DELIVERABLE

D3.1 – Protocol for the analysis of biomarkers

Dissemination level	PU – Public
Type of Document	Report or Demonstrator
Contractual date of delivery	31/12/2021
Deliverable Leader	REM
Status & version	Final, V1.2 – 6/11/2021
WP responsible	WP3 (REM)
Keywords:	BioMarkers, Microbiome, Vitamins

Deliverable Leader:	REM Analytics
Contributors:	HUAFE, SVI, HUA
Reviewers:	CU
Approved by:	OSR, ALPES

Document History			
Version	**Date**	**Contributor(s)**	**Description**
V0.1	06/09/2021	REM	DRAFT
v0.2	06/10/2021	REM, HULAFE, SVI, OSR, HUA	DRAFT
v1.0	01/11/2021	REM, HULAFE, SVI, OSR, HUA	First complete version
V1.2	05/11/2021	REM, HULAFE, SVI, OSR, HUA	Complete version with minor modifications
V2.1	24/01/2022	CU, HUA	Included Modifications required by PO
V2.2	07/02/2022	ALPES	Revisions and checks

Executive Summary

The following delivery describes the development of methods to measure biomarkers in the Nutrishield studies.

It is the ambition of NUTRISHIELD to integrate several biomarker measurements to better understand the relationship between diet and health. This understanding will then be integrated into a personalized nutrition algorithm.

Several biomarkers will therefore be measured from several samples. Each of these analyses requires its own protocol development, optimization and possibly calibration and validation.

In this deliverable, are described the protocols used for the analysis of metabolites in urine and human milk. Significant development was required to adapt chromatographic and mass spectrometry techniques to measurer the panel of selected metabolites from two different matrixes.

Vitamins will be measured in human milk. To accurately measure a large panel of lipid and water-soluble vitamins from human milk, Ultra Performance Liquid Chromatography was used. Following optimization, they had to be validated using reference material.

The measurement of microbiome has not yet been attempted using ATGC in a large study. A tailored assay focusing on the most relevant micro-organisms in human milk, new-born gut, and teenage gut microbiome had to be established. It included the development of novel DNA extraction procedures for human milk, as well as several steps of calibration and optimization.

All main steps in method optimization are now complete, and the NUTRISHIELD analytical partners are ready to begin analysing the samples coming from the Clinical studies carried out as part of WP5, as planned within WP3.

Table of Contents

Table of Figures

List of Tables

Definitions, Acronyms and Abbreviations

Acronym	Title
CTCE	Cycling temperature capillary electrophoresis
ATGC	Advanced Testing for genetic composition
PCR	Polymerase Chain Reaction
MS	Mass Spectrometry
GC	Gas Chromatography
LC	Liquid Chromatography
HPLC	High Performance Liquid Chromatography
EDTA	Ethylenediaminetetraacetic acid
UPLC	Ultra-Performance Liquid Chromatography
SDS	Sodium dodecyl sulphate
EtOH	Ethanol
SCFAs	Short Chain Fatty acids
BCAAs	Branched Chain Amino Acids
BAs	Bile Acids
OTU	Operational Taxonomic Units

1. Metabolic Analysis in Urine

Four methods are developed to measure urine samples: A) Metabolic fingerprinting is carried out by untargeted liquid chromatography coupled to mass spectrometry (LC-MS). B) 27 nutrition and microbiota biomarkers (BM) are measured by LC-MS, in addition, a semi-quantification of 195 BM is carried out. C) Microbiota biomarkers, i.e., short chain fatty acids (SCFAs) and branched chain amino acids (BCAAs) are measured by gas chromatography coupled to mass spectrometry method. D) Microbiota biomarkers, i.e., bile acids (BAs) are measured by LC-MS.

1.1. Sample collection

Mother's first morning urine is collected in sterile polypropylene containers and infants' urine is collected by placing sterile cotton pads in the diaper and, after 1h, squeezing with a sterile polypropylene syringe. The process is repeated until collecting a minimum of 1 mL. Urine samples are aliquoted to avoid freeze-thawing cycles and stored at -80 °C until analysis.

1.2. Method validation

Targeted methods were validated based on the US Food and Drug Administration (FDA) guidelines for bioanalytical method validation, including the bioanalytical parameters: linearity range, selectivity and specificity, sensitivity, accuracy, precision, recovery, and stability were assessed for method validation. The linear range was selected according to the expected concentrations ranges. The calibration curves included a blank without analytes or IS, a zero calibrator (i.e., blank with IS) and, at least, 6 standards covering the selected concentration ranges. Accuracy, precision, and recovery were assessed by replicate (n=3) analysis of standards at three concentration levels and replicate (n=3) analysis of spiked samples at three concentration levels (low, medium, and high) on three validation days. Precision was estimated as the percentage of relative standard deviation (RSD) of replicate standards within one validation batch (intra-day) and between validation batches (inter-day). Selectivity and specificity were demonstrated by analysing blank samples from multiple sources. Carry-over between samples was assessed by the analysis of zero-injections after the analysis of high concentrated standards and spiked samples (n=3). Autosampler sample stability was assessed by comparing concentrations observed in a freshly prepared sample and in the same processed sample after 20 h stored in the autosampler (sealed vial, 25 °C). Analytes' freeze-thaw stability and long-term stability were established by comparing concentrations observed in sample extracts after three freeze-thaw cycles and in a sample stored for one year (-80 °C), respectively, to a freshly prepared sample.

1.3. Metabolic fingerprinting by untargeted LC-MS

1.3.1. Sample processing and analysis

Urine samples are thawed on ice, homogenized on a Vortex® mixer during 30 s, and centrifuged (16,000 × g for 15 min at 4 °C). A total of 50 µL of supernatant are withdrawn and spiked with 50 µL of internal standard (IS) solution in 96-well plates containing phenylalanine-D5, caffeine-D9, leukine enkephalin, betaine-D11, 2-dGC13N15, 8OHdG-13C-15N, taxifolin, p-Tyrosine-D2, creatinine-D3 and reserpine in H_2O, 0.1% v/v HCOOH at 2 µM. Blanks are cotton pad H_2O-extracts (new-born urine collection) and a quality control (QC) is prepared by mixing 5 µL of each final sample extract.
For the analysis, a 1290 Infinity UPLC system coupled to a 6550 Spectrometer iFunnel quadrupole time-of-flight (qTOF) MS system from Agilent Technologies (Santa Clara, CA, USA) are used. Two columns are used in both positive and negative electrospray ionization modes (ESI + and ESI -):

I. Synergi-Hydro C18 column (150 x 2.1 mm, 4 µm) from Phenomenex (Torrance, CA, USA) employing a stepwise gradient with solvent A (H_2O with 0.2% v/v HCOOH) and solvent B (CH_3CN) as mobile phase components are used as follows: 1% B is held for 2 min followed by a linear gradient from 30 to 80% B in 8 min and 98% B for 2 min before returning to initial conditions in 0.1 min and column equilibration with 1% B during 7 min. The flow rate is set to 0.25 mL min^{-1}, column and autosampler to 55 and 4 °C, respectively, and the injection volume is 2 µL.
II. UPLC BEH C18 (100 × 2.1 mm, 1.7 µm, Waters, Wexford, Ireland) column running a binary mobile phase gradient (mobile phase A: H_2O, 0.1% v/v HCOOH and mobile phase B: CH_3CN, 0.1% v/v HCOOH). The flow rate is set to 0.5 mL min^{-1}, column and autosampler to 55 and 4 °C, respectively, and the injection volume is 3 µL.

For the data analysis, an automated MS/MS-based annotation is carried out after blank clean-up, followed by a QC-SVR batch effect correction, concatenation of data sets and normalization with creatinine values obtained with the Urinary Creatinine Detection Kit (Arbor Assays™, Ann Arbor, MI).

1.4. Nutrition and microbiota BM

1.4.1. LC-MS method for 27 BM and semi-quantification of 195 BM

This method has been developed for the quantification of 19 nutrition BM including fruit, vegetables, meat, fish, seeds, milk, dairy products, coffee, and sugar sweetened beverages BM and 8 microbiota BM. In addition, 195 nutrition biomarkers are semi-quantified.

1.4.1.1. Sample processing and analysis

The method consists in centrifuging the urine samples (10000 x g for 10 min at 4 °C) and diluting at 1:20 to a final volume of 100 µL by the addition of a 10 µM IS mixture containing caffeine-D9, phenylalanine-D5, betaine-D11, taxifolin and tyrosine-D2 in a 96-well plate.

Targeted metabolomic profiling of urine is carried out on an Agilent 1290 Infinity UHPLC system (Santa Clara, CA, U.S.A.) coupled to a Sciex QTRAP 6500 mass spectrometer equipped with Ion Drive Turbo V ion source (Framingham, MA, U.S.A.). The column used was a 100 × 2.1 mm (1.6 µm inner diameter) Luna Omega Polar C18, equipped with a fully porous polar C18 security guard cartridge from Phenomenex (Torrance, CA, U.S.A.). Chromatographic conditions are as follows: column temperature, 40 °C; autosampler temperature, 4 °C; injection volume, 2 µL; and flow rate, 0.5 mL min^{-1}. Two different mobile phase combinations are employed depending upon the ionization mode of data acquisition. In the negative ion mode, 0.1% formic acid and 10 mM ammonium formate in water and pure acetonitrile are delivered as aqueous (A) and organic (B) mobile phases, respectively. The gradient program are as follows: 0−8 min, 5−20% B; 8−10 min, 20−100% B; 10−12 min, 100% B; 12−12.1 min, 100−5% B; and 12.1−14 min, 5% B. On the other hand, water and acetonitrile, both containing 0.5% formic acid, are used as mobile phases when the mass spectrometer operated in the positive ion mode. In that case, the gradient profile was as follows: 0−5 min, 5−50% B; 5−5.1 min, 50− 100% B; 5.1−7 min, 100% B; 7−7.1 min, 100−5% B; and 7.1−9 min, 5% B. MS detection is performed using the scheduled multiple reaction monitoring (sMRM) mode. The mass spectrometer operated in positive and negative ionization modes in separate runs, using the following parameters: ion spray voltage, +5500/−4000 V; source temperature, 600 °C; curtain gas, 30 psi; ion source gases 1 and 2, 50 psi each; collision-activated dissociation gas, 3 psi; and entrance potential, (±)10 V. The sMRM transitions were optimized by infusing individual solutions of commercial standards dissolved in the mobile phase (1:1, v/v, proportion A/B, 500 µg L^{-1}) into the mass spectrometer using a syringe pump at a flow rate of 5 µL min^{-1}. The method also included 195 metabolites identified in urine samples by product ion scan monitoring, for which authentic standards were not available. The measurement parameters used for the studied analytes are summarized in Table 1.

For quantification, an external calibration is carried out using standard solutions obtained from serial dilutions of a working solution containing mixtures of pure analytical standards in ultrapure water. A quality control (QC) sample is prepared by mixing 20 µL of each urine sample. Aliquots of the QC sample are analysed every 10 samples in the randomized analytical batch for monitoring the instrument's performance. QC RSD < 25% is the batch acceptance criteria. In addition, calibration blanks (addition of water instead of sample) and process blanks (addition of water instead of sample to a polypropylene container, cotton or gauze pad and squeezed before adding it to the tubes) are injected. For semi-quantification, the area of the peak is used. Urinary metabolite concentrations are normalized to creatinine quantified with a commercial kit.

1.4.1.1. Method validation results

The method was validated as described in section 1.2. Table 2 summarizes the employed concentration intervals, which were chosen considering the expected wide inter- and intra-individual variability.

Table 1. Measurement parameters and main figures of merit of the LC-MS method

Category	Metabolite	Parent Ion (Da)	Daughter Ion (Da)	DP (V)	CE (V)	CXP (V)	RT ± s (min)	Calibration range (µM)	R^2	LOD (mM)	LOQ (mM)
Microbiota	Phenylpropionylglycine	208	105	36	25	14	7.6 ± 0.02	0.008 – 10	0.998	0.09	0.3
	3-IPA	190	130	1	19	14	9.8 ± 0.02	0.008 – 10	0.993	0.15	0.5
	L-Kynurenine	209	192	1	13	22	2 ± 0.02	0.008 – 10	0.997	0.16	0.5
	3-IAA	176	130	41	19	14	9.5 ± 0.013	0.016 – 20	0.997	0.09	0.3
	L-Tyrosine	182	165	20	13	8	1.3 ± 0.02	0.016 – 20	-	2	7
	Hippuric Acid	178	134	65	16	-9	4.3 ± 0.03	0.2 – 300	0.998	1.5	5
	Ferullic Acid Sulphate	273	193	60	20	-17	5.3 ± 0.03	0.03 – 45	0.997	0.08	0.3
Fruits	Proline betaine	144	84	101	29	10	0.9 ± 0.02	0.03 – 45	0.999	1.3	4
	Hesperetin	301	151	200	30	-9	9.8 ± 0.02	0.008 – 10	0.997	0.09	0.3
	Phloretin	273	167	130	22	-11	9.9 ± 0.02	0.008 – 10	0.998	0.03	0.112
Vegetables	Quercetin	301	151	200	30	-9	9.8 ± 0.03	0.008 – 10	0.999	0.2	0.8
	Kaempferol	285	239	140	38	-13	10 ± 0.014	0.008 – 10	0.997	0.4	1.4
Seeds	O-DMA	259	149	121	15	22	10 ± 0.03	0.008 – 10	0.996	0.03	1.10
	Daidzein	253	223	105	44	-25	9.8 ± 0.012	0.008 – 10	0.996	0.5	1.7
	Equol	241	121	95	20	-13	10 ± 0.014	0.008 – 10	0.996	0.03	0.10
	Glycitein	283	268	150	22	-23	9.8 ± 0.02	0.008 – 10	0.998	0.5	1.6
	Genistein	269	133	40	40	-9	10 ± 0.01	0.008 – 10	0.997	0.3	1.2
Meat	1-Methylhistidine	170	124	36	19	6	0.7 ± 0.02	0.016 – 20	0.998	0.6	2
	3-Methylhistidine	170	96	36	21	12	0.6 ± 0.02	0.016 – 20	0.998	0.4	1.2
	Anserine	241	109	81	31	14	0.7 ± 0.02	0.016 – 20	0.996	0.2	0.7
Fish	TMAO	76	58	81	25	8	1.16 ± 0.013	0.05 – 60	0.999	0.2	0.7
Dairy products	Isovalerylglycine	158	74	40	22	-27	3.3 ± 0.03	0.03 – 45	-	0.3	1.3
	Isobutirylglycine	144	74	30	14	-9	1.9 ± 0.03	0.03 – 46	0.997	0.3	0.9
Milk	Galactitol	181	101	65	20	-17	0.7 ± 0.02	0.05 – 60	0.998	0.2	0.5

Coffee	Gallic Acid	169	125	95	20	-13	1.9 ± 0.012	0.008 – 10	0.994	0.5	1.6
Sugar sweetened beverages	Citrulline	176	131	71	23	14	0.7 ± 0.02	0.016 – 20	0.995	0.5	1.8
	Taurine	126	109	131	21	12	0.7 ± 0.03	0.016 – 20	0.993	0.4	1.4

Note: 3-IPA = 3-indolepropionic acid; 3-IAA = indole-3-acetic acid; O-DMA = O-desmethylangolensin; TMAO = trimethylamine N-oxide; DP = decluttering potential; CE = collision energy; CXP = cell exit potential; RT = retention time; R = coefficient of determination; Limit of quantification (LOQ) = concentration of analyte that can be measured with an imprecision of less than 20% and a deviation from target of less than 20% and taking into account the preconcentration factor achieved during sample processing; Limit of detection (LOD) = 3/10*LOQ

Table 2. Calculated intra- and inter- day accuracy (i.e., recovery) and precision (i.e., RSD) of the LC-MS method in standard solutions and spiked urine samples

	Concentration levels (µM)			Accuracy ± RSD											
				Standard solutions						Spiked urine samples					
				Intra-day (N = 3)			Inter-day (N = 3)			Intra-day (N = 3)			Inter-day (N = 3)		
Metabolite	Low	Medium	High	Low	Medium	High	Low	Medium	High	Low	Medium	High	Low	Medium	High
Phenylpropionylglycine	1.7	3	4	101 ± 11	104 ± 2	115 ± 8	103 ± 6	103 ± 11	98 ± 12	106 ± 1.2	107 ± 2	104 ± 2	105 ± 13	107 ± 17	103 ± 11
3-IPA	0.7	1.0	1.7	94 ± 5	98 ± 3	99 ± 3	94 ± 5	116 ± 7	85.8 ± 1.2	106 ± 16	113 ± 13	112 ± 2	102 ± 12	110 ± 21	110 ± 6
L-Kynurenine	1.7	3	4	106 ± 3	110 ± 5	108 ± 16	93 ± 6	88 ± 2	88 ± 4	108 ± 8	99 ± 3	106 ± 4	107 ± 11	98 ± 16	106 ± 15
3-IAA	2	3	5	106 ± 11	95 ± 4	116 ± 8	103 ± 11	106 ± 10	115 ± 13	98 ± 16	93 ± 13	94 ± 8	97 ± 5	92 ± 5	94 ± 17
L-Tyrosine	2	3	5	102 ± 17	87 ± 8	97 ± 5	101 ± 7	105 ± 14	95 ± 5	83 ± 1.3	88 ± 3	86 ± 19	82 ± 20	87 ± 20	86 ± 1.4
Hippuric Acid	13	20	33	98.9 ± 1.2	104 ± 12	119.2 ± 1.3	101 ± 10	101 ± 11	116.4 ± 1.3	112 ± 20	107 ± 3	102 ± 19	120 ± 2	112 ± 1.4	105 ± 8
Ferullic Acid Sulphate	7	10	17	92 ± 9	115 ± 8	103 ± 6	103 ± 11	98 ± 12	112 ± 5	99 ± 5	105 ± 21	98 ± 10	99 ± 20	105 ± 6	98 ± 16
Proline betaine	7	10	17	98 ± 3	89 ± 4	105 ± 5	103 ± 10	117 ± 4	82 ± 9	102 ± 13	83 ± 17	121 ± 2	106 ± 9	86 ± 5	122 ± 5
Hesperetin	1.3	2	3	97.2 ± 1.3	96 ± 7	93 ± 2	105 ± 8	103 ± 7	111 ± 11	100 ± 5	121 ± 3	971 ± 19	99 ± 6	120 ± 10	96 ± 1.3
Phloretin	2	3	5	101 ± 9	104 ± 2	111 ± 18	101 ± 10	107 ± 2	109 ± 14	115 ± 5	120 ± 12	114 ± 21	114 ± 12	119 ± 6	114 ± 5
Quercetin	1.0	1.5	3	100 ± 14	105 ± 10	101 ± 6	105 ± 7	106 ± 3	98 ± 4	112 ± 9	106 ± 19	97 ± 6	109 ± 4	104 ± 1.2	96 ± 12
Kaempferol	1.7	3	4	99 ± 4	104 ± 18	100 ± 7	86 ± 6	93.2 ± 1.2	116 ± 2	103 ± 21	102 ± 14	117 ± 1	102 ± 3	101 ± 16	117 ± 4

