

Supplementary Information

Comparative characterization of the infant gut microbiome and their maternal lineage by a multi-omics approach.

Authors: Tomás Clive Barker-Tejeda^{1,2†}; Elisa Zubeldia-Varela^{1,2†}; Andrea Macías-Camero^{1,2†}; Lola Alonso³; Isabel Adoración Martín-Antoniano^{2,3,4}; María Fernanda Rey-Stolle¹; Leticia Mera-Berriatua²; Raphaëlle Bazire^{5,6}; Paula Cabrera-Freitag⁷; Meera Shanmuganathan⁸; Philip Britz-McKibbin⁸; Carles Ubeda^{9,10}; M. Pilar Francino^{10,11}; Domingo Barber²; María Dolores Ibáñez-Sandín^{5,6}; Coral Barbas¹; Marina Pérez-Gordo^{2*}; Alma Villaseñor^{1,2*}

Affiliations:

¹Centro de Metabolómica y Bioanálisis (CEMBIO), Facultad de Farmacia, Universidad San Pablo-CEU, CEU Universities, Boadilla del Monte, Spain.

²Departamento de Ciencias Médicas Básicas, Instituto de Medicina Molecular Aplicada (IMMA) Nemesio Díez, Facultad de Medicina, Universidad San Pablo-CEU, CEU Universities, Boadilla del Monte, Spain.

³Genetic and Molecular Epidemiology Group, Spanish National Cancer Research Centre (CNIO), Madrid, Spain.

⁴Instituto de Estudios de las Adicciones IEA-CEU, Universidad San Pablo-CEU, CEU Universities, Madrid, Spain

⁵Department of Allergy, Hospital Infantil Niño Jesús, Fib-HNJ, Madrid, Spain.

⁶Instituto de Investigación Sanitaria-La Princesa, Madrid, Spain.

⁷Pediatric Allergy Unit, Allergy Service, Hospital General Universitario Gregorio Marañón, and Gregorio Marañón Health Research Institute, Madrid, Spain.

⁸Department of Chemistry and Chemical Biology, McMaster University, Hamilton, ON, Canada.

⁹Fundació per al Foment de la Investigació Sanitària i Biomèdica de la Comunitat Valenciana (FISABIO), Valencia, Spain

¹⁰CIBER en Epidemiología y Salud Pública, Madrid, Spain

¹¹Joint Research Unit in Genomics and Health, Fundació per al Foment de la Investigació Sanitària i Biomèdica de la Comunitat Valenciana (FISABIO) and Institut de Biologia Integrativa de Sistemes (Universitat de València / Consejo Superior de Investigaciones Científicas), València, Spain

[†]: These Authors have equally contributed

^{*}: Corresponding authors

Corresponding authors:

*Alma Villaseñor, PhD

CEMBIO, Centro de Metabolómica y Bioanálisis, Facultad de Farmacia, Universidad San Pablo CEU. Avda. Montepríncipe s/n, 28668 Boadilla del Monte, Madrid, SPAIN.

E-mail: alma.villasenor@ceu.es

*Marina Pérez-Gordo, PhD

Instituto de Medicina Molecular Aplicada (IMMA) Nemesio Díez, Departamento de Ciencias Médicas Básicas, Facultad de Medicina, Universidad San Pablo CEU, 28660 Madrid, Spain

E-mail: marina.perezgordo@ceu.es

Supplementary Figure Titles:

Figure S1: Justification of sample numbers and outliers in the different steps.

Figure S2: Graphical representation of age distributions and age outliers in all techniques.

Figure S3: Quality Assurance and normalizations in GC-QTOF-MS analysis.

Figure S4: MSI-CE-MS workflow and multivariate models.

Figure S5: Correlations between the metabolites that were measured in both techniques (MSI-CE-MS and GC-QTOF-MS).

Figure S6: Supplementary statistics figures for metabolomics.

Figure S7: Taxonomic composition of faecal samples using 16S rRNA and shotgun gene sequencing.

Figure S8: Correlations between metabolites and bacterial taxa that are known producers of these metabolites.

Figure S9: Cytoscape representation of the links between ASVs, KOs, and metabolites obtained from mixOmics and DIABLO.

Figure S10: Experimental design of the QCs preparation and worklists for GC-QTOF-MS equipment used for large-scale sample analysis.

Supplementary Tables:

Table S1: Main characteristics for each age group, with their descriptive statistics.

Table S2: Pathway analysis of the 88 significant metabolites associated with age after application of the linear mixed effect model with the correction for multiple test comparisons FDR p value < 0.05. Significant and relevant metabolic pathways for the host-microbiota crosstalk are shown according to their FDR p-value ≤ 0.05 .

Table S3: Correlations between gamma-aminobutyric acid (GABA) and bacterial phyla, genera, and species using spearman's rank correlation coefficient.

Table S4: Correlations between short-chain fatty acids (SCFA) and bacterial genera and species using spearman's rank correlation coefficient.

Table S5: Correlations between the matched bacterial genera between 16S rRNA gene sequencing and shotgun sequencing using two-sided Pearson correlation test.

Table S6: Description of the KOs from Heatmap and circosplot from Figure 9.

Supplementary Data files:

Supplementary Data 1: Significant identified metabolites detected in the GC-QTOF-MS analysis, with their chemical information, quality parameters and statistics. The statistical test applied was linear mixed-effects model with the correction for multiple test comparisons, FDR p-value < 0.05.

Supplementary Data 2: Significant metabolites detected in the MSI-CE-TOF-MS analysis, with their chemical information, quality parameters and statistics. The statistical test applied was linear mixed-effects model with the correction for multiple test comparisons, FDR p-value < 0.05.

Supplementary Data 3: Comparisons between the bacterial phyla and genera detected by 16S rRNA gene and shotgun sequencing. Data used for the stacked bars in Figure 7A-D.

Supplementary Data 4: Significant results at phylum level for 16S rRNA gene and shotgun sequencing. The statistical test applied was linear mixed-effects model with the correction for multiple test comparisons, FDR p-value < 0.05.

Supplementary Data 5: Significant results at genus level for 16S rRNA gene and shotgun sequencing. The statistical test applied was linear mixed-effects model with the correction for multiple test comparisons, FDR p-value < 0.05.

Supplementary Data 6: Significant results at species level for shotgun sequencing. The statistical test applied was linear mixed-effects model with the correction for multiple test comparisons, FDR p-value < 0.05.

Supplementary Data 7: Significant results from the linear mixed effects model for the KEGG Orthologs (KOs).

Supplementary Data 8: Significant results from the linear mixed effects model for the pathways at BRITE level B.

Supplementary Data 9: Significant results from the linear mixed effects model for the pathways at BRITE level C.

Supplementary Data 10: Detailed information for each sample, including epidemiological variables, techniques in which it was measured and models where it is included.

