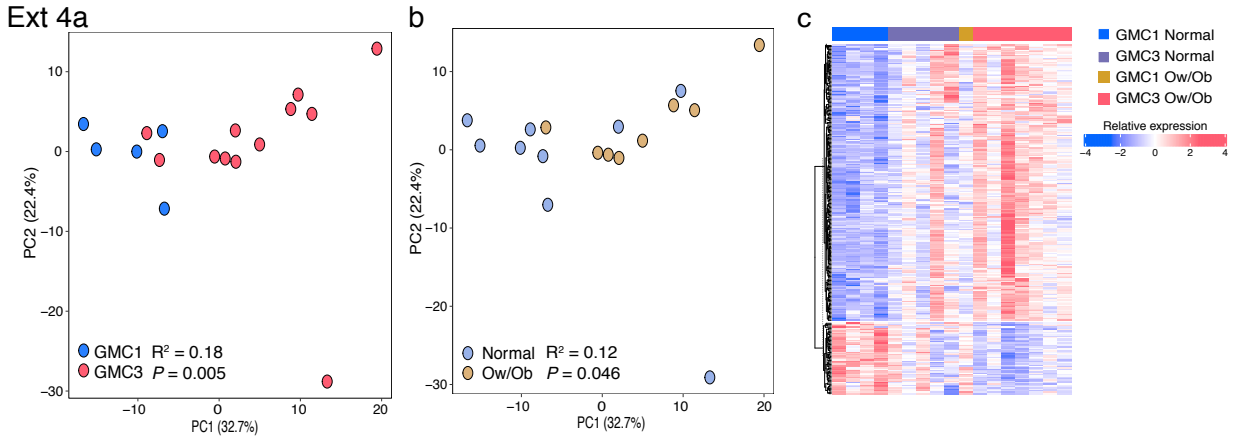
Extended Data Fig. 2. GMC3 and GMC1 microbiomes exhibit distinct functional capacities.

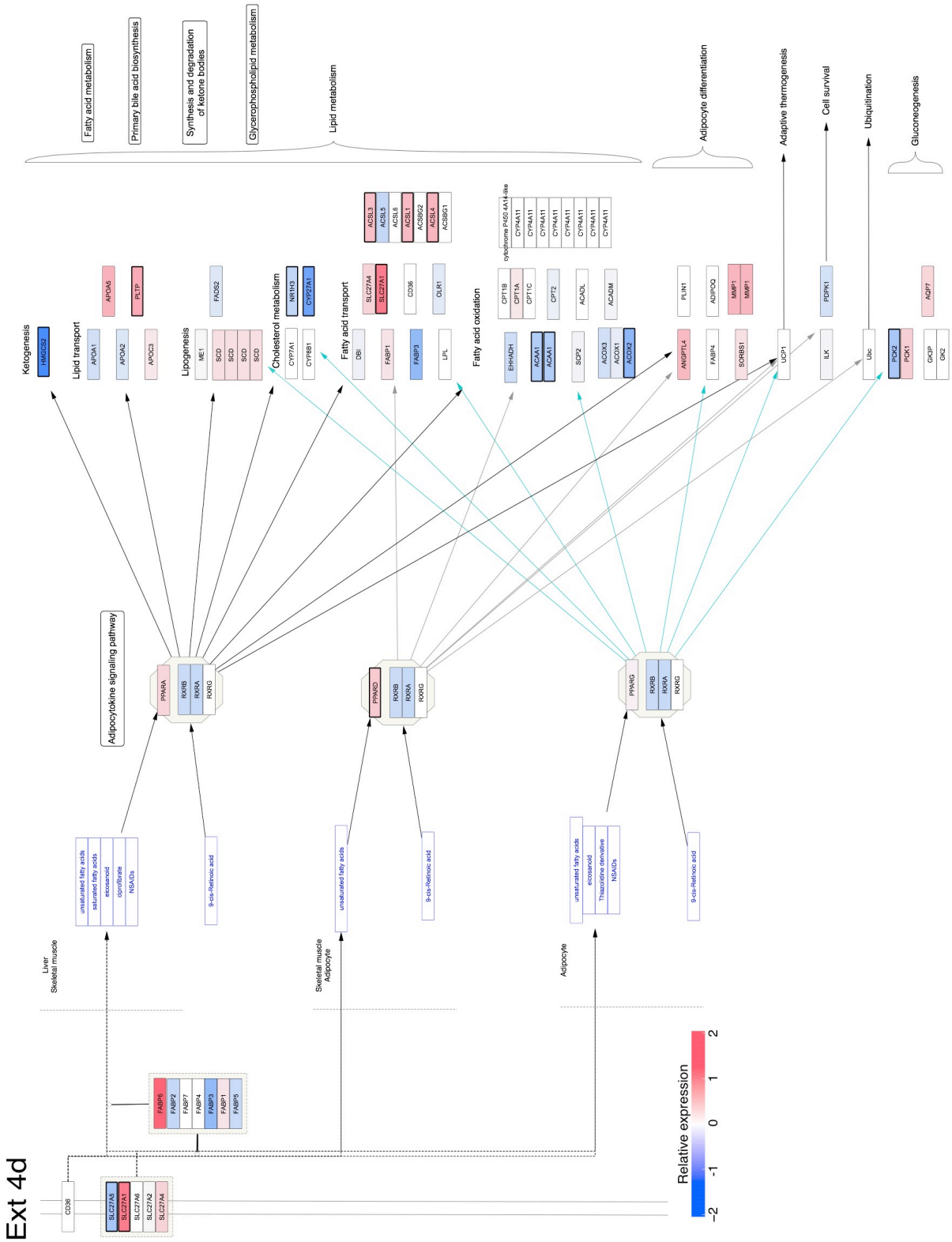
Comparative analysis of microbial metabolic pathways by GMC status (Supplementary Table 9A) indicated that GMC3 (n=23) microbiota are enriched for pathways that degrade **(a)** D-glucarate, D-galactarate, **(b)** myo/chiro/scillo-inositol, **(c)** pyruvate and **(d)** glycerol, whereas GMC1 (n=20) is enriched in pathways that degrade **(e)** methylgloxal, **(f)** glycol, and **(g)** ketogluconate. GMC1 is also enriched in capacity for synthesis of **(h)** L-phenylalanine, **(i)** arginine and polyamines, **(j)** thiamine and **(k)** phylloquinol, and for **(l)** allantoin degradation (ALLANCOM; Supplementary Table 9B). Each dot represents an independent fecal sample. *P* values are two-tailed, from zero-inflated compound Poisson (ZICP) models Benjamini-Hochberg FDR adjusted for multiple comparisons; $P_{FDR} < 0.2$ considered significant. Boxplots show normalized abundance (nCPM) and indicate the median (center), the 25th and 75th percentiles, and the smallest and largest values within 1.5 × the interquartile range (whiskers).


