

Development of gut microbiota and metabolic changes in preterm babies

Thesis submitted by

Natalia Przysiężna

For the degree of Doctor of Philosophy of Imperial College

Supervisors:

Professor Neena Modi

Professor Elaine Holmes

Professor Julian Marchesi

Academic Neonatal Medicine

Department of Medicine

Faculty of Medicine

Imperial College London

2015

Statement of originality

I certify that the research conducted and described in this thesis as well as the thesis itself are the product of my own work. Any results of cooperation as well as quotations from the work of other people are fully acknowledged in accordance with the standard referencing practices of the discipline.

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Abstract

Gut microbiota are an important component of the human organism, contributing to metabolism and believed to play a role in the maintenance of health. Preterm babies are exposed to different early life interventions than full-term ones. They are exposed to prolonged courses of antibiotics and often fed parenterally, with delayed or limited enteral nutrition intake.

The aim of this research was to assess how early life interventions (type of parenteral nutrition, type of milk, mode of delivery) and other factors (gestational age at birth, postnatal age, sex of the baby) affect the development of gut microbiota and the consequent metabolic impact in preterm babies.

Two thousand stool and urine samples were collected from 160 babies. Gut microbiota were identified by 16S rRNA gene sequencing, while the activity of bacteria was assessed by metabolic profiling using ^1H NMR. A new DNA and faecal water extraction method was developed and validated.

Metabolic differences in stool and urine in relation to early life interventions were observed. Higher postnatal age was associated with vit. B₅ and vit. B₁, while lower was associated with increased bile acids and sugars. Increasing GA was correlated with higher concentration of GABA and glutamate. Breast-fed babies had higher concentration of lactose than formula-fed ones. There was no metabolic difference in relation to mode of delivery or sex of the baby. Microbial profiles were mostly affected by route of nutrition (oral – more *Staphylococcaceae*, intravenous – more *Enterobacteriaceae*), type of milk (breast milk – more *Enterobacteraceae*) and mode of delivery (Caesarean section – more *Staphylococcus* and *Clostridium*). The most abundant phylum was *Proteobacteria*. Average abundance of *Bifidobacteria* was 10%.

Those results contribute to current understanding of the establishment of gut microbiota in preterm babies. Future work should focus on repeating the analysis on a larger sample set from a trial designed for this purpose.

Acknowledgments

I would like to start by thanking my academic supervisors: Professor Neena Modi, Professor Elaine Holmes, and Professor Julian Marchesi for surviving with me those three years and guiding me into the right direction not only with my research but also career wise. I am also extremely grateful to Dr Sabita Uthaya who was a great teacher and I valued greatly the scientific discussions we had.

Many thanks to my co-workers from both CSM and Neonatal Medicine, especially to my two post docs Dr James Parkinson and Dr Matthew Hyde who were acting as my protectors on numerous occasions. I would also like to thank Dr Sundhiya Manila for making sure I understood statistics, Gonçalo Correia for teaching me how to create models in Matlab, Daniel Homola for helping with the microbiota vs. metabolites regressions, and last but not least to Abellona U who was a great lab partner and an amazing programmer. Big thanks to the people who were not only my colleagues but also dear friends and often had to listen to numerous rant marathons – Nick Andreas, Kayleigh Davis – this one is for you!

I am also grateful to the Westminster Medical School Research Trust for financing my bursary.

Finally, I would like to thank my family and friends for their support during those three years of my life. Especially to my mum, Leo Marques, Gosia Furmanik and Pawel Michałowski for always being there for me. Your emotional support and listening to my constant complains when no one else would make this whole PhD experience a much more enjoyable.

“A successful man is one who can lay a firm foundation with the bricks others have thrown at him.”

David Brinkley

Table of Contents

Chapter 1 – Background

1.1	Introduction	14
1.2	Gut microbiota.....	14
1.2.i	Importance of gut microbiota	15
1.2.i.a	Nutrition	15
1.2.i.b	Gut maturation	16
1.2.i.c	Immunological system development	16
1.2.i.d	Eubiosis / dysbiosis balance	17
1.2.i.e	Human gene expression alteration	17
1.2.ii	Development of gut microbiota.....	18
1.2.iii	Factors affecting gut colonisation	19
1.2.iii.a	Mode of delivery	23
1.2.iii.b	Type of diet	23
1.2.iii.c	Antibiotics.....	25
1.2.iii.d	Gestational age.....	25
1.2.iii.e	Other factors	26
1.2.iii.f	Conclusion on systematic review	26
1.2.iv	Association of the gut microbiota with functional clinical outcomes.....	26
1.2.iv.a	Sepsis.....	27
1.2.iv.b	Necrotizing enterocolitis	28
1.2.iv.c	Allergy, atopy and asthma	28
1.2.iv.d	Obesity.....	29
1.2.iv.e	Diabetes.....	30
1.2.iv.f	Neurodevelopmental disorders.....	30
1.2.iv.g	Conclusion from the systematic review	30
1.3	Preterm infants	31
1.3.i	Definition and epidemiology.....	31
1.3.ii	Types of prematurity	32
1.3.iii	Factors contributing to prematurity	33
1.3.iv	Characteristics	35
1.3.v	Mode of delivery	36
1.3.v.a	Vaginal	36
1.3.v.b	Caesarean section	37
1.3.vi	Type of diet.....	37
1.3.vi.a	Enteral nutrition	37
1.3.vi.b	Parenteral nutrition.....	39
1.3.vi.c	Total parenteral nutrition.....	43

1.4	Conclusions	44
Chapter 2 – Materials and Methods		
2.1	Study design	45
2.1.i	NEON study.....	45
2.1.ii	Sample collection	46
2.1.iii	Sample analysis strategy	46
2.2	Metabonomics	47
2.2.i	NMR spectroscopy	47
2.2.i.a	Theory.....	47
2.2.i.b	One-dimensional NMR spectroscopy	49
2.2.i.c	Two-dimensional NMR spectroscopy	50
2.2.ii	Sample preparation for NMR analysis	50
2.2.ii.a	Urine.....	51
2.2.ii.b	Stool.....	51
2.2.iii	Pre-processing of ¹ H-NMR spectra	52
2.2.iv	Multivariate statistical analysis	52
2.2.iv.a	Unsupervised multivariate statistical analysis	52
2.2.iv.b	Supervised multivariate statistical analysis.....	53
2.2.v	Metabolite identification	54
2.3	Gut microbiota profiling.....	55
2.3.i	Theory	55
2.3.i.a	What is 16S rRNA	55
2.3.i.b	Comparison of the analytical techniques	56
2.3.ii	DNA extraction from stool samples.....	56
2.3.iii	Quantitative Polymerase Chain Reaction (qPCR) analysis	57
2.3.iv	Pre-processing of the sequenced data.....	59
2.3.v	Analysing the microbial profiles.....	59
2.4	Regression of metabolites against gut microbiota.....	61
Chapter 3 - Method development		
3.1	Introduction	62
3.2	Faecal water extraction.....	62
3.2.i	Solvents	62
3.2.ii	Stool sample mass	67
3.2.iii	Type of homogenisation.....	69
3.2.iv	Solvent evaporation	72
3.2.v	Filtration	77
3.2.vi	Preparation of lipid layer for lipidomics.....	79
3.2.vii	Final protocol	80