O-DMA	1.7	3	4	101 ± 9	100 ± 5	100 ± 5	93 ± 8	101 ± 11	112 ± 2	89 ± 6	97 ± 11	107 ± 10	87 ± 20	96 ± 1	106 ± 10
Daidzein	0.7	1.0	1.7	100 ± 14	100 ± 3	103 ± 11	106 ± 10	115 ± 13	106 ± 6	101 ± 19	115 ± 8	98 ± 7	97 ± 7	113 ± 14	96 ± 8
Equol	1.0	1.5	3	99 ± 4	101 ± 10	100 ± 8	114.6 ± 1.3	104 ± 11	117 ± 6	115 ± 17	103 ± 16	99 ± 5	112 ± 16	102 ± 3	98 ± 15
Glycitein	0.7	1.0	1.7	117 ± 4	117 ± 4	96 ± 16	102 ± 17	107 ± 4	119 ± 4	107 ± 15	99 ± 6	98 ± 17	103 ± 2	97 ± 6	96 ± 11
Genistein	0.7	1.0	1.7	89 ± 4	94 ± 9	99 ± 15	98.9 ± 1.2	90 ± 8	108 ± 5	90 ± 13	112 ± 5	99 ± 12	86 ± 21	109 ± 5	97 ± 21
1-Methylhistidine	2	3	5	105 ± 10	89 ± 4	101.2 ± 1.0	106 ± 11	102 ± 3	99 ± 3	84 ± 4	101 ± 8	103 ± 11	98 ± 6	110 ± 11	109 ± 9
3-Methylhistidine	2	3	5	104 ± 18	110 ± 6	87 ± 16	94 ± 16	98 ± 4	108 ± 5	85 ± 10	102 ± 15	103 ± 6	84 ± 11	101 ± 16	103 ± 15
Anserine	2	3	5	107 ± 5	89 ± 10	96 ± 9	106 ± 17	95 ± 8	108.0 ± 1.1	122 ± 21	102 ± 12	118 ± 16	121 ± 17	101 ± 3	118 ± 19
TMAO	7	10	17	100 ± 10	96.6 ± 1.1	100 ± 3	93 ± 6	97.2 ± 1.3	87 ± 14	105 ± 7	98 ± 14	87 ± 1	109 ± 19	105 ± 4	88 ± 6
Isovalerylglycine	1.7	3	4	114 ± 4	96 ± 7	101 ± 10	90 ± 5	89 ± 4	105 ± 14	107 ± 16	122 ± 20	121 ± 2	106 ± 20	121 ± 4	120 ± 5
Isobutirylglycine	1.7	3	4	106 ± 10	89.0 ± 1.3	117 ± 4	110 ± 3	97 ± 9	92 ± 4	81 ± 9	89 ± 2	89 ± 5	80 ± 19	88 ± 13	89 ± 16
Galactitol	1.7	3	4	108 ± 6	111 ± 2	119 ± 11	95 ± 5	98 ± 2	99 ± 9	112 ± 2	108 ± 21	118 ± 2	123 ± 6	116 ± 16	112 ± 18
Gallic Acid	1.7	3	4	112 ± 5	104 ± 6	96 ± 14	103 ± 10	94 ± 6	110 ± 9	102 ± 19	102 ± 21	100 ± 1.1	101 ± 20	102 ± 15	99 ± 2
Citrulline	2	3	5	107 ± 5	89 ± 10	109 ± 7	95 ± 4	103 ± 20	111.4 ± 1.1	117 ± 10	107 ± 18	92 ± 3	117 ± 15	107 ± 16	92 ± 5
Taurine	2	3	5	110 ± 4	101 ± 12	110 ± 6	107 ± 6	110 ± 4	115 ± 10	108 ± 7	107 ± 4	109 ± 5	113 ± 14	110 ± 12	113 ± 12

Note: 3-IPA = 3-indolepropionic acid; 3-IAA = indole-3-acetic acid; O-DMA = O-desmethylangolensin; TMAO = trimethylamine N-oxide.

1.4.2. GC-MS method for SCFAs and BCAAs

This method involves an initial SCFAs and BCAAs 1-step derivatization to propyl-esters, specifically tailored to the determination of these microbiota biomarkers in a small volume of urine samples.

1.4.2.1. Sample processing and analysis

The determination of SCFAs (i.e., acetic acid, propionic acid, butyric acid, valeric acid, caproic acid, heptanoic acid, isobutyric acid, 2-methylbutyric acid, and isovaleric acid) and BCAAs (i.e., valine, leucine, and isoleucine) is performed as follows; 300 µL of sample and 500 µL of 5 mM aqueous NaOH containing IS (5 µL mL-1 caproic acid-d3) are mixed in a 15 mL falcon tube. An aliquot of 500 µL propanol/pyridine solvent mixture (v/v = 3:2) and 100 µL of propyl chloroformate are added and vortexed briefly. Derivatization is carried out during 1 min in an ultrasonic water bath prior to a two-step extraction by adding 300 and 200 µL of n-hexane, respectively, followed by centrifugation (2000 x g for 5 min at 25 °C). The upper n-hexane layers containing the extracted derivatives are collected and pooled followed by thorough mixing during 3 s prior to analysis.

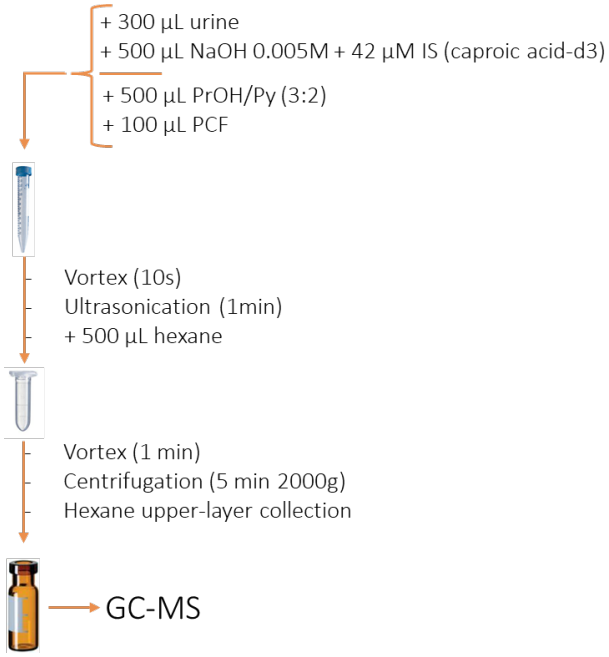


Figure 1. Experimental procedure for the SCFAs and BCAAs derivatization

GC-MS analysis is conducted using an Agilent 7890B gas chromatography system coupled to an Agilent 5977A quadrupole mass spectrometric detector (Agilent Technologies, Santa Clara, CA, USA) operating in selected ion monitoring (SIM) mode. Separations are performed using an HP-5 MS capillary column coated with 5% phenyl-95% methylpolysiloxane (30 m x 250 µm i.d., 0.25 µm film thickness, Agilent J & W Scientific, Folsom, CA, USA). One microliter of derivatives is injected in split mode with a ratio of 10:1, and the solvent delay time is set to 2.36 min. The initial oven temperature is held at 50 °C for 2 min, ramped to 70 °C at a rate of 10 °C min^{-1}, to 85 °C at a rate of 3 °C min^{-1}, to 110 °C at a rate of 5 °C min^{-1}, to 290 °C at a rate of 30 °C min^{-1}, and finally held at 290 °C for 8 min. Helium is used as a carrier gas at a constant

flow rate of 1 mL min^{-1} through the column. The temperatures of the front inlet, transfer line, and electron impact (EI) ion source are set at 260, 290, and 230 °C, respectively and the electron energy is -70 eV. The measurement parameters used for the studied analytes are summarized in Table 3.

Table 3. Measurement parameters and main figures of merit of the GC-MS method for the SCFAs and BCAAs analysis

Compound class	Metabolite	m/z	RT ± s (min)	Calibration range (µM)	R^2	LOD (µM)	LOQ (µM)
SCFA	Acetic acid	61	2.75 ± 0.02	0.7 - 655	0.997	0.5	1.7
	Propionic acid	75	3.98 ± 0.02	0.6 - 547	0.994	0.6	2
	Isobutyric acid	89	4.763 ± 0.012	0.6 - 609	0.996	0.12	0.4
	Butyric acid	89	5.59 ± 0.02	0.18 - 171	0.997	0.110	0.4
	2-Methylbutyric acid	103	6.733 ± 0.014	0.13 - 121	0.996	0.10	0.3
	Isovaleric acid	85	6.83 ± 0.03	0.07 - 65	0.998	0.05	0.15
	Valeric acid	103	8.24 ± 0.03	0.10 - 98	0.997	0.05	0.15
	Caproic acid	117	11.22 ± 0.02	0.10 - 98	0.998	0.04	0.13
	Caproic acid-d3 (IS)	120	11.281 ± 0.014	-	-	-	-
	Heptanoic acid	131	14.31 ± 0.03	0.05 - 52	0.998	0.02	0.05
BCAA	Valine	158	17.843 ± 0.012	1.0 - 908	0.998	0.5	1.7
	Leucine	172	18.164 ± 0.012	0.9 - 844	0.997	0.5	1.7
	Isoleucine	172	18.240 ± 0.015	0.8 - 751	0.998	0.4	1.3

Note: RT = retention time; R = coefficient of determination; Limit of quantification (LOQ) = concentration of analyte that can be measured with an imprecision of less than 20% and a deviation from target of less than 20% and considering the preconcentration factor achieved during sample processing; Limit of detection (LOD) = 3/10*LOQ

For quantification, an external calibration is carried out using standard solutions obtained from serial dilutions of a working solution containing mixtures of pure analytical standards in ultrapure water. A quality control (QC) sample is prepared by mixing 20 µL of each urine sample. Aliquots of the QC sample are analysed every 10 samples in the randomized analytical batch for monitoring the instrument's performance. QC RSD < 25% is the batch acceptance criteria. In addition, calibration blanks (addition of water instead of sample into the tubes) and process blanks (addition of water instead of sample to a polypropylene container, cotton or gauze pad and squeezed before adding it to the tubes) are injected. Urinary SCFAs and BCAAs concentrations are normalized to creatinine quantified with a commercial kit.

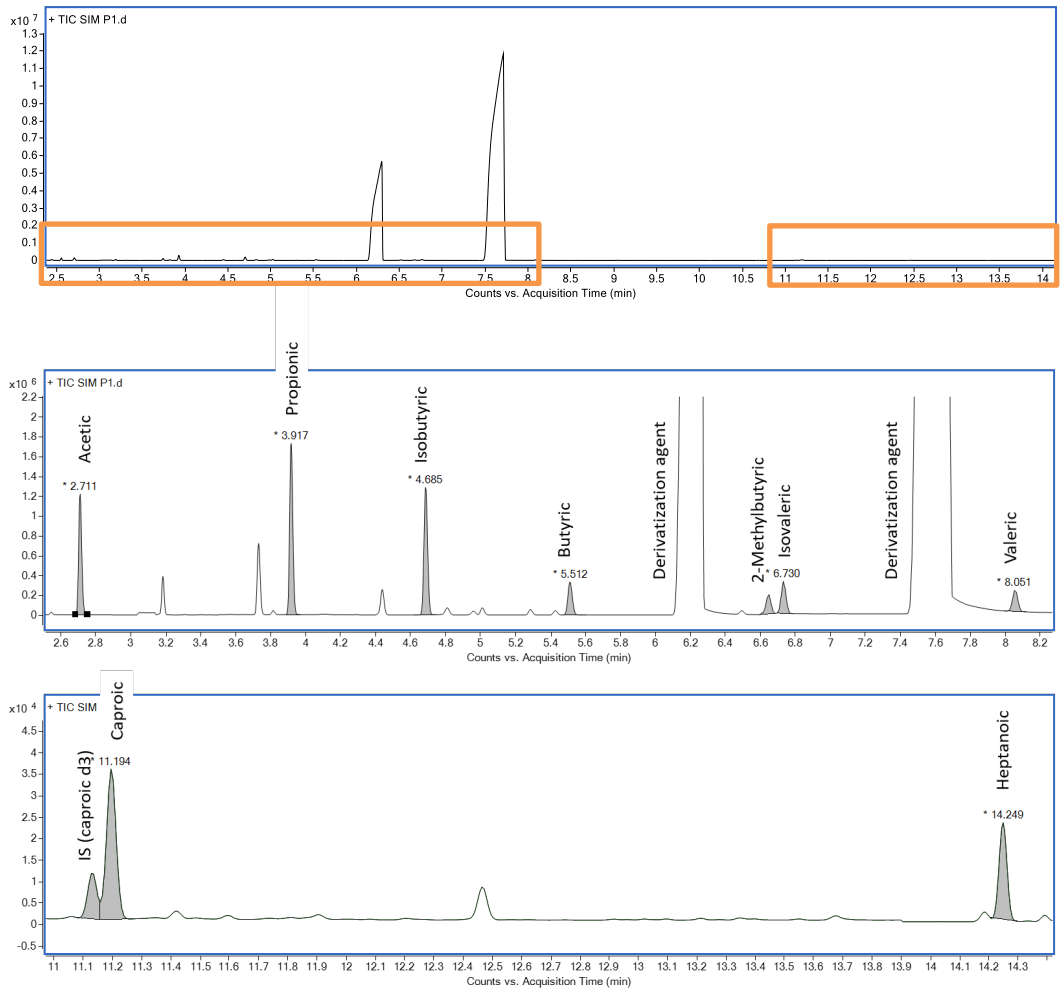


Figure 1. Chromatogram of a standard solution. On the top is the general overview of the output signal. The 2 orange rectangles represent the zoom area. The middle panel represents a zoomed view of the rightmost rectangle drawn on the top panel. The bottom panel represents a zoomed view of the rightmost rectangle drawn on the top panel. Are labelled the various SCFA that are being targeted in this analysis

1.4.2.2. Method validation results

The method was validated as described in section 1.2. Table 4 summarizes the employed concentration intervals, which were chosen considering the expected wide inter- and intra-individual variability.

Table 4. Calculated intra- and inter- day accuracy (i.e. recovery) and precision (i.e. RSD) of the LC-MS method in standard solutions and spiked urine samples

Compound class	Metabolite	Concentration levels (µM)			Accuracy ± RSD											
					Standard solutions						Spiked urine samples					
					Intra-day (N = 3)			Inter-day (N = 3)			Intra-day (N = 3)			Inter-day (N = 3)		
		Low	Medium	High	Low	Medium	High	Low	Medium	High	Low	Medium	High	Low	Medium	High
SCFA	Acetic acid	15	101	151	99 ± 3	100 ± 7	120 ± 2	116 ± 3	113 ± 8	105 ± 6	105 ± 11	106 ± 2	100 ± 5	102 ± 17	100 ± 7	106 ± 8
	Propionic acid	13	84	126	112 ± 5	99 ± 11	115 ± 9	115 ± 6	115 ± 2	115 ± 4	114 ± 3	101 ± 7	102 ± 6	106 ± 4	102 ± 3	109 ± 7
	Isobutyric acid	14	94	140	108 ± 2	96 ± 3	116 ± 5	102 ± 3	108 ± 5	101 ± 9	93 ± 4	88 ± 3	94 ± 6	92 ± 11	85 ± 3	93 ± 5
	Butyric acid	4	26	39	116 ± 7	118 ± 6	112 ± 5	118 ± 2	96 ± 1	117 ± 8	107 ± 4	106 ± 3	106 ± 4	94 ± 6	98 ± 7	100 ± 5
	2-Methylbutyric acid	3	19	28	115 ± 6	119 ± 4	104 ± 8	99 ± 5	109 ± 2	110 ± 5	111 ± 4	97 ± 3	102 ± 5	99 ± 4	100 ± 7	97 ± 4
	Isovaleric acid	2	10	15	111 ± 9	103 ± 5	118 ± 2	95 ± 4	99 ± 7	116 ± 4	114 ± 10	110 ± 6	112 ± 4	94 ± 12	93 ± 17	95 ± 16
	Valeric acid	2	15	23	101 ± 4	116 ± 11	112 ± 4	120 ± 10	107 ± 8	100 ± 6	101 ± 3	97 ± 4	99 ± 1	96 ± 10	96 ± 5	98 ± 3
	Caproic acid	2	15	23	100 ± 11	117 ± 2	115 ± 9	120 ± 7	111 ± 3	110 ± 7	102 ± 8	101 ± 2	100 ± 1	90 ± 10	97 ± 4	98 ± 2
	Heptanoic acid	1	8	12	95 ± 4	98 ± 5	110 ± 2	100 ± 11	97 ± 5	97 ± 9	93 ± 5	99 ± 2	99 ± 3	92 ± 11	97 ± 4	100 ± 3
	Caproic acid	4	16	24	110 ± 13	92 ± 9	102 ± 3	110 ± 5	90 ± 8	119 ± 4	108 ± 6	99 ± 3	97 ± 6	108 ± 5	99 ± 9	97 ± 9
	Heptanoic acid	3	11	16	119 ± 9	100 ± 10	107 ± 1	115 ± 9	102 ± 3	108 ± 5	112 ± 5	110 ± 5	92 ± 4	112 ± 7	110 ± 9	92 ± 1
BCAA	Valine	35	140	209	101 ± 5	86 ± 4	95 ± 5	118 ± 3	116 ± 10	112 ± 1	105 ± 6	115 ± 4	113 ± 9	106 ± 8	103 ± 10	94 ± 2
	Leucine	32	130	195	115 ± 9	104 ± 2	95 ± 4	117 ± 3	116 ± 1	93 ± 9	103 ± 5	108 ± 7	112 ± 5	108 ± 3	99 ± 3	97 ± 8
	Isoleucine	29	116	173	103 ± 11	103 ± 2	103 ± 11	90 ± 8	116 ± 3	90 ± 5	106 ± 2	103 ± 3	94 ± 6	112 ± 3	110 ± 10	92 ± 10

1.4.3. LC-MS method for Bile Acids

A method for the determination of urinary BAs as microbiota biomarkers is currently being developed. Tables 5 and 6 show the metabolites that are covered by the method and the IS used, respectively.