Ext 3a
Super Pathway
Sub Pathway
Enriched in GMC1
Enriched in GMC3
Fold Change
2.5
5
14
Carbohydrate
Amino Acid
Lipids
Cofactors and Vitamins
Peptide
Nucleotide
Xenobiotics
b
P = 0.047
Lactate (scaled intensity)
c
P = 0.029
Pyruvate (scaled intensity)
d
P = 0.006
Glutamine (scaled intensity)
e
P = 0.024
5−aminovalerate (scaled intensity)
GMC1
GMC3

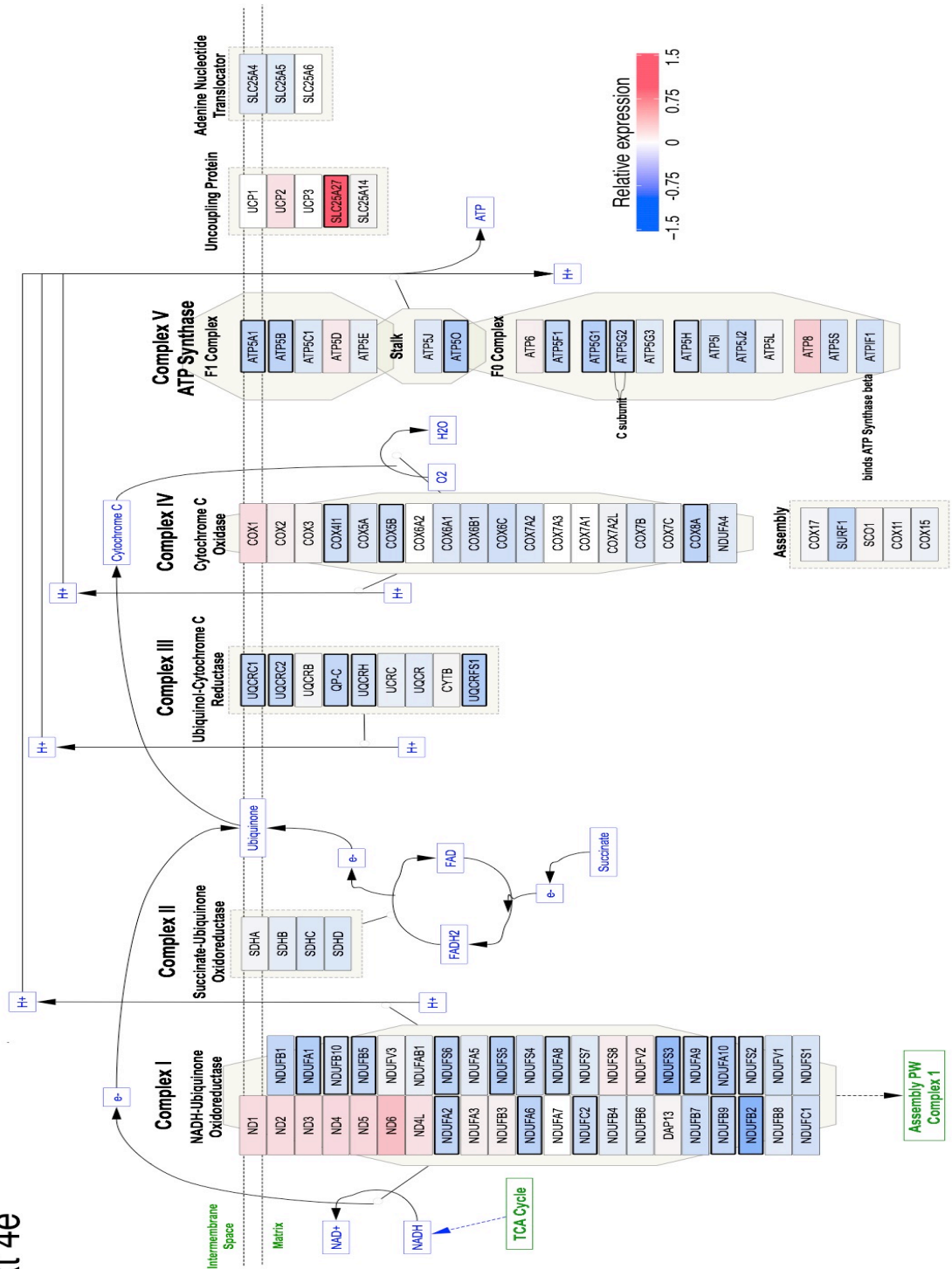
Extended Data Fig. 3. GMC3 and GMC1 are metabolically distinct.

(a) Comparison of fecal samples from GMC3 (n=32) and GMC1 (n=28) by untargeted mass spectrometry identifies widespread metabolic differences. GMC3 subjects exhibit evidence of increased fermentation and protein metabolism, the later reflective of greater incidence of formula feeding. In contrast, GMC1 subjects, who are enriched in oligosaccharides associated with breast feeding, exhibit evidence of greater fatty acid β-oxidation. GMC1 also contains greater concentrations of the methylglyoxal degradation products **(b)** lactate and **(c)** pyruvate, whereas GMC3 is enriched in **(d)** the GABA precursor glutamine and **(e)** the weak GABA agonist 5-aminovalerate. Each dot represents an independent fecal sample in **b**-**e.** *P* values are two-tailed, from Welch's two-sided t-test FDR BH-adjusted for multiple comparisons; $P_{FDR} < 0.05$ considered significant.




Ext 4d
Fatty acid metabolism
Primary bile acid biosynthesis
Synthesis and degradation of ketone bodies
Glycerophospholipid metabolism
Lipid metabolism
Adipocyte differentiation
Adaptive thermogenesis
Cell survival
Ubiquitination
Gluconeogenesis
Ketogenesis
Lipid transport
Lipogenesis
Cholesterol metabolism
Fatty acid transport
Fatty acid oxidation
Adipocytokine signaling pathway
Liver
Skeletal muscle
Skeletal muscle
Adipocyte
Adipocyte
Relative expression
−2
-1
0
1
2