3.3	DNA extraction from stool pellet.....	80
3.3.i	Untreated stool vs. faecal pellet.....	82
3.3.ii	Type of homogenisation.....	83
3.3.iii	Solvents.....	85
3.3.iv	Presence of sodium azide.....	87
3.3.v	Presence of the lipid layer	89
3.3.vi	DNA yield variation in preterm, full-term and adult stool samples	90
3.3.vii	Concentration of DNA.....	91
3.3.viii	Purity improvement.....	92
3.3.ix	Final improvements to QIAamp DNA Stool Mini Kit.....	95
3.3.x	Final DNA extraction protocol.....	96

Chapter 4 – The urinary metabonome of the preterm baby

4.1	Introduction	97
4.2	Experimental design	99
4.2.i	Urine sample preparation.....	99
4.2.ii	Nuclear Magnetic Resonance spectra acquisition	99
4.2.iii	Sample pre-processing.....	99
4.3	Results.....	101
4.3.i	Metabolic profiles associated with randomised PN feeding group	101
4.3.i.a	<i>Intralipid vs. SMOF lipid</i>	101
4.3.i.b	<i>Recommended daily intake (RDI) vs. increasing concentration of amino acid</i>	102
4.3.ii	Metabolic profiles associated with gestational age (GA)	102
4.3.iii	Metabolic profiles associated with postnatal age (PNA).....	105
4.3.iv	Metabolic profiles associated with postmenstrual age (PMA)	109
4.3.v	Metabolic profiles associated with sex	111
4.3.vi	Metabolic profiles associated with mode of delivery.....	111
4.3.vii	Metabolic profiles associated with diet	113
4.3.vii.a	<i>Parenteral nutrition vs. enteral nutrition</i>	114
4.3.vii.b	<i>Human milk vs. formula milk</i>	117
4.4	Discussion	120
4.5	Conclusion.....	126

Chapter 5 – The stool metabonome of the preterm baby

5.1	Introduction	127
5.2	Experimental design	128
5.2.i	Stool sample preparation.....	128
5.2.ii	Nuclear Magnetic Resonance spectra acquisition	128
5.2.iii	Sample pre-processing.....	128
5.3	Results.....	130

5.3.i	Metabolic profiles associated with randomised PN feeding group	130
5.3.i.a	<i>Intralipid vs. SMOF lipid</i>	130
5.3.i.b	<i>Recommended daily intake vs. increasing concentration of amino acid</i>	131
5.3.ii	Metabolic profiles associated with gestational age (GA)	132
5.3.iii	Metabolic profiles associated with postnatal age (PNA).....	135
5.3.iv	Metabolic profiles associated with sex	138
5.3.v	Metabolic profiles associated with mode of delivery.....	139
5.3.vi	Metabolic profiles associated with diet	140
5.3.vi.a	<i>Parenteral nutrition vs. enteral nutrition</i>	141
5.3.vi.b	<i>Human milk vs. formula milk.....</i>	144
5.4	Discussion	147
5.5	Conclusion	152
Chapter 6 – The development of gut microbiota in the preterm baby		
6.1	Introduction	153
6.2	Experimental design	154
6.2.i	DNA extraction.....	154
6.2.ii	Sequencing.....	155
6.2.iii	Sample pre-processing.....	155
6.2.iv	Data analysis.....	156
6.2.iv.a	<i>Analysed factors</i>	156
6.2.iv.b	<i>Reporting gut microbiota</i>	157
6.3	Results.....	159
6.3.i	Randomised groups.....	160
6.3.ii	Gut microbiota associated with gestational age (GA).....	161
6.3.iii	Gut microbiota associated with postnatal age (PNA)	161
6.3.iv	Gut microbiota associated with sex	163
6.3.v	Gut microbiota associated with mode of delivery	163
6.3.vi	Gut microbiota associated with occurrence of labour	164
6.3.vii	Gut microbiota associated with diet	164
6.3.vii.a	<i>Parenteral nutrition vs. enteral nutrition</i>	164
6.3.vii.b	<i>Breast milk vs. other milk.....</i>	166
6.3.viii	Regression of the stool metabolites against gut microbiota.....	167
6.4	Discussion	167
6.5	Conclusion	172
Chapter 7 – General discussion		173
References		180
Appendix		196

Table of Figures:

Figure 1: Flowchart presenting subgroups of premature birth.....	32
Figure 2: Potential sites of bacterial infections in the uterus.....	34
Figure 3: Potential pathways from chorodecidual bacterial colonisation to occurrence of premature birth	35
Figure 4: Types and subtypes of possible enteral nutrition.....	38
Figure 5: Study flowchart representing the four randomised groups.....	46
Figure 6: Flowchart presenting the sample analysis strategy.....	47
Figure 7: Principles of NMR. The segment A shows how the nuclei align themselves when the external magnetic field is applied. The segment B shows how the radio frequency 'pushes' the magnetic field of the nuclei.....	49
Figure 8: The result of the STORM analysis of the unknown peak of the chemical shift of 2.04ppm.....	54
Figure 9: The simplified structure of the prokaryotic (70S) ribosome	55
Figure 10: The simplified structure of the 16S rRNA gene with conserved regions dark shaded, and the variable regions light shaded.....	56
Figure 11: Flowchart presenting the statistical analysis options. The approach used for our analysis is presented in the thick-outlined boxes.....	60
Figure 12: Flowchart presenting the statistical approach when working with comparing > 2 groups	60
Figure 13: Comparison of CPMG one-dimensional ¹ H NMR spectra of faecal water extracted in D2O or PBS.....	64
Figure 14: Comparison of standard one-dimensional ¹ H NMR spectra of faecal water extracted in D2O, PBS or phosphate buffer.....	65
Figure 15: Comparison of standard one-dimensional ¹ H NMR spectra of faecal water extracted in D2O and H2O. Spectra are scaled to TSP.....	65
Figure 16: Comparison of standard one-dimensional ¹ H NMR spectra of faecal water extracted by two different modes of bead beating	67
Figure 17: Shape of stool pellet after extracting faecal water from 500mg of faeces	67
Figure 18: Comparison of CPMG one-dimensional ¹ H NMR spectra of faecal water extracted by bead beating and vortexing	70
Figure 19: Comparison of standard one-dimensional ¹ H NMR spectra of faecal water extracted by two different modes of bead beating (presented in duplicates).	71
Figure 20: Comparison of standard one-dimensional ¹ H NMR spectra of faecal water extracted by two different bead beaters (Precellys 24 and Mini beadbeater).	72
Figure 21: Comparison of standard one-dimensional ¹ H NMR spectra of faecal water extracted before and after speed vacuuming.....	73
Figure 22: Comparison of standard one-dimensional ¹ H NMR spectra of faecal water extracted before and after speed vacuuming.....	74
Figure 23: Comparison of standard one-dimensional ¹ H NMR spectra presenting the influence of speed vacuuming of faecal water on the metabolic profile.....	75
Figure 24: Comparison of standard one-dimensional ¹ H NMR spectra presenting the influence of freeze-drying of faecal water on the metabolic profile.....	76
Figure 25: Comparison of standard one-dimensional ¹ H NMR spectra presenting the influence of drying samples with Nitrogen flow on the metabolic profile	77
Figure 26: Standard one-dimensional ¹ H NMR spectra presenting the influence of filtration of solvent through Spin-x centrifuge tube filters.....	78
Figure 27: Standard one-dimensional ¹ H NMR spectra presenting the influence of filtration of 25% acetonitrile solution through Spin-x centrifuge tube filter	79
Figure 28: Flowchart presenting the final protocol for faecal water extraction.....	80
Figure 29: Gel electrophoresis of PCR products of 16S rRNA gene extracted from stool samples.	87
Figure 30: Picture of faecal water presenting the lipid layer floating on top of faecal water.....	89
Figure 31: Gel electrophoresis of PCR products of 16S rRNA gene extracted from stool samples.	91
Figure 32: Graph showing correlation of 260/280nm ratio and DNA concentration	94
Figure 33: Graph showing correlation of 260/230nm ratio and DNA concentration (sample dilution).	95
Figure 34: Flowchart presenting the analysis pipeline for urine samples obtained from premature babies characterising changes in the urinary metabonome associated with postnatal age (PNA).	98
Figure 35: Scores scatter plot of the principal component analysis (PCA) model of urine NMR spectral data generated from samples collected at multiple time points (whole data set).	100
Figure 36: The O-PLS models for urine samples comparing the types of lipids	101
Figure 37: The O-PLS models for urine samples comparing the concentrations of amino acids.	102

Figure 38: PCA Scores Scatter Plot of whole data set of NMR urine spectra. Samples coloured according to gestational age (in weeks).....	103
Figure 39: Orthogonal Partial Least Squares (O-PLS) loadings plot of urine metabolic profiles coming from samples collected at multiple time points and created in relation to the gestational age.....	104
Figure 40: PCA Scores Scatter Plot of whole data set of NMR urine spectra. Samples coloured according to postnatal age (in weeks).....	106
Figure 41: Orthogonal Partial Least Squares (O-PLS) loadings plot of urine metabolic profiles coming from samples collected at multiple time points and created in accordance to the postnatal age.....	108
Figure 42: PCA Scores Scatter Plot of whole data set of NMR urine spectra. Samples coloured according to postmenstrual age (in weeks).	109
Figure 43: Orthogonal Partial Least Squares (O-PLS) loadings plot of urine metabolic profiles coming from samples collected at multiple time points and created in accordance to the postmenstrual age.....	110
Figure 44: The O-PLS models for urine samples comparing sex of baby.	111
Figure 45: The O-PLS models for urine samples comparing mode of delivery.....	112
Figure 46: The O-PLS models of urine samples created for presence of labour.....	113
Figure 47: PCA Scores Scatter Plot of whole sample set of NMR urine spectra. Samples are coloured according to feeding route).	115
Figure 48: Orthogonal Partial Least Squares (O-PLS) loadings plot of urine metabolic profiles coming from samples collected at multiple time points and created in accordance to the route of nutrition.....	116
Figure 49: PCA Scores Scatter Plots of whole sample set of NMR urine spectra (principal component 3 vs. principal component 1) coloured according to whether human milk was given or formula milk was given.....	117
Figure 50: Orthogonal Partial Least Squares (O-PLS) loadings plot of urine metabolic profiles coming from samples collected at multiple time points and created in accordance to the type of milk.....	118
Figure 51: Scores scatter plot of the principal component analysis (PCA) model of faecal water NMR spectral data generated from samples collected at multiple time points (whole data set).....	129
Figure 52: The O-PLS models for stool samples comparing the types of lipids.....	130
Figure 53: The O-PLS models for stool samples concentration of amino acids.....	131
Figure 54: PCA Scores Scatter Plot of whole data set of NMR stool spectra. Samples coloured according to gestation age (in weeks).....	132
Figure 55: Orthogonal Partial Least Squares (O-PLS) loadings plot of stool metabolic profiles coming from samples collected at multiple time points and created in accordance to the gestational age.....	134
Figure 56: PCA Scores Scatter Plot of whole data set of NMR stool spectra. Samples coloured according to postnatal age (in weeks).....	135
Figure 57: Orthogonal Partial Least Squares (O-PLS) loadings plot of stool metabolic profiles coming from samples collected at multiple time points and created in relation to postnatal age.....	137
Figure 58: O-PLS models for stool samples comparing sex of baby.....	138
Figure 59: The O-PLS models for stool samples comparing mode of delivery.....	139
Figure 60: The O-PLS models for stool samples comparing presence of labour.....	140
Figure 61: PCA Scores Scatter Plot of whole sample set of NMR stool spectra. Samples are coloured according to feeding route.....	141
Figure 62: PCA Scores Scatter Plot of whole sample set of NMR stool spectra. Samples are coloured according to feeding route.....	142
Figure 63: Orthogonal Partial Least Squares (O-PLS) loadings plot of stool metabolic profiles coming from samples collected at multiple time points and created in accordance to the route of feeding.....	143
Figure 64: PCA Scores Scatter Plot of whole sample set of NMR stool spectra. Samples are coloured according to type of enteral nutrition.....	144
Figure 65: Orthogonal Partial Least Squares (O-PLS) loadings plot of stool metabolic profiles coming from samples collected at multiple time points and created in accordance t type of milk.....	145
Figure 66: The most significant for this study members of the Actinobacteria phylum, organized according to their taxonomic level.	157
Figure 67: The most significant for this study members of the Firmicutes phylum, organized according to their taxonomic level.....	158
Figure 68: The most significant for this study members of the Proteobacteria phylum, organized according to their taxonomic level.	158
Figure 69: The Principal Coordinates Analysis plots presenting the separation based on the relative abundance of (A) phyla and (B) genera.	159
Figure 70: The relative abundance of the Proteobacteria (blue), Actinobacteria (red), and Firmicutes (green) phyla among all the samples collected from premature babies.....	160