Table 5. Bile acids to be determined in urine

Abbreviation	Bile acid	Molecular formula
TUDCA	Tauroursodeoxycholic Acid	C26 H44 N Na O6 S
TLCA	Taurolithocholic Acid	C26 H44 N Na O5 S
TDCA	Taurodeoxycholic Acid	C26 H44 N Na O6 S
GDCA	Glycodeoxycholic Acid	C26 H43 N O5
TCDCA	Taurochenodeoxycholic Acid	C26 H44 N Na O6 S
GCDCA	Glycochenodeoxycholic Acid	C26 H42 N Na O5
TCA	Taurocholic Acid	C26 H44 N Na O7 S
GCA	Glycocholic Acid	C26 H42 N Na O6
UDCA	Ursodeoxycholic Acid	C24 H40 O4
LCA	Lithocholic Acid	C24 H40 O3
DCA	Deoxycholic Acid	C24 H40 O4
CDCA	Chenodeoxycholic Acid	C24 H40 O4
CA	Cholic Acid	C24 H40 O5
a-MCA	alpha-muricholic Acid	C24 H40 O5
T-a-MCA	tauro-alpha-muricholic Acid	C26 H44 N Na O7 S
b-MCA	beta-muricholic Acid	C24 H40 O5
HCA	hyocholic Acid	C24 H40 O5
GHCA	glycohyocholic Acid	C26 H42 N Na O6
THCA	taurohyocholic Acid	C26 H44 N Na O7 S
GUDCA	glycoursodeoxycholic acid	C26 H43 N O5
GLCA	glycolithocholic acid	C26 H43 N O4
HDCA	hyodeoxycholic acid	C24 H40 O4
GHDCA	glycohyodeoxycholic acid	C26 H42 N Na O5
THDCA	taurohyodeoxycholic acid	C26 H44 N Na O6 S
MCA	murocholic acid	C24 H40 O4
DHCA	dehydrocholic acid	C27 H38 O5
GDHCA	glycodehydrocholic acid	C26 H37 N O6
TDHCA	taurodehydrocholic acid	C26 H38 N Na O7 S
TUDCA-S	Tauroursodeoxycholic Acid-3-Sulfate Sodium Salt	C26 H43 N O9 S2 2Na
TLCA-S	Taurolithocholic Acid-3-Sulfate Sodium Salt	C26 H43 N O8 S2 2Na
TDCA-S	Taurodeoxycholic Acid-3-Sulfate Sodium Salt	C26 H43 N O8 S2 2Na
GDCA-S	Glycodeoxycholic Acid-3-Sulfate Sodium Salt	C26 H41 N O8 S 2Na
TCDCA-S	Taurochenodeoxycholic Acid-3-Sulfate Sodium Salt	C26 H43 N O9 S2 2Na

GCDCA-S	Glycochenodeoxychlolic Acid-3-Sulfate Sodium Salt	C26 H41 N O8 S 2Na
TCA-S	Taurocholic Acid-3-Sulfate Sodium Salt	C26 H43 N O10 S2 2Na
GCA-S	Glycocholic Acid-3-Sulfate Sodium Salt	C26 H43 N O9 S 2Na
UDCA-S	Ursodeoxycholic Acid-3-Sulfate Sodium Salt	C24 H38 O7 S 2Na
LCA-S	Lithocholic Acid-3-Sulfate Sodium Salt	C24 H38 O6 S 2Na
DCA-S	Deoxycholic Acid-3-Sulfate Sodium Salt	C24 H38 O7 S 2Na
CDCA-S	Chenodeoxycholic Acid-3-Sulfate Sodium Salt	C24 H38 O7 S 2Na
CA-S	Cholic Acid-3-Sulfate Sodium Salt	C24 H38 O8 S 2Na
GUDCA-S	Glycoursodeoxycholic Acid-3-Sulfate Sodium Salt	C26 H41 N O8 S 2Na
GLCA-S	Glycolithocholic Acid-3-Sulfate Sodium Salt	C26 H41 N O7 S 2Na

Table 6. Internal standards for the bile acid determination

Abbreviation	Bile acid	Molecular formula
LCA-d4	Lithocholic Acid-d4	C24 D4 H36 O3
CA-d4	cholic acid-d4	C24 D4 H36 O5
GCDCA-d4	glycochenodeoxycholic-d4	C26 D4 H29 N O5
GCA-d4	glycocholic acid-d4	C26 D4 H39 N O6
LCA-S-d4	Lithocholic Acid-[d4]-3-Sulfate Sodium Salt	C24 D4 H34 O6 S 2Na
GLCA-S-d4	Glycolithocholic Acid-[d4]-3-Sulfate Sodium Salt	C26 D4 H37 N O7 S 2Na
TCA-S-d4	Taurocholic Acid-[d4]-3-Sulfate Sodium Salt	C26 D4 H39 N O10 S2 2Na

2. Fatty Acid Analysis in Human Milk (HM)

This method involves the derivatization of 36 fatty acids to fatty acid methyl esters (FAMEs) prior to a *n*-hexane extraction and followed by GC-MS analysis.

2.1. Sample collection

Milk expression is accomplished with breast milk pumps following the standard operating procedure routinely employed at the hospital and the HM bank. Prior to extraction, both, removable parts of the breast milk pump and collection bottles, are sterilized. In addition, mothers have to wash their hands with soap and water and nipples with water. After extraction, bottles are maintained at -6 ºC, brought to the hospital for aliquoting (in order to avoid freeze-thaw cycles) and stored at -80 ºC.

2.2. Sample processing and analysis

For the determination of 36 FAMEs, a HM aliquot was defrosted on ice and gently shaken to avoid phase separation. Then, 250 µL of HM and 600 µL of n-hexane containing two IS (12 µM lauric acid-D23 and 26 µM nonadecanoic acid) were mixed in a 15 mL test tube equipped with Teflon-lined screw caps. An aliquot of 2 mL of CH_3OH, 2 mL of CH_3OH/HCl (3N), and 1 mL of *n*-hexane were added and vortexed vigorously. Derivatization was carried out in a water bath at 90 ºC for 60 min, with occasional additional shaking. After cooling down to room temperature, 2 mL of water were added and shaken vigorously prior to centrifugation (1200 x g for 5 min at 4 ºC). The upper hexane layer containing the extracted derivatives was transferred into GC-MS vials.

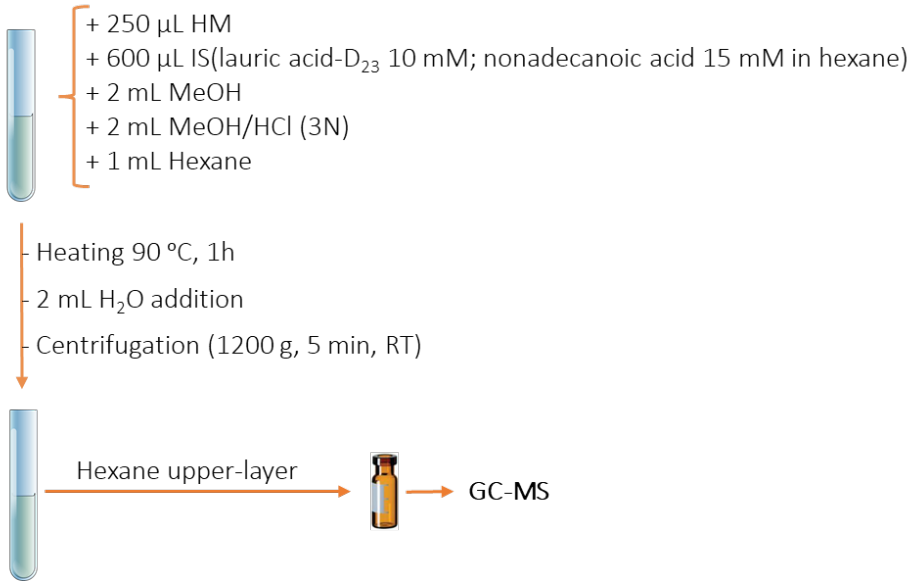


Figure 2. Experimental procedure for the FAMEs derivatization

GC-MS analysis was conducted using an Agilent 7890B GC system coupled to an Agilent 5977A quadrupole MS detector operating in selected ion monitoring (SIM) mode. Separations were performed using a Zebron™ ZB-WAXplus™ column (30 m x 250 µm i.d., 0.25 µm film thickness, Phenomenex, Torrance, CA, USA). Two microliters of derivatives were injected in split mode with a ratio of 40:1, and the solvent delay time was set to 2.6 min. The initial oven temperature was held at 60 ºC for 2 min, ramped to 150 ºC at a rate of 13 ºC min^{-1} and held for 15 min and to 240 ºC at a rate of 2 ºC min^{-1} and held for 2 min. Helium was used as a carrier gas at a constant flow rate of 1 mL min^{-1} through the column. The temperatures of the front inlet, transfer line, and electron impact (EI) ion source were set at 250, 290, and 230 ºC, respectively, and the electron energy was -70 eV. The measurement parameters used for the studied analytes are summarized in Table 7.

Table 7. Measurement parameters and main figures of merit of the GC-MS method for FAMEs determination

FAME	m/z	RT ± s (min)	Calibration range (mM)	R^2	LOD (mM)	LOQ (mM)
Hexanoate (C6:0)	133.04	5.53 ± 0.02	10.1 - 60.0	0.997	0.09	0.3
Octanoate (C8:0)	127.07	6.66 ± 0.02	6 - 41	0.994	0.04	0.13
Decanoate (C10:0)	186.13	8.73 ± 0.03	6 - 41	0.996	0.06	0.19
Undecanoate (C11)	200.16	9.762 ± 0.012	0.018 - 3.302	0.996	0.07	0.2
Laurate (C12:0)	214.19	11.08 ± 0.03	3 - 21	0.998	0.07	0.2
Tridecanoate (C13:0)	228.21	12.907 ± 0.016	0.007 - 1.312	0.997	0.010	0.03
Myristate (C14:0)	242.22	15.54 ± 0.02	0.014 - 5.544	0.998	0.06	0.19
Myristoleate (C14:1)	240.19	16.78 ± 0.02	0.007 - 1.309	0.997	0.008	0.03
Pentadecanoate (C15:0)	256.23	19.39 ± 0.02	0.006 - 1.260	0.998	0.02	0.06
cis-10-Pentadecenoic (C15:1)	254.2	21.23 ± 0.03	0.007 - 1.201	0.998	0.02	0.07
Palmitate (C16:0)	270.25	25.143 ± 0.013	2 - 13	0.997	0.02	0.07
Palmitoleate (C16:1)	268.23	26.61 ± 0.03	0.006 - 1.112	0.998	0.014	0.05
Heptadecanoate (C17:0)	284.27	31.04 ± 0.03	0.006 - 1.134	0.999	0.02	0.05
cis-10-Heptadecenoic (C17:1)	282.25	32.32 ± 0.02	0.006 - 1.105	0.999	0.015	0.05
Stearate (C18:0)	298.28	36.39 ± 0.02	0.011 - 4.112	0.998	0.012	0.04
Oleic (C18:1n9c)	296.26	37.172 ± 0.016	6 - 40	0.997	0.02	0.05
Elaidic (C18:1n9t)	296.26	37.37 ± 0.03	0.006 - 1.100	0.997	0.012	0.04
Linoleic (C18:2n6c)	294.25	39.33 ± 0.03	1.9 - 10.2	0.999	0.02	0.06
Linolelaidic (C18:2n6t)	294.25	39.7 ± 0.03	1.12 - 8.40	0.999	0.010	0.03
gamma-Linolenic (C18:3n6)	292.23	40.69 ± 0.02	0.006 - 1.176	0.999	0.02	0.06
Linolenic (C18:3n3)	314.3	42.3 ± 0.015	0.01 - 1.92	0.997	0.014	0.05
Eicosanoic (C20:0)	326.31	45.673 ± 0.014	0.005 - 0.987	0.998	0.0108	0.04
cis-11-Eicosenoic (C20:1)	324.29	46.36 ± 0.02	0.005 - 1.032	0.999	0.010	0.03
cis-11,14-Eicosadienoic (C20:2)	322.28	48.28 ± 0.03	0.005 - 1.022	0.999	0.014	0.05
cis-8,11,14-Eicosatrienoic (C20:3n6)	320.26	49.31 ± 0.03	0.005 - 1.032	0.998	0.010	0.03
Arachidonic (C20:4n6)	247.16	49.8 ± 0.02	0.005 - 0.966	0.998	0.02	0.06
Heneicosanoate (C21:0)	340.32	50.13 ± 0.02	0.005 - 1.013	0.998	0.02	0.06
cis-11,14,17-Eicosatrienoic (C20:3n3)	340.32	50.918 ± 0.013	0.005 - 1.009	0.997	0.015	0.05
cis-5,8,11,14,17-Eicosapentaenoic (C20:5n3)	247.16	52.755 ± 0.012	0.009 - 1.744	0.998	0.04	0.14
Docosanoate (C22:0)	354.35	53.757 ± 0.015	0.005 - 0.954	0.995	0.02	0.06
Erucic acid (C22:1)	352.32	54.45 ± 0.02	0.005 - 0.992	0.999	0.00	0.014
cis-13,16-Docosadienoic (C22:2)	350.31	56.23 ± 0.03	0.005 - 0.801	0.998	0.005	0.02
Tricosanoate (C23:0)	368.37	57.49 ± 0.02	0.005 - 0.966	0.994	0.02	0.06
cis-4,7,10,13,16,19-Docosahexaenoic (C22:6n3)	241.17	61.03 ± 0.02	0.009 - 1.623	0.995	0.006	0.02
Lignocerate (C24:0)	382.38	61.4 ± 0.02	0.004 - 0.886	0.998	0.02	0.07
Nervonic acid (C24:1)	380.35	61.75 ± 0.015	0.004 - 0.955	0.995	0.02	0.07

Note: RT = retention time; R = coefficient of determination; Limit of quantification (LOQ) = concentration of analyte that can be measured with an imprecision of less than 20% and a deviation from target of less than 20% and considering the preconcentration factor achieved during sample processing; Limit of detection (LOD) = 3/10*LOQ

For quantification, an external calibration line was employed using standard solutions obtained from different volumes of the Supelco 37-component FAME mix after evaporation and reconstitution in n-hexane containing derivatized IS compounds. This procedure was used to remove the 37-component FAME mix solvent (i.e., dichloromethane) and consequently, the most volatile FAME (i.e., FAME of butyric acid) was lost and could no longer be quantified.

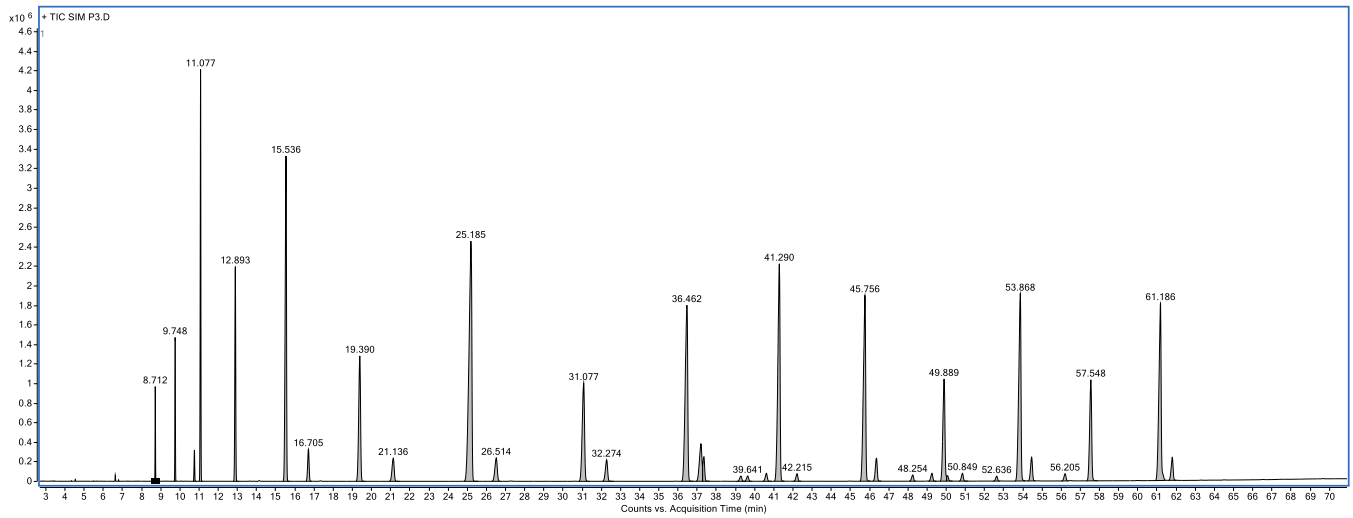


Figure 3: Chromatogram of a standard solution of FAMEs. The analytes' elution order follows the Table 7.

2.3. Method validation results

The method was validated as described in section 1.2. Table 8 summarizes the employed concentration intervals, which were chosen considering the expected wide inter- and intra-individual variability.