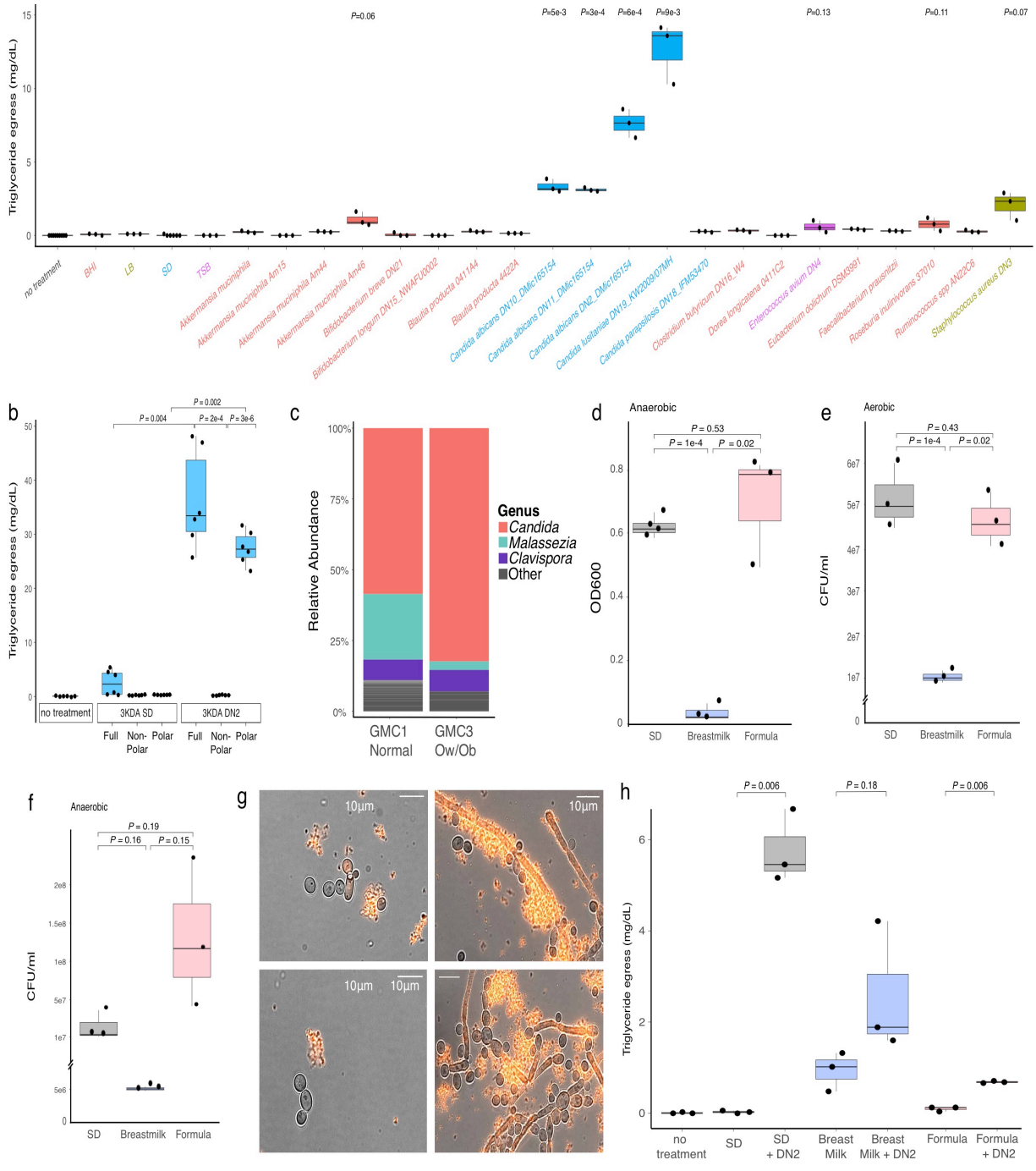
Ext 4e


Complex I
NADH-Ubiquinone Oxidoreductase
ND1
ND2
ND3
ND4
ND5
ND6
ND4L
NDUFA2
NDUFA3
NDUFB3
NDUFA6
NDUFA7
NDUFC2
NDUFB4
NDUFB6
DAP13
NDUFB7
NDUFB9
NDUFB2
NDUFB8
NDUFC1
NDUFB1
NDUFA1
NDUFB10
NDUFB5
NDUFV3
NDUFAB1
NDUFS6
NDUFA5
NDUFS5
NDUFS4
NDUFA8
NDUFS7
NDUFS8
NDUFV2
NDUFS3
NDUFA9
NDUFA10
NDUFS2
NDUFV1
NDUFS1
Complex II
Succinate-Ubiquinone Oxidoreductase
SDHA
SDHB
SDHC
SDHD
Complex III
Ubiquinol-Cytochrome C Reductase
UQCRC1
UQCRC2
UQCRB
QP-C
UQCRH
UCRC
UQCR
CYTB
UQCRFS1
Complex IV
Cytochrome C Oxidase
COX1
COX2
COX3
COX4I1
COX5A
COX5B
COX6A2
COX6A1
COX6B1
COX6C
COX7A2
COX7A3
COX7A1
COX7A2L
COX7B
COX7C
COX8A
NDUFA4
Assembly
COX17
SURF1
SCO1
COX11
COX15
Complex V
ATP Synthase
F1 Complex
ATP5A1
ATP5B
ATP5C1
ATP5D
ATP5E
Stalk
ATP5J
ATP5O
F0 Complex
ATP6
ATP5F1
ATP5G1
ATP5G2
ATP5G3
ATP5H
ATP5I
ATP5J2
ATP5L
ATP8
ATP5S
ATPIF1
C subunit
binds ATP Synthase beta
Uncoupling Protein
UCP1
UCP2
UCP3
SLC25A27
SLC25A14
Adenine Nucleotide Translocator
SLC25A4
SLC25A5
SLC25A6
Relative expression
−1.5 -0.75 0 0.75 1.5
Intermembrane Space
Matrix
TCA Cycle
NAD+
NADH
FAD
FADH2
Succinate
Ubiquinone
Cytochrome C
H2O
O2
ATP
H+
e-
Assembly PW Complex 1

Extended Data Fig. 4. Cell-free fecal products of 1-month-old GMC3 infants who become OW/OB at 2 years or GMC1 infants with normal BMIs at age 2 years induce divergent transcriptional responses in Caco-2 enterocytes.

(a-b) Principal-component analysis of Euclidean distances of top 10,000 variably expressed genes (by coefficient of variation) and **(c)** heat map of most significantly differentially expressed ($log_2FC > |1|$, $P_{FDR} < 0.05$) genes in Caco-2 enterocytes treated with cell-free fecal products from GMC1 Normal BMI (n=4), GMC3 Normal BMI (n=5), GMC1 OW/OB (n=1) or GMC3 OW/OB (n=7) subjects, as determined by RNA sequencing, indicate a transcriptional shift in enterocyte expression based on GMC class and 2-year OW/OB phenotypes. Pathway analysis indicates that enterocytes treated with cell-free fecal products of GMC3 OW/OB (n=7) infants exhibit altered expression of **(d)** PPAR signaling and lipid metabolism and **(e)** reduced expression of mitochondrial oxidative phosphorylation genes compared with cells treated with GMC1 Normal BMI cell-free fecal extracts (n=4). Genes in **d** and **e** with $log_2FC > |1|$ and $P_{FDR} < 0.15$ are indicated within bolded solid black boxes. Each dot represents an independent fecal sample in **a** and **b.** PERMANOVA test was used for significance for **a** and **b**. *DESeq2* was used to calculate pairwise significance using a two-sided FDR and log_2FC.

Ext 5a


Triglyceride egress (mg/dL)
P=0.06
P=5e-3
P=3e-4
P=6e-4
P=9e-3
P=0.13
P=0.11
P=0.07
no treatment
BHI
LB
SD
TSB
Akkermansia muciniphila
Akkermansia muciniphila Am15
Akkermansia muciniphila Am44
Akkermansia muciniphila Am46
Bifidobacterium breve DN21
Bifidobacterium longum DN15_NWAFU002
Blautia producta 0411A4
Blautia producta 4422A
Candida albicans DN10_DMic165154
Candida albicans DN11_DMic165154
Candida albicans DN2_DMic165154
Candida lusitaniae DN19_KW2009/07MH
Candida parapsilosis DN18_IFM53470
Clostridium butyricum DN16_W4
Dorea longicatena 0411C2
Enterococcus avium DN4
Eubacterium dolichum DSM3991
Faecalibacterium prausnitzii
Roseburia inulinivorans 37010
Ruminococcus spn AN2206
Staphylococcus aureus DN3
b
Triglyceride egress (mg/dL)
P = 0.002
P = 0.004
P = 2e-4
P = 3e-6
no treatment
3KDA SD
3KDA DN2
Full
Non-Polar
Polar
c
Relative Abundance
Genus
Candida
Malassezia
Clavispora
Other
GMC1 Normal
GMC3 Ow/Ob
d
Anaerobic
P = 0.53
P = 1e-4
P = 0.02
OD600
SD
Breastmilk
Formula
e
Aerobic
P = 0.43
P = 1e-4
P = 0.02
CFU/ml
SD
Breastmilk
Formula
f
Anaerobic
P = 0.19
P = 0.16
P = 0.15
CFU/ml
SD
Breastmilk
Formula
g
10μm
h
P = 0.006
P = 0.18
P = 0.006
Triglyceride egress (mg/dL)
no treatment
SD
SD + DN2
Breast Milk
Breast Milk + DN2
Formula
Formula + DN2

Extended Data Fig. 5. *Candida albicans* promotes robust triglyceride egress from enterocytes in a formula-dependent manner irrespective of oxygen availability.