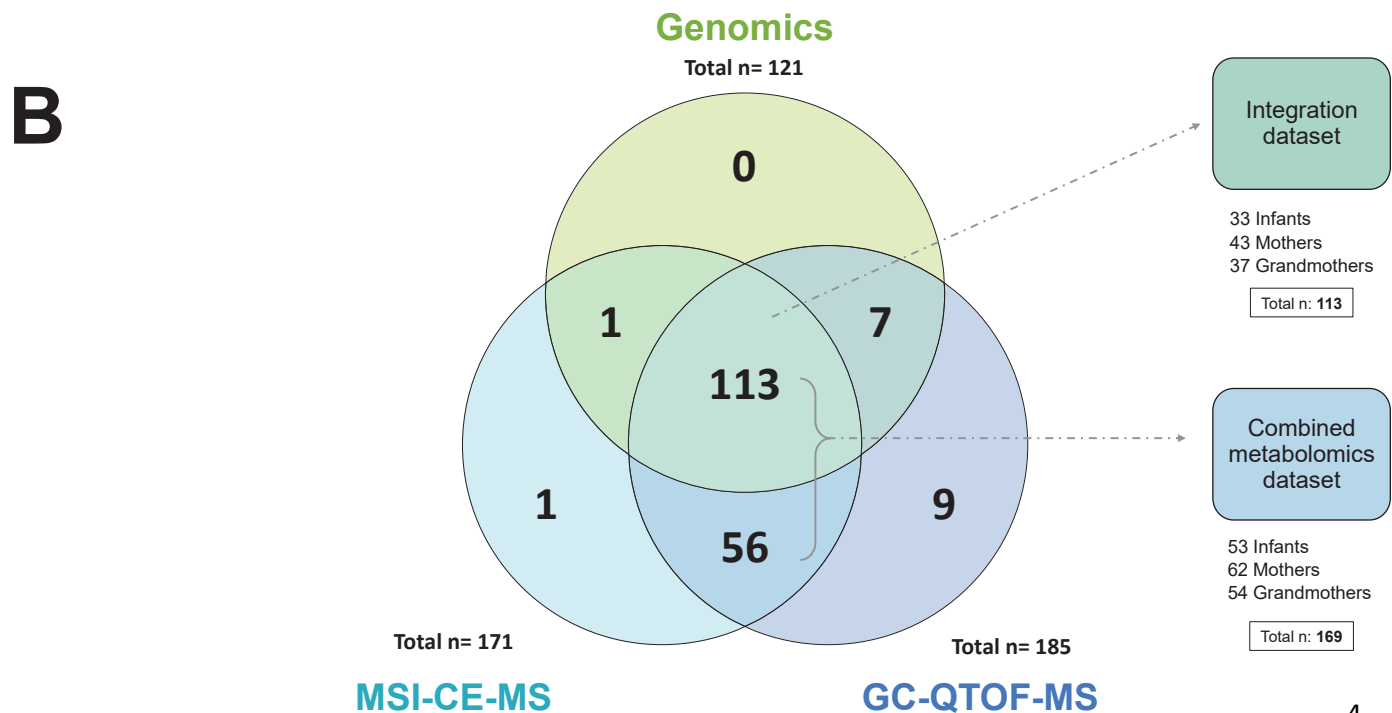
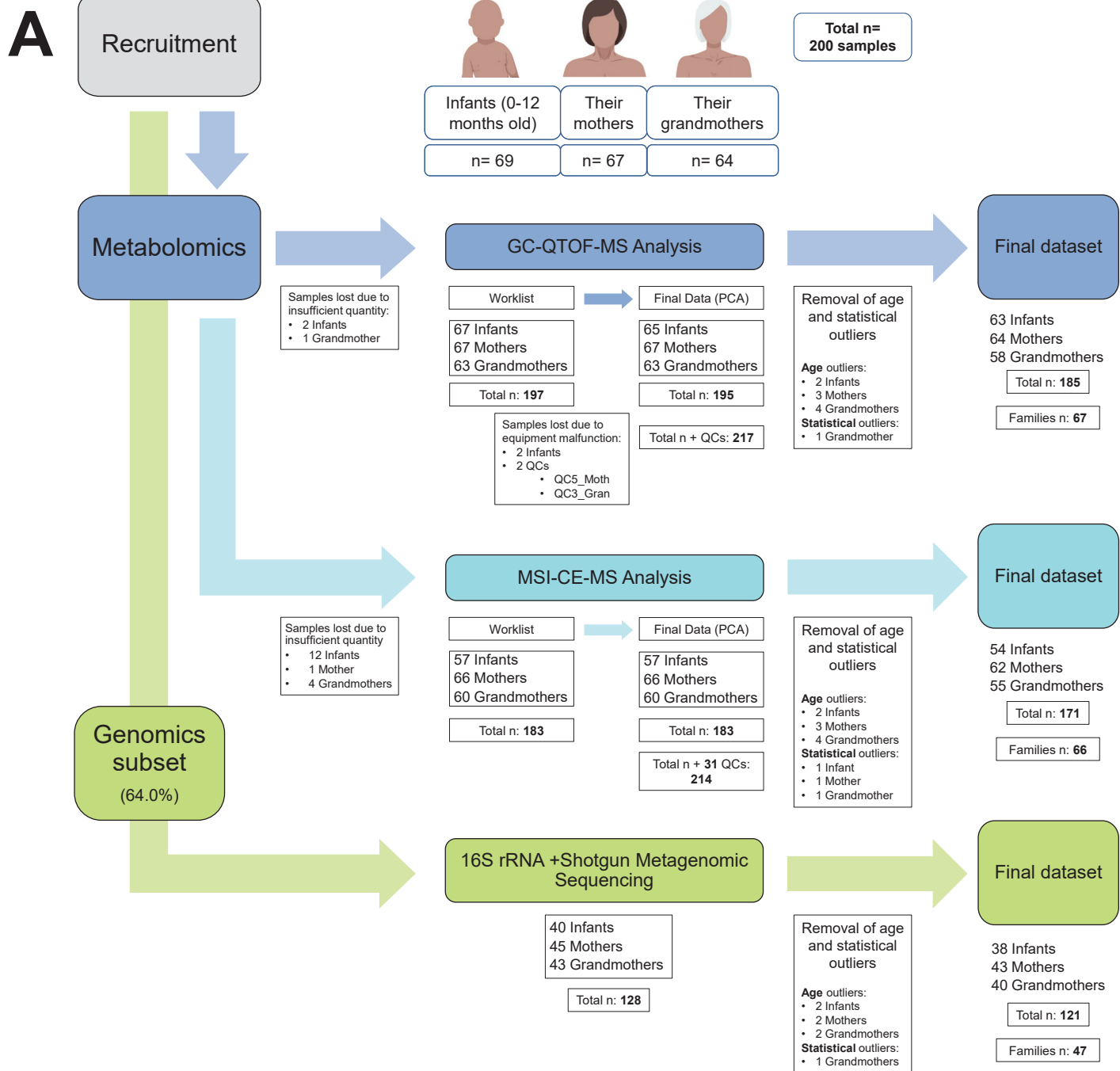
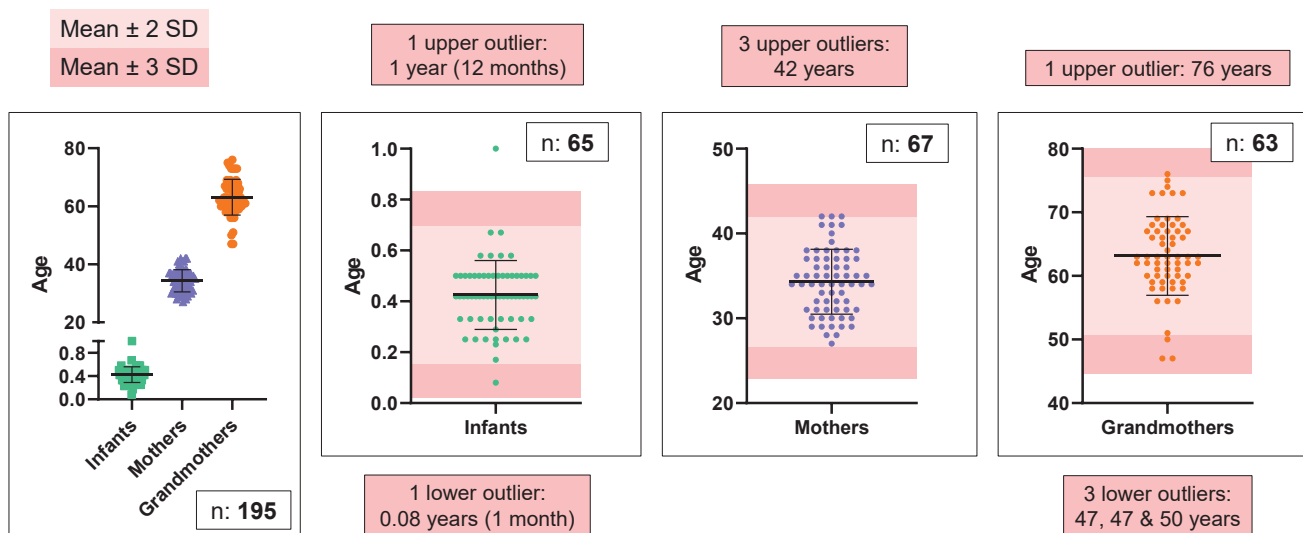


Figure S1: Justification of sample numbers and outliers in the different steps. **A:** Workflow of the study, from recruitment to the final dataset of each experiment. The number of samples and the explanation for the lost samples is given in each step. Age outliers are further detailed in Figure S2. **B:** Venn diagram of the samples measured in all techniques. The 113 common samples were the ones selected for mixOmics integration, and the 169 common in both metabolomics platforms were used for the combined metabolomics dataset. Source data are provided as a Source Data file. Icons were created using biorender.com.

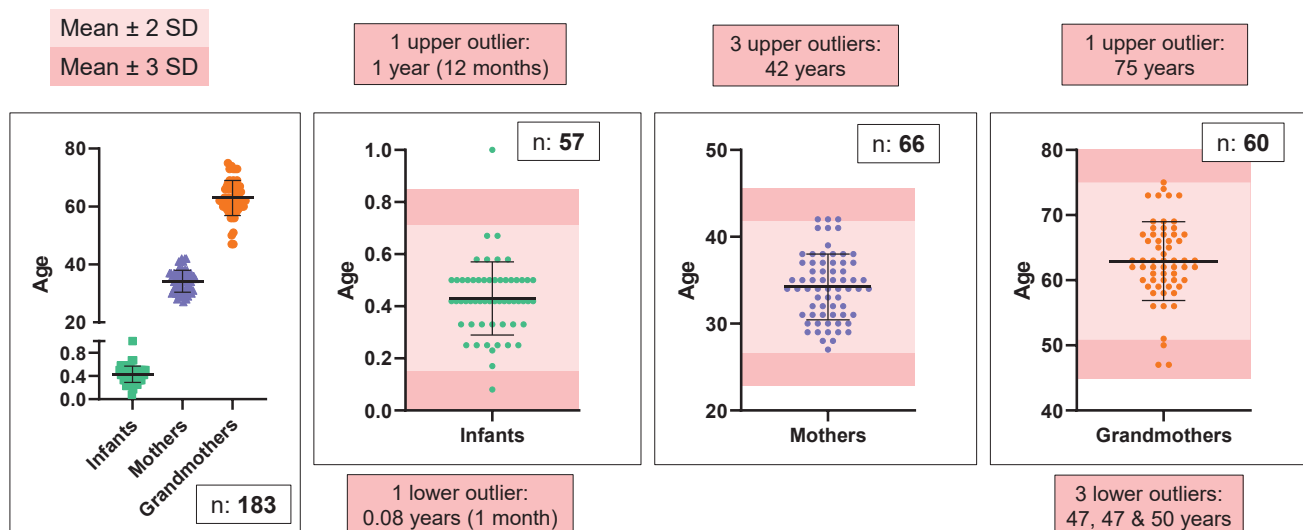
A

GC-QTOF-MS Analysis



B

MSI-CE-MS Analysis



C

16S rRNA +Shotgun Metagenomic Sequencing

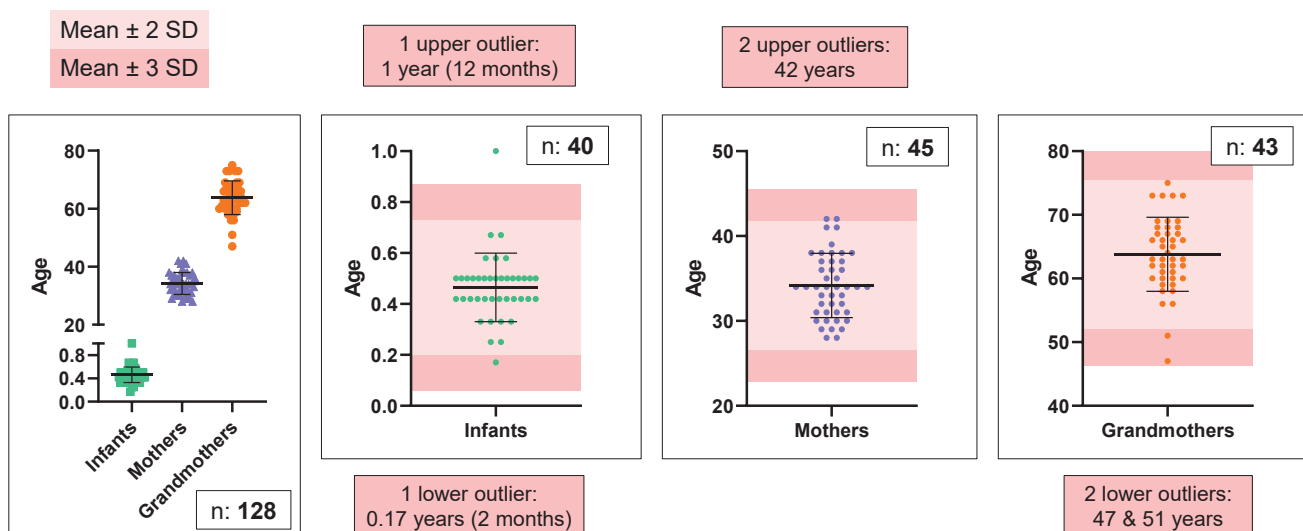
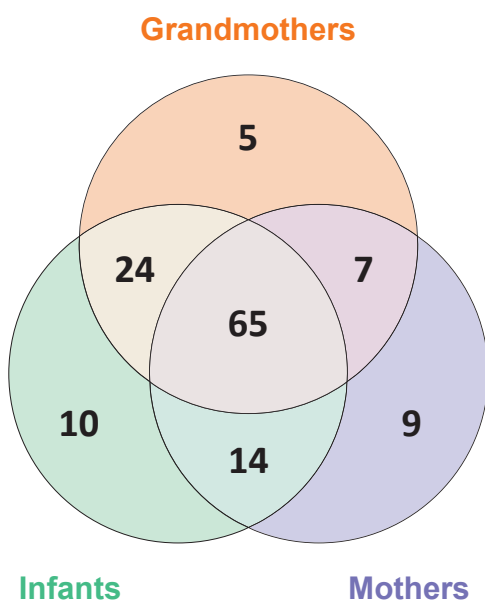


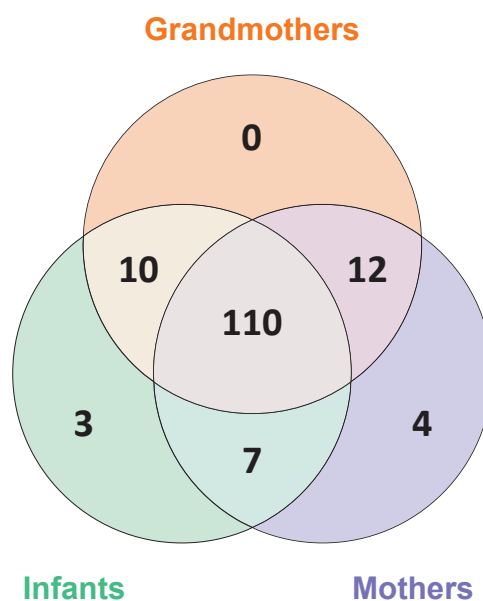
Figure S2: Graphical representation of age distributions and age outliers in all techniques. A: Age outliers in GC-QTOF-MS dataset. **B:** Age outliers in MSI-CE-MS dataset. **C:** Age outliers in Genomics (16S rRNA sequencing + shotgun metagenomics) dataset. Numbers of samples used A and B (n) are shown in each figure. Figures represent all samples in each age group, with the central line representing mean age of each group, error bars showing $\pm$ standard deviation (SD), lighter pink shaded area $\pm$ 2SD and darker pink $\pm$ 3SD. Samples lying outside the Mean $\pm$ 2SD area were considered age outliers and were excluded from further analyses. Source data are provided as a Source Data file. QC: quality control sample.