Table 8. Calculated intra- and inter- day accuracy (i.e., recovery) and precision (i.e., RSD) of the LC-MS method in standard solutions and spiked HM samples

FAME	Concentration levels (mM)			Accuracy ± RSD											
				Standard solutions						Spiked HM samples					
				Intra-day (N = 3)			Inter-day (N = 3)			Intra-day (N = 3)			Inter-day (N = 3)		
	Low	Medium	High	Low	Medium	High	Low	Medium	High	Low	Medium	High	Low	Medium	High
Hexanoate (C6:0)	2	3	6	97 ± 6	101 ± 4	118 ± 7	99 ± 5	118 ± 9	107 ± 7	90 ± 7	110 ± 3	100 ± 10	80 ± 9	95 ± 2	91 ± 9
Octanoate (C8:0)	1.0	2	4	92 ± 4	106 ± 7	116 ± 6	104 ± 12	107 ± 10	115 ± 17	103 ± 3	111.4 ± 1.1	117 ± 3	109 ± 4	121 ± 3	117 ± 4
Decanoate (C10:0)	1.0	2	4	99 ± 9	112 ± 9	111 ± 5	111 ± 9	99 ± 8	114 ± 15	118 ± 10	115 ± 10	109 ± 2	106 ± 9	111 ± 9	99 ± 2
Undecanoate (C11)	0.2	0.8	1.5	110 ± 9	97 ± 9	108 ± 6	100 ± 6	121 ± 10	116 ± 16	98 ± 4	115 ± 10	98 ± 3	98 ± 4	111 ± 4	103 ± 13
Laurate (C12:0)	0.8	1.2	2	105 ± 4	91.9 ± 1.3	108 ± 4	105 ± 12	103 ± 4	110 ± 9	116 ± 2	119 ± 7	118 ± 3	116 ± 3	119 ± 14	118 ± 9
Tridecanoate (C13:0)	0.2	0.8	1.5	108 ± 2	99 ± 4	103.3± 1.4	109 ± 10	118.1± 1.1	113 ± 16	104 ± 1	117 ± 11	118 ± 11	113 ± 7	120 ± 6	121 ± 8
Myristate (C14:0)	0.4	1.0	1.7	104 ± 8	106 ± 10	95 ± 6	103 ± 15	106 ± 6	106 ± 4	97 ± 9	108 ± 5	111 ± 9	108 ± 4	112 ± 6	117 ± 4
Myristoleate (C14:1)	0.2	0.8	1.5	99 ± 8	108 ± 5	92 ± 3	105.2± 1.3	117 ± 6	121 ± 2	117 ± 2	117.6 ± 1.2	113 ± 4	116 ± 4	121 ± 3	107 ± 7
Pentadecanoate (C15:0)	0.2	0.8	1.5	103 ± 9	112 ± 7	105 ± 5	113 ± 11	118 ± 10	120 ± 10	103 ± 3	117 ± 6	120 ± 3	117 ± 3	120 ± 4	103 ± 12
cis-10-Pentadecenoic (C15:1)	0.2	0.8	1.5	92.3 ± 1.2	85 ± 4	93 ± 4	110 ± 12	117 ± 15	119 ± 8	102 ± 9	117 ± 7	119 ± 9	120 ± 7	119 ± 14	114 ± 7
Palmitate (C16:0)	0.5	0.9	1.3	101 ± 7	116.2 ± 1.4	104 ± 2	81 ± 11	87 ± 14	84 ± 13	103 ± 8	98 ± 7	94.3 ± 1.1	84 ± 3	84 ± 11	83 ± 12
Palmitoleate (C16:1)	0.2	0.8	1.5	97 ± 5	105.0 ± 1.1	98 ± 7	99 ± 10	105 ± 14	103 ± 15	88 ± 6	96 ± 2	100.3 ± 1.4	103 ± 6	106.1 ± 1.0	104.8 ± 1.3
Heptadecanoate (C17:0)	0.2	0.8	1.5	110 ± 8	95 ± 5	98.3 ± 1.0	103 ± 13	110 ± 13	112 ± 18	92 ± 12	102 ± 3	106 ± 9	109 ± 11	110 ± 14	108.1 ± 1.2
cis-10-Heptadecenoic (C17:1)	0.2	0.8	1.5	97 ± 2	116.4 ± 1.3	117 ± 9	101 ± 12	109 ± 13	111 ± 7	91 ± 3	98 ± 7	103 ± 4	110 ± 6	111 ± 3	110 ± 8
Stearate (C18:0)	0.4	1.0	1.7	110 ± 5	112 ± 5	98 ± 8	80 ± 11	87 ± 9	90 ± 16	108 ± 4	97 ± 9	95 ± 7	81.9 ± 1.2	81 ± 6	82 ± 9
Oleic (C18:1n9c)	1.0	2	4	115 ± 9	105.8 ± 1.1	119 ± 4	113 ± 5	115 ± 14	108 ± 15	115 ± 9	111 ± 5	112.3 ± 1.0	113.5 ± 1.0	117 ± 5	111 ± 11
Elaidic (C18:1n9t)	0.2	0.8	1.5	109 ± 4	102 ± 3	108 ± 5	103 ± 8	107 ± 16	105 ± 11	97 ± 6	106 ± 2	103 ± 6	115 ± 4	114 ± 6	122 ± 8
Linoleic (C18:2n6c)	0.5	0.9	1.3	112 ± 2	107.2 ± 1.2	99 ± 3	107 ± 2	84 ± 16	99 ± 14	116 ± 4	109 ± 7	120 ± 2	84 ± 14	83.6 ± 1.4	97 ± 8
Linolelaidic (C18:2n6t)	0.5	0.9	1.3	101.7 ± 1.4	93 ± 7	108 ± 5	96 ± 3	103 ± 4	105.7± 1.2	86 ± 5	96.2 ± 1.4	101 ± 4	93 ± 7	99 ± 8	100 ± 13
gamma-Linolenic (C18:3n6)	0.2	0.8	1.5	113 ± 10	116 ± 10	108.0± 1.1	99 ± 15	107 ± 7	109 ± 14	86.5±1.3	99 ± 4	102 ± 5	108 ± 8	109 ± 7	108 ± 3
Linolenic (C18:3n3)	0.2	0.8	1.5	114 ± 5	89 ± 4	100 ± 5	118 ± 4	109.3± 1.0	81 ± 14	95 ± 9	88 ± 8	98 ± 9	94 ± 3	114 ± 2	104 ± 9

Eicosanoic (C20:0)	0.3	0.9	1.5	107 ± 4	117 ± 3	102 ± 6	101 ± 12	107 ± 4	108.5± 1.4	88 ± 10	98 ± 3	100 ± 5	107 ± 5	108 ± 4	106 ± 7
cis-11-Eicosenoic (C20:1)	0.5	0.9	1.3	103 ± 2	105 ± 8	96 ± 10	108 ± 13	117 ± 19	120 ± 15	110 ± 12	117 ± 9	115 ± 9	103 ± 7	115 ± 12	121 ± 13
cis-11,14-Eicosadienoic (C20:2)	0.13	0.5	0.9	99 ± 3	106 ± 7	110 ± 7	96 ± 15	102 ± 4	98 ± 6	111 ± 5	104 ± 3	100.7 ± 1.1	104 ± 2	105.7 ± 1.4	105 ± 10
cis-8,11,14-Eicosatrienoic (C20:3n6)	0.5	0.9	1.3	110 ± 5	104 ± 11	90 ± 5	96 ± 10	103 ± 11	105 ± 16	83 ± 9	94 ± 6	98 ± 2	100 ± 2	102 ± 5	101 ± 10
Arachidonic (C20:4n6)	0.5	0.9	1.3	93 ± 5	110 ± 13	104 ± 4	96 ± 8	104 ± 19	106 ± 20	98 ± 2	96 ± 8	99.3 ± 1.2	102 ± 7	103 ± 6	103 ± 13
Heneicosanoate (C21:0)	0.5	0.9	1.3	92 ± 3	119 ± 9	108 ± 9	105 ± 2	113 ± 10	115 ± 5	102 ± 3	99.8 ± 1.0	106 ± 11	87 ± 14	103 ± 8	107 ± 10
cis-11,14,17-Eicosatrienoic (C20:3n3)	0.2	0.8	1.5	110 ± 4	90 ± 5	95 ± 5	87 ± 16	94 ± 16	96 ± 14	98 ± 4	97 ± 9	101 ± 6	105 ± 7	107 ± 2	106 ± 3
cis-5,8,11,14,17-Eicosapentaenoic (C20:5n3)	0.2	0.8	1.5	101.2 ± 1.0	110 ± 3	103 ± 10	96 ± 9	106 ± 17	109 ± 7	95 ± 8	98 ± 2	100 ± 7	86 ± 6	89 ± 4	93.2 ± 1.2
Docosanoate (C22:0)	0.2	0.8	1.5	110 ± 5	116 ± 7	85.8 ± 1.2	99 ± 3	106 ± 3	108 ± 16	87 ± 8	97 ± 5	101 ± 7	105 ± 14	107 ± 4	105 ± 8
Erucic acid (C22:1)	0.5	0.9	1.3	95 ± 4	88 ± 2	88 ± 4	105 ± 5	111 ± 18	116 ± 8	104 ± 12	119.2 ± 1.3	101 ± 10	101 ± 11	103 ± 10	101 ± 10
cis-13,16-Docosadienoic (C22:2)	0.5	0.9	1.3	106 ± 10	89.0 ± 1.3	92 ± 9	93 ± 2	101 ± 11	104 ± 2	115 ± 8	103 ± 6	103 ± 11	98 ± 12	101 ± 9	100 ± 5
Tricosanoate (C23:0)	0.4	1.0	1.7	108 ± 6	111 ± 2	100 ± 10	93 ± 8	102 ± 17	105 ± 10	89 ± 4	94 ± 5	98 ± 3	96.6 ± 1.1	100 ± 14	100 ± 3
cis-4,7,10,13,16,19-Docosahexaenoic (C22:6n3)	0.5	0.9	1.3	112 ± 5	104 ± 6	114 ± 4	89 ± 4	98.9 ± 1.2	104 ± 18	110 ± 6	93 ± 6	97.2 ± 1.3	96 ± 7	99 ± 4	101 ± 10
Lignocerate (C24:0)	0.2	0.8	1.5	95 ± 4	90 ± 8	107 ± 6	96 ± 16	106 ± 11	107 ± 5	89 ± 10	103 ± 11	106 ± 10	115 ± 13	117 ± 4	117 ± 4
Nervonic acid (C24:1)	0.4	1.0	1.7	94 ± 6	102 ± 3	94 ± 9	99 ± 15	103 ± 20	110 ± 4	101 ± 12	100 ± 8	114.6 ± 1.3	104 ± 11	103 ± 7	119 ± 11

3. Questionnaires for nutritional, psychological, financial, socio-economical assessment.

3.1. Dietary assessment

Dietary assessment of Studies I & II has been described in detail in D2.7.
Briefly, in Study I three methods of dietary assessment are employed, namely a 4-day food record, which is collected through the NUTRISHIELD app, accompanied with food pictures uploading, an extensive semi-quantitative food frequency questionnaire (FFQ) and a short questionnaire assessing diet quality (KIDMED). Below, the methods are thoroughly described.

Regarding the 4-day food record, all children with the help of their caregivers (for ages <9 years) are asked to record all foods and drinks consumed for four days in the NUTRISHIELD app and upload pictures of all of them. Detailed instructions are incorporated in the app, explaining how dietary intake should be recorded as well as instructions on how the pictures should be taken. For more details on this, please see Deliverable 6.4. Nutrient intake is calculated using a standard food analysis program [the Nutritionist Pro™ Diet Analysis software (Axxya Systems, Woodinville, WA, USA)].

The second method of dietary assessment is the completion of a semi-quantitative FFQ. This questionnaire comprises 69 questions on the consumption of foods that are commonly eaten by various populations throughout a year, including dairy products, cereals, fruits, vegetables, meat, fish, legumes, added fats, alcoholic beverages, stimulants and sweet[1]. Using a 6-grade scale ("never/rarely", "1-3 times/month", "1-2 times/week", "3-6 times/week", "1 times/day", "≥2 times/day"), participants are required to indicate the absolute frequency of consuming a certain amount of food, expressed in g, millilitres or in other common measures, such as slice, tablespoon or cup, depending on the food. The previous month is set as the timeframe. The FFQ is completed by the children with the help of the caregiver when needed. From the answers to the FFQ, total energy intake, and intake of macro-nutrients is calculated.

Diet quality is assessed by the KIDMED questionnaire. The KIDMED was originally developed in an attempt to combine the Mediterranean diet characteristics as well as the general dietary guidelines for children in a single index[2]; it is based on the principles for sustaining a healthy, Mediterranean-type pattern (e.g. daily fruit and vegetable consumption, weekly fish and legumes intake), as well as on those that undermine it (e.g. frequent fast-food intake, increased consumption of sweets). The index comprises 16 questions in the form of "yes or no": questions denoting a negative connotation are assigned a value of –1 and those with a positive aspect +1. Total score ranges from -4 to 12 and it is divided into three levels indicating different levels of diet quality: (1) ≥8, optimal Mediterranean diet adherence; (2) 4-7,

[1] Bountziouka V, Bathrellou E, Zazpe I, Ezquer L, Martinez-Gonzalez MA, Panagiotakos DB. Repeatability of food frequency assessment tools in relation to the number of items and response categories included. Food and nutrition bulletin. 2012;33:288-95

[2] Serra-Majem L, Ribas L, Ngo J, Ortega RM, Garcia A, Perez-Rodrigo C, et al. Food, youth and the Mediterranean diet in Spain. Development of KIDMED, Mediterranean Diet Quality Index in children and adolescents. Public health nutrition. 2004;7:931-5

improvement is needed to adjust intake closer to the Mediterranean diet; (3) ≤3, very low diet quality. It has been used so far in a variety of settings and countries[3].

For Study II, a semi-quantitative food frequency questionnaire and a 24-hour recall is delivered to all mothers to assess energy intake as well as macro- and micro-nutrient intake and consumption of foods of specific food groups.

Regarding the 24-hour recall, trained researchers ask for all foods and beverages participants consumed the previous day, using the multiple-pass method[4]. Recall data are analysed in terms of nutrients using the dietary analysis software Nutritionist Pro™ (2007, Axxya Systems, Texas, USA). Additionally, dietary intake is grouped into food groups, namely fruits, vegetables, bread/starch, meat/high fat, meat/medium fat, meat/lean fat, meat/very lean fat, milk/non-fat fat, milk/low fat, milk/full fat and other carbohydrates.

The FFQ is administered by trained personnel and it comprises 142 questions on the consumption of foods that are commonly eaten by the Spanish population throughout a year, including dairy products, cereals, fruits, vegetables, meat, fish, legumes, added fats, alcoholic beverages, stimulants and sweets. Using a 9-grade scale ("never or less than 1 time/month", "1-3 times/month", "1 time/week", "3-4 times/week", "5-6 times/week", "1 time/day", "2-3 times/day", "4-5 times/day", "≥6 times/day") participants are required to indicate the absolute frequency of consuming a certain amount of food, expressed in g, millilitres or in other common measures, such as slice, tablespoon or cup, depending on the food. The previous month is set as the timeframe. The FFQ is a questionnaire easy to use and is not expected to increase the burden of the lactating mothers.

Based on the FFQ-responses, adherence to the Mediterranean Diet is evaluated by using the MedDietScore, a composite score calculated for each participant[4]. For food groups presumed to be part of the Mediterranean pattern (i.e. those with a recommended intake of 4 servings per week or more, such as non-refined cereals, fruits, vegetables, legumes, olive oil, fish, and potatoes) higher scores are assigned when the consumption is according to the rationale of the Mediterranean pattern, while lower scores are assigned when participants report no, rare, or moderate consumption. For the consumption of foods presumed to be eaten less frequently within the Mediterranean diet (i.e., consumption of meat and meat products, poultry, and full fat dairy products), scores are assigned on a reverse scale. As the sample of the study is lactating mothers, the original score had been modified by removing the component regarding alcohol consumption. Thus, the range of this modified MedDietScore is between 0 and 50, with higher values of the score indicating greater adherence to the Mediterranean diet.

[3] Mariscal-Arcas M, Rivas A, Velasco J, Ortega M, Caballero AM, Olea-Serrano F. Evaluation of the Mediterranean Diet Quality Index (KIDMED) in children and adolescents in Southern Spain. Public health nutrition. 2009;12:1408-12.
González-Valero G, Ubago-Jiménez JL, Ramírez-Granizo IA, Puertas-Molero P. Association between Motivational Climate, Adherence to Mediterranean Diet, and Levels of Physical Activity in Physical Education Students. Behavioral Sciences. 2019;9:37.
Çağiran Yilmaz F, Çağiran D, Özçelik AÖ. Adolescent Obesity and Its Association with Diet Quality and Cardiovascular Risk Factors. Ecology of Food and Nutrition. 2019

[4] Panagiotakos DB, Pitsavos C, Stefanadis C. Dietary patterns: a Mediterranean diet score and its relation to clinical and biological markers of cardiovascular disease risk. Nutrition, metabolism, and cardiovascular diseases : NMCD. 2006;16:559-68

3.2. Psychological Assessment

The Perceived Stress Scale for Children[5], for Study I, and the Perceived Stress Assessment Questionnaire[6] , for Study II, is administered. The Perceived Stress Scale is a measure of the degree to which an individual's situations in life are appraised as stressful. Also, it should be noted that for Study I a validation of the Perceived Stress Scale for Children is planned, which includes the administration of the questionnaire during Phase I, following a re-administration with 15-days apart from the first time and statistical analyses will be performed for correlations between the administrations.

3.3. Socio-Economical Assessment

The questions used for assessing the societal and financial aspects are according to those proposed by WHO and used in large-scale epidemiological studies, such as NHANES. In detail, in both studies, basic sociodemographic information, such as age, origin, education and sex, is collected. Also, parents, for Study I, and mothers, for Study II, are asked to self-characterize their annual family and household income.

In Study I the age (in years) of children, the number of people living with (in the family and in the household), whether they attend school (yes/no), and whether the school is private (yes/no) are recorded. Furthermore, information is collected regarding parents' marital status (1; married, 2; unmarried, 3; cohabiting, 4; single mother, 5; divorced, 6; separated, 7; widowed), parent's origin (1; White, non-Hispanic; 2; Black, non-Hispanic; 3, Hispanic; 4, Asian; 5, other), years of education described by the number of educational years (mother and father separately), the maternal and the paternal occupation [1, unemployed; 2, employee in private sector; employee in public sector; 3, self-employed; 4, working without pay in family business; 5, on leave (excluding maternity); 6, pensioner]. Finally, the mother or the father is asked to self-characterise their annual family and household income (ranging from 1, 0-4,999 € to 9, over 50,000 €) and whether the child has active health insurance (yes/no).

In Study II, participants are asked to self-characterize their annual household income and their family income (ranging from 1, 0-4,999 € to 9, over 50,000 €), whether they have active health insurance (yes/no), their employment status (1, unemployed; 2, employee in private sector; employee in public sector; 3, self-employed; 4, working without pay in family business; 5, on leave (excluding maternity); 6, pensioner), whether they are on maternity leave (yes/no) and their marital status (married, unmarried, cohabiting, single mother, divorced, separated, widowed). Education level is described by the number of years on education, while information about the number of people living with the family and in the household is also included. Participants' age and origin are recorded (1; White, non-Hispanic; 2; Black, non-Hispanic; 3, Hispanic; 4, Asian; 5, other).

[5] White B. The Perceived Stress Scale for Children: A Pilot Study in a Sample of 153 Children. International Journal of Pediatrics and Child Health. 2014;2:45-52.

[6] Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav. 1983;24:385-96

3.4. Other Assessments

For Study I, physical activity level of children is evaluated using the self-administered IPAQ questionnaire[7]. Participants also report mean nocturnal sleep duration (in hours/day) and whether they have a TV set in their room (yes/no). Smoking habits of the family are also recorded [10]. Finally, intake of dietary supplements (e.g. vitamins) is recorded and all children are also asked whether they are medically diagnosed with dietary allergies or other diseases: 1, lactose intolerance; 2, celiac disease; 3, medical diseases under restrictive dietary schemes (sub-question-please refer); 4, other (please refer)] and whether they follow a vegetarian (all sub-types) or vegan diet (yes/no). If the answer to the former question is positive, they are asked for further details. Finally, anthropometric assessment is performed and medical history is recorded.

For Study II, physical activity levels of mothers are evaluated using the validated short version of the International Physical Activity Questionnaire (IPAQ)[7]. They report the time spent on vigorous, moderate activities, walking (days/week, minutes/day) and sitting (hours/day) and their mean nocturnal sleep duration (in hours/day). All questions regarding sedentary and physical activity behaviour refer to the last week, except for sleep, which refers to the last month. Furthermore, the mothers are asked about their smoking habits[8]. Finally, anthropometric assessment is performed and medical history is recorded.