(a) Members of the *Candida* genus, including the dominant fungus in GMC3 *C. albicans*, promote the greatest egress of accumulated triglycerides from enterocytes (21 strains from 12 genera). Strains are color coded by growth media. *P* values indicate significant ($P < 0.05$) comparison with relevant uninoculated growth media. **(b)** Egress of accumulated triglycerides from enterocytes is promoted by the polar 3kDa fraction of aerobically grown *C. albicans* DN2. **(c)** Relative abundance of fungal genera differs between GMC3 OW/OB (n=11) and GMC1 Normal BMI (n=59) participants. **(d)** Effects of breast milk and infant formula compared to SD media on anaerobic growth and **(e-f)** 24 h survival of *C. albicans* DN2. (**g**) Additional phase contrast and fluorescence micrographs of *C. albicans* DN2 grown in breast milk (left) and infant formula (right) at 60x with propidium iodide stain for DNA in red. **(h)** Promotion of triglyceride egress by *C. albicans* is cell density dependent. OD_{600} normalization of *C. albicans* DN2 grown in formula to that of breast milk cultures markedly reduces its supernatant's ability to promote triglyceride egress from enterocytes. Each dot represents a biological replicate in **a**, a technical triplicate from two experiments in **b** and an independent milk donor or formula brand in **d**-**f** and **h**. All *P* values were calculated using Welch's two-sided t-test. Boxplots indicate the median (center), the 25th and 75th percentiles, and the smallest and largest values within 1.5 × the interquartile range (whiskers).

Supplemental Files

Supplementary Table 1 | Association between maternal and early life factors and GMCs

Supplementary Table 2A | Unadjusted and adjusted association between 1-month GMC1-3 and obese/overweight at age 2

Supplementary Table 2B | Unadjusted and adjusted association between 1-month GMC1-3 and BMI z-score at age 2

Supplementary Table 3A | Unadjusted and adjusted association between 6-month GMC4-7 and obese/overweight at age 2

Supplementary Table 3B | Unadjusted and adjusted association between 6-month GMC4-7 and BMI z-score at age 2

Supplementary Table 4 | Differentially abundant (FDR <0.05) bacterial OTUs between GMC3 (n=78) and GMC1 (n=141). Taxa ordered by decreasing log2 Fold Change (Log2FC) of OTU relative abundance in GMC3 versus GMC1.

Supplementary Table 5 | Differentially abundant (FDR <0.05) fungal OTUs between GMC3 (n=32) and GMC1 (n=68). Taxa ordered by decreasing log2 Fold Change (Log2FC) of OTU relative abundance in GMC3 versus GMC1.

Supplementary Table 6 | Differentially abundant (FDR <0.05) bacterial OTUs between GMC3 (n=78) and GMC2 (n=130). Taxa ordered by decreasing log2 Fold Change (Log2FC) of OTU relative abundance in GMC3 versus GMC2.

Supplementary Table 7 | Differentially abundant (FDR <0.05) fungal OTUs between GMC3 (n=32) and GMC2 (n=50). Taxa ordered by decreasing log2 Fold Change (Log2FC) of OTU relative abundance in GMC3 versus GMC2.

Supplementary Table 8 | Distribution of microbiologically representative infant fecal samples that underwent shotgun metagenomic and untargeted metabolomic profiling.

Supplementary Table 9A | Differential abundance of MetaCyc pathways between GMC3 (n=23) and GMC1 (n=20). Pathways ordered by increasing BH-adjusted P value in GMC3 versus GMC1.

Supplementary Table 9B | Aggregation of MetaCyc pathway variants

Supplementary Table 10 | Differential abundance of metabolites detected in GMC3 (n=32) and GMC1 (n=28) stool samples.

Supplementary Table 11 | Differentially expressed (log2FC >|1| and FDR <0.15) genes in Caco-2 enterocytes treated with GMC3 Ow/Ob (n=7) versus GMC1 Normal (n=4) cell-free fecal products. Genes ordered by decreasing log2 Fold Change (Log2FC) of expression in GMC3 versus GMC1.

Supplementary Table 12 | Growth conditions for microorganisms whose 24 h supernatant was tested in enterocyte triglyceride trafficking assay.

Methods

Study population, subsample criteria of subjects for stool microbiome analysis and OW/OB definition

The Wayne County Health, Environment, Allergy and Asthma Longitudinal Study (WHEALS) US birth cohort recruited pregnant women (n=1258) aged 21 - 49 years between August 2003 and November 2007[58]. Women were considered to be eligible if they lived in a predefined cluster of contiguous zip codes in Wayne County, Michigan (including the city of Detroit), had no intention of moving out of the area in the subsequent two years and provided informed written consent. For this study, we selected WHEALS children who had a stool sample collected during a 1 and/or 6 month home visit, and had completed their 24-month clinic visit with height and weight measurements (n=543 subjects; n=756 samples). The age of infants at stool collection ranged from 1–11 months and samples were stored at −80 °C. Samples were shipped to the University of California, San Francisco (UCSF) on dry ice, where they were also stored at −80 °C until processed. At the study's 2-year clinic visit, trained field staff measured child height and weight. Overweight or obesity (OW/OB) at age 2 years was defined using the 2000 age and sex adjusted CDC growth charts[59] as BMI at or above the 85th percentile and normal BMI as BMI between the 5th to <85th percentile.

Since age strongly influences microbiome composition during early life, DMM modeling was only applied to samples that were collected within a standard deviation of the mean collection age for each home visit to control for age-specific microbiome differences. Samples were stratified by time of sample collection (n=403 1-month, n=353 6-month).

In the DMM analytical dataset, stool specimens from the 1-month visit were collected at a mean ± 1 standard deviation (SD) of 39 ± 19 days (n = 349, median age 35 days; range 21 - 58 days) and stool specimens from the 6-month visit were collected at a mean ± 1 SD of 205 ± 33 days (n = 287, median age 206 days; range 174 - 238 days).

DNA extraction

Fecal DNA was extracted from stool samples using the modified cetyltrimethylammonium bromide (CTAB) method previously used for fungal and bacterial profiling[60,61]. Briefly, 500 µl modified CTAB extraction buffer was added to 25 mg of stool in a 2 ml Lysing Matrix E tube (MP Biomedicals) prior to incubation at 65 °C for 15 min. Samples were bead-beaten (5.5 m/s, 30 sec) in a Fastprep-24 (MP Biomedicals) and then centrifuged (16000 × g, 5 min) before the top aqueous phase was transferred to a 2 ml polypropylene 96-well plate (USA Scientific). A further 500 µl modified CTAB extraction buffer was added to each LME tube, similarly bead-beaten and centrifuged to collect a total of 1 ml aqueous phase per sample. After adding 1 ml of phenol:chloroform:isoamyl alcohol (25:24:1) to the collected aqueous supernatant, samples were centrifuged (3200 × g, 20 min, 4 °C) and the resulting top aqueous phase was transferred to a new 2 ml polypropylene 96-well plate (USA Scientific). Polyethylene glycol/NaCl (2 v/v) was added to the collected aqueous supernatant and incubated at room temperature for 2 h. Samples were then centrifuged (3200 × g, 60 min, 4 °C), washed with ice cold 70% EtOH and resuspended in 30 µl of TE buffer (Invitrogen).

PCR conditions and library preparation for bacterial and fungal biomarker sequencing

The V4 region of the 16S rRNA bacterial gene was amplified using primers designed by Caporaso *et al*[62]. PCR was performed in 25 µl reactions using 0.025 U Takara Hot Start ExTaq (Takara Mirus Bio Inc.), 1x Takara buffer with $MgCl_2$, 0.4 pmol/µl of F515 and R806 primers, 0.56 mg/ml of bovine serum albumin (BSA; Roche Applied Science), 200 µM of dNTPs and 10 ng of gDNA. Reactions were performed in triplicate with the following: initial denaturation (98 °C, 2 min), 30 cycles of 98 °C (20 s), annealing at 50 °C (30 s), extension at 72 °C (45 s) and final extension at 72 °C (10 min). Amplicons from technical triplicates were pooled and verified using a 2% TBE agarose e-gel (Life Technologies), cleaned up and normalized using SequalPrep Normalization Plates (Applied Biosystems), and quantified using the Qubit dsDNA HS Assay Kit (Invitrogen). Samples were pooled in equal moles (5 ng), purified using AMPure SPRI beads (Beckman Coulte), quantified using KAPA SYBR (KAPA Biosystems), denatured and diluted to 2 nM, and 5 pmol was loaded onto the Illumina Nextseq cartridge with 40% (v/v) of denatured 12.5 pM PhiX spike-in control.