A Before Normalizations



Total features: 134

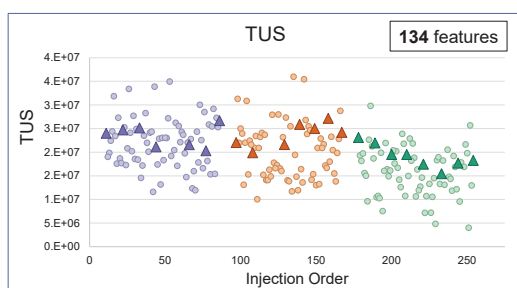
C After QC-SVRC + QC-Norm Normalizations



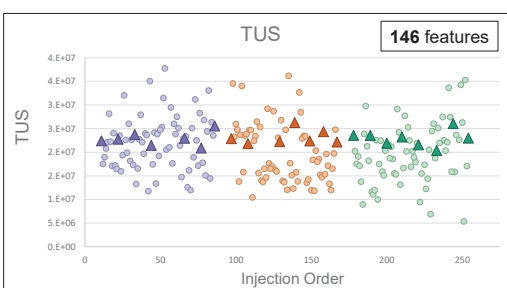
Total features: 146

B Before Normalizations

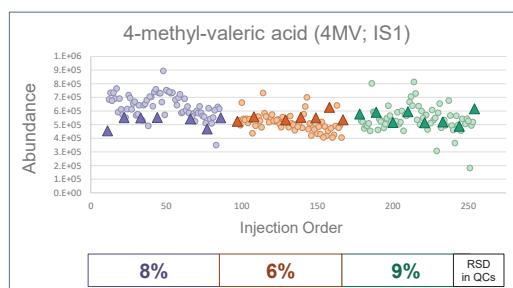
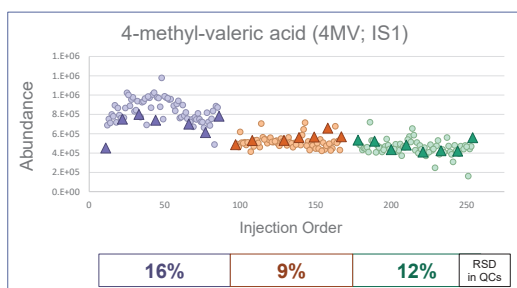
(i)



D After QC-SVRC + QC-Norm Normalizations



(ii)



(iii)

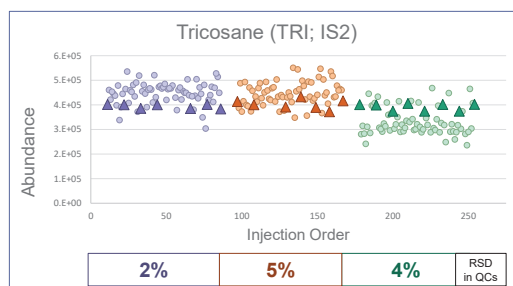
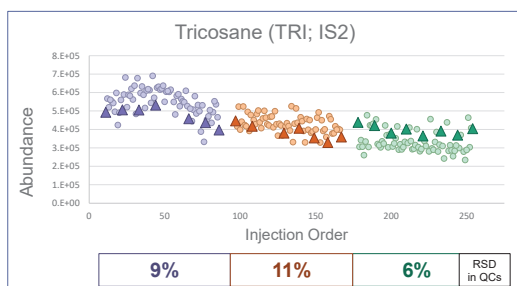


Figure S3: Quality Assurance and normalizations in GC-QTOF-MS analysis. **A:** Venn diagram of the number of features that passed the QA in each batch before normalizations, after the analysis of n = 67, n = 63, n = 65 of biologically independent samples for mothers, grandmothers, and infants, respectively. **B:** Quality parameters before normalizations, plotted by injection order: **(i)** Total Useful Signal (TUS), the sum of the abundances of all features that pass the QA in at least one batch; Internal Standards (IS): **(ii)** 4-methylvaleric acid (4MV), IS1, the IS that is added before the derivatization and measures the variability of the sample preparation as well as the analytical variability; **(iii)** tricosane (TRI), IS2, the IS that is added after derivatization and measures mostly the variability of the instrumental performance. **C:** Venn diagram of the number of features that passed the QA in each batch after the normalizations after the analysis of n = 67, n = 63, n = 65 of biologically independent samples for mothers, grandmothers, and infants, respectively. **D:** Quality parameters after the normalizations, plotted by injection order: **(i)** TUS; Internal Standards (IS): **(ii)** 4MV, IS1; **(iii)** TRI, IS2. Source data are provided as a Source Data file.

Graphs in **B** and **D** are plotted by injection order and show the three batches in the order they were measured: n = 67, n = 63, n = 65 of biologically independent samples for mothers, grandmothers, and infants, respectively. In general, the figures show that the normalization strategy was successful, as it reduced the analytical variability between the three batches – as is evidenced by the reduction in the RSD for the IS – and allowed a higher number of features to pass the QA.

RSD: relative standard deviation (also known as CV: coefficient of variance). 4MV: 4-methylvaleric acid; TRI: tricosane; QCs: quality control samples.

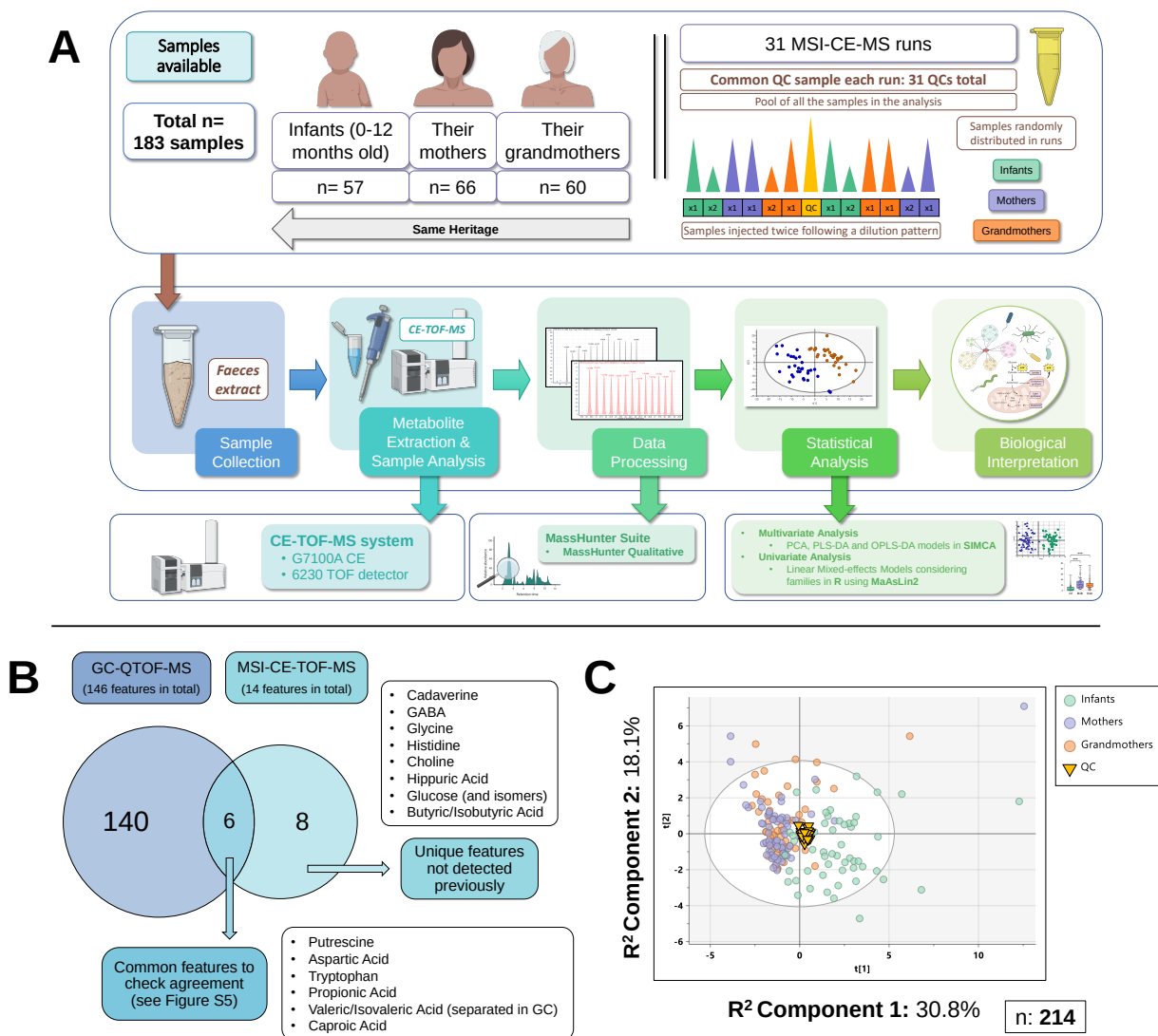


Figure S4: MSI-CE-MS workflow and multivariate models. **A:** Experimental design of the study, analysis, and workflow schematic. **B:** Venn diagram of the features detected in each metabolomics technique, features from GC-QTOF-MS were after the analysis of $n = 63$, $n = 64$, $n = 58$ of biologically independent samples for infants, mothers, and grandmothers, respectively, whereas features from MSI-CE-MS were obtained after the analysis of $n = 54$, $n = 62$, $n = 55$ of biologically independent samples for infants, mothers, and grandmothers, respectively. The 14 metabolites detected in a targeted search in MSI-CE-TOF-MS are shown, and further details about these metabolites are given in Supplementary Data 2. **C:** Unsupervised multivariate PCA-X model of all samples, including QCs, which are shown tightly clustered in the center. Number of samples used to build the model (n) is shown in the figure. The full explanation for sample numbers is presented in Figure S1 and S2. Source data are provided as a Source Data file. Icons were created using biorender.com.