3.5. Data Collection

3.1.1 In the Paediatric Unit of the San Raffaele Hospital a researcher with the help of a Paediatrician collects data about dietary information (FFQ, KIDMED), physical activity, lifestyle and stress assessment through the App developed from WP3 (Clinical Trial Application), with the respective ID generated by the DASHBOARD after completing the general information and the medical history of the patient. Stress assessment is also collected after two weeks via phone from the Paediatrician.

[7] Papathanasiou G, Georgoudis G, Papandreou M, Spyropoulos P, Georgakopoulos D, Kalfakakou V, et al. Reliability measures of the short International Physical Activity Questionnaire (IPAQ) in Greek young adults. Hellenic journal of cardiology : HJC = Hellenike kardiologike epitheorese. 2009;50:283-94

[8] GATS Collaborative Group List. Global health promotion. 2016;23:76-8

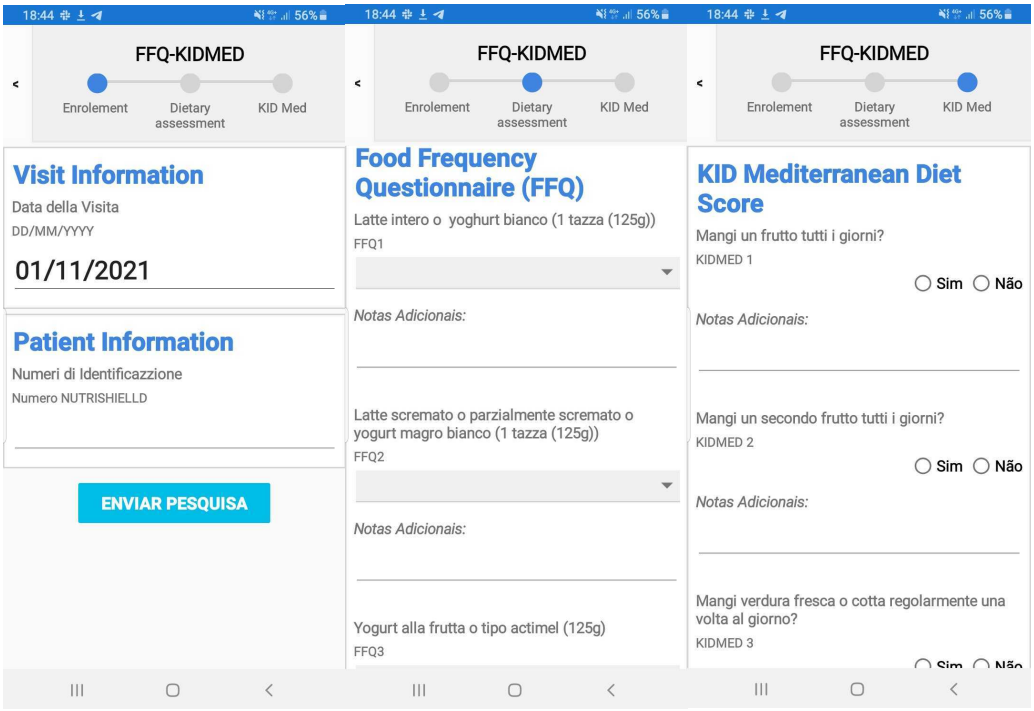


Figure 4: Screen shots of CTA as it is seen when collecting FFQ and Kidmed questionnaires. From left to right the Enrolment questions, the FFQ questions and the KIDMED questions.

For practical proposes, it was often not practical to collect data using a mobile device. In such cases, the information is collected on Excel, or on paper and then transcribed into Excel. In order to normalize an introduce the data in the system, a Web-API has been made available.

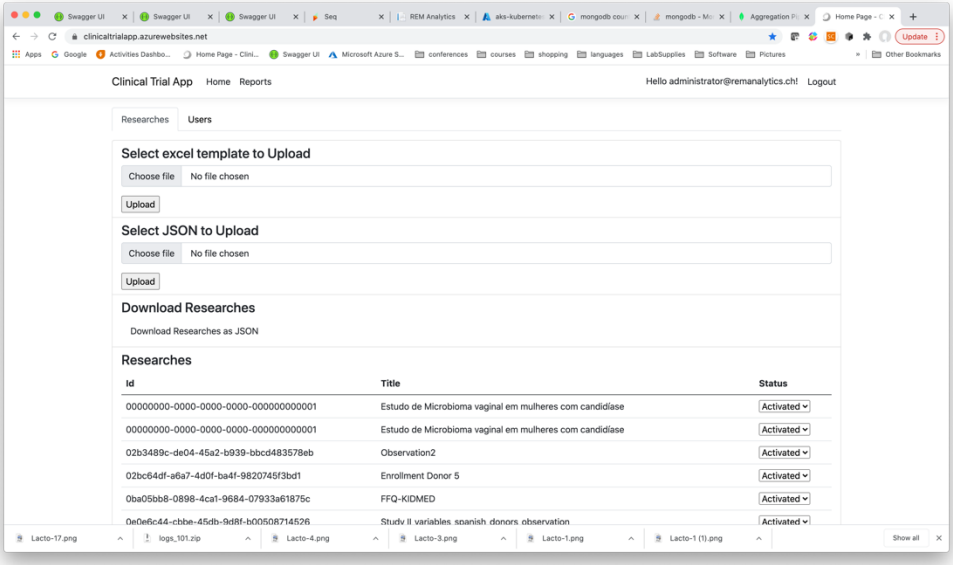


Figure 5: Web-API for the upload of excel sheets containing patient questionnaire answers.

As can be seen in Figure 5, a web-page is available to upload excel sheets. To access this page, APP users must log-in using the same username and password they have for the CTA.

3.6. Data Analysis

In order to facilitate monitoring of the Clinical trial, as well as the acquisition of key statistics, a Report generating interface has been set-up. A report is a code written in .NET

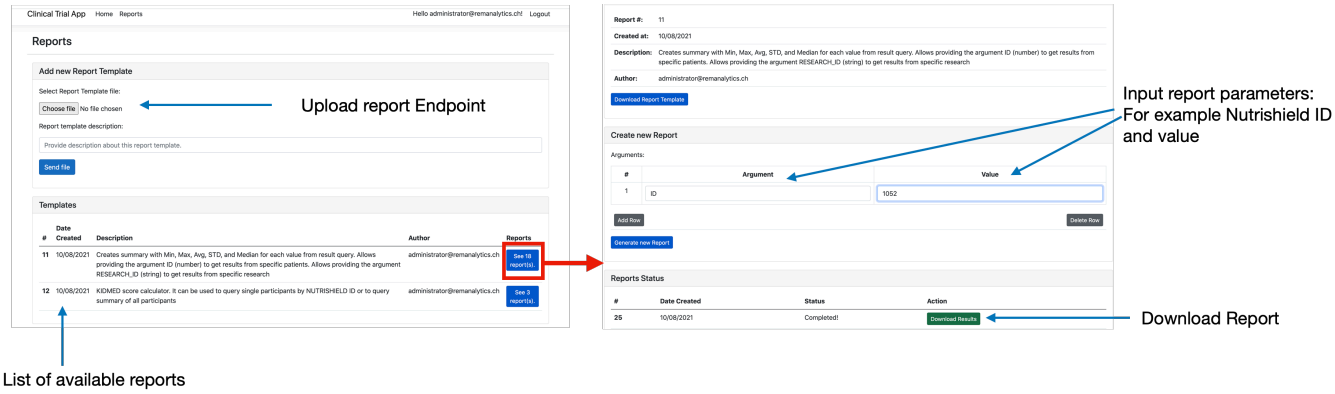


Figure 6: Current layout of CTA report generating interface. On the left is the endpoint to upload new reports generating code, as well as the list of available reports to aa given user. By clocking ono the link enclosed in the red rectangle, one accesses the page on the right. It has the interface to "use" a report. A report can take as input several parameters, in the example shown, it will take a Nutrishield ID, and generate the summary nutritional statistics for that specified patient. This report is available to the Nutritionists supervising the Clinical trial in HUA. Once computed, the reports can be downloaded.

This report system enables fast, and controlled data queries to the data being collected. A given user is only given access to the reports it is supposed to access. The data that is therefore delivered is only the "relevant" data. This ensures confidentiality, while at the same time guaranteeing that the people that need it have access to the data required.

Reports are compiled and uploaded as .NET executables, thus not exposing the underlying data structure.

PatientId	1030
MacroNutrientsCalculation__TotalProteinIntakeG	130.948436
MacroNutrientsCalculation__TotalCarbohydrateIntakeG	259.64620249999996
MacroNutrientsCalculation__TotalFatIntakeG	178.135406
MacroNutrientsCalculation__TotalAlcoholIntakeG	0.0
MacroNutrientsCalculation__TotalEnergyIntakeG	3165.5972079999997
PercentualTotalEnergyIntake__Carbohydrate	32.80849526197838
PercentualTotalEnergyIntake__Protein	16.546443201184424
PercentualTotalEnergyIntake__Fat	50.64506153683719
PercentualTotalEnergyIntake__Alcohol	0.0
FoodGroupIntakeScore__NonRefinedCereals	0.85
FoodGroupIntakeScore__RefinedCereals	0.168
FoodGroupIntakeScore__Potatoes	1.7
FoodGroupIntakeScore__Fruits	0.0
FoodGroupIntakeScore__FruitJuices	0.0
FoodGroupIntakeScore__Vegetables	0.42
FoodGroupIntakeScore__Legumes	3.2

FoodGroupIntakeScore__Fish	0.63
FoodGroupIntakeScore__Nuts	0.0
FoodGroupIntakeScore__RedMeatAndProducts	1.0010000000000001
FoodGroupIntakeScore__Poultry	0.105
FoodGroupIntakeScore__FullFatDairyProducts	2.64
FoodGroupIntakeScore__LowFatDairyProducts	2.64
FoodGroupIntakeScore__OliveOil	1.0
FoodGroupIntakeScore__VegetableOils	1.0
FoodGroupIntakeScore__Sweeteners	0.0
FoodGroupIntakeScore__Olives	0.0
FoodGroupIntakeScore__Alcohol	0.21
FoodGroupIntakeScore__ColdCuts	0.05600000000000001
FoodGroupIntakeScore__Eggs	1.0
FoodGroupIntakeScore__Sweets	0.64
FoodGroupIntakeScore__RegularSoftDrinks	0.0
FoodGroupIntakeScore__DietSoftDrinks	0.0
FoodGroupIntakeScore__TotalSoftDrinks	0.0
FoodGroupIntakeScore__SaltySnacks	0.0
FoodGroupIntakeScore__CoffeTea	0.0
FoodGroupIntakeScore__Sauces	0.21
FoodGroupIntakeScore__TotalEnergyIntakeG	17.470000000000002

Table 9: Example of a report generated and downloaded through the Web-Interface. In this example the report is the Macro Nutrient intake for a given patient.

4. Vitamin Analysis in Human Milk

4.1. Sample collection

4.1.1. Collection process

Samples are collected in 8 mL opaque sterile containers without any fixative.

4.1.2. Storage

Samples are immediately frozen at -80 ºC, and kept frozen.

4.1.3. Shipping and handling

Shipping is realized using dry ice, to maintain sample integrity throughout.

4.2. Water Soluble vitamins

4.2.1.1. Target Compounds

Vitamin Group	Compound of Interest	Concentration Range (µg/100ml)
Vitamin B	Thiamine (B1)	0.25 - 50
	Thiamine monophosphate (B1)	0.5 - 50
	Riboflavin (B2)	2- 100
	Flavin adenine dinucleotide – FAD (B2)Pantothenic acid (B5)	10 - 400
	Nicotinamide (B3)	15 - 300
	Folic acid (B9)	1 - 20

Table 10: Target water-soluble vitamins.

4.2.2. Sample preparation

4.2.2.1. Nicotinamide, Riboflavin & Flavin adenine dinucleotide

Frozen milk samples (stored at -80°C) were defrosted and homogenized by vortex and sonication prior to analysis. Measures were taken to protect the samples from UV light during the preparation such Amber glassware or aluminium foil protection and avoiding UV light sources.
To the sample (1ml), TCA (100µl; 10%) 40 µL was added and100 µL mixed for 10 minutes. After centrifugation (10 minutes;16000g), 500µl µL of the supernatant was neutralized by adding 1M NaOH solution, washed with hexane/ethyl acetate (200µl µL; 8:2 v:v) and centrifuged (10 minutes; 16000g). After removal of the upper layer, the remaining sample was analysed for riboflavin, FAD & nicotinamide or stored at -20°C before further treatment for the analysis of B1 and B6 forms (see below).

4.2.2.2. Thiamine, Thiamine monophosphate & Pyridoxal

To 80µl of the aqueous lower phase (of section 4.2.2.1), potassium ferricyanide solution (30µl) followed by semicarbazide/glycine (40µl was added and mixed by vortex. Prior to injection orthophosphoric acid (2.86 mol/L; 10µl) was added to neutralize the mixture.

4.2.3. Chromatography conditions

4.2.3.1. Chromatographic conditions for analysis of Nicotinamide, Riboflavin & Flavin adenine dinucleotide (FAD)

LC System	Waters Acquity UPLC I-Class Plus
Column	Acquity UPLC HSS T3 1,8um; 2,1x100mm
Column Temp	40 °C
Flow rate	0.35 mL/min
Mobile Phase A	Ammonium formate (25 mM)
Mobile Phase B	Acetonitrile

Table 11: Liquid chromatography equipment and set-up for Water-Lipid soluble vitamins

Time (min)	*%A*	*%B*	*Curve*
0	99	1	6
2	99	1	6
4	95	5	6
5	80	20	6
6	75	25	6
7	70	30	6
7.1	5	95	6
9	5	95	6
9.1	99	1	6
14	99	1	6

Table 12: Chromatographic Gradient of liquid phases passed through the chromatography.

PDA System	**Waters Acquity UPLC PDA detector**
FLR System	Waters Acquity UPLC FLR detector
Wavelength (PDA)	270nm
Wavelength (FLR)	435 excitation /513 nm emission

Table 13: Detection system used

4.2.3.2. Chromatographic conditions for analysis of Thiamine, Thiamine monophosphate & Pyridoxal

LC System	**Waters Acquity UPLC I-Class Plus**
Column	Acquity UPLC BEH 1,7um; 2,1x50mm
Column Temp	35˚C
Flow rate	0.30ml/min
Mobile Phase A	Na_2HPO_4 (25mM)/Methanol 19:1 (v:v) ; pH 6.5
Mobile Phase B	MeOH

Table 14: Liquid chromatography equipment and set-up for Water-soluble vitamins

TIME (MIN)	**%A**	**%B**	**CURVE**
0	100	0	6
1	100	0	6
4	70	30	5
5	70	30	6
7	30	70	6
7.1	100	0	6
9	100	0	6

Table 15: Chromatographic Gradient

FLR System	Waters Acquity UPLC FLR detector
Wavelength (FLR)	365 excitation /452nm emission

Table 16: Detection system used

4.3. Lipid soluble vitamins

4.3.1. Target compounds

VITAMIN GROUP	COMPOUND OF INTEREST	CONCENTRATION RANGE (µG/100ML)
VITAMIN A	Retinol forms	15-150
	β-Carotene	0.3-10
	β-Cryptoxanthin	1-30
	Lutein	14-50
	Lycopene	14-50
	Zeaxanthin	3-60
VITAMIN E	α-Tocopherol	100-4000
	γ-Tocopherol	20-1000
VITAMIN K	Phylloquinone	0.05-0.5
	Menaquinone-4	0.05-0.5
VITAMIN D	Cholecalciferol (D3)	0.005-1
	Calcifediol (25OHD3)	0.005-1

Table 17: List of targeted Lipid-soluble vitamins. These are the vitamins for which an analytical process has been designed and optimized within the scope of NUTRISHIELD.

4.3.2. Sample preparation

4.3.2.1. Retinol

All operations were conducted in subdued light or away from natural light using amber glassware or aluminium foil protection to minimize light induced degradation. Prior to analysis all samples were thawed at room temperature one hour before use and Milk was homogenized by vortex mix.
Human milk (200µl) was added to an amber glass tube followed by ascorbic acid (5mg), ethanol (1ml) and KOH (30%, 0.5ml). The tube was then flushed with nitrogen, sealed and its contents mixed by vortex. All samples were heated sonication in a water bath at 80°C for 30 minutes with intermittent mixing by vortex at ten-minute intervals. After cooling in an ice bath, all samples were extracted in the same amber tube with hexane/ethyl acetate (9:1 v/v containing 350mg/L BHT; 2ml) and mixed vigorously on an electric shaker for 5 min. The upper organic phase was collected in a separate tube and the extraction process repeated two more times. The combined organics were washed with water and dried under nitrogen. The residue was reconstituted in ethanol (200µl).

4.3.2.2. Carotenoids & Vitamin E

Milk samples (stored at -80°C) were defrosted and homogenized by vortex/sonication prior to analysis. Measures were taken to protect the samples from UV light during the preparation such as using amber glassware or foil protected containers and avoiding UV light sources.
To a brown glass tube containing the sample (1ml), internal standard containing δ-tocopherol solution (20µl) and BHT (80g/L in water; 30µl) were added followed by methanol (4 mL) and potassium hydroxide (70% 1ml).The contents of the tube were mixed, purged with nitrogen, sealed and heated with intermittent shaking at 45 ˚C with sonication for 10 minutes. After a further 25 minutes in the warm bath, the mixture was allowed to cool to room temperature and extracted then extracted with hexane/ethyl acetate (9:1 v/v 2ml) by shaking for 5 minutes. The above step was repeated two times, and the combined organics neutralized by consecutive water washes and evaporated to dryness under vacuum. The residue was initially reconstituted with THF/EtOH (100µl; 100µ µL);
1:1) before further dilution with acetonitrile (200µl) to obtain conditions closer to the starting mobile phase.200 µL). A final centrifugation step (5 minutes at 13000g) was performed to remove any insoluble components.

4.3.2.3. Vitamin D3

Internal standard (40µl) was evaporated under nitrogen in an amber glass tube. For calibration standards an aliquot of work or intermediate solution was also evaporated along with the internal standard. Human milk (1ml) and ascorbic acid (~5mg) were added to the dried residue and the contents mixed by vortex. Ethanol (1ml) was then added, and the contents mixed by vortex. The mixture was subsequently extracted with hexane/ethyl acetate (9:1 v/v; 2.5ml) for 5 mins on an electric shaker. The extraction process was repeated two more times, whereupon the combined organic phase was dried under nitrogen. The residue was reacted with PTAD solution (160µl; 1mg/ml) for one hour at room temperature before quenching with water (110µl) and centrifugation (5min; 16000g) to remove any suspension.