The internal transcribed spacer region 2 (ITS2) of the fungal rRNA gene was amplified using the primer pair fITS7 (5'- GTGAATCATCGAATCTTTG-3') and ITS4 (5'-TCCTCCGCTTATTGATATGC-3')[63]. PCR was performed in triplicate in 25 µl reactions with 1x Takara buffer (Takara Mirus Bio), 200 nM of each primer, 200 µM dNTPs, 2.75 mM of $MgCl_2$, 0.56 mg ml^{-1} of BSA (Roche Applied Science), 0.025 U Takara Hot Start ExTaq and 50 ng of gDNA. Reactions were conducted under the following conditions: initial denaturation (94 °C for 5 min) followed by 30 cycles of 94 °C (30 sec), annealing

at 54 °C (30 sec), extension at 72 °C (30 sec) and a final extension at 72 °C (7 min). PCR amplicons were verified, purified, quantified and pooled as described above for bacterial library preparation. ITS2 PCR was performed on all stool samples, which produced ITS2 amplicons in n=186 1-month and n=180 6-month samples; samples without fungal data had no detectable ITS2 amplicons. The amplicon library was purified, quantified, denatured and diluted similar to the 16S amplicon library described above. 10 pmol of the ITS2 amplicon library was loaded onto the Illumina MiSeq cartridge with 25% (v/v) of denatured 10 pM PhiX spike-in control.

Biomarker sequence data processing

Paired-end sequences were assembled using FLASH v1.2.7[64] requiring a minimum base pair overlap of 25 bp and demultiplexed by barcode using QIIME v1.9.1[65]. Quality filtering was performed using USEARCH v8.0.1623[66] to remove reads with >2 expected errors. Quality reads were dereplicated at 100% sequence identity, clustered at 97% sequence identity into operational taxonomic units (OTUs), filtered of chimeric sequences by UCHIME[67], and mapped back to resulting OTUs using UPARSE[68]; sequence reads that failed to cluster with a reference sequence were clustered *de novo.* Taxonomy was assigned to the OTUs using the Greengenes v13_5 database[69]. Sequences were aligned using PyNAST[70], and FastTree 2.1.3[71] was used to build a phylogenetic tree. Resulting sequence reads were normalized by multiply rarefying to 60,000 reads per sample as described previously[61] to ensure reduced data were representative of the fuller data for each sample.

Fungal sequences were quality trimmed (Q score, <25) and removed of adaptor sequences using cutadapt[72]. Paired-end sequences were assembled, demultiplexed by barcode, clustered into OTUs at 97% identity and filtered of chimeras using similar methods as described for 16S amplicons. Taxonomy was assigned using UNITE v7.0[73]. Resulting sequence reads were normalized by multiply rarefying to 1,000 reads per sample to ensure reduced data were representative of the fuller data for each sample.

Prediction of microbiota age using random forests

Random forest models were used to regress the relative abundances of all 16S rRNA-derived bacterial OTUs in infant stool samples against their chronological age using *randomForest* in R as previously described[39]. Default parameters were used with the following exceptions: ntree = 10,000, importance = TRUE. Tenfold cross-validation was performed using the *rfcv* function over 100 iterations to estimate the minimum number of features needed to accurately predict microbiota age. The features most important for prediction were identified over 100 iterations of the *importance* function, and a sparse model consisting of the 50 most important features was constructed and trained on a set of n=255 normal BMI infants (n=356 fecal samples) randomly selected from the larger normal BMI infant set (including 50% of normal BMI GMC1 infants [n=54]; excluded GMC2 and GMC3 infants). This model was validated in the remaining n=54 normal BMI GMC1 infants, and then applied to all remaining GMC1-3 infants to predict microbiota age. Microbiota-for-age z-scores (MAZ) was computed as previously described[39], enabling comparisons of microbiota maturity as the metric accounts for differing variance in predicted microbiota age throughout infant development.

Metagenomic processing and data analysis

DNA was extracted from n=28 GMC1 (n=15 Normal, n=13 OW/OB) and n=32 GMC3 (n=15 Normal, n=17 OW/OB; **Supplementary Table 8**) stool samples using the modified CTAB methods described above – samples selected had the highest posterior probability of GMC membership and sufficient remaining material for paired metagenomic and metabolomic profiling. Extracted DNA was sent to the Vincent J. Coates Genomic Sequencing Laboratory at the California Institute for Quantitative Biosciences for library preparation and 150-bp paired-end sequencing on an Illumina HiSeq 4000 (www.qb3.berkeley.edu/gsl). Only samples with >25,000 raw reads in each direction (>50,000 total raw reads) were included in this study (n=43, indicated within parentheses in **Supplementary Table 8**). The median number of raw reads per sample was 13,541,440 (IQR 6,300,000). The median number of reads following Q15 quality trimming and filtering human DNA using Bbduk v38.73 (https://sourceforge.net/projects/bbmap/) was 13,367,212 (IQR 2,073,330). All analyses were performed on trimmed and filtered reads. HUMAnN2 v2.8.1[74] was used to identify genes, level4ECs and functional MetaCyc pathways from the short-reads, and to normalize outputs into copies per million (CPM). Zero-inflated compound Poisson regression (*MaAsLin2*[75] package) was used to determine pathways that differed in relative abundance between GMCs, which were corrected for multiple testing using Benjamini-Hochberg (BH) False Discovery Rate (FDR). Significantly different pathways (P_{FDR} <0.25) were visualized using BioCyc's Pathway Collage and overlaid with $\log_2$ FC paired level4EC and metabolite (see *Metabolomic profiling*) values.

Metabolomic profiling

Stool samples (200 mg) sent for metagenomic profiling (n=60, **Supplementary Table 8**) were provided to Metabolon (Durham, NC) for Ultrahigh Performance Liquid Chromatography/Tandem Mass Spectrometry (UPLC-MS/MS) and Gas Chromatography-Mass Spectrometry (GC-MS) using their standard protocol (http://www.metabolon.com/). Identified compounds were compared to Metabolon's in-house library of purified standards, which includes more than 3,300 commercially available compounds.

In vitro exposure of enterocytes to cell-free fecal products

Fecal samples from 4 GMC1 Normal BMI, 1 GMC1 OW/OB, 5 GMC3 Normal BMI and 7 GMC3 OW/OB infants with metagenomic and metabolic profiles were used (biological replicates) to prepare cell-free fecal products and for microbial isolation (see *Microbial isolation and supernatant*); excluded samples from these groups had insufficient material. Stool samples were homogenized 1 g/ml in pre-warmed phosphate-buffered saline (PBS) containing 20% fetal bovine serum (FBS). Samples were vortexed, incubated (37 °C, 10 min) and centrifuged (14,000 rpm, 30 min). Supernatant was filtered through a 0.2 µm filter before being used in the enterocyte assay described below.