R^2 : percentage of the variability of the samples that the selected component is able to explain. Values of R^2 closer to 1 indicate higher quality of the model. All multivariate models were built with univariate scaling (UV) and without any transformation of the data.

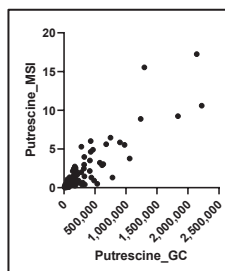
QC: quality control sample; PCA: Principal Component Analysis; PLS-DA: Partial Least Square Discriminant Analysis, OPLS-DA; Orthogonal PLS-DA.

Correlation

Putrescine

$$\rho = 0.9248$$

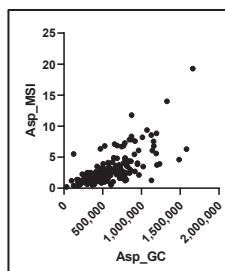
$$p < 0.001$$



Aspartic Acid

$$\rho = 0.6959$$

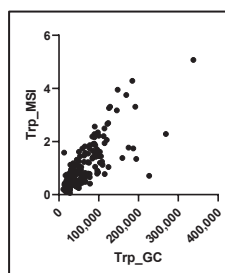
$$p < 0.001$$



Tryptophan

$$\rho = 0.7578$$

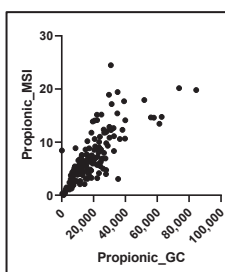
$$p < 0.001$$



Propionic Acid

$$\rho = 0.8200$$

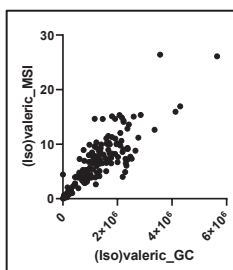
$$p < 0.001$$



Valeric / Isovaleric Acid

$$\rho = 0.8648$$

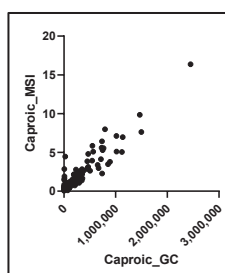
$$p < 0.001$$



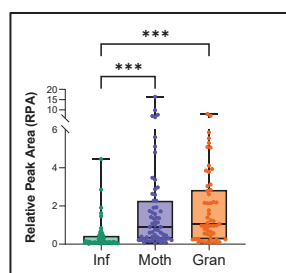
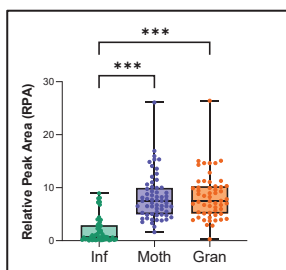
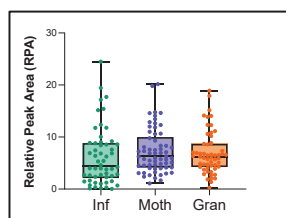
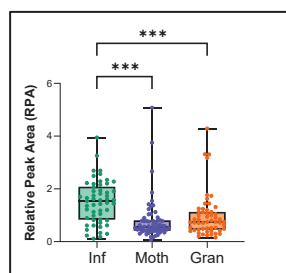
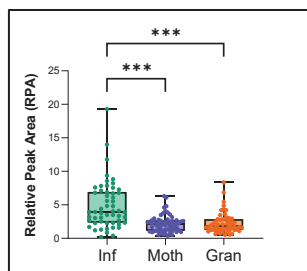
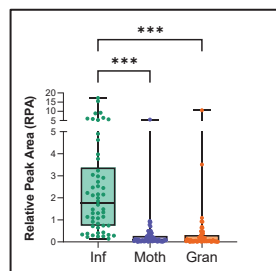
Caproic Acid

$$\rho = 0.8556$$

$$p < 0.001$$



MSI-CE-MS



GC-QTOF-MS

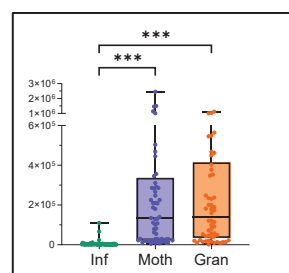
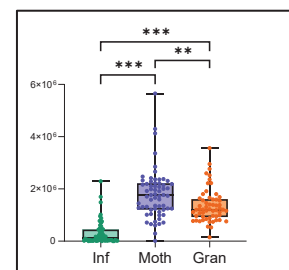
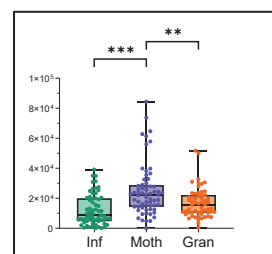
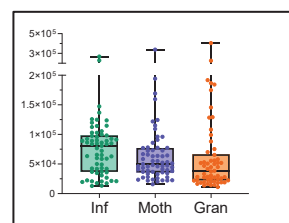
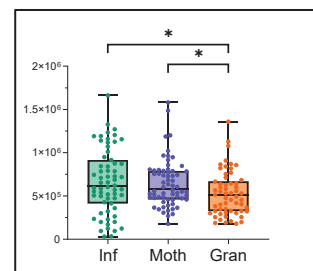
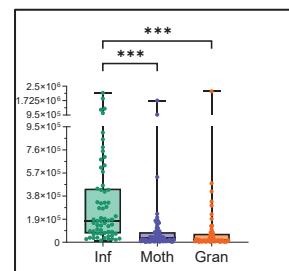


Figure S5: Correlations between the metabolites that were measured in both techniques (MSI-CE-MS and GC-QTOF-MS) using spearman's rank correlation coefficient. Correlations were calculated with the $n = 169$ of the common biologically independent samples between analytical techniques. ρ : spearman rho correlation value. p : p -value of the correlation. MSI-CE-MS plots with the $n = 171$ of the total biologically independent samples including the Infants, Mothers and Grandmothers groups. GC-QTOF-MS plots with the $n = 185$ of the total biologically independent samples including the Infants, Mothers and Grandmothers groups. Statistical test was linear mixed-effects model with the correction for multiple test comparisons, FDR p -value < 0.05 . *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$, exact p -values and FDR p -values are provided in Supplementary Data 1 and 2. Source data are provided as a Source Data file.

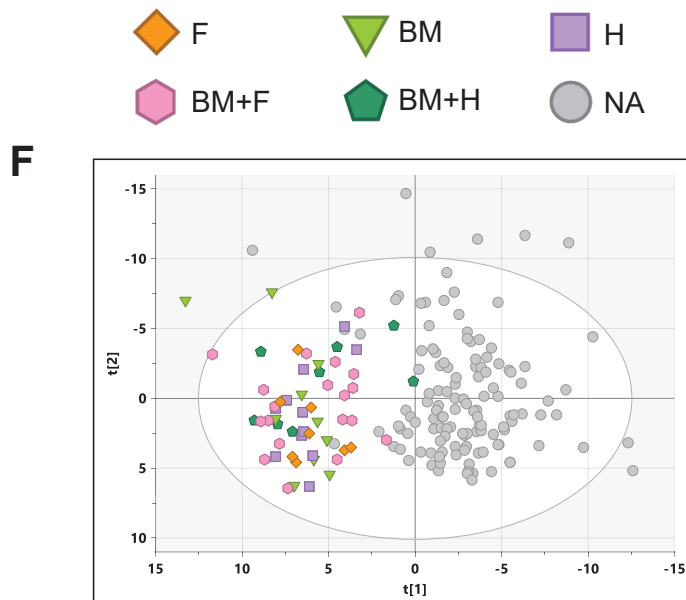
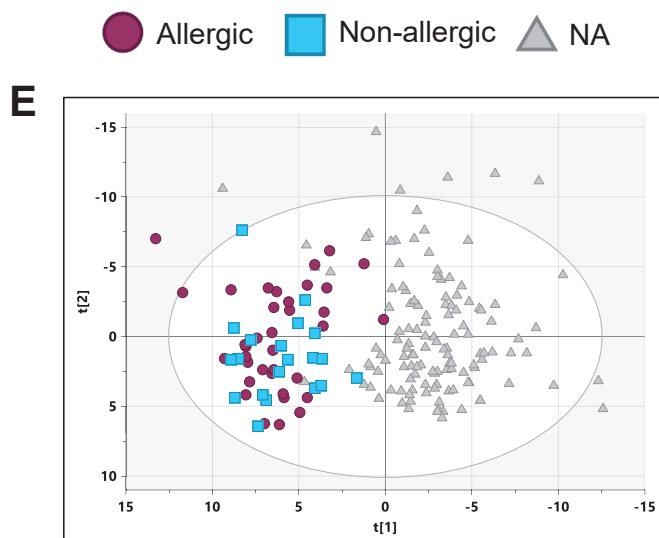
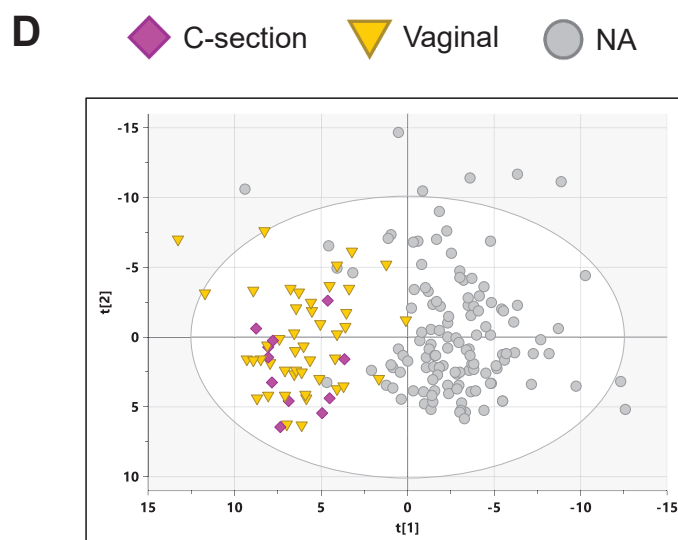
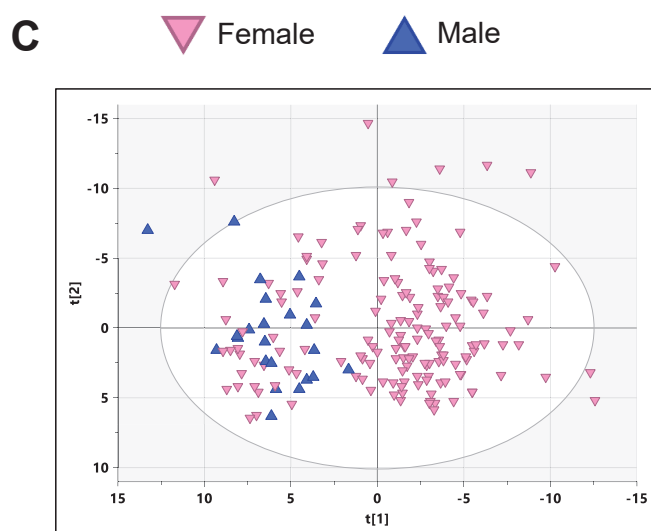
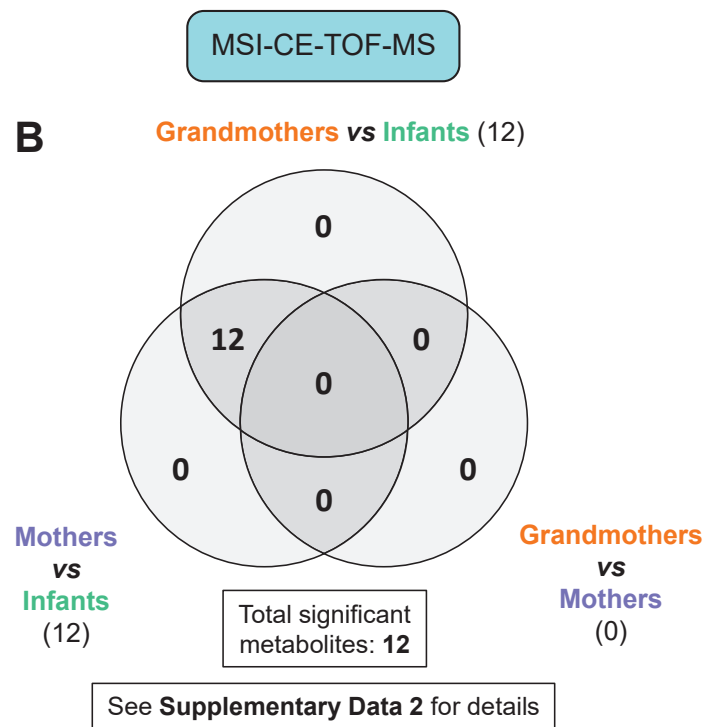
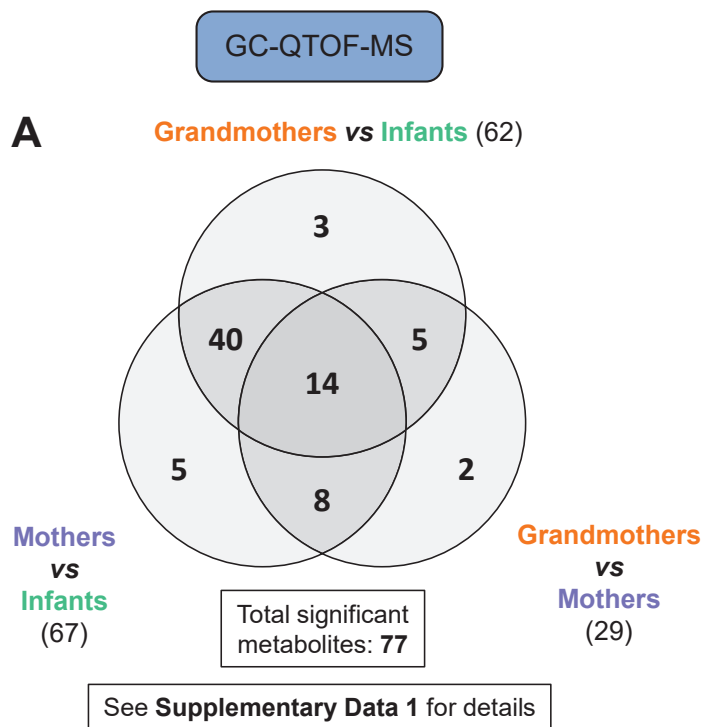