Figure 71: The relative abundance of the (A) <i>Bifidobacteria</i> genus and (B) <i>Enterobacteriaceae</i> family among samples collected from babies born at different gestational age	161
Figure 72: The relative abundance of Firmicutes	162
Figure 73: The relative abundance of the (A) <i>Staphylococcus</i> , and (B) <i>Clostridia</i> grouped according to the postnatal age of the baby at the time of sample collection.	163
Figure 74: The (A) phyla, (B) classes, and (C) genera found to be significantly associated with the mode of delivery of the baby..	164
Figure 75: The relative abundance of (A) <i>Proteobacteria</i> phylum, and (B) <i>Firmicutes</i> phylum grouped according to the route of nutrition that the baby was on during the sample collection.	165
Figure 76: The (A) phyla and (B) families found to be significantly associated with the route of nutrition.....	166
Figure 77: The (A) phylum and (B) family found to be significantly associated with the breast milk	166
Figure 78: PCA Scores Scatter Plot of NMR spectra of urine samples collected in the 1 st week of life. Samples are coloured according to type of intravenous lipid solution.....	196
Figure 79: PCA Scores Scatter Plot of NMR spectra of urine samples collected in the 3 rd week of life. Samples are coloured according to type of intravenous lipid solution.....	197
Figure 80: PCA Scores Scatter Plot of NMR spectra of the last urine samples collected. Samples are coloured according to type of intravenous lipid solution.....	197
Figure 81: PCA Scores Scatter Plot of NMR spectra of urine samples collected in the 1 st week of life. Samples are coloured according to concentration of amino acids in parenteral nutrition.	198
Figure 82: PCA Scores Scatter Plot of NMR spectra of urine samples collected in the 3 rd week of life. Samples are coloured according to concentration of amino acids in parenteral nutrition.	198
Figure 83: PCA Scores Scatter Plot of NMR spectra of the last urine samples collected. Samples are coloured according to concentration of amino acids in parenteral nutrition.	199
Figure 84: PCA Scores Scatter Plot of NMR spectra of urine samples collected in the 1 st week of life. Samples are coloured according to the gestational age at birth (in weeks).	199
Figure 85: PCA Scores Scatter Plot of NMR spectra of urine samples collected in the 3 rd week of life. Samples are coloured according to the gestational age at birth (in weeks).	200
Figure 86: PCA Scores Scatter Plot of NMR spectra of urine samples collected from babies born in 24 th week of gestation. Samples are coloured according to the postnatal age given in weeks.	200
Figure 87: PCA Scores Scatter Plot of NMR spectra of urine samples collected from babies born in 28 th week of gestation. Samples are coloured according to the postnatal age given in weeks.	201
Figure 88: PCA Scores Scatter Plot of whole data set of NMR urine spectra. Samples are coloured according to the sex of the baby	201
Figure 89: PCA Scores Scatter Plot of whole data set of NMR urine spectra. Samples are coloured according to the mode of delivery	202
Figure 90: PCA Scores Scatter Plot of whole data set of NMR urine spectra. Samples are coloured according to the presence of labour during delivery	202
Figure 91: PCA Scores Scatter Plot of NMR spectra of stool samples collected in the 1 st week of life. Samples are coloured according to type of intravenous lipid solution.....	203
Figure 92: PCA Scores Scatter Plot of NMR spectra of stool samples collected in the 3 rd week of life. Samples are coloured according to type of intravenous lipid solution.....	203
Figure 93: PCA Scores Scatter Plot of NMR spectra of the last stool samples collected. Samples are coloured according to type of intravenous lipid solution.....	204
Figure 94: PCA Scores Scatter Plot of NMR spectra of stool samples collected in the 1 st week of life. Samples are coloured according to concentration of amino acids in parenteral nutrition.	204
Figure 95: PCA Scores Scatter Plot of NMR spectra of stool samples collected in the 3 rd week of life. Samples are coloured according to concentration of amino acids in parenteral nutrition..	205
Figure 96: PCA Scores Scatter Plot of NMR spectra of the last stool samples collected. Samples are coloured according to concentration of amino acids in parenteral nutrition.	205
Figure 97: PCA Scores Scatter Plot of NMR spectra of stool samples collected in the 1 st week of life. Samples are coloured according to the gestational age at birth (in weeks).	206
Figure 98: PCA Scores Scatter Plot of NMR spectra of stool samples collected in the 3 rd week of life. Samples are coloured according to the gestational age at birth (in weeks).	206
Figure 99: PCA Scores Scatter Plot of NMR spectra of stool samples collected from babies born in 24 th week of gestation. Samples are coloured according to the postnatal age given in weeks.	207
Figure 100: PCA Scores Scatter Plot of NMR spectra of stool samples collected from babies born in 28 th week of gestation. Samples are coloured according to the postnatal age given in weeks.	207

Figure 101: PCA Scores Scatter Plot of whole data set of NMR stool spectra. Samples are coloured according to the sex of the baby.	208
Figure 102: PCA Scores Scatter Plot of whole data set of NMR stool spectra. Samples are coloured according to the mode of delivery.	208
Figure 103: PCA Scores Scatter Plot of whole data set of NMR stool spectra. Samples are coloured according to the presence of labour during delivery.	209

Table of Equations:

Equation 1: DNA concentration recalculation.	82
--	----

Table of Tables:

Table 1: Studies included in the systematic review.	20
Table 2: Comparison of the analytical tool for the bacterial identification.	58
Table 3: The complete reagent recipe (master mix) for 1X PCR reaction.	57
Table 4: Results showing the influence of mass of stool sample and the shape of pellet and volume of retrieved faecal water when centrifuged at $< 16,000 \times g$	68
Table 5: Results showing the correlation of volume of retrieved faecal water and the time of centrifuging at the speed $< 16,000 \times g$	68
Table 6: Comparison of DNA yields after extraction from untreated stool and stool pellet obtained from faecal water extraction.	83
Table 7: Comparison of DNA yield from pellets that were bead beaten and those that were vortexed.	84
Table 8: Comparison of DNA yield from pellets that were bead beaten for different time.	84
Table 9: Comparison of DNA yield from pellets that were bead beaten with different programs.	85
Table 10: Comparison of DNA yield from pellets extracted either in D2O, H2O, PBS or phosphate buffer.	86
Table 11: Comparison of DNA yield from pellets extracted either in 25% acetonitrile or H2O.	87
Table 12: Comparison of DNA yield from pellets extracted either in phosphate buffer or D2O.	88
Table 13: Comparison of DNA yields after extraction from stool pellet with and without lipid layer.	89
Table 14: Comparison of DNA yield from different types of stool samples including: preterm, full-term, 1-year-old and adult.	90
Table 15: The association of DNA yield with the elution buffer volume and number of elutions.	91
Table 16: The association of DNA purity with the improvements of DNA extraction protocol.	93
Table 17: The association of DNA concentration with the DNA purity.	94
Table 18: The summary of the models created for the analysis of the changes between the urinary metabolic profiles.	119
Table 19: The list of discriminant urinary metabolites identified to be associated with each analysed factor.	120
Table 20: The summary of produced models testing the differences in metabolic profile of faecal water.	146
Table 21: The list of discriminant faecal metabolites identified to be associated with each analysed factor.	147
Table 22: The comparison of the characteristics of samples that were included in the modeling and those that were excluded from the modeling.	155

Abbreviations:

1D – one dimensional	NMR – nuclear magnetic resonance
¹ H – proton	O-PLS – orthogonal partial least squares analysis
2D – two dimensional	OTU – operational taxonomic unit
BM – breast milk (mother's milk)	PBS – phosphate buffered saline
BV – bacterial vaginosis	PC – principal component
COSY – correlation spectroscopy	PC1 – principal component 1
CS – Caesarean section	PCA – principal component analysis
D ₂ O – heavy water	PCR – polymerase chain reaction
DGGE – degradation gradient gel electrophoresis	PLS – partial least squares analysis
dH ₂ O – deionised water	PMA – postmenstrual age
DM – donor's milk	PN – parenteral nutrition
DNA – deoxyribonucleic acid	PNA – postnatal age
DXP – deoxyxylulose 5-phosphate	PNALD - parental nutrition-associated liver disease
EN – enteral nutrition	PPROM – preterm premature rupture of membrane
FID – free induction decay	QC – quality control
FM – formula milk	qPCR – quantitative polymerase chain reaction
GA – gestational age	RDI – recommended daily intake
GABA – γ-aminobutyric acid	RNA – ribonucleic acid
HM – human milk	SCFA – short-chain fatty acid
HMBC – heteronuclear multiple-bond correlation	SD – standard deviation
HSQC – heteronuclear single-quantum correlation	SMOF – Soya oil, medium-chained triglycerides, olive oil, fish oil
IPP – isopenteryl pyrophosphate	TGGE – temporal temperature gradient gel electrophoresis
JRES – J-resolved	TOCSY – total correlation spectroscopy
LCT – long chain triglycerides	TPN – total parenteral nutrition
MCT – medium chain triglycerides	T-RFLP – terminal restriction enzyme fragment polymorphism
MS – mass spectrometry	TSP – trimethyl silyl propionic acid
NEC – necrotizing enterocolitis	VD – vaginal delivery
NEON – Nutritional Evaluation and Optimisation in Neonates	

Chapter 1 – Background

1.1 Introduction

The gut microbiota is believed to have an important function for human health. There are many reports suggesting an association between changes in microbial profiles and functional clinical outcomes such as obesity, diabetes, allergy and asthma. What is more, it is believed that the pattern of colonisation of the gut immediately after birth may affect the long-term health. There have been a number of studies analysing the development of gut microbiota in full term neonates, however the detailed knowledge regarding the colonisation of preterm babies is lacking. Preterm babies undergo different early-life interventions than infants born at full term. This thesis focuses on understanding the intestinal colonisation in preterm babies and assessing how early-life interventions, such as mode of delivery route of nutrition and type of diet, affect this process.

1.2 Gut microbiota

The gastro-intestinal microbiota is a substantial component of the human organism. In the adult human gut there are around 10^{14} bacteria composed of at least 500 species, of which only 30-40 are of high abundance and represent 99% of the total gut microbial population [1-4]. The composition of microbiota in different parts of the gastro-intestinal tract differs. The concentration of bacteria increases from stomach to the colon. In the human gut there are 10^3 colony forming units (live bacterial cells) per ml of sample in the stomach and duodenum, 10^4 - 10^8 in the jejunum and ileum and 10^9 - 10^{12} in the colon [5]. In fact, one-third of dry stool weight is believed to be bacteria, although the majority of them are dead. There are approximately 10^{10} live bacteria per gram of faeces which means that in the body there are 10x more bacterial cells than human ones [3, 6, 7].

1.2.i Importance of gut microbiota

The gut microbiota has a number of important functions in the human body. It supports nutrition, assists with the gut maturation and immunological system development, inhibits the growth of pathogenic bacteria and when off balance can contribute to development of diseases [4].

The method most utilised to-date to assess the influence of the gut microbiota on the host is experimentation involving germ-free animals [3, 8]. There are two main approaches. The first analyses the importance of commensal bacteria by comparing the development and metabolism of conventional animal with germ-free ones. The second approach is to introduce gut microbiota from conventional to germ-free animals and observe what changes the animal undergoes.