Caco-2 enterocytes between passage numbers 24 and 30 were seeded at 20,000 cells/cm^2 into tissue culture-treated 96-well flat bottom plates (Corning) and incubated for 14 days in Modified Eagle Medium (MEM, Gibco) media supplemented with 10% FBS (Gibco), 1x Non-Essential Amino Acids Solution (NEAA, Gibco), 100 U/ml penicillin

and 100 µl/ml streptomycin (Gibco). Media was changed every other day. Oleic acid (Sigma Aldrich) was used as the substrate to measure fatty acid uptake and triglyceride egress by enterocytes. Differentiated enterocytes were co-incubated with 500 µM oleate-albumin and 5% v/v cell-free fecal products for 24 h at 37 °C. Control exposures included no treatment and sterile 20% FBS in PBS.

Caco-2 enterocyte RNA sequencing and data analysis

RNA was extracted from Caco-2 enterocytes treated with cell-free fecal products using the RNAqueous kit (ThermoFisher) and quantified using Qubit RNA HS assay (ThermoFisher). Extracted RNA was sent to the Vincent J. Coates Genomic Sequencing Laboratory at the California Institute for Quantitative Biosciences for library preparation and 150-bp paired-end sequencing on an Illumina NovaSeq 6000 (www.qb3.berkeley.edu/gsl). Demultiplexed paired-end reads were quality filtered and Q20 trimmed, removed of PCR duplicates and Illumina adapters using *HTStream* (https://github.com/s4hts/HTStream) and aligned to the human genome (Hg38 release) using *STAR*[76] with ENCODE recommended parameters. Features were assigned to transcripts using *STAR* and normalized using *DESeq2*. Differential expression was evaluated using *DESeq2* genes with at least 20 reads per gene in respective sample grouping. Log-normalized read counts were obtained from *DESeq2* package, genes were filtered for presence in 75% of samples per comparison group, top variable genes were identified by the coefficient of variance and used to calculate principal components of Euclidean distances. Differential gene expression was mapped onto WikiPathways (www.wikipathways.org/instance/WP3942 and www.wikipathways.org/instance/WP111) using *RCy3*[77].

Enterocyte triglyceride accumulation and egress

Five µl conditioned media from enterocytes treated with cell-free fecal products was incubated with 200 µl Infinity triglyceride reagent (ThermoFisher Scientific) for 5 min at 37 °C and measured at absorbance 540 nm to quantify triglyceride egress from enterocytes. To quantify triglyceride accumulation in enterocytes, cells were washed twice with PBS before 250 µl hexane:isopropanol (60:40, HIP) washes were added to each well thrice. HIP washes were pooled into 1.5 ml tubes and dried down using a rotating speedvac at room temperature (Savant SC110, ThermoFisher Scientific). 100 µl 1% Triton-X was added to each 1.5 ml tube, incubated on a shaking incubator at room temperature for 4 h, and 5 µl was incubated with 200 µl Infinity triglyceride reagent for 5 min at 37 °C and measured at absorbance 540 nm. Cytotoxicity was quantified using the Pierce LDH cytotoxicity kit (ThermoFisher Scientific); all enterocytes exhibited ≥95% viability following treatment. Triglyceride quantities were normalized by total protein concentrations. To quantify total protein per well, HIP-extracted cells were dried at room temperature for 1 h before 300 µl 0.1M NaOH/0.1% sodium dodecyl sulfate was added to each well and incubated for 1 h at room temperature on a rotating incubator at 55 rpm. Ten µl from each sample was added to 200 µl protein assay dye (Bio-Rad), incubated at room temperature for 5 min and quantified at absorbance 595 nm.

Microbial isolation and supernatant

Microbial strains were isolated from the same n=17 stool samples extracted for cell-free fecal products (with metagenomic and metabolic profiles and sufficient remaining material). 25mg stool was incubated at 37 °C on Sabouraud Dextrose (SD) with 0.05

mg/ml chloramphenicol, Brain Heart Infusion (BHI), Luria-Bertani (LB) and Tryptic Soy Broth (TSB) agar. Colony species identification was confirmed by Sanger sequencing (Quintara Biosciences) amplicons generated using primer pairs 27 F and 1492 R[114] for the full-length bacterial 16S rRNA gene and ITS1 F and ITS4 R[115] for the fungal ITS2 gene. Taxonomy of sequences was assigned by SINA[78] against the curated SILVA[79] database. Isolated strains and additional strains from ATCC and lab collections were grown in media conditions detailed in **Supplementary Table 12**. Microbial cells were propagated for experiments by inoculation of single colonies into relevant sterile media and incubated overnight at 37 °C. To generate cell-free culture supernatants, overnight cultures were normalized spectrophotometrically to OD_{600} 0.1 in fresh media and grown at 37 °C under continuous OD_{600} measurement using a Cytation3 plate reader and Gen5 2.06 software (BioTek) every 30 min for 24 h. Cells were then normalized to OD_{600} 0.3 as needed, centrifuged for 10 min at 13,000 rpm and filtered through a 0.2µm filter. Cell-free supernatant was further centrifuged for 30 min at 16,000 rpm and room temperature through a pre-rinsed Amicon Ultra 3kDa filter (Millipore) as needed to obtain the <3kDa fraction. Methyl Tertiary Butyl Ether (MTBE) liquid-liquid extractions[80] were used to further separate the 3kDa fraction into polar and non-polar fractions. Briefly, solvents comprised of 3:1 v/v MTBE:methanol (solvent 1) and 3:1 v/v H_2O:methanol (solvent 2) were vortexed with the complete 3kDa fraction at a 3:2:1 v/v solvent 1:solvent 2:3kDa fraction ratio, then centrifuged at 20,000 x g and room temperature for 20 minutes. Upper (non-polar) and lower (polar) phases were separated into individual 2ml Eppendorf tubes and dried overnight at room temperature in a rotating speedvac (Savant SC110, ThermoFisher Scientific). The complete 3kDa

fraction was used as a positive control. Dried fractions were resuspended in SD media equivalent to the volume of original full 3kDa fraction subjected to MTBE extraction. 20% v/v cell-free supernatant and 3kDa fractions were used in the enterocyte assay described above, with filtered blank media as a negative control.

Fungal growth and morphology

Breast milk media was prepared by centrifuging human breast milk from three different donors (Innovative Research) at 3000 rpm for 10 min at 4 °C. Formula milk media was prepared from three common formula brands (Tippy Toe's Soy, Similac Pro-Total Comfort, Holle Bio Stage 1). Five grams of formula powder was mixed with 30 ml molecular grade water and centrifuged at 3000 rpm for 60 min at 4 °C. Both breast milk and formula were filtered through a 0.2 µm PES filter and diluted 1:2 in sterile PBS prior to use. Liquid *Candida albicans* DN2 cultures were grown for 24 h at 37 °C in SD media. Cultures were normalized to OD_{600} 0.1, washed with PBS at 3000 rpm for 5 min at room temperature to remove residual SD media, and incubated in SD, breast milk or formula media. Fungal cultures were incubated in a Cytation3 spectrophotometer (BioTek) at 37 °C for 24 h and OD_{600} was recorded every 30 min. Anaerobic 24 h OD_{600} readings were measured after resuspension with a pipette to homogenize floc structures that aggregated in wells. Growth curves were modeled by logistic regression in *growthcurver*[81]. Colony forming units (CFU) were counted from serial dilutions of 24 h cultures, grown on SD agar plates. 24 h cultures were also stained with propidium iodide and transferred onto glass slides for visualization by phase contrast and fluorescence microscopy.

Statistical analysis

Except where indicated, all analyses were performed using the R statistical programming language. Faith's phylogenetic diversity was calculated in QIIME and Student's or Welch's t-tests or Wilcoxon tests were calculated in R, depending on the data distribution. Distance matrices based on unweighted and weighted UniFrac[82], Bray-Curtis and Canberra algorithms were calculated in QIIME to assess compositional dissimilarity between samples and were visualized using PCoA plots in R and Emperor[83]. PERMANOVA was performed using *Vegan::Adonis*[84] in R to determine factors that significantly ($p<0.05$) explained variation in microbiota β-diversity.