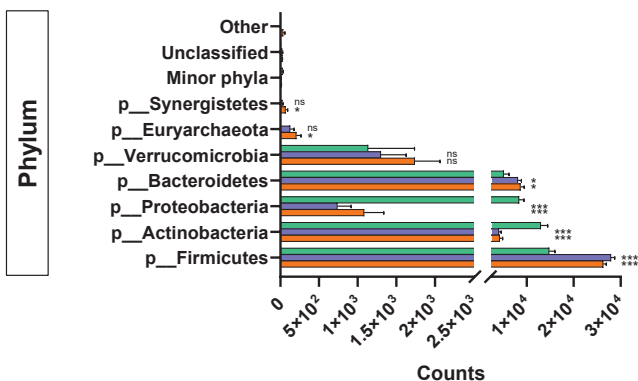
Figure S6: Supplementary statistics figures for metabolomics. **A-B:** Venn diagrams of the significant metabolites according to univariate statistics for the three comparisons for **(A)** GC-QTOF-MS data, and **(B)** MSI-CE-TOF-MS data. The statistical test was linear mixed-effects model with the correction for multiple test comparisons, FDR p-value < 0.05. Exact p-values, FDR p-values, and detailed information are provided in Supplementary Data 1 and 2. **C-F:** PCA models built with the combined metabolomics data, such as the one shown in Figure 2A. For these models, the number of samples are n = 33, n = 43, n = 37 of biologically independent samples for infants, mothers, and grandmothers, respectively. In these models, the samples are colored according to different covariables: sex **(C)**, mode of birth **(D)**, cow's milk allergy **(E)** and type of feeding of the infant **(F)**. The feeding types are named as follows: formula (F), breast milk (BM), hydrolyzed formula (H), combination of BM and F (BM+F) and combination of BM and H (BM+H). Source data are provided as a Source Data file.

The samples do not cluster according to these covariables, which proves that they have no impact on the differences according to the main variable (the age group).