1.2.i.a Nutrition

The main input of microbiota into nutrition is the enhancement of nutritional value extraction from food. Germ-free animals require greater calorie density food to maintain body weight than conventional animals [3]. Additionally, humans are not able to digest all the polysaccharides that come from the consumption of plants [4]. After screening the genome of gut microbiota it was found that at least 81 families of the glycoside hydrolase enzymes are coded in the microbial genome. The gut bacteria create a trophic chain which enables the full digestion and energy retrieval from the plant polysaccharides and supporting the human organism in the metabolism of arabinose, fructose, galactose, glucose, mannose, starch, sucrose and xylose [4]. The gut microbiota play an important role in nutrition not only by digestion of sugars but also by producing short chain fatty acids, GABA, essential amino acids and vitamins [3, 4, 9]. The analysis of gut microbiome identified a number of genes coding the enzymes used for the synthesis of DXP (deoxyxylulose 5-phosphate) and IPP (isopenteryl pyrophosphate). DXP is used for the synthesis of thiamine (vitamin B₁) and pyridoxal phosphate (vitamin B₆) which explains why germ-free animals require certain vitamins (like vitamin K and some B vitamins) in their diet in contrast to conventional animals [3]. At the same time, IPP has over 25000 derivatives and is found in every known type of cell [4].

1.2.i.b Gut maturation

The gut is a large organ with a maximal surface of approximately 8000m² which makes it the largest surface exposed to exterior world [10]. There are many reasons why an infant's gut changes over the time. Firstly, at some point it must adapt itself to a potentially omnivore diet. In infancy the main food source is human milk and at this point the gut must be programmed to harvest energy from this lipid-rich source. During weaning infants receive a more diverse diet that is shifted towards a high concentration of carbohydrates [11]. One study on mice found that the structure of intestinal epithelial glycans change over time. At the stage of breastfeeding mice have intestinal epithelial glycans terminated mainly with sialic acid [12]. Once mice are weaned the expression of intestinal epithelial glycans shifts towards fucose terminated [12-14]. This change is triggered by modification of expression of glycosyltransferases [13, 14]. What is important is that such change of intestinal epithelial glycans is not observed in germ-free mice. However, when adult germ-free mice were colonised by microorganisms the intestinal epithelial glycans terminated with sialic acid changed to terminated with fucose [13, 14]. This suggests that presence of bacteria is involved in this aspect of gut development. It is suggested that perhaps bacteria use fucose as an energy source while colonising intestines during weaning [11].

The studies on germ-free animals have also shown a decreased intestinal weight and surface and thinner villi in those animals when compared to conventional ones. Additionally, the caecum has been shown to be up to eight times larger with higher pancreatic protein content in germ-free rodents [8, 15, 16]. Another important aspect of gut maturation that is related to gut microbiota is the development of the villus capillary network. Germ-free mice have poorly developed villus capillaries; however colonisation by bacteria results in a rapid angiogenesis of the gut [8, 17].

1.2.i.c Immunological system development

Another important change in early life development is the development of the immunological system. Soon after birth infants have rather well developed innate immunity while the adaptive immunity is facilitated by antibodies received with mother's milk. In infants that are not breast-fed the responsibility for fighting with antigens lies on macrophages, neutrophils and natural killer cells [18].

During weaning the infant must be able to produce their own antibodies in order to fight potential diseases [11].

One study on mice analysed the influence of gut microbiota on the expression of Angiogenin-4, a Paneth cell protein that is believed to participate in epithelial host defence. The expression of this protein is developmentally regulated postpartum and increases to the adult level during weaning. In germ-free mice levels of Angiogenin-4 were low and never reached levels that are found in conventional adults. However, after colonising the germ-free adult mice with gut bacteria the expression of Angiogenin-4 increased suggesting that presence of gut microbiota is crucial for the expression of this protein [19]. Additionally, other components of the immune system (mucosal immunoglobulin A and intraepithelial lymphocytes) are reduced in germ-free animals in comparison to conventional ones [20-23]. It is yet unknown if the host defense reacts to the increasing number of bacteria in the gut or is it triggered by 'beneficial' bacteria [11].

1.2.i.d Eubiosis / dysbiosis balance

Although the gut bacteria normally habitate the lumen of the gut and should not be able to penetrate the deeper layers, there are few pathways for that the microorganisms may use for such transfer. Firstly, tight junctions (the intercellular connection that holds cells attached one to another) may be used as a gate for pathogens. Secondly, some bacteria can use the trans-cellular transfer in order to spread from the lumen of the gut [10]. Such a situation is sometimes referred to as 'leaky gut'.

Since preterm infants have immature gut and immune system, the exposure to pathogenic bacteria in the gut may lead to a pro-inflammatory and tissue-damaging outcomes [24]. The incidence of necrotizing enterocolitis (NEC) and sepsis are not uncommon in the premature neonates thus the understanding of the gut colonisation may be crucial in the fight with those clinical outcomes [25-33].

1.2.i.e Human gene expression alteration

Due to the fact that there are more bacterial cells in human organism than the host's ones it is not surprising that the number of bacterial genes is also higher than the human genes. It is estimated that there are 100-350 times more bacterial genes than human ones and on top of that, there are studies

showing that bacterial metabolites can lead to alterations in human gene expression [3, 34]. One of the examples of human gene expression alteration is related to improvement of extraction of nutritious value from food. A series of studies analysing the influence of microbiota on mRNA profiles found that there was an increased expression of genes coding proteins involved in metabolism of either glucose or fatty acids [3, 34].

Bacterial gene expression seems to remain stable between humans, which suggests that there may be a core set of functions/metabolic pathways that is performed by the microbial community regardless of the slight alterations in the bacterial profile [35]. This is apparent especially during pregnancy as the microbial shift from first trimester to third trimester is enormous while the genetic profile of those bacteria remains stable and regardless of the dysbiosis of the gut microbiota of pregnant women do not develop any serious diseases [35]. Such microbial change may be caused by altered immune system during the pregnancy (immune system is less effective in order to minimise the risk of attacking the fetus).

1.2.ii Development of gut microbiota

There are a number of studies analysing the sequence of bacteria colonising the gut of newborn full-term infants. One of the studies focused on analysing the development of the gut microbiota with correlation to microbial changes in maternal milk. The first bacteria to colonise infant's gut were members of *Enterococcus* and *Streptococcus* with *E. faecalis* and *S. salivarius* being the most abundant species [36]. This occurred within the first day of life. On 10th, 30th and 90th day postpartum the most abundant genera were bifidobacteria (*B. brave*, *B. bifidum* and *B. pseudocatenulatum*), streptococci (*S. salivarius*, *S. vestibularis*), lactobacilli (*L. gasseri*) and enterococci (*E. faecalis*) respectively. Members of *Staphylococcus* genus were found only on 1st and 10th day postpartum [36]. What is important is that vertical transfer of bifidobacterial species was found between mother and infant [36]. A Dutch cohort of approximately 1000 infants found that between 3rd and 6th week postpartum the most abundant microbial genus was *Bifidobacterium*. Bacteria belonging to this group were found in almost 99% of infants [37]. *E. coli* and *B. fragilis* were found in above 80% of infants. *Clostridium difficile* and bacteria belonging to lactobacilli were less commonly found [37].