Dirichlet Multinomial Mixture (DMM) modeling[35], which uses an unsupervised Bayesian approach to cluster samples, was used to identify clusters of subjects based on bacterial taxon relative abundance. Samples were stratified by time of sample collection (1-month, 6-month; n=349 and n=287, respectively), with rarefied counts collapsed at the genus-level to avoid extreme sparsity. The best-fitting DMM model was determined using the Laplace approximation to the negative log model evidence, testing up to 10 underlying microbiota classes. Each sample was assigned to a particular Gut Microbiota Class (GMC) based upon the maximum posterior probability of membership; GMCs were examined for good separation and interpretability.

Unadjusted and adjusted risk ratios (RRs) and corresponding 95% confidence intervals were calculated for OW/OB using log-binomial regression with maximum likelihood estimation, using PROC GENMOD in SAS version 9.4. Linear regression was used to test if BMI Z-scores were significantly different between GMCs. To determine which

OTUs differed in relative abundance between GMCs, unnormalized read counts were transformed using *DESeq2*[85] to identify log-Fold Change (FC) enrichment and corrected for multiple hypothesis testing using BH FDR (P_{FDR} <0.05). Taxon fold change in relative abundance between GMCs was $\log_{10}$ transformed for illustration on a phylogenetic tree using iTOL v5.6.1[86]. Metabolites exhibiting significantly different (P_{FDR} <0.05) scaled intensities between GMCs were illustrated using Cytoscape v3.7.2[87].

Data Availability

All raw sequences are deposited in the SRA Bioproject PRJNA648818.

References

1. Kumar, S. & Kelly, A. S. Review of Childhood Obesity: From Epidemiology, Etiology, and Comorbidities to Clinical Assessment and Treatment. *Mayo Clin. Proc.* **92**, 251–265 (2017).
2. Ley, R., Turnbaugh, P., Klein, S. & Gordon, J. Microbial ecology: human gut microbes associated with obesity. *Nature* **444**, 1022–3 (2006).
3. Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457**, 480–484 (2009).
4. Le Chatelier, E. *et al.* Richness of human gut microbiome correlates with metabolic markers. *Nature* **500**, 541–546 (2013).
5. Liu, R. *et al.* Gut microbiome and serum metabolome alterations in obesity and after weight-loss intervention. *Nat. Med.* **23**, 859–868 (2017).
6. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–131 (2006).
7. Turnbaugh, P. J., Bäckhed, F., Fulton, L. & Gordon, J. I. Diet-Induced Obesity Is Linked to Marked but Reversible Alterations in the Mouse Distal Gut Microbiome. *Cell Host Microbe* (2008). doi:10.1016/j.chom.2008.02.015
8. De Vadder, F. *et al. Microbiota-Produced Succinate Improves Glucose Homeostasis via Intestinal Gluconeogenesis*. *Cell Metabolism* **24**, (2016).
9. Samuel, B. S. *et al.* Effects of the gut microbiota on host adiposity are modulated by the short-chain fatty-acid binding G protein-coupled receptor, Gpr41. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 16767–16772 (2008).
10. Hoyles, L. *et al.* Molecular phenomics and metagenomics of hepatic steatosis in

non-diabetic obese women. *Nat. Med.* **24**, 1070–1080 (2018).

11. Cani, P. D. *et al.* Changes in gut microbiota control metabolic endotoxemia-induced inflammation in high-fat diet-induced obesity and diabetes in mice. *Diabetes* **57**, 1470–1481 (2008).
12. Dominguez-Bello, M. G. *et al.* Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 11971–11975 (2010).
13. Trasande, L. *et al.* Infant antibiotic exposures and early-life body mass. *Int. J. Obes.* **37**, (2013).
14. Ajslev, T. A., Andersen, C. S., Gamborg, M., Sørensen, T. I. A. & Jess, T. Childhood overweight after establishment of the gut microbiota: the role of delivery mode, pre-pregnancy weight and early administration of antibiotics. *Int. J. Obes.* **35**, 522–529 (2011).
15. Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486**, 222–227 (2012).
16. Baumann-dudenhoeffer, A. M., Souza, A. W. D., Tarr, P. I., Warner, B. B. & Dantas, G. Infant diet and maternal gestational weight gain predict early metabolic maturation of gut microbiomes. *Nat. Med.* **24**, (2018).
17. De Vadder, F. *et al.* Microbiota-generated metabolites promote metabolic benefits via gut-brain neural circuits. *Cell* **156**, 84–96 (2014).
18. Perry, R. J. *et al.* Acetate mediates a microbiome-brain-β-cell axis to promote metabolic syndrome. *Nature* **534**, 213–217 (2016).
19. Miyamoto, J. *et al.* Gut microbiota confers host resistance to obesity by

metabolizing dietary polyunsaturated fatty acids. *Nat. Commun.* **10**, 4007 (2019).

20. Arenz, S., Rückerl, R., Koletzko, B. & Von Kries, R. Breast-feeding and childhood obesity - A systematic review. *International Journal of Obesity* **28**, 1247–1256 (2004).
21. Bäckhed, F. *et al.* Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host Microbe* **17**, 690–703 (2015).
22. Stewart, C. J. *et al.* Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* **562**, 583–588 (2018).
23. Rackaityte, E. *et al.* Viable bacterial colonization is highly limited in the human intestine in utero. *Nat. Med.* **26**, 599–607 (2020).
24. Sprockett, D., Fukami, T. & Relman, D. A. Role of priority effects in the early-life assembly of the gut microbiota. *Nat. Rev. Gastroenterol. Hepatol.* **15**, 197–205 (2018).
25. Blanton, L. V *et al.* Gut bacteria that prevent growth impairments transmitted by microbiota from malnourished children. *Science (80-.).* **351**, (2016).
26. Dogra, S. *et al.* Dynamics of Infant Gut Microbiota Are Influenced by Delivery Mode and Gestational Duration and Are Associated with Subsequent. *MBio* **6**, 1–9 (2015).
27. Smith, M. I. *et al.* Gut microbiomes of Malawian twin pairs discordant for kwashiorkor. *Science (80-.).* **339**, 548–554 (2013).
28. Durack, J. *et al.* Delayed gut microbiota development in high-risk for asthma infants is temporarily modifiable by Lactobacillus supplementation. *Nat. Commun.* **9**, 1–9 (2018).

29. Stokholm, J. *et al.* Maturation of the gut microbiome and risk of asthma in childhood. *Nat. Commun.* **9**, 1–10 (2018).

30. Cox, L. M. *et al.* Altering the Intestinal Microbiota during a Critical Developmental Window Has Lasting Metabolic Consequences. *Cell* **158**, 705–721 (2014).

31. Cho, I. *et al.* Antibiotics in early life alter the murine colonic microbiome and adiposity. *Nature* **488**, 621–626 (2012).

32. Tun, H. M. *et al.* Roles of Birth Mode and Infant Gut Microbiota in Intergenerational Transmission of Overweight and Obesity From Mother to Offspring. *JAMA Pediatr.* 1–10 (2018). doi:10.1001/jamapediatrics.2017.5535

33. Kalliomäki, M., Collado, M. C., Salminen, S. & Isolauri, E. Early differences in fecal microbiota composition in children may predict overweight. *Am. J. Clin. Nutr.* **87**, 534–8 (2008).

34. Luoto, R. *et al.* Initial dietary and microbiological environments deviate in normal-weight compared to overweight children at 10 years of age. *J. Pediatr. Gastroenterol. Nutr.* **52**, 90–95 (2011).