A

16S rRNA gene sequencing

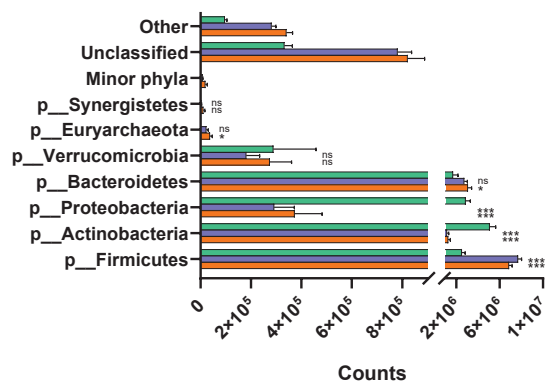
Infants Mothers Grandmothers



B

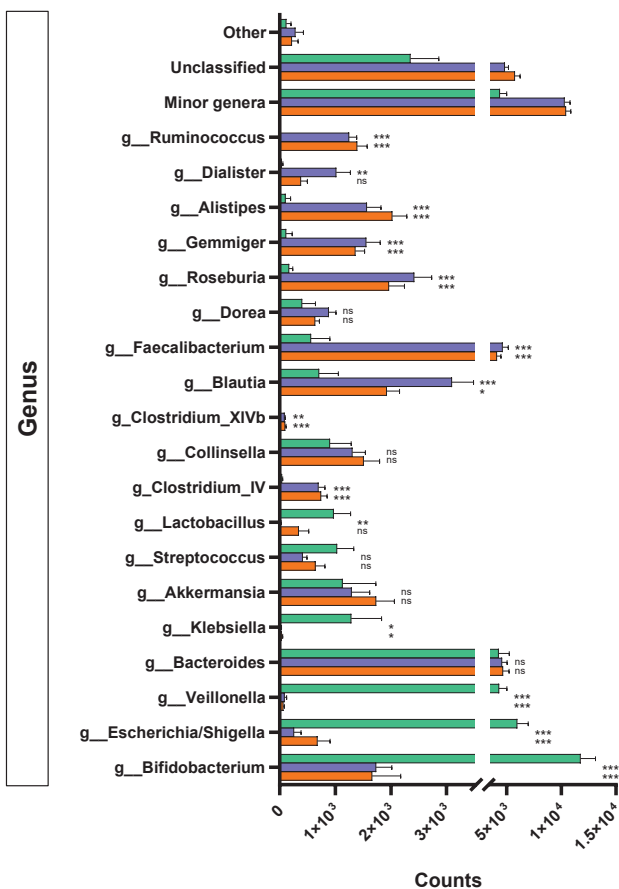
Shotgun sequencing

Infants Mothers Grandmothers



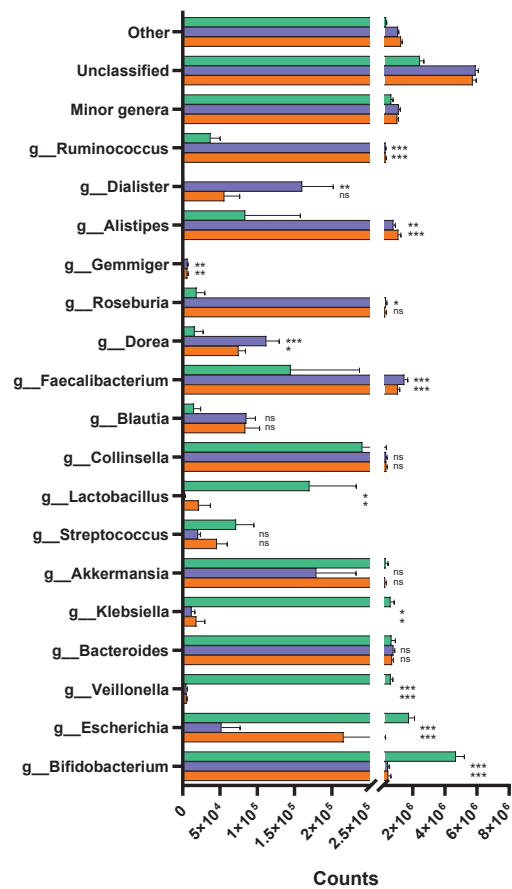
C

Infants Mothers Grandmothers



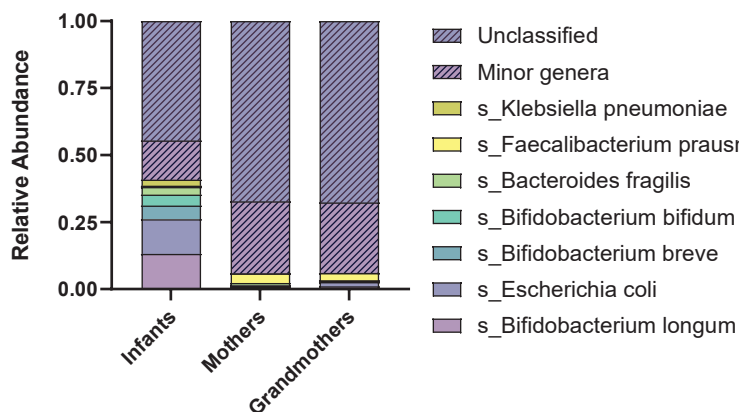
D

Infants Mothers Grandmothers



Shotgun sequencing

E



F

Infants Mothers Grandmothers

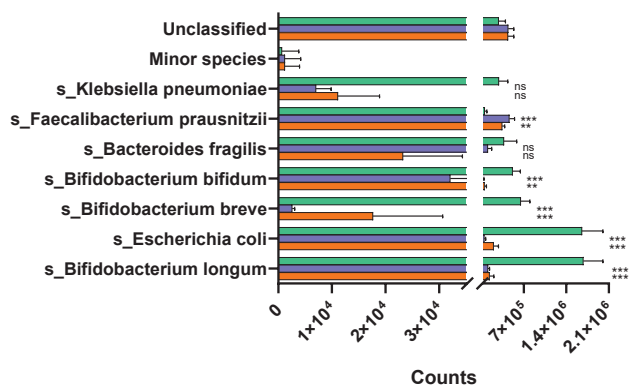
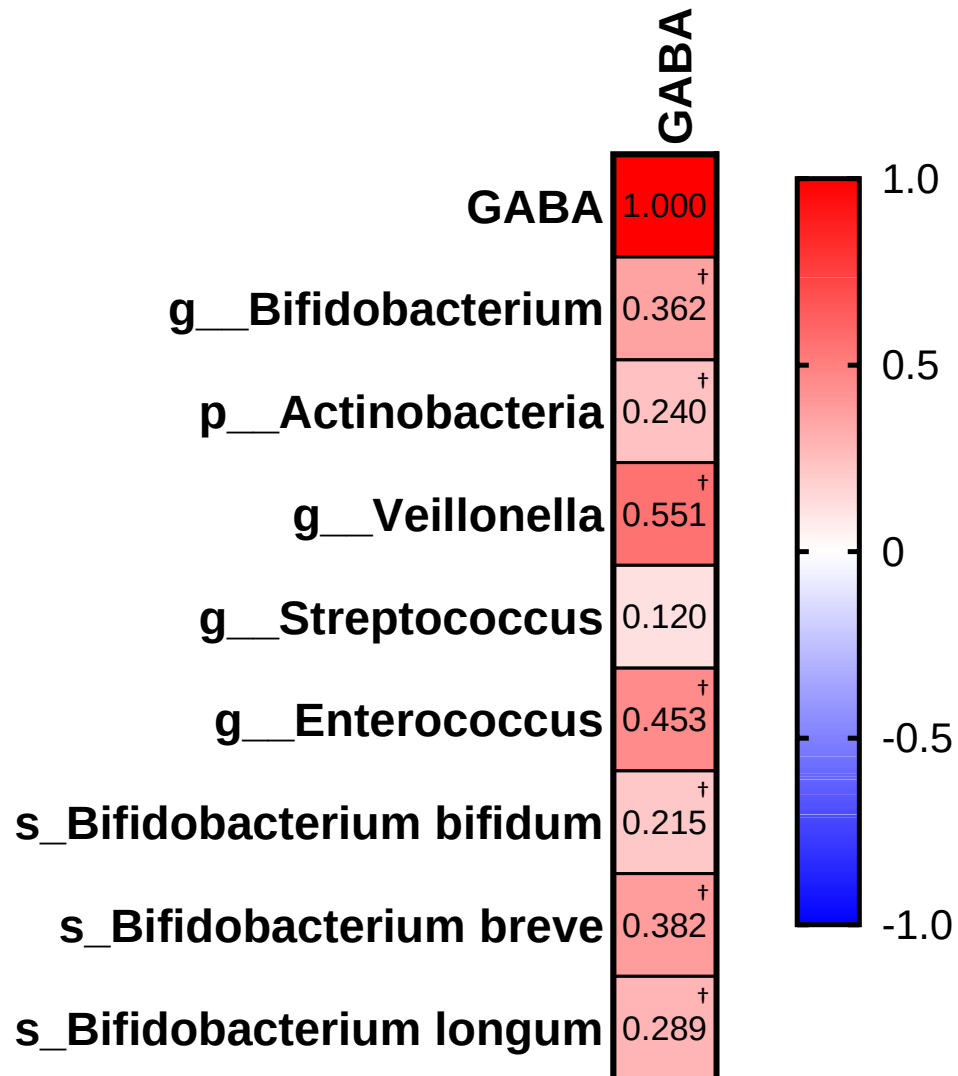


Figure S7: Taxonomic composition of faecal samples using 16S rRNA and shotgun gene sequencing. **A-B:** Relative abundance of the main **phyla** for each age group using 16S rRNA gene and shotgun sequencing, respectively. The number of samples per group are $n = 38$, $n = 43$, $n = 40$ of biologically independent samples for Infants, Mothers and Grandmothers groups, respectively. Data are represented as bars of the Mean + Standard Error of the Mean (SEM). Statistical test was linear mixed-effects model with the correction for multiple test comparisons, FDR p -value < 0.05 . *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$, exact p -values and FDR p -values are provided in Supplementary Data 4. **C-D:** Relative abundance of the main **genera** for each age group using 16S rRNA gene and shotgun sequencing, respectively. The number of samples per group are $n = 38$, $n = 43$, $n = 40$ of biologically independent samples for Infants, Mothers and Grandmothers groups, respectively. Data are represented as bars of the Mean + Standard Error of the Mean (SEM). Statistical test was linear mixed-effects model with the correction for multiple test comparisons, FDR p -value < 0.05 . *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$, exact p -values and FDR p -values are provided in Supplementary Data 5. **E:** Main **species** for each age group using shotgun sequencing. Data are represented using stacked bars. Detailed information can be found in Supplementary Data 6. **F:** Relative abundance of the main **species** for each age group using 16S rRNA gene and shotgun sequencing, respectively. The number of samples per group are $n = 38$, $n = 43$, $n = 40$ of biologically independent samples for Infants, Mothers and Grandmothers groups, respectively. Data are represented as bars of the Mean + Standard Error of the Mean (SEM). Statistical test was linear mixed-effects model with the correction for multiple test comparisons, FDR p -value < 0.05 . *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$, exact p -values and FDR p -values are provided in Supplementary Data 6. Source data are provided as a Source Data file.

All comparisons in sections A, B, C, D, and F, are from the adults compared to the Infants, since no significant results were obtained between Mothers and Grandmothers.

A: GABA



B: SCFAs

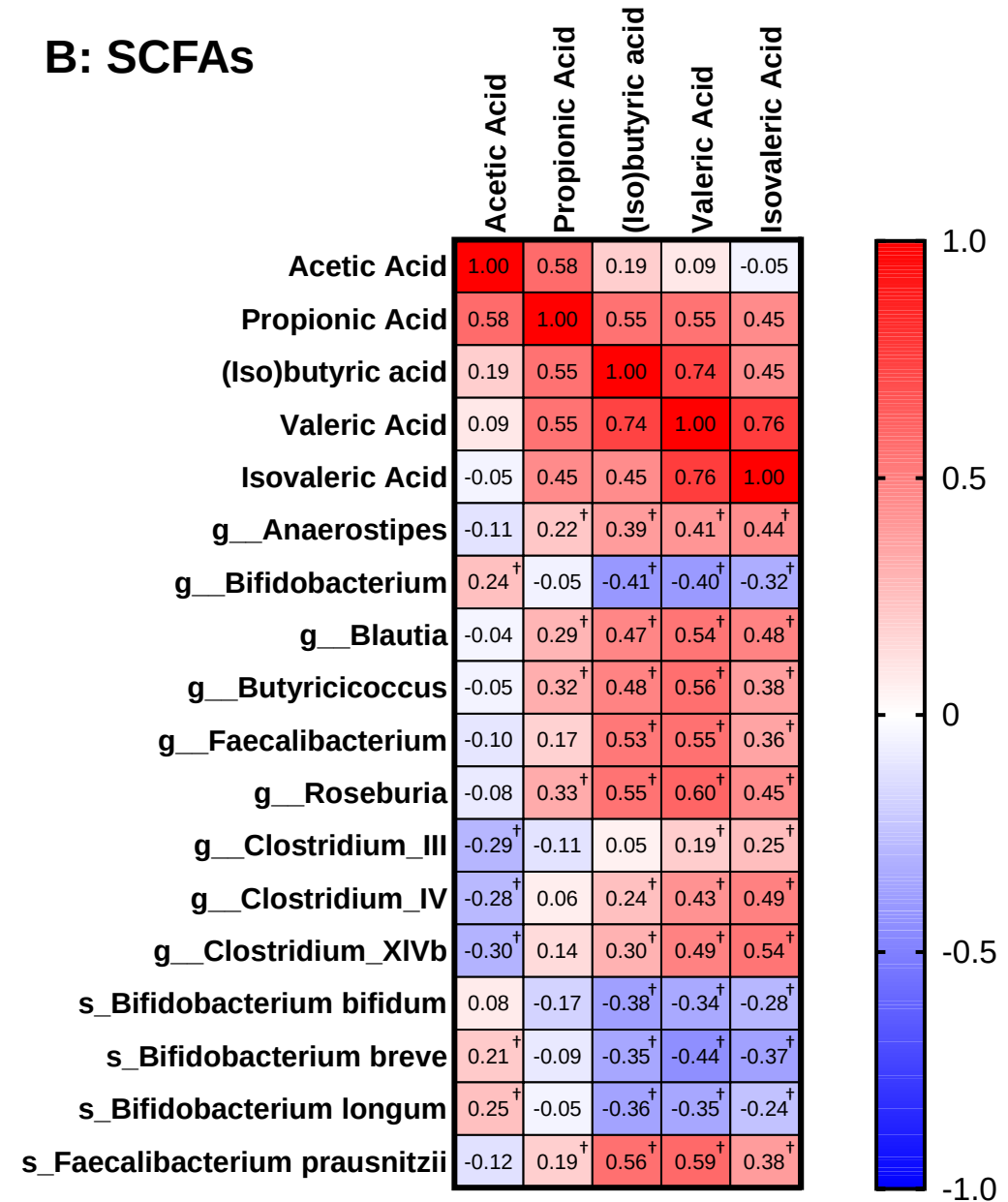


Figure S8: Correlations between metabolites and bacterial taxa that are known producers of these metabolites using spearman's rank correlation coefficient. **A:** Correlation between gamma-aminobutyric acid (GABA) and bacterial taxa (phyla, genera and species). The number of samples are $n = 33$, $n = 43$, $n = 37$ of biologically independent samples for infants, mothers, and grandmothers, respectively. Statical test was spearman's rank correlation coefficient, significant correlations (p -values < 0.05) are indicated with the symbol † . **B:** Correlation between short-chain fatty acids (SCFAs) and bacterial taxa (genera and species). The number of samples are $n = 33$, $n = 43$, $n = 37$ of biologically independent samples for infants, mothers, and grandmothers, respectively. Statical test was spearman's rank correlation coefficient, significant correlations (p -values < 0.05) are indicated with the symbol † . For both, A and B, the correlations are shown as heatmaps, where the Spearman rho (ρ) correlation parameter for each comparison is shown with a color code. Red boxes indicate a positive correlation and blue boxes a negative one, exact p -values are given in Table S3 for GABA and Table S4 for SCFAs. Source data are provided as a Source Data file.