4.3.2.4. Vitamin K

Milk samples (stored at -80°C) were defrosted and homogenized by vortex/sonication prior to analysis. Measures were taken to protect the samples from UV light during the preparation such as using amber glassware or foil protected containers and avoiding UV light sources.
K1 standard solution (10µl), K2 standard solution (10µl) and internal standard solution (100µl) were evaporated and then reconstituted in pooled breast milk (0.5ml). The second and third standards were prepared in a similar fashion with 40µl and 100µl of standard solutions respectively. Water (300µl) at 40ºC was then added to the mixture followed by phosphate buffer (300µl; pH 7.9-8) and mixed. Subsequently, lipase (type VII; 100mg) was added, mixed by vortex and incubated for two hours at 37ºC. During this incubation period, the samples were removed from the oven and mixed on four occasions. Upon cooling, MeOH/EtOH (5:95 v:v; 0.5ml) and potassium carbonate (50mg) were added and mixed by vortex. The samples were then extracted with hexane/Ethyl acetate (9:1 v:v; 1.5ml) on an electric mixer for 15 min. The extraction process was repeated two times and, on each occasion, the supernatant transferred to a brown tube for evaporation under nitrogen. The residue dissolved in EtOH (200ul) prior to analysis.
Alternatively, an ongoing development method using LC-MS/MS might be used for this analysis.

4.3.3. Chromatography conditions

4.3.3.1. Chromatographic conditions for analysis of Vitamin A

LC System	**Perkin Elmer Flexar**
Column	Nucleodur 100-5 250/4 C18
Column Temp	30˚C
Flow rate	1.0ml/min
Mobile Phase	Methanol (0.5% H_2O)

Table 18: Liquid chromatography equipment and set-up for Retinol

Detector System	**Flexar UV/Vis UHPLC detector**
Wavelength (UV)	325nm

Table 19: Detection system used

4.3.3.2. Chromatographic conditions for analysis of Carotenoids & Vitamin E

LC System	Waters Acquity UPLC I-Class Plus
Column	Acquity UPLC HSS T3 1,8 µm; 2,1x100 mm
Column Temp	35 °C
Flow rate	0.5ml/min
Mobile Phase A	Ammonium acetate (50 mM)
Mobile Phase B	Acetonitrile (76) : Methanol (15) : MTBE (9)

Table 20: Liquid chromatography equipment and set-up for Lipid soluble vitamins

TIME (MIN)	%A	%B	CURVE
0	35	65	6
7	25	75	6
12.6	22	78	6
18.2	0	100	6
22.4	0	100	6
22.54	35	65	6
28	35	65	6

Table 21: Chromatographic Gradient

PDA System	Waters Acquity UPLC PDA detector
FLR System	Waters Acquity UPLC FLR detector
Wavelength (PDA)	295 nm, 450 nm, 472 nm
Wavelength (FLR)	296 nm excitation /330 nm emission

Table 22: Detection system used

4.3.3.3. Chromatographic conditions for analysis of Vitamin D3

LC System	Waters Acquity UPLC H-Class
Column	Acquity UPLC BEH Phenyl 130Å 2,1 x 50 mm x 1,7 µm column
Column Temp	40̊ C
Flow rate	0.6 ml/min
Mobile Phase A	Water (0.1% formic acid)
Mobile Phase B	Acetonitrile (0.1% formic acid)

Table 23: Liquid chromatography equipment and set-up for Vitamin D

TIME (MIN)	%A	%B	CURVE
0	80	20	6
0.25	80	20	6
4.75	0	100	6
6.5	0	100	6
6.6	80	20	6
9.0	80	20	6

Table 24: Chromatographic Gradient

MS System	**Waters TQ-S**
Desolvation Temperature	500°C
Desolvation (L/Hr)	1000
Cone (L/Hr)	150
Source Temperature	120

Table 25: Detection system used

	PARENT (M/Z)	DAUGHTER (M/Z)	DWELL (S)	CONE (V)	COLLISION (V)
25OHD3-PTAD	558.4	298.1	0.06	30	15
25OHD3-PTAD DEUTERATED	561.4	301.1	0.06	30	15
D3-PTAD	560.3	298.1	0.06	43	8
D3-PTAD DEUTERATED	563.27	301.15	0.06	4	16

Table 10: MS transitions

4.3.3.4. Chromatographic conditions for analysis of Vitamin K

LC System	Perkin Elmer Series 200 LC Pump /Waters ™ 717plus Autosampler
Column	EC Nucleodur 100-5 C18 5µm; 4,0x125mm
Reduction Column	Zinc powder 4,0x50mm
Column Temp	25̊ C
Flow rate	0.7ml/min
Mobile Phase	Dichloromethane/MeOH/solution acetate & zinc chloride (100:900:5)

Table 26: Liquid chromatography equipment and set-up for Vitamin K

FLR System	Thermo FL3000
Wavelength (FLR)	243nm excitation /430nm emission

Table 27: Detection system used

4.4. Data Integration

4.4.1. Water-soluble Vitamins

4.4.1.1. Integration procedure for Nicotinamide, Riboflavin & Flavin adenine dinucleotide

Quantification was performed by standard addition based on a calibration curve consisting of five concentration levels prepared in the matrix. Calibration coefficients greater than 0.999 were obtained for all compounds. As quality control samples, pooled human milk was analysed in triplicate.

4.4.1.2. Integration procedure for Thiamine, Thiamine monophosphate & Pyridoxal

Quantification was performed by standard addition based on a calibration curve consisting of five concentration levels prepared in the matrix. Calibration coefficients greater than 0.98 were obtained for all compounds. As quality control samples, pooled human milk was analysed in triplicate.

4.4.2. Lipid soluble Vitamins

4.4.2.1. Integration procedure for Vitamin A

Quantification was performed using linear regression ($r^2 = 0.999$) based on the external calibration of four standards. As quality control samples, pooled human milk was analysed in triplicate.

4.4.2.2. Integration procedure for Carotenoids and Vitamin E

Quantification was performed using linear regression based on a calibration curve consisting of five concentration levels. Peak areas were corrected by internal standards (δ-tocopherol for Vitamin E forms and trans-β-apo-8'-carotenal for carotenoid forms). Calibration coefficients greater than 0.98 were obtained for the different carotenoid and tocopherol forms. As quality control samples, pooled human milk was analysed in triplicate.

4.4.2.3. Integration procedure for Vitamin D

Quantification was performed by standard addition based on a calibration curve consisting of five concentration levels prepared in the matrix with correction by isotopic internal standards. Calibration

coefficients greater than 0.99 was obtained for Cholecalciferol (D3) and Calcifediol (25OHD3). As quality control samples, pooled human milk was analysed in triplicate.

4.4.2.4. Integration procedure for Vitamin K

Quantification was performed by standard addition based on a calibration curve consisting of three concentration levels prepared in the matrix with internal standard correction. Calibration coefficients greater than 0.99 was obtained for Phylloquinone (K1) and Menaquinone-4 (K2). As quality control samples, pooled human milk was analysed in triplicate.

4.5. Measurement of Vitamins D3 in Human milk

The measurement of Vitamin D3 in Human milk provides opportunities to determine nutritional status of mother-infant dyads from a non-invasive analysis, or for interventional or observational studies building knowledge on the composition of human milk. This approach could also explore the effect of high-dose maternal Vitamin D3 supplementation alone as possible option to prevent deficiency in the breastfeeding mother-infant dyads.

5. Microbiome Analysis

5.1. Sample collection

5.1.1. Sample tube preparation

5.1.1.1. Storage Buffer

As was described in D2.1, a buffer capable if immediately fixing all bacterial activities, while at the same time preserving the DNA needs to be used to store all microbiome samples. The same buffer, also needs to enable the easy extraction of bacterial DNA in downstream processing.
After several experimentations, the following buffer recipe was chosen:

1. 1 mL of 500 mM EDTA solution
2. 25 mL of 10% SDS Solution
3. 500 mL Ethanol
4. Enough water to dissolve EDTA & SDS completely (~30 mL).

The ethanol concentration guarantees that all bacterial activities are blocked upon mixture with the sample. EDTA prevents enzymatic activities such as DNase (DNA breaking enzymes) from happening. SDS also helps neutralizing micro-organisms, viruses, as well as contribute in DNA extraction.
The described buffer is known as the “Fixation” buffer, since it fixes the sample in time preventing any modifications to the bacterial relative abundance. This buffer guarantees no change in bacterial composition at room temperature for over 1 month.

5.1.1.2. Collection Tube

To collect the microbiome, 2 mL sample tubes are used. In each tube 1 mL of Fixation buffer. The fixation buffer contains the following: In brief, the buffer contains the following: 1% SDS, 0.1M EDTA in 95% Ethanol.
Tubes are then sealed, and submitted UV treatment. UV treatment is necessary to guarantee that there is no DNA contamination present in the tubes prior to sample collection. The following specifications were followed:

- UV wavelength: 254 nm
- UV intensity: 54µW/cm^2
- UV duration: 180 min

In a previous experiment, the described UV treatment was shown to completely degrade any DNA present in the sample. This was tested in similar conditions, with equivalent tubes, sealed and in similar buffers.

5.1.1.3. Human Milk

Approximately 1 mL of freshly obtained HM sample is taken with a sterilized pipette and placed into a collection tube containing the storage buffer. The storage buffer tubes are prepared at REM, and shipped to HULAFE.

5.1.1.4. Baby Stool

Approximately 1 g of freshly obtained stool sample is taken with sterilized tweezers and placed into a collection tube containing the storage buffer.

5.1.1.5. Teenager Stool

After the recruitment, stool samples are collected by the families in the home environment to avoid as much as possible children' discomfort. Once the samples are in the hospital, a researcher prepares the aliquots for the analysis in the provided ethanol vials, with the respective ID generated from the dashboard. Afterwards, the aliquots are barcoded in the Clinical Trial App and stored at San Raffaele Scientific Institute at -20°C until the shipment to REM Analytics for analysis.

5.1.2. Sample storage and logistics

All samples are stored in -20 °C following collection. The fixative guarantees stability for over 30 days at room temperature (RT). However, for long period storage freezing is preferable. Once sufficient number of samples are collected, they are shipped at room temperature to REM Analytics for analysis. Rem receives them and stores them at -80 °C until processing.

5.1.3. Sample Identification

All samples are identified by a unique QR Core. These QR codes are placed on the collection tube either before shipping them to the relevant hospital or following sample collection Using this QR Code, sample collector can register the sample on the CTA.

5.2. Assay design

5.2.1. Concept

Advanced Testing for Genetic Composition (ATGC) represents a new paradigm in microbiome analysis. It provides accurate and quantitative results on microbiome composition within a selected "Target."
The target can be seen as the "micro-organisms space" that is of interest in a specific experiment. The analysis set-up has the following step:

- Target definition
- Sequence Mapping
- OTU Definition
- Primer Design
- Optimization and Calibration

Target definition: is the process of defining a manageable set of microorganisms that will be probed in the study. This list can be quite comprehensive. In our case, we will begin by focusing on the Lactobacillus, Bifidobacterium, Enterococcus and Yeasts. With the possibility of adding the Akkermansia later on. Each of these genera have their own complexities, which will be explored through sequence mapping.

5.2.2. Mapping the sequences

The first step in designing an ATGC assay is to map the sequences. This enables a clear vision of the genetic distances between target organisms. On the genetic map, it is also possible to select the OTUs that one wishes to discriminate between. The mapping process relies on a proprietary bio-informatic pipeline of REM analytics. It can process a large number of genetic sequences from a variety of sources, both public and private. The results are then displayed in an interactive 3D map.

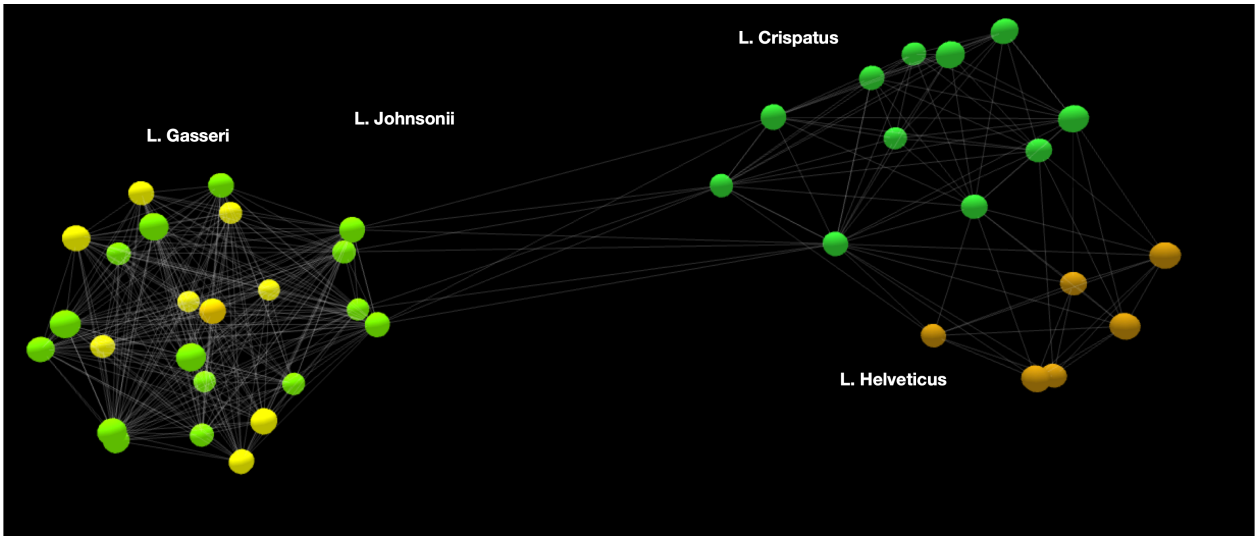


Figure 7: Example of genetic map generated using REM's pipeline. This map displays a sub-set of lactobacillus species. From this map, it is clear that there are some comparisons which are more appropriate than others. It is easy to compare L. gasseri *to L. johnsonii. However, a direct comparison between L.* gassseri *and L.* helveticus *would be impractical. It is also clear from this map, that a grouping of L. gasseri with L. johnsonii to be compared with a group including L. crispatus and L. helveticus would be possible.*

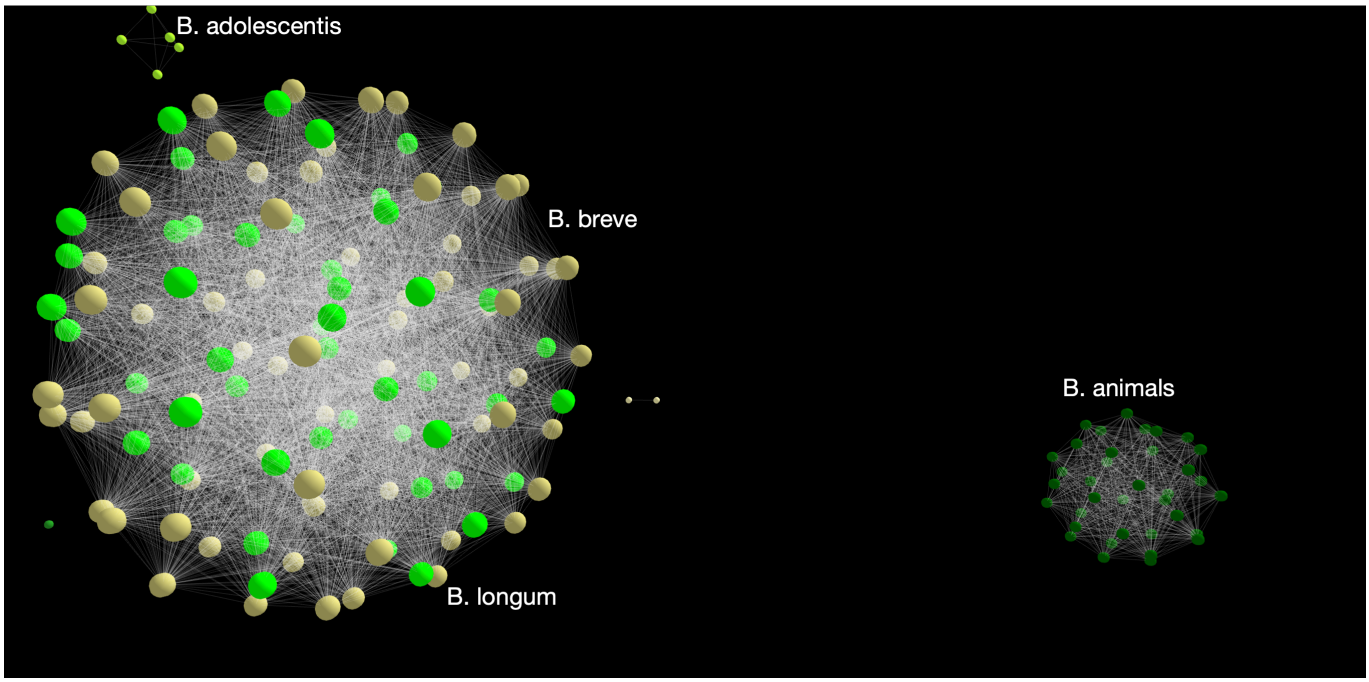


Figure 8: A Bifidobacterium Map. It is immediately clear by comparison with Figure 7, that the Bifidobacterium Genus is different from the Lactobacillus. With B. breve and B. longum forming a clique, and B. animalis behaving almost like a separate genus, rather than a species within a genus. From this map, it was inferred that Bifidobacterium must be analysed at the sub-species level, and that species level analysis will not provide sufficient information.

5.2.3. Selecting the OTUs

From maps such as shown in Figure 7 and Figure 8, it is possible to intelligently define the OTUs of interest. An OTU, or Operational Taxonomic Unit, is a group of micro-organisms that is not further resolved. It is part of the Assay definition, and our genetic maps help in the process of selecting them. For example, an OTU can be a single strain (that needs to be discriminated from the rest of the species), or a subspecies, or an entire species.

Looking at the map in Figure 8, we see that individual Bifidobacterium species behave almost like genus. They have several sequences clustering together. Thus, further detailed maps had to be established for individual species. Figure 9 shows a map of only Bifidobacterium longum. It shows that subspecies form distinct clusters, and should therefore be considered as OTUs.

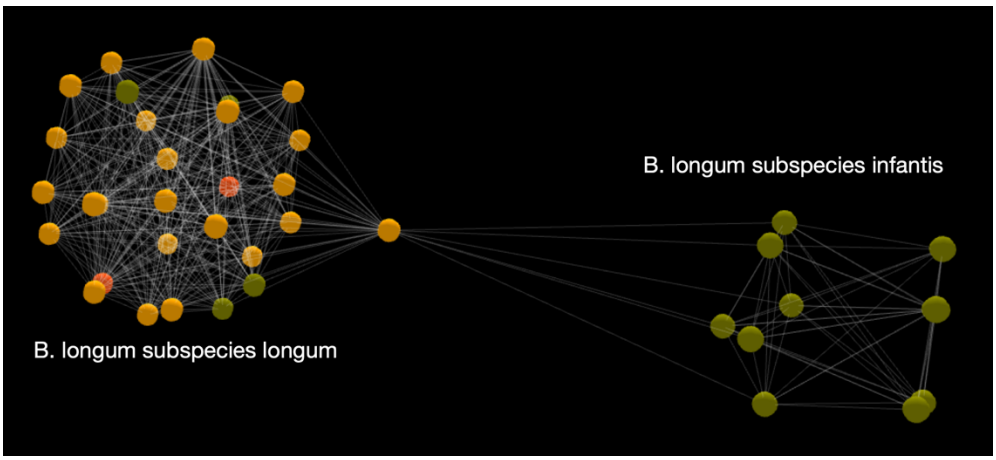


Figure 9: A genetic map of Bifidobacterium longum only. It is clear that the subspecies longum and infantis separate. Thus, it makes sense to define OTUs at the subspecies level, and to discriminate between the two clusters visible here.