1.2.iii Factors affecting gut colonisation

There is a number of endogenic and exogenic factors that can affect the microbiota composition of the gut. Endogenic factors include genetics, sex, gestational age, postnatal age, maturity of the gut and immunity [29, 38]. Exogenic factors also called interventions that can affect gut microbiota are mode of delivery, route of feeding and type of diet, exposure to antibiotics and environment [5, 26, 29, 38-44]. An interesting finding is that identical twins who share the same habitat and diet can still have different gut bacteria [8].

There are two types of mechanisms that can control the microbiota composition. The first one is the host-bacterial interaction which involves immunological reactions. The other type of control mechanism is a bacterial-bacteria interaction. This type of interaction is based on ecological changes caused by production of metabolites (short-chained fatty acids, bacteriocins), consumption of oxygen, competition over food or epithelial receptors [45]. A change in gut environment can lead to shift of the bacterial population.

In order to get an unbiased insight into the influence of factors affecting gut colonisation on changes in microbial profiles in newborn babies, a systematic review was completed. Literature on association of early life events and gut microbiota was assessed through PubMed (www.ncbi.nlm.nih.gov) in accordance with PRISMA guidelines 5 (all studies that were retrieved after search were screened for inclusion and exclusion criteria). The search was conducted on the 1st August 2013, using the following Medical subject heading (MeSH) key words, limited to human studies: (gut OR gastrointestinal* OR enteric OR bowel) AND (biota OR microflora OR bacteria OR microbes OR microbiome) AND (neonat* OR infant* OR baby OR babies) AND (development OR colonisation OR colonization). Forward citations were traced and authors were contacted where additional data were required. Data on study design, location, population, outcome, adjustment variables, method of outcome measurement, and potential sources of bias (including recruitment selection and exclusion criteria) were extracted for the following factors: 1) mode of delivery, 2) gestational Age, 3) nutrition, 4) exposure to antibiotics. Studies were excluded if microbiota data were not presented in relation to

at least one of these four parameters, babies had NEC/sepsis or if infants were administered probiotics and/or prebiotics.

This literature review located 2179 papers of which 874 studies were excluded as either they were reviews or were not analysing gut microbiota in humans. The remaining 1305 papers were screened for inclusion and exclusion criteria. Fifty one publications that met inclusion criteria were included after abstract screening. Fourteen studies were excluded as they did not meet exclusion criteria or they did not analyse influence of any of the factors, did not show details of infants' characteristics and presented general data for infants. This left thirty six publications to be discussed in the review. Due to the fact that some of the publications were based on the same study and the results were complementary, they were merged which gave final number of studies equal thirty three.

The influence of mode of delivery on gut microbiota was analysed in fourteen publications [46-60]; type of diet in twenty three experiments [39, 41, 50, 51, 54, 57, 58, 60-77]; exposure to antibiotics in ten [41, 49, 51, 57, 58, 60, 64, 78-81] and gestation age in three publications [61, 64, 79]. Eleven studies analysed gut microbiota of preterm infants [39, 41, 46, 48, 50, 51, 61, 64, 71, 79, 80], eleven of full-term infants [47, 52-54, 61, 63-65, 69, 72, 76, 79, 81, 82] and in fourteen publications there was no information about gestational age of infants [49, 56-59, 62, 63, 66-68, 70, 73-76, 78]. Majority of studies did not focus on any particular class of bacteria. Only eight papers had any class/species of interest including *Clostridium* [64, 80], *Lactobacilli* and *Bifidobacteria* [51, 52, 67, 68] and *Enterobacteriaceae* [59]. A summary of the studies included for this systematic review is given in Table 1.

Table 1: Studies included in the systematic review

Author	Year and location of sample collection	Species studies	Population details:	Analytical technique	Parameters evaluated
Blakey JL [41]	1979-80 Australia	Prevalence of bacterial species	28 preterm infants aged 25-36 weeks GA. Samples were collected daily from 1 st till 8 th day of life then twice a week for 6 th weeks. Samples were analysed immediately.	Culture, Gas-liquid chromatography	Diet: EN, TPN Antibiotics
Stark PL and Lee A [61]	Prior to 1982 Australia	Prevalence and counts of bacterial species	11 preterm infants aged 30-35 weeks GA and 15 full term infants. 3-6 samples per individual were collected in 1 st week of life then weekly till 1 st month. Samples were stored at -60°C.	Culture	Diet: human milk, formula milk