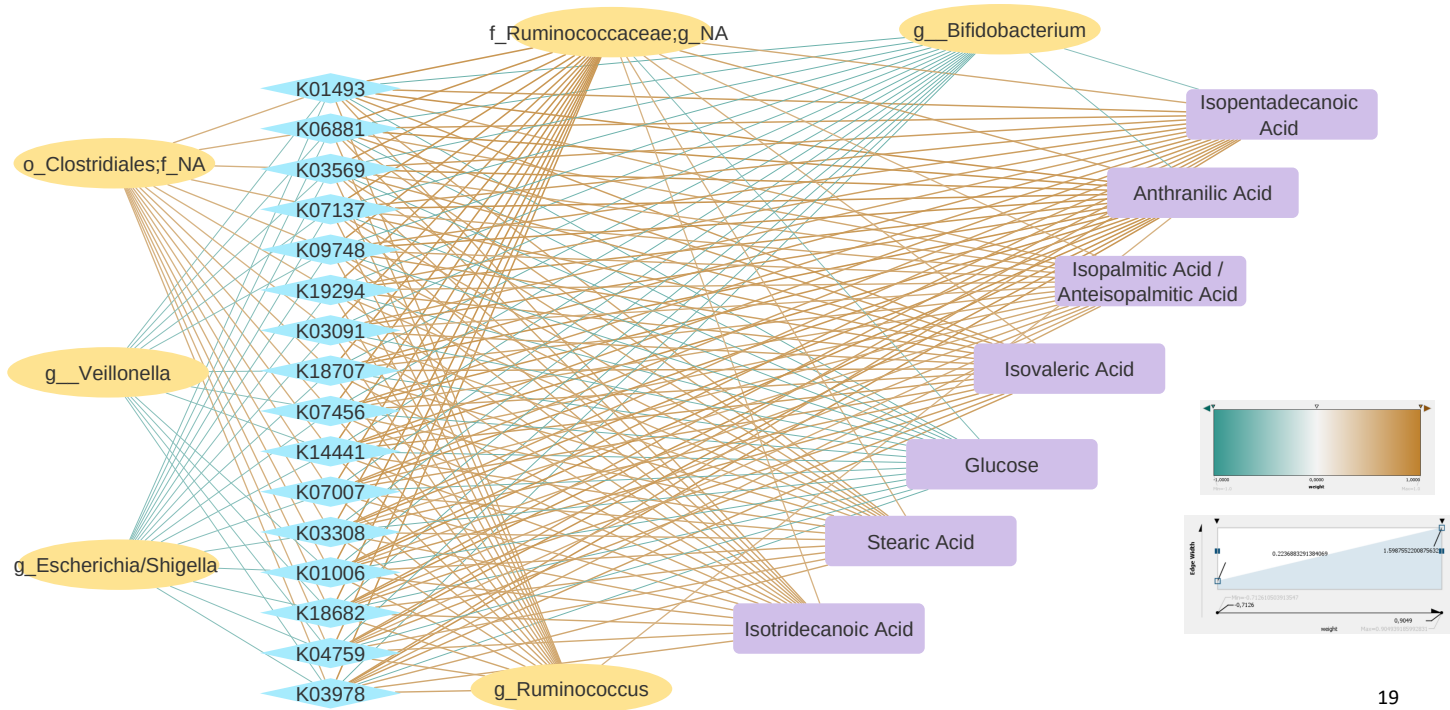


Figure S9: Cytoscape representation of the links between ASVs, KOs, and metabolites obtained from mixOmics and DIABLO. ASVs are represented by the yellow ovals, KOs by the blue diamonds, and metabolites by the purple rectangles. KOs names are described in Figure 9. The blue-green edges indicate a negative correlation, while the brown edges represent a positive correlation. The width of the edge increases with the correlation weight.

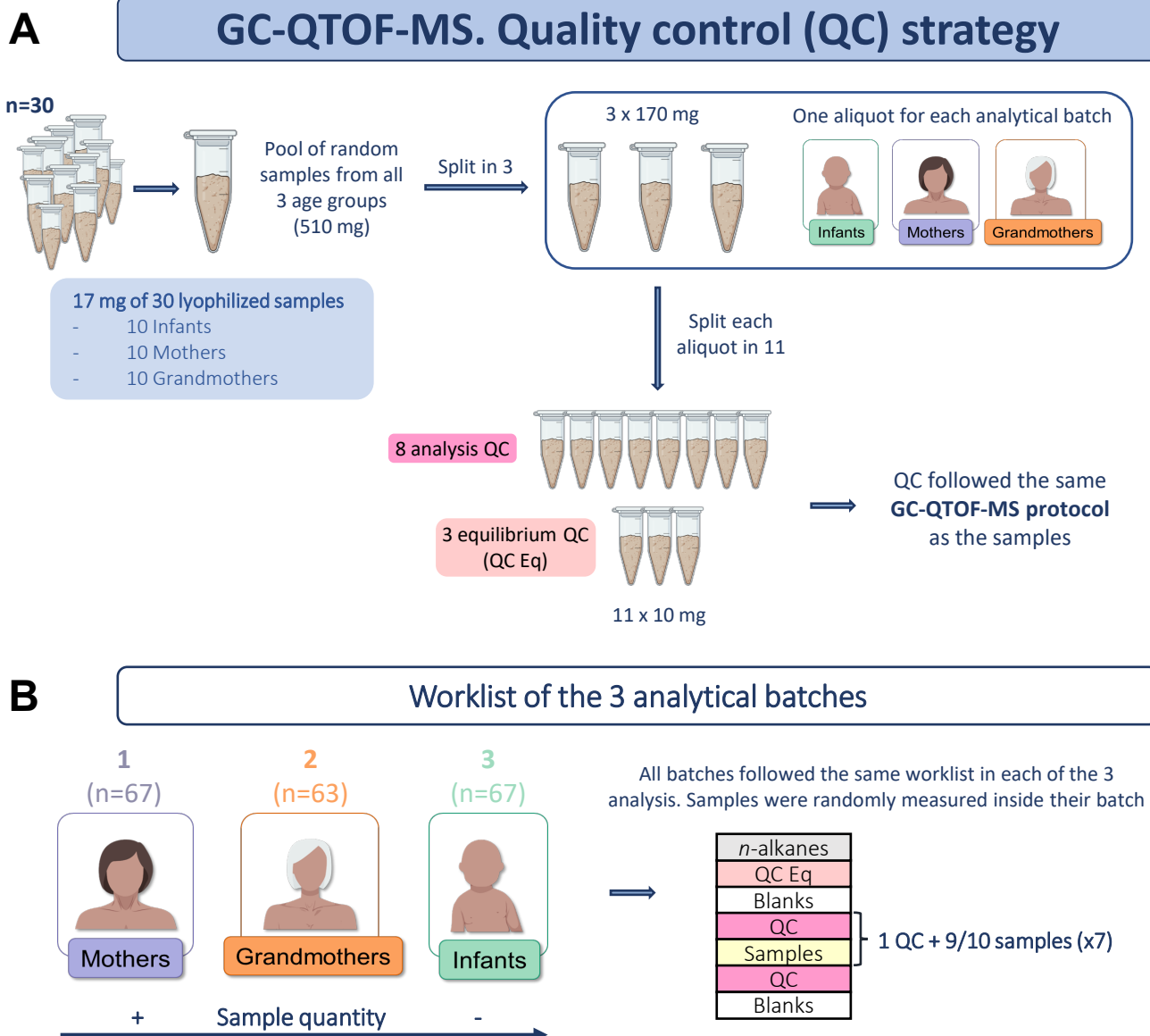


Figure S10: Experimental design of the QCs preparation and worklists for GC-QTOF-MS equipment used for large-scale sample analysis. A: Scheme of the QC strategy used in the analysis to ensure the subsequent joining of the batches through a normalization approach. **B:** Outline of the three analytical batches (Infants, Mothers and Grandmothers) and their corresponding worklists carried out on the GC-QTOF-MS equipment. QCs: quality control samples. Icons were created using biorender.com.

Table S1: Main characteristics for each age group, with their descriptive statistics.

		Infants		Mothers		Grandmothers	
		N	%	N	%	N	%
N		69		67		64	
Age (mean years $\pm$ SD)		0.42 $\pm$ 0.14		34.31 $\pm$ 3.8		63.09 $\pm$ 6.09	
Age (mean months $\pm$ SD)		5.03 $\pm$ 1.71		-		-	
Age range (years)		0.08-1		27-42		47-76	
Age range (months)		1-12		-		-	
Sex							
	Female	38	55%	67	100%	64	100%
	Male	31	45%	-	-	-	-
Mode of birth							
	C-section	16	23%	-		-	
	Vaginal	53	77%	-		-	
Antibiotics during delivery							
	No	55	80%	-		-	
	Yes	10	14%	-		-	
	NA	4	6%	-		-	
Cow's milk allergy (Infants)							
	Allergic	46	67%	-		-	
	Non-allergic	23	33%	-		-	
Infant feeding							
	Breastmilk	12	17%	-		-	
	Formula	8	12%	-		-	
	Hydrolysed formula	12	17%	-		-	
	Breastmilk + Formula	23	33%	-		-	
	Breastmilk + Hydrolysed formula	14	20%	-		-	
Smoking habits							
	Never	-		48	72%	36	56%
	Ever	-		19	28%	28	44%

Table S2: Metabolic Pathway Analysis results using MetaboAnalyst v6.0.

Pathway name	Match Status	p	-log(p)	Holm p	FDR	Impact
Biosynthesis of unsaturated fatty acids	8/36	2.24E-06	5.650	1.79E-04	1.79E-04	0.000
Butanoate metabolism	5/15	2.61E-05	4.583	2.06E-03	1.04E-03	0.143
Glutathione metabolism	6/28	6.08E-05	4.216	4.74E-03	1.22E-03	0.126
Alanine, aspartate and glutamate metabolism	6/28	6.08E-05	4.216	4.74E-03	1.22E-03	0.594
Glyoxylate and dicarboxylate metabolism	5/32	1.25E-03	2.903	9.50E-02	1.93E-02	0.138
Glycine, serine and threonine metabolism	5/33	1.44E-03	2.840	1.08E-01	1.93E-02	0.355
Propanoate metabolism	4/22	2.26E-03	2.646	1.67E-01	2.58E-02	0.041
Citrate cycle (TCA cycle)	3/20	1.47E-02	1.834	1.00E+00	1.00E-01	0.167
Tryptophan metabolism	2/41	2.99E-01	0.524	1.00E+00	8.25E-01	0.145

Labels of the colors: In blue: $p < 0.001$, in light green: $0.001 < p < 0.01$, in green: $0.01 < p < 0.05$, in red: $p > 0.05$. Pathways in bold font are those selected for discussion in the manuscript. Analysis was carried out using Metaboanalyst v6.0

Table S3: Correlations between gamma-aminobutyric acid (GABA) and bacterial phyla, genera, and species.