The currently selected OTUs are shown in Table 28. As can be seen, the Bifidobacterium are split into sub-species level. This is due to the large genetic differences observed within the species. However, the lactobacillus and the Yeasts are not. The list in Table 28 is susceptible to change as observations become available. Not enough information exits on the Lactobacillus species to split the OTUs below the species level

GENUS	SPECIES	SUBSPECIES
Bifidobacterium	animalis	animalis
Bifidobacterium	animalis	lactis
Bifidobacterium	animalis	
Bifidobacterium	longum	longum
Bifidobacterium	longum	infantis
Bifidobacterium	longum	suillum
Bifidobacterium	longum	
Bifidobacterium	Breve	
Lactobacillus	johnsonii	
Lactobacillus	crispatus	
Lactobacillus	gasseri	
Lactobacillus	iners	
Lactobacillus	helveticus	
Lactobacillus	rhamnosus	
Lactobacillus	salivarius	
Lactobacillus	paracasei	
Lactobacillus	delbrueckii	
Lactobacillus	plantarum	
Lactobacillus	brevis	
Lactobacillus	fermentum	

Saccharomyces	cerevisiae
Saccharomyces	boulardii
Malassezia	furfur
Malassezia	globosa

Table 28: Currently selected OTUs. Initial focus is on the Lactobacillus, Bifidobacterium and Yeasts. Akkermansia will be added as soon as possible.

5.2.4. Reference Material

In order to calibrate the analytical process, a large set of reference material is being collected progressively. These are both mixes of bacterial species and strains, as well as several individual strains. These have been sourced from:

- Commercial suppliers, such as probiotics companies
- Bacterial repositories, such as DSMZ
- Bacterial culture laboratories such as IHMA

In total, over 100 different bacterial strains have been collected in as part of the reference library at REM. Most of the species and strains are in the Bifidobacterium, Lactobacillus and Saccharomyces genus. However, several other strains are available in the Candida, and Akkermansia genera. More reference material is being collected regularly, and the library is constantly increasing in size.

5.2.5. Validation and calibration

The following experiments have been realized as part of the initial validation process before running the Pre-Pilot samples:

Analysis of artificial mixed samples: These are samples constructed by producing "representative" samples using reference material.
Analysis of spiked samples: Reference material is mixed into human milk, and into stool samples, and the complete process of DNA extraction, PCR and CTCE is performed. It enables the validation of the complete process to extract and detect the desired micro-organisms.
Dilution series: In order to demonstrate the linearity of the analysis, the only solution is a dilution series of one bacterial type into another. If the results are linear on the log scale, then it demonstrates the quantitative precision of the whole process.

The last step of dilution series is the only way to ensure that the process is quantitative. It is very difficult to have absolute values for bacterial amounts with confidence intervals in the order of 3% (which is the theoretical limit of ATGC). Consequent, it is through the establishment of dilution series of several bacteria into others, that the quantitative precision can be established. For each of the primers, 2 bacteria that the primer can distinguish are selected.

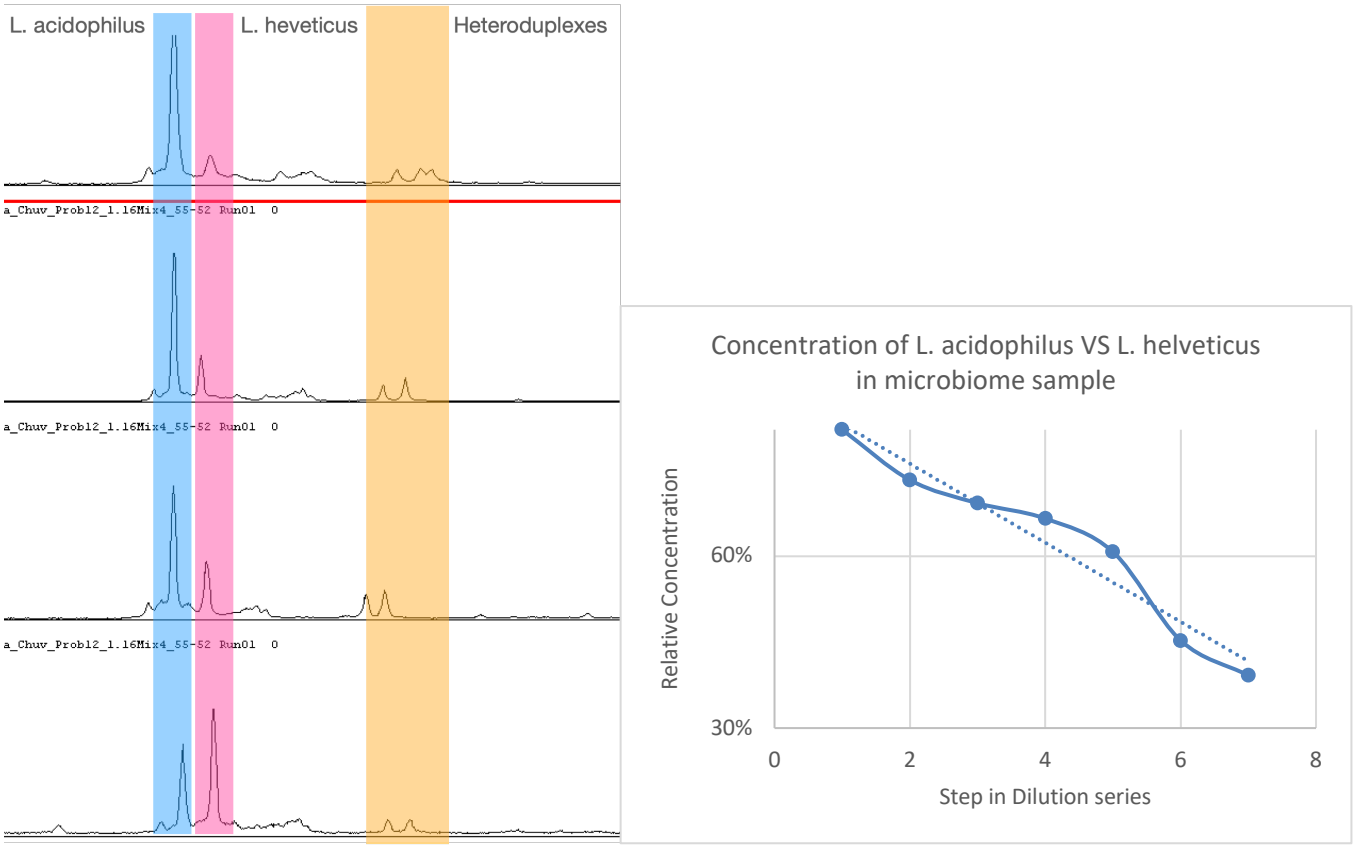


Figure 10: Dilution of reference material: Lactobacillus acidophilus, into a microbiome sample. The sample contains L. helveticus, normally found in several microbiomes. On the left, the raw data coming from the instrument, showing 4 steps of the dilution series. On the right, the measured relative abundance of L. acidophilus against L. helveticus, measured after signal processing. The relative abundance is plotted on the log scale. The linearity of the measures demonstrates the quantitative precision of the method.

5.3. Sample analysis

5.3.1. DNA extraction

This protocol was developed to process microbiome samples from human milk and can be applied to human gut microbiome. Its main particularity is the ability to extract DAN without separating the fat layer. Significant evidence exists that many micro-organisms live specifically in the lipid layer of milk, and thus removing it (as is suggested by most commercial kits) is not appropriate for NUTRISHIELD's objectives. This protocol does not require centrifugation or spin columns. Binding of DNA is performed with magnetic beads. This protocol is designed for high-throughput automation with automated liquid handling.

5.3.1.1. Materials and Buffers:

Lysis Buffer 1:
1% SDS
5% Triton X-100
250 mM NaCl
200 mM Tris-HCl pH 7.6
25 mM EDTA
80 mM DTT

Wash Buffer:
70% EtOH

Elution Buffer:
10 mM Tris-HCl pH 9
0.5 mM EDTA

Treatment Conditions:
A: Heating at 80C 15 min

Materials:
1. OH-1000 magnetic beads Beaver
2. Heat block
3. Vortex
4. Deep well and normal 96 well plates

5.3.1.2. Process

I Sample Lysis
1. Take 750 µL Human Milk sample already fixed in 50% EtOH and stored at -20 °C freezer and put in a 1.5 mL Eppendorf tube/Deep well
2. Add 1 Volume of Lysis buffer (750 µL)
3. Vortex vigorously to make sure fat goes into solution and sample is homogenous
4. Heat sample at 80 °C for 15 min without shaking (only once all samples are vortexed with Lysis buffer)
5. Let cool to room temp all samples (Takes 20 mins)

II Binding
1. Add 60 µL of Magnetic Beads OH-1000
2. Invert tubes 6X to mix
3. Let stand 20 min, occasionally inverting tubes every 5 min

III Washing
1. Engage Magnet to draw beads (20 mins)
2. Remove supernatant

3. Add 500 µL 70% EtOH
4. Mix by pipetting 8 times
5. Engage Magnet to draw beads (2 mins)
6. Remove supernatant
7. Repeat steps 3-6 once more
8. Add 100 µL 70% EtOH, mix and transfer all to 96 well 150 µL well plate
9. Engage magnet to draw beads
10. Remove all supernatant completely

IV Elution

11. Add 60 µL TE
12. Mix by pipetting 8 times
13. Heat the beads resuspended in TE 10 min at 65C 300 rpm (Tape the plate to prevent evaporation)
14. Transfer eluted DNA to new well
15. Store at -20 °C

5.3.1.3. DNA extraction Automation

The process described for DNA extraction has been designed with automation in mind. This is the reason why no centrifugation step was included, and magnetic beads are used. In order to account for the large number of samples

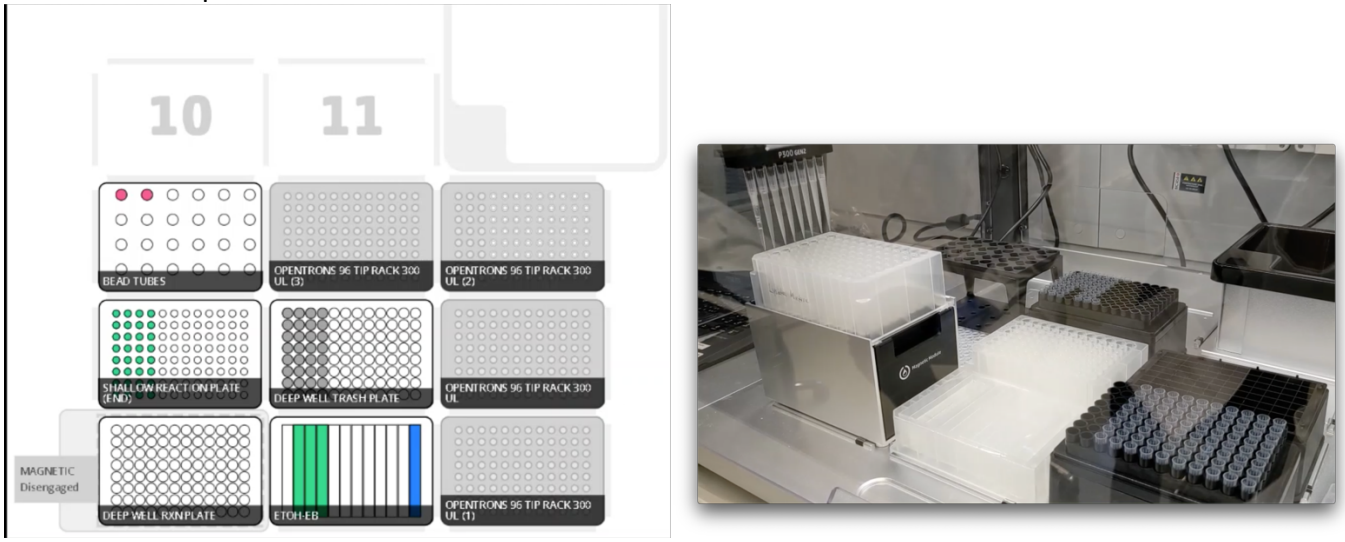


Figure 11: Left: Layout of samples and consumables on robot. Right: Picture of automatic pipetting robot in operations extracting DNA from human milk.

Figure 11 shows the current automation solution used for DNA extraction. The platform is called Opentron, and all protocols are written in Open Source Python code. Which enables sharing across laboratories, and easy replication of results. With the current set-up it is possible to extract up to 48 samples a day, from either human milk, or human stool samples.

5.3.2. PCR conditions

5.3.2.1. Calibration and optimization

As is described in section 5.2, to fully characterizes the microbiome of any of the samples, a set of primers that amplify the selected fragments are needed. Each primer must be carefully optimized so that its PCR conditions are known

5.3.2.2. PCR Preparation and recipe

All PCR are performed using the 3-primer system described in Refinetti et al.[9] In brief, one of the specific primers (forward or reverse) has an added "GC-tail" at the end. A third primer which amplifies on the complement of this GC-tail is also used. This third primer is also FAM labelled on its 5' end, and is added GC-FAM primer.
PCR recipe has been standardized for all primers used in the context of NUTRISHIELD. The recipe is as follows:

- 400 nM of primer without GC-tail
- 200nm of primer with GC-tail
- 100nm of GC-FAM primer
- 400µM of dNTPs
- 1U Pfu with SSo7D fusion protein.
- Into 1X Thermopol Buffer

The DNA concentration is variable, as is described later.

5.3.2.3. Primer Specific optimization.

Although all PCR will follow the same recipe, each primer requires different PCR conditions. These are mostly the Annealing temperature. Each designed primer must therefore have its PCR condition optimized. This is done by running each of them on an annealing temperature gradient, and selecting the temperature with highest performance.

[9] Refinetti, P., Morgenthaler, S., & Ekstrøm, P. O. (2016). Cycling temperature capillary electrophoresis: A quantitative, fast and inexpensive method to detect mutations in mixed populations of human mitochondrial DNA. Mitochondrion, 29(C), 65–74. http://doi.org/10.1016/j.mito.2016.04.006

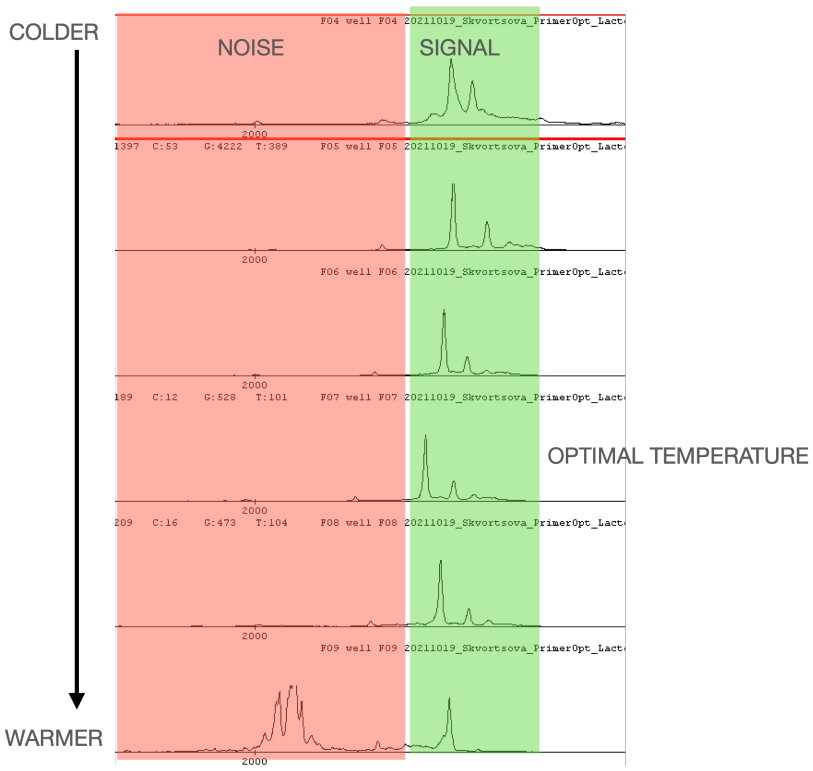


Figure 12: Primer specific PCR condition optimization. This example shows aa primer specific for the discrimination between L. helveticus and L. crispatus. As can be seen, as the temperature increases, the SNR gradually improves, reaches a maximum, and then decreases.

5.3.2.4. Sample Specific Optimization

PCR in general, can only amplify within a given copy number range. This range is about 2 orders of magnitude (factor 100). Which mean that aa primer that amplifies Organisms group A, which is 1000 times les abundant than Organism group B, will need a different DNA concentration in its PCR reaction.
The range of 2 orders of magnitude provides latitude for intra-sample variability in given micro-organisms. However, a sample-primer DNA concentration optimization is required in order to provide maximum. For each sample type, and each primer, an optimal concentration of extracted DNA/PCR reaction must therefore be defined.
To define it, a mix of DNA is produced by selecting several representative samples from the Pre-Pilot. A dilution series is then prepared by serially diluting the DNA, and running each primer on this dilution series. The concentration where the optimal products are observed is then selected for use during the analysis.