35. Holmes, I., Harris, K. & Quince, C. Dirichlet Multinomial Mixtures: Generative Models for Microbial Metagenomics. *PLoS One* **7**, (2012).

36. Fujimura, K. E. *et al.* Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation. *Nat. Med.* **22**, 1187–1191 (2016).

37. Pallister, T. *et al.* Untangling the relationship between diet and visceral fat mass through blood metabolomics and gut microbiome profiling. *Int. J. Obes.* (2017). doi:10.1038/ijo.2017.70

38. Le Roy, C. I. *et al.* Dissecting the role of the gut microbiota and diet on visceral fat

mass accumulation. *Sci. Rep.* **9**, 1–10 (2019).

39. Subramanian, S. *et al.* Persistent gut microbiota immaturity in malnourished Bangladeshi children. *Nature* **510**, 417–421 (2014).
40. Del Chierico, F. *et al.* Gut microbiota profiling of pediatric nonalcoholic fatty liver disease and obese patients unveiled by an integrated meta-omics-based approach. *Hepatology* **65**, 451–464 (2017).
41. Walker, A. *et al.* Distinct signatures of host-microbial meta-metabolome and gut microbiome in two C57BL/6 strains under high-fat diet. *ISME J.* **8**, 2380–2396 (2014).
42. Midtvedt, A.-C., Carlstedt-Duke, B., Norin, K. E., Saxerholt, H. & Midtvedt, T. Development of Five Metabolic Activities Associated with the Intestinal Microflora of Healthy Infants. *J. Pediatr. Gastroenterol. Nutr.* **7**, 559–567 (1988).
43. Gameiro, A. *et al.* The neurotransmitters glycine and GABA stimulate glucagon-like peptide-1 release from the GLUTag cell line. *J. Physiol.* **569**, 761–772 (2005).
44. Meng, F. *et al.* New inducible genetic method reveals critical roles of GABA in the control of feeding and metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 3645–3650 (2016).
45. Hwang, I. *et al.* GABA-stimulated adipose-derived stem cells suppress subcutaneous adipose inflammation in obesity. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 11936–11945 (2019).
46. Brown, A. J. P., Brown, G. D., Netea, M. G. & Gow, N. A. R. Metabolism impacts upon candida immunogenicity and pathogenicity at multiple levels. *Trends Microbiol.* **22**, 614–622 (2014).

47. He, X. *et al.* Metabolic phenotype of breast-fed infants, and infants fed standard formula or bovine MFGM supplemented formula: a randomized controlled trial. *Sci. Rep.* **9**, 1–13 (2019).

48. Ballard, O. & Morrow, A. L. Human Milk Composition. Nutrients and Bioactive Factors. *Pediatric Clinics of North America* **60**, 49–74 (2013).

49. Rogier, E. W. *et al.* Secretory antibodies in breast milk promote long-term intestinal homeostasis by regulating the gut microbiota and host gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 3074–3079 (2014).

50. Hu, P., Zhao, F., Zhu, W. & Wang, J. Effects of early-life lactoferrin intervention on growth performance, small intestinal function and gut microbiota in suckling piglets. *Food Funct.* **10**, 5361–5373 (2019).

51. Marcobal, A. *et al.* Consumption of Human Milk Oligosaccharides by Gut-Related Microbes. *J. Agric. Food Chem* **58**, 5334–5340 (2010).

52. Bournat, J. C. & Brown, C. W. Mitochondrial dysfunction in obesity. *Curr. Opin. Endocrinol. Diabetes Obes.* **17**, 446–452 (2010).

53. Muoio, D. M. & Newgard, C. B. Obesity-related derangements in metabolic regulation. *Annu. Rev. Biochem.* **75**, 367–401 (2006).

54. D'Aquila, T., Zembroski, A. S. & Buhman, K. K. Diet induced obesity alters intestinal cytoplasmic lipid droplet morphology and proteome in the postprandial response to dietary fat. *Front. Physiol.* **10**, 1–13 (2019).

55. Araújo, J. R. *et al.* Fermentation Products of Commensal Bacteria Alter Enterocyte Lipid Metabolism. *Cell Host Microbe* **27**, 358-375.e7 (2020).

56. Monteiro-Sepulveda, M. *et al.* Jejunal T Cell Inflammation in Human Obesity

Correlates with Decreased Enterocyte Insulin Signaling. *Cell Metab.* **22**, 113–124 (2015).

57. Walther, T. C. & Farese, R. V. Lipid droplets and cellular lipid metabolism. *Annu. Rev. Biochem.* **81**, 687–714 (2012).

58. Aichbhaumik, N. *et al.* Prenatal exposure to household pets influences fetal immunoglobulin E production. *Clin. Exp. Allergy* **38**, 1787–1794 (2008).

59. Ogden, C. L. *et al.* Centers for Disease Control and Prevention 2000 growth charts for the United States: Improvements to the 1977 National Center for Health Statistics version. *Pediatrics* **109**, 45–60 (2002).

60. DeAngelis, K. M. *et al.* Selective progressive response of soil microbial community to wild oat roots. *ISME J.* **3**, 168–178 (2009).

61. Fujimura, K. E. *et al.* SUPPLEMENTARY Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation. *Nat. Med.* **22**, 1187–1191 (2016).

62. Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6**, 1621–4 (2012).

63. Ihrmark, K. *et al.* New primers to amplify the fungal ITS2 region - evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiol. Ecol.* **82**, 666–677 (2012).

64. Magoc, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).

65. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–6 (2010).

66. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–1 (2010).

67. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194–200 (2011).

68. Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**, 996–8 (2013).

69. DeSantis, T. Z. *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72**, 5069–5072 (2006).

70. Caporaso, J. G. *et al.* PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26**, 266–267 (2010).

71. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).

72. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10 (2011).

73. Kõljalg, U. *et al.* Towards a unified paradigm for sequence-based identification of fungi. *Mol. Ecol.* **22**, 5271–5277 (2013).

74. Franzosa, E. A. *et al.* Species-level functional profiling of metagenomes and metatranscriptomes. *Nat. Methods* **15**, 962–968 (2018).

75. Mallick, H. *et al.* Multivariable Association in Population-scale Meta-omics Studies.

76. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–

21 (2013).

77. Gustavsen, J. A., Pai, S., Isserlin, R., Demchak, B. & Pico, A. R. RCy3: Network Biology using Cytoscape from within R. *bioRxiv* 1–20 (2019). doi:10.1101/793166

78. Pruesse, E., Peplies, J. & Glöckner, F. O. SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**, 1823–1829 (2012).

79. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).

80. Salem, M. A., Jüppner, J., Bajdzienko, K. & Giavalisco, P. Protocol: A fast, comprehensive and reproducible one-step extraction method for the rapid preparation of polar and semi-polar metabolites, lipids, proteins, starch and cell wall polymers from a single sample. *Plant Methods* **12**, 1–15 (2016).

81. Sprouffske, K. & Wagner, A. Growthcurver: An R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinformatics* **17**, 172 (2016).

82. Lozupone, C. & Knight, R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **71**, 8228–35 (2005).

83. Vázquez-Baeza, Y., Pirrung, M., Gonzalez, A. & Knight, R. EMPeror: a tool for visualizing high-throughput microbial community data. *Gigascience* **2**, 16 (2013).

84. Jari Oksanen, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, D. M., Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, E. S. & and Helene Wagner. vegan: Community Ecology Package. R package version 2.5-6. (2019). Available at: https://cran.r-

project.org/package=vegan.

85. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

86. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).

87. Shannon, P. *et al.* Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
Germaine Yong
9E1938FD7F7B4CA...
Author Signature

12/1/2020
Date