	GABA vs. g__Bifidobacterium	GABA vs. p__Actinobacteria	GABA vs. g__Veillonella	GABA vs. g__Streptococcus	GABA vs. g__Enterococcus	GABA vs. s_Bifidobacterium bifidum	GABA vs. s_Bifidobacterium breve	GABA vs. s_Bifidobacterium longum
Spearman rho (ρ)	0.3624	0.2404	0.5511	0.1202	0.4527	0.2154	0.3819	0.2889
95% confidence interval	0.1770 to 0.5230	0.04439 to 0.4186	0.3962 to 0.6755	-0.07985 to 0.3109	0.2797 to 0.5973	0.01807 to 0.3966	0.1988 to 0.5392	0.09623 to 0.4606
P (two-tailed)	0.0002	0.0140	0.0000	0.2243	0.0000	0.0281	0.0001	0.0029
P value summary	***	*	***	ns	***	*	***	**
Significant? (alpha = 0.05)	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes

Spearman rho (ρ) color scale: ρ = positive values are displayed in red while negative values would be in blue, $\rho=0$ would be in white. Significance color scale cell, for p values < 0.05, cells are in green and p values > 0.05 are in red.

Table S4: Correlations between short-chain fatty acids (SCFA) and bacterial genera and species.

		Producer of	Acetic Acid		Propionic Acid		Butyric/Isobutyric Acid		Valeric Acid		Isovaleric Acid	
			ρ	p -value	ρ	p -value	ρ	p -value	ρ	p -value	ρ	p -value
SCFAs	Acetic Acid	-	1.0000		0.5799	1.69E-11	0.1921	4.15E-02	0.0890	3.49E-01	-0.0466	6.24E-01
	Propionic Acid	-	0.5799	1.69E-11	1.0000		0.5515	2.44E-10	0.5457	4.09E-10	0.4516	5.15E-07
	Butyric/Isobutyric Acid	-	0.1921	4.15E-02	0.5515	2.44E-10	1.0000		0.7358	1.64E-20	0.4505	5.52E-07
	Valeric Acid	-	0.0890	3.49E-01	0.5457	4.09E-10	0.7358	1.64E-20	1.0000		0.7590	2.05E-22
	Isovaleric Acid	-	-0.0466	6.24E-01	0.4516	5.15E-07	0.4505	5.52E-07	0.7590	2.05E-22	1.0000	
Genera	g_Anaerostipes	Butyrate	-0.1136	2.31E-01	0.2175	2.07E-02	0.3872	2.27E-05	0.4148	4.94E-06	0.4445	8.13E-07
	g_Bifidobacterium	Acetate	0.2434	9.37E-03	-0.0521	5.84E-01	-0.4067	7.80E-06	-0.3967	1.36E-05	-0.3248	4.48E-04
	g_Blautia	Propionate	-0.0424	6.55E-01	0.2893	1.88E-03	0.4717	1.34E-07	0.5392	7.18E-10	0.4824	6.30E-08
	g_Butyricicoccus	Butyrate	-0.0508	5.93E-01	0.3169	6.28E-04	0.4831	5.97E-08	0.5608	1.04E-10	0.3802	3.27E-05
	g_Faecalibacterium	Butyrate	-0.0998	2.93E-01	0.1734	6.62E-02	0.5262	2.16E-09	0.5547	1.83E-10	0.3555	1.12E-04
	g_Roseburia	Propionate and butyrate	-0.0846	3.73E-01	0.3287	3.78E-04	0.5478	3.40E-10	0.6021	1.73E-12	0.4546	4.25E-07
	g_Clostridium_III	Valerate	-0.2895	1.87E-03	-0.1135	2.31E-01	0.0531	5.76E-01	0.1936	3.99E-02	0.2528	6.91E-03
	g_Clostridium_IV	Valerate	-0.2777	2.90E-03	0.0649	4.95E-01	0.2440	9.21E-03	0.4254	2.64E-06	0.4940	2.70E-08
	g_Clostridium_XIVb	Valerate	-0.2987	1.31E-03	0.1391	1.42E-01	0.3015	1.18E-03	0.4885	4.05E-08	0.5404	6.48E-10
Species	s_Bifidobacterium bifidum	Acetate	0.0767	4.20E-01	-0.1726	6.75E-02	-0.3760	4.05E-05	-0.3411	2.18E-04	-0.2836	2.33E-03
	s_Bifidobacterium breve	Acetate	0.2077	2.73E-02	-0.0885	3.51E-01	-0.3484	1.56E-04	-0.4405	1.05E-06	-0.3715	5.11E-05
	s_Bifidobacterium longum	Acetate	0.2467	8.44E-03	-0.0451	6.35E-01	-0.3625	7.96E-05	-0.3469	1.67E-04	-0.2395	1.06E-02
	s_Faecalibacterium prausnitzii	Butyrate	-0.1199	2.06E-01	0.1857	4.90E-02	0.5600	1.13E-10	0.5859	9.32E-12	0.3849	2.56E-05

Spearman rho (ρ) color scale: ρ = positive values are displayed in red while negative values would be in blue, ρ =0 would be in white. Significance color scale cell, for p values < 0.05, cells are in green and p values > 0.05 are in red.

Table S5: Correlations between the matched bacterial genera between 16S rRNA gene sequencing and shotgun sequencing.

Genera	p.value_Pearson	ρ Pearson
g_Caryophanon	0.00E+00	1.000
g_Hafnia	3.88E-94	0.989
g_Megasphaera	2.16E-85	0.984
g_Acidaminococcus	8.86E-84	0.983
g_Slackia	8.34E-77	0.977
g_Klebsiella	7.63E-71	0.971
g_Parasutterella	1.29E-69	0.969
g_Bilophila	6.51E-68	0.967
g_Collinsella	1.10E-67	0.967
g_Catenibacterium	1.34E-66	0.965
g_Citrobacter	3.43E-66	0.965
g_Dialister	3.87E-64	0.962
g_Parabacteroides	2.32E-63	0.960
g_Desulfovibrio	2.12E-62	0.959
g_Butyricimonas	1.89E-61	0.957
g_Prevotella	2.95E-61	0.957
g_Hungatella	4.03E-61	0.956
g_Alistipes	2.35E-57	0.949
g_Streptococcus	6.25E-57	0.948
g_Odoribacter	1.31E-56	0.947
g_Senegalimassilia	9.13E-56	0.945
g_Enterococcus	3.94E-55	0.944
g_Bifidobacterium	4.79E-55	0.943
g_Paraprevotella	3.46E-54	0.941
g_Lactococcus	1.52E-52	0.937
g_Megamonas	1.47E-48	0.925
g_Eggerthella	8.24E-47	0.919
g_Akkermansia	1.56E-46	0.918
g_Faecalibacterium	5.81E-46	0.916
g_Flavonifractor	3.41E-44	0.910
g_Lactobacillus	3.75E-42	0.901
g_Haemophilus	4.28E-40	0.892
g_Methanobrevibacter	6.62E-40	0.891
g_Butyricicoccus	1.23E-38	0.885
g_Faecalitalea	6.09E-38	0.881
g_Veillonella	1.69E-37	0.879
g_Gordonibacter	2.13E-36	0.873
g_Ruminococcus	7.62E-33	0.851
g_Romboutsia	1.63E-30	0.835
g_Granulicatella	3.62E-30	0.832
g_Fusicatenibacter	2.04E-29	0.826
g_Phascolarctobacterium	1.59E-27	0.810

g_Gemmiger	9.88E-27	0.803
g_Holdemanella	5.53E-26	0.796
g_Coprobacillus	4.59E-25	0.787
g_Adlercreutzia	1.02E-24	0.784
g_Turicibacter	1.65E-24	0.782
g_Coprococcus	2.11E-24	0.781
g_Olsenella	6.06E-24	0.776
g_Bacteroides	8.11E-24	0.775
g_Butyrvibrio	2.86E-23	0.769
g_Blautia	5.91E-21	0.741
g_Enterobacter	1.12E-20	0.738
g_Intestinibacter	7.92E-20	0.727
g_Actinomyces	1.44E-18	0.709
g_Oscillibacter	1.18E-17	0.696
g_Barnesiella	2.39E-16	0.675
g_Roseburia	2.65E-15	0.657
g_Succinilasticum	3.08E-15	0.656
g_Cloacibacillus	8.96E-15	0.648
g_Christensenella	7.00E-14	0.631
g_Sutterella	9.76E-13	0.607
g_Dorea	1.58E-11	0.581
g_Anaerostipes	1.81E-10	0.555
g_Faecalicoccus	2.80E-08	0.493
g_Terrisporobacter	6.89E-08	0.481
g_Anaerotruncus	8.53E-07	0.444
g_Sporobacter	1.81E-05	0.391
g_Anaerovorax	1.25E-04	0.353
g_Eisenbergiella	1.23E-03	0.300
g_Asaccharobacter	6.37E-03	0.255
g_Subdoligranulum	8.58E-03	0.246
g_Pseudoflavonifractor	1.04E-02	0.240
g_Mogibacterium	7.31E-02	0.169
g_Intestinimonas	3.68E-01	0.086
g_Alloprevotella	5.32E-01	0.059
g_Coprobacter	5.62E-01	0.055
g_Peptococcus	5.83E-01	0.052
g_Acetanaerobacterium	7.04E-01	0.036
g_Salmonella	7.24E-01	-0.034
g_Anaeroglobus	7.50E-01	-0.030
g_Enterorhabdus	8.69E-01	-0.016
g_Eubacterium	9.40E-01	-0.007
g_Hespellia	9.60E-01	0.005

Pearson correlation (ρ) color scale: ρ = positive values are displayed in red while values near to 0 or negative are in white to light blue. Significance color scale cell, for p values < 0.05, cells are in green and p values > 0.05 are in white.

Table S6: Description of the KOs from Heatmap and circosplot from Figure 9.

Order	feature	Description
1	K06881	bifunctional oligoribonuclease and PAP phosphatase NrnA [EC:3.1.3.7 3.1.13.3]
2	K01493	dCMP deaminase [EC:3.5.4.12]
3	K18682	ribonuclease Y [EC:3.1.-.-]
4	K18707	threonylcarbamoyladenosine tRNA methylthiotransferase MtaB [EC:2.8.4.5]
5	K15633	2,3-bisphosphoglycerate-independent phosphoglycerate mutase [EC:5.4.2.12]
6	K14441	ribosomal protein S12 methylthiotransferase [EC:2.8.4.4]
7	K07137	uncharacterized protein
8	K01006	pyruvate, orthophosphate dikinase [EC:2.7.9.1]
9	K22132	tRNA threonylcarbamoyladenosine dehydratase
10	K03569	rod shape-determining protein MreB and related proteins
11	K03978	GTP-binding protein
12	K09748	ribosome maturation factor RimP
13	K07456	DNA mismatch repair protein MutS2
14	K11749	regulator of sigma E protease [EC:3.4.24.-]
15	K01808	ribose 5-phosphate isomerase B [EC:5.3.1.6]
16	K19294	alginate O-acetyltransferase complex protein AlgI
17	K03499	trk system potassium uptake protein
18	K03308	neurotransmitter:Na ⁺ symporter, NSS family
19	K07007	3-dehydro-bile acid Delta4,6-reductase [EC:1.3.1.114]
20	K07095	uncharacterized protein
21	K11175	phosphoribosylglycinamide formyltransferase 1 [EC:2.1.2.2]
22	K04759	ferrous iron transport protein B
23	K02238	competence protein ComEC
24	K07699	two-component system, response regulator, stage 0 sporulation protein A
25	K03091	RNA polymerase sporulation-specific sigma factor