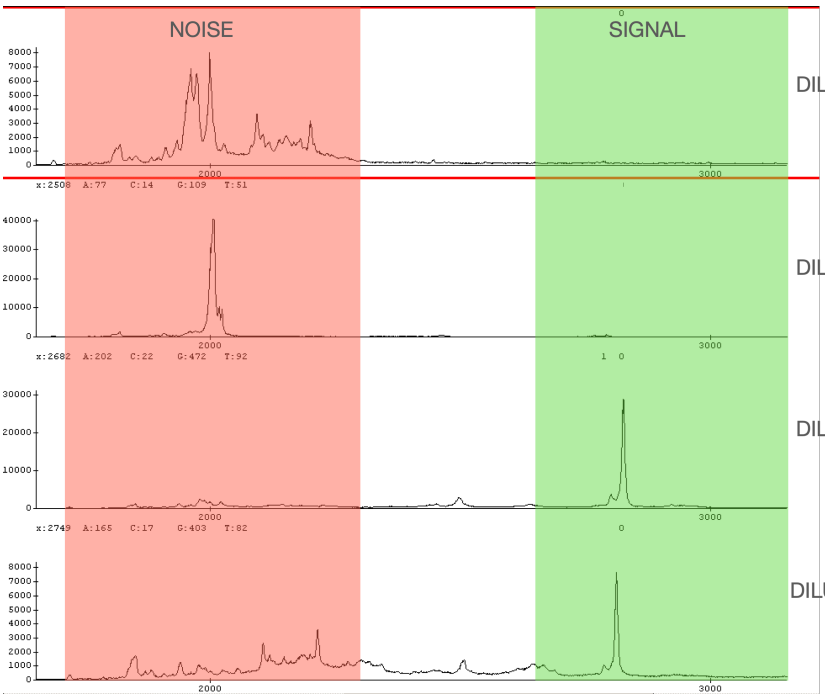


Figure 13: Example of DNA concentration calibration. Are shown 4 electropherogram, of the same primer on the same sample. In green is the area where the actual product, while the area in red is the PCR noise generated by sub-optimal conditions. The only difference between the 4 signals is the initial DNA concentration. As can be seen, too high concentration also prevents PCR, ass does too low (not visible in this experiment).

Results such as shown in Figure 13, clearly demonstrate the need for careful optimization of DNA concentration in PCR reactions. There is no "one concentration fits all." This is due to the very different concentration of different micro-organisms in samples. It is clearly the case that some micro-organisms are more than 1000 times more abundant than others. Thus, specific assay portions that target these micro-organisms must have different DNA concentration.
For each of the microbiome sample types: Huma milk, new-born stools and teenager stools, the DNA concentration for each of the assay's element must be optimized.

5.3.3. Cycling Temperature Capillary Electrophoresis

5.3.3.1. Set-up

CTCE is performed using MegaBace system. These are sanger sequencers that have been modified and adapted for CTCE. They use fused silica capillary with an internal diameter of 75µm, and an effective length of 32cm. The detection is achieved through an Argon laser with wavelength of 488nm.

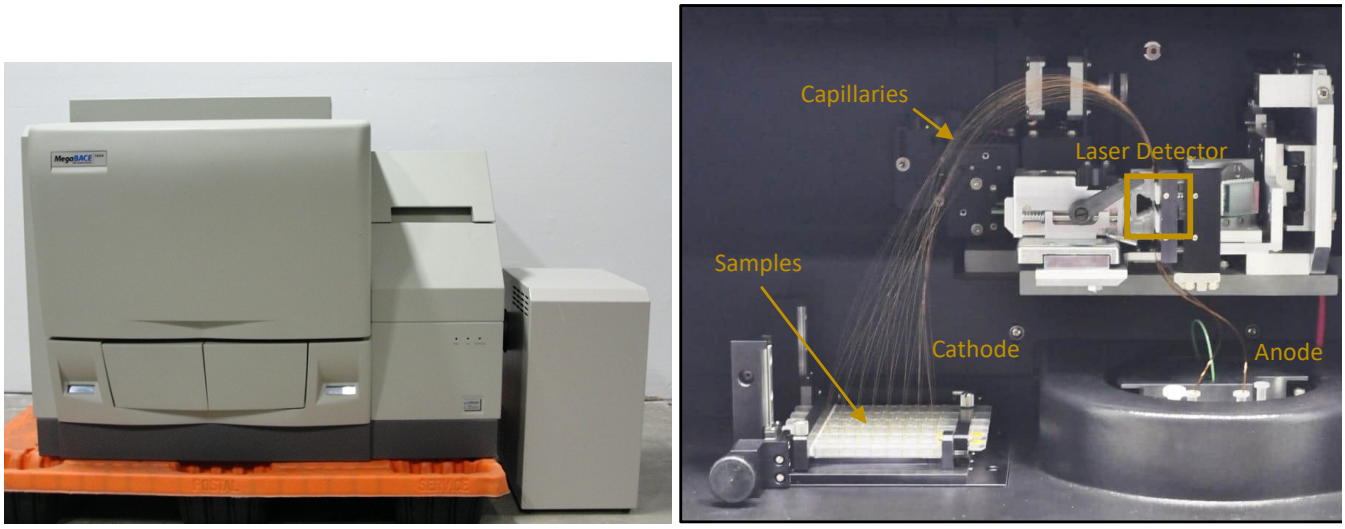


Figure 14: MegaBace 1000 system. On the left is an outside view. On the right, is a view of the inside, with annotations.

5.3.3.2. CTCE conditions

CTCE is required to separate the variants after PCR. A general CTCE program has been developed for all fragments. The following conditions are constant:

- Injection voltage: 10kV
- Injection time: 30sec
- Electrophoresis voltage: 9kV

The gel used has been optimized for separation under CTCE conditions, and it is a mix of 2% Linear Polyacrylamide with 1% Poly-dimethyl acrylamide, with 7M Urea.
During the electrophoresis, the temperature is cycled 20 times, between TL and TH. TL and TH are the lower and upper bound of the temperature cycles, and are adapted for each primer.

5.3.3.3. Cycling condition optimization

For individual primers, the cycling temperature must be optimized, in order to have optimal separation between the variants. This is achieved by testing several cycling conditions on PCR product that has been amplified on reference material. The material must contain 2 varieties that are resolved by a given primer.

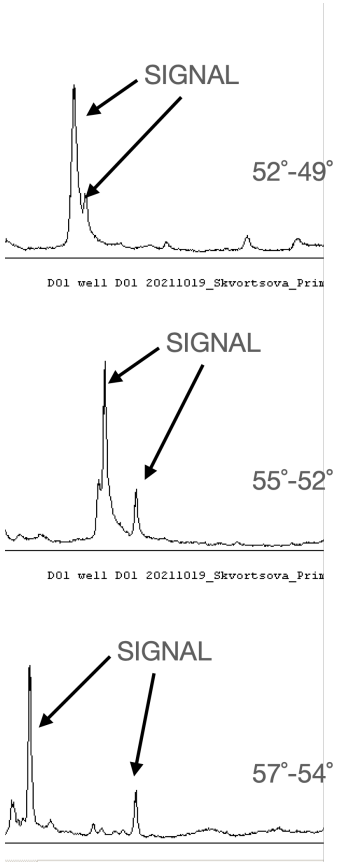


Figure 15: Optimisation of CTCE condition. The example shown is again a primer that resolves L. crispatus from L. helveticus. As can be seen, as the temperature change, the distance between the 2 signal peaks (one for L. helveticus and the second for L. crispatus) move further apart. If the distances between the 2 peaks is not sufficient, it becomes difficult to discriminate between the 2 types.

5.4. Data processing

5.4.1. Signal Processing

All data output from CTCE must be analysed. The signa must be interpreted to determine the relative abundance of each 2 groups present in the microbiome.

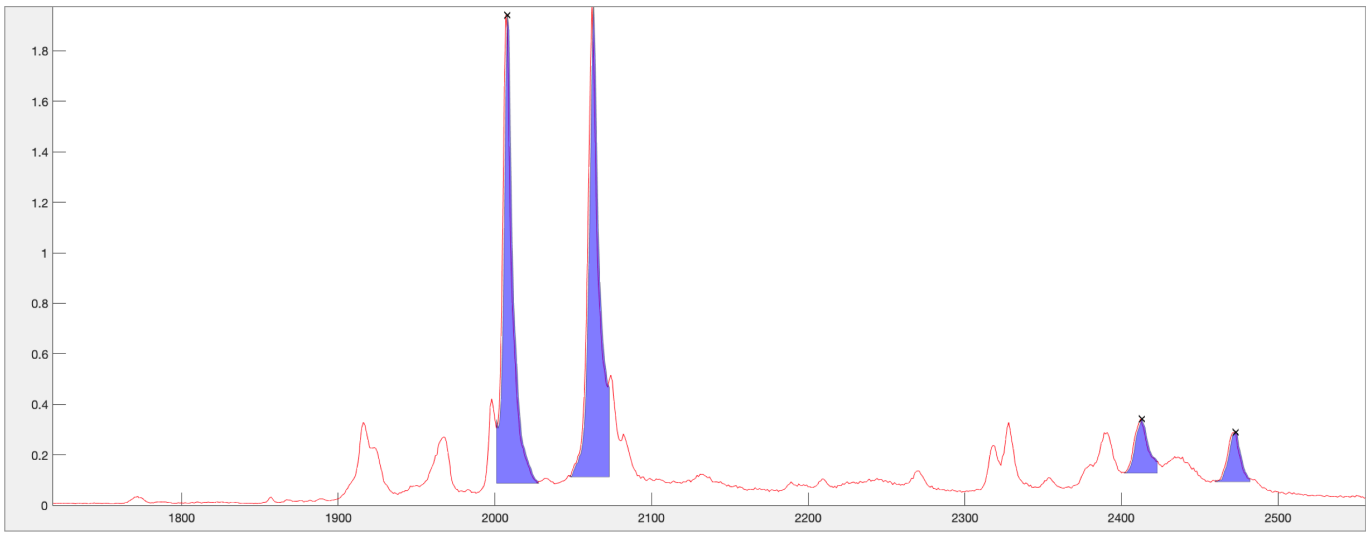


Figure 16: Signa processing interface. The system used is an adaptation of an open source software: LabelStudio. It enables the manual annotation of signals, as well as the implementation of Machine Learning algorithms. This combination means that, initially, results must be analysed by hand. However, once a sufficient number of signals have been annotated, the machine learning algorithm is capable of coring additional signals. This provides the means that although the analysis of the first few samples is very labour intensive, it quickly becomes automated as a sufficiently large training set becomes available.

5.4.2. Microbiome Profiling

The output from signal processing is, for each primer, the relative abundance between the 2 groups that are resolved. In order to compute the relative abundance of the different OTU, a complex system of equations must be resolved. This is achieved through a linear solver, integrated in REM's analytical pipeline. It makes used of theoretical PCR simulation, as well as observed data. The output is the relative abundance of each OTU in a given sample. Redundancy between primers, also means that in case a single primer fails, good results can be obtained.

6. Genotype Analysis

6.1. Sequencing

Samples required for performing whole genome sequencing are currently being collected on-site at OSR, anonymised, and are stored until they are ready for DNA extraction (i.e. once the recruitment phase is achieved). DNA extraction will take place according to the protocol required by sequencing provider. Quotation has been obtained from Novogene Cambridge lab, and we are currently seeking another quotation from a German provider (CeGat Laboratory). PCR quality control will be performed before sampled are shipped, and again upon receival by the sequencing facility. Samples will be sent over dry-ice for sequencing, where a sequencing depth of 30x will be targeted. QC and library preparation will follow the Illumina NovoSeq protocol (Paired End, 150bp). Raw sequencing reads will be retrieved via the provider's Cloud storage facility, where reads QC and variant calling is performed at CU.

6.2. Bio-informatic Analysis

Raw sequencing reads, in fastq format, will undergo necessary quality control using FastQC to control for:

- Basic Statistics
- Per base sequence quality
- Per sequence GC content
- Overrepresented sequences (usually adapters)
- Per base N content (N represents an unknown base in the read)

Following quality control, the reads will be trimmed (or filtered) accordingly using Trimmomatic, to remove low quality bases, reads below a certain length, and adapters. This step will be performed while keeping in mind the balance between obtaining high-quality reads and retaining enough coverage to perform a reliable downstream analysis (variant calling). The chosen Trimmomatic parameters will thus depends on the amount and the quality of the raw reads, as assessed by FastQC.

The processed reads will then be aligned against the latest human reference genome (GRch38, in fasta format) using the Burrow Wheeler Aligner (BWA) tool. First, the genome will be indexed with BWA, and the reads will be aligned with the "mem" algorithm (BWA-MEM), which is the recommended algorithm for short Illumina reads. The output of the alignment will be in SAM format, one file per sample (specifications available at the following address: https://samtools.github.io/hts-specs/SAMv1.pdf). As these files are text-based and usually large, they will be converted to binary BAM files, using samtools. The BAM format is standard, and most bioinformatics tools accept this format as input.

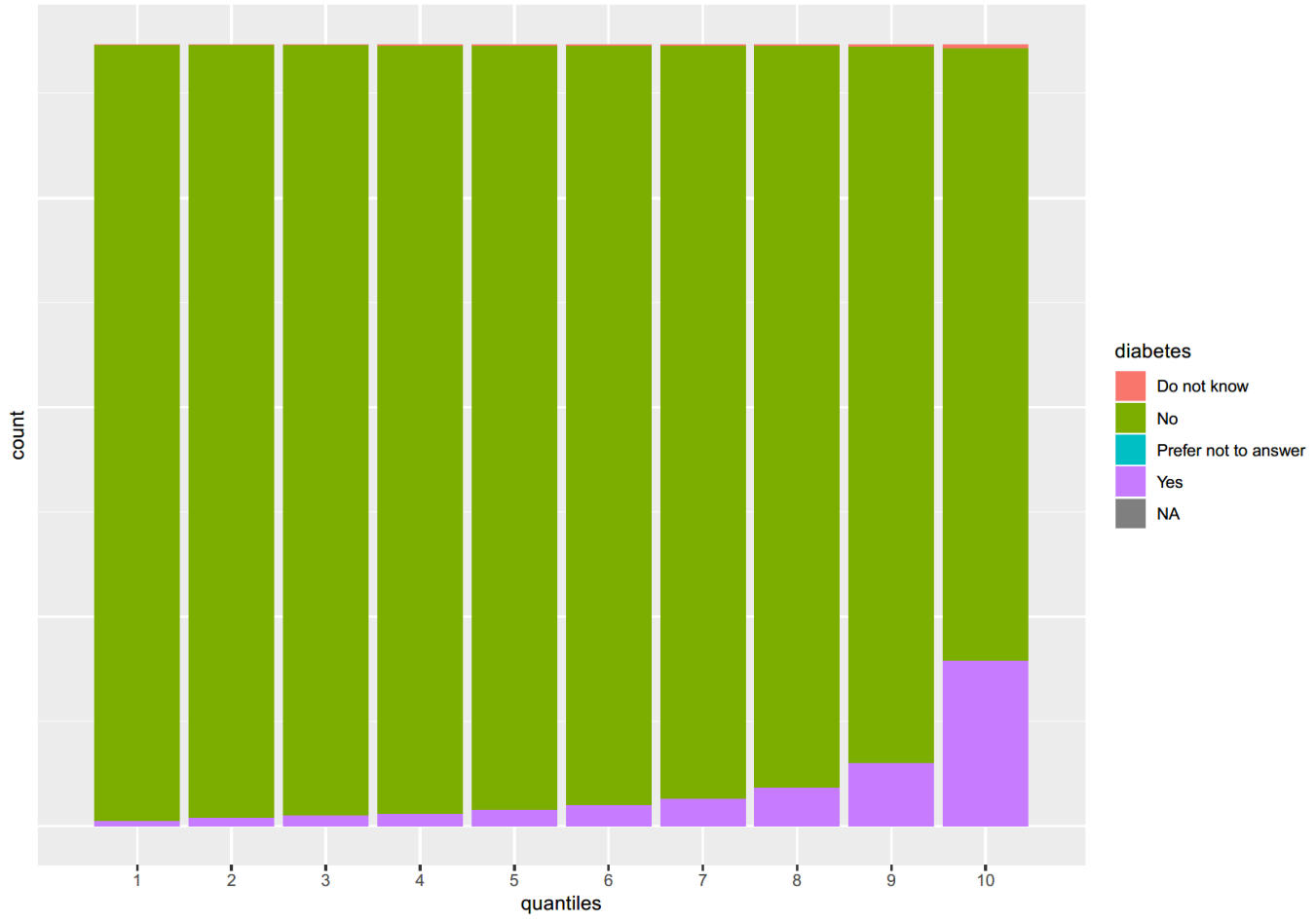
Variant calling will be performed with the GATK software suite. We will follow the best practices for "Germline short variant discovery", i.e. Single Nucleotide Polymorphisms (SNPs) and small insertions/deletions (InDels). In brief, the BAM files will be marked for duplicates to avoid having sequencing errors detected as SNPs, the base quality scores will be recalibrated and finally *HaplotypeCaller* will be run on the processed BAM files. This will result in a VCF file (specifications: https://samtools.github.io/hts-specs/VCFv4.3.pdf) containing the list of raw variants called by GATK. This VCF file will then be filtered to remove low-quality variants (for example, variants from regions with a low coverage, or supported by low-quality bases). Then the VCF file will be annotated to give information and context to each variant.

The genotype files, one per sample, can then be used as input to the PRS models previously developed at CU for Nutrishield (one for Body Mass Index, the other for Diabetes type 2). In short, the risk score is obtained by summing all the known disease-causing alleles carried by the individual (potentially weighted by their effect size) and dividing this number by the total number of non-missing variants detected in the individual. Plink will be used to calculate the PRS score, thus the VCF will first be converted to the plink format (.bed, .bim and .fam files) and the *–score* function from Plink will be used to calculate the PRS score, using the equation below:

$$PRS_j = \frac{\sum_i^N (ES_i * EA_{ij})}{P * S_j}$$

Equation 3: Plink formula to compute the Polygenic Risk Score for sample j. With N being the total number of variants, ES_i the effect size for SNP i, EA_{ij} the number of effect alleles observed in sample j, P the ploidy (here 2) and S_j the number of non-missing SNPs in sample j.

This score can then be compared to the scores obtained from the UK Biobank set, to estimate the relative risk this individual has of developing diabetes or a low/high BMI compared to the rest of the UK Biobank individuals (~300.000 individuals), which were split in 10 quantiles according to their genetic risk. Since, the UK Biobank also contain anthropometric data, we were able to confirm that individuals in a high PRS quantile have a higher chance of becoming obese or diabetic respective of the model chosen, as shown below by the proportion of individuals with diabetes (diagnosed by a doctor) in each PRS quantile.



It is important to keep in mind that the genetic risk is indicative, and only explain a part of the overall risk of developing obesity or diabetes, as the environment, microbiome, lifestyle, and other factors are also playing an important role in the pathogenesis of these complex diseases. In Nutrishield, the PRS score will serve to identify high-risk individuals in the trials.

Genotypes files (in VCF format) will also be intersected with annotated lists of variants prepared using the VarGen tool, developed by CU as part of Nutrishield. The VarGen package generated lists of annotated variants related to obesity, diabetes type 1 and type 2, which will help to identify the most deleterious variants and explore the reasons why a certain individual was placed in a certain risk quantile from the PRS analysis.

The PRS score, and list of annotated variants (which resulted in placing the individual in a given risk quantile), will be provided to the Nutrishield dashboard using a REST-API provided by CU, and the results will appear in a separate tab for each given patient registered on the platform, in tabular form. Since the

raw sequencing reads will be sent to CU in an anonymised form, linking the genotype to a given patient will be deciphered using the samples unique identifier on the dashboard side. This will allow clinicians and nutrition expect to gain more details about individual pathways, molecular functions and genes impacted by genetic disposition of the individual.