Development of gut microbiota and metabolic changes in preterm babies

Stark PL and Lee A [62]	Prior to 1982 Australia	Prevalence and counts of bacterial species	14 infants of undisclosed GA. 5 samples per individual were collected within 1 st week of life, then at 4 th week and every 6 weeks f life 6 th month of life. Samples were stored at -60°C.	Culture	Diet: human milk, formula milk
Yoshioka H [63]	Prior to 1983 Japan	Prevalence and counts of bacterial species	13 infants of undisclosed GA. Samples were collected daily for 6 days after 1 st month of life for 6 days and analysed within 2h.	Culture	Diet: human milk, formula milk
Lundequist B [72]	Prior to 1985 Sweden	Prevalence of bacterial species	22 full-term infants. Samples were collected on 5 th day, 3 rd and 8 th week of life and analysed within 5h.	Culture	Diet: human milk, formula milk
Rotimi VO [46]	1982-83 Nigeria	Prevalence of bacterial species	23 preterm infants aged 24-36 weeks GA. Samples were collected on 1 st , 2 nd , 3 rd and 6 th day of life and analysed within 30min.	Culture	Mode of delivery
Merida V [64]	Prior to 1986 Belgium	Prevalence of <i>Clostridium</i> species	32 preterm infants aged 29-35 weeks GA and 31 full-term infants. No data was provided on sample collection timing.	Culture	Diet: human milk, formula milk, PN Antibiotics
Neut C [73]	Prior to 1987 France	Prevalence of bacterial species	15 infants of undisclosed GA. Samples were collected at 0h and 6 th h then on 1 st , 2 nd , 4 th , 7 th and 14 th day of life and were analysed within 24h.	Culture	Diet: human milk, formula milk
Bezirtzoglou E and Romond C [74]	Prior to 1990 France	Counts of bacterial species	19 infants of undisclosed GA. Samples were collected at 0h and 6 th h then on 1 st , 2 nd , 4 th , 7 th and 14 th day of life. Then they were analysed within few hours.	Culture	Diet: human milk, formula milk
Adlerberth I [49]	1982-83 Sweden, Pakistan	Prevalence of bacterial species	39 infants of undisclosed GA. Samples were collected daily till undisclosed time point. No data on storage method was provided.	Culture	Mode of delivery
Bennet R [78]	1987-88 Sweden, Ethiopia	Prevalence of bacterial species	60 infants of undisclosed GA. Samples were collected from 2 nd till 8 th week of life and analysed within 24h.	Culture	Antibiotics
Kleessen, B [76]	Prior to 1995 Germany	Prevalence and counts of bacterial species	39 full-term infants. Samples were collected on 7 th , 15 th , 30 th , 60 th and 90 th day of life and analysed immediately.	Culture	Diet: human milk, formula milk
Gewolb IH [39]	1999 USA	Prevalence of bacterial species	29 preterm infants of undisclosed GA. Samples were collected on 10 th , 20 th and 30 th day of life and analysed immediately.	Culture	Diet: human milk, formula milk
Gronlund MM [47]	1995-96 Finland	Prevalence of bacterial species	64 full-term infants. Samples were collected on 1 st , 10 th , 30 th , 60 th and 180 th day of life and stored at 4°C for up to 35h.	Culture	Mode of delivery
Bonang G [65]	2000 Netherlands	Counts of bacterial species	51 full-term infants. Samples were collected in 1 st week and then in 2 nd , 4 th and 6 th month of life and then stored at -20°C.	Culture	Diet: human milk, formula milk
Harmsen HJ [66]	Prior to 2000 Netherlands	Prevalence of bacterial species	12 infants of undisclosed GA. 6 samples per individual were collected till 20 th day of life and stored at 4°C.	Culture, RAPD, qPCR, FISH	Diet: human milk, formula milk
Favier CF [81]	Prior to 2003 Netherlands	Counts of bacterial species	5 full-term infants. Samples were collected daily for first 2 weeks and then weekly till 160 th day of life. No data on storage method was provided.	PCR-DGGE	Antibiotics
Hällström M [48]	Prior to 2004 Finland	Prevalence of bacterial species	42 preterm infants aged 25-31 weeks GA. Samples were collected twice weekly for first 2 months of life. No data on storage method was provided.	Gas-liquid chromatography	Mode of delivery
Ahrne S [67]	1998-2000 Sweden	Prevalence of <i>Lactobacillus</i> species	112 infants of undisclosed GA. Samples were collected 1 st , 2 nd , 4 th and 8 th week and then in 6 th , 12 th and 18 th month of life stored at 4°C.	Culture, RAPD, qPCR	Mode of delivery Diet: human milk

Development of gut microbiota and metabolic changes in preterm babies

Adlerberth I [49]	1998-2000 Sweden	Prevalence of bacterial species	117 infants of undisclosed GA. Samples were collected in 1 st , 2 nd , 4 th and 8 th week and then in 6 th , 12 th and 18 th month of life stored at 4°C.	Culture, qPCR	Mode of delivery Antibiotics
Bezirtzoglou E [68]	Prior to 2006 Greece	Counts of <i>Bifidobacterium</i> species	26 infants of undisclosed GA. Samples were collected on 4 th day of life. No data on storage method was provided.	FISH	Diet: human milk, formula milk
Magne F [40]	2000-01 France	Presence of bacterial species	16 preterm infants aged 27-36 weeks GA. One sample per infant was collected between 5 th and 56 th day of life and stored at -20°C.	PCR-TTGE, sequencing	Mode of delivery Diet: human milk, formula milk, mixed milk
Butel MJ [42]	2002-05 France	Presence of <i>Bifidobacterium</i> species	52 preterm infants aged 30-36 weeks GA. Samples were collected twice weekly till 34 th day and stored at -80°C.	PCR-TTGE	Mode of delivery Diet: formula milk, mixed Antibiotics
Chen J [52]	2006 China	Counts of bacterial species	40 full-term infants. Samples were collected daily till 7 th day of life. Samples were stored at 4°C for several hours then at -80°C.	qPCR	Mode of delivery
Biasucci G [53, 55]	2003 Italy	Presence of bacterial species	46 full-term infants. Samples were collected on 3 rd day of life stored at -20°C.	PCR-DGGE, PCR-TTGE	Mode of delivery
Mitsou EK [54]	Prior to 2008 Greece	Prevalence and counts of bacterial species	97 full-term infants. Samples were collected on 4 th , 30 th and 90 th day of life and analysed immediately.	Culture, qPCR	Mode of delivery Diet: formula milk, mixed
Dominguez-Bello MG [56]	Prior to 2010 Venezuela	Abundance of bacterial species	10 infants of undisclosed GA. One sample per infant was collected within 24h of life and stored in liquid nitrogen.	qPCR	Mode of delivery
Fallani M [57, 58]	Prior to 2010 Sweden, UK, Germany, Spain, Italy	Abundance of bacterial species	606 infants of undisclosed GA. Samples were collected at 6 th week of life and stored at 4°C.	FISH	Mode of delivery Diet: human milk, formula milk, mixed milk Antibiotics
Roger LC and McCartney AL [69]	2005-07 UK	Counts of bacterial species	14 full-term infants. Samples were collected fortnightly for 10 weeks from 1 st month of life. Then monthly till 12 th month of life, then on 15 th and 18 th month. Samples were stored at -20°C. Findings only till 6 th month were included in the review.	FISH	Diet: human milk, formula milk
Bezirtzoglou E [70]	Prior to 2011 Greece	Abundance of bacterial species	12 infants of undisclosed GA. Samples were collected from 11 th till 36 th day of life and stored firstly at 0-4°C and then at -70°C.	FISH	Diet: human milk, formula milk
Morowitz MJ [71]	Prior to 2011 USA	Abundance of bacterial species	1 preterm infant aged 28 weeks GA. Samples were collected daily from 5 th till 21 st day and stored at -80°C.	Sequencing	Diet: human milk, PN supplemented with formula milk, TPN
Arboleya S [79]	Prior to 2012 Spain	Counts of bacterial species	20 preterm infants aged 30-35 weeks GA and 21 full-term infants. Samples were collected on 2 nd , 10 th , 30 th and 90 th day. Samples were stored at -20°C and within a week at -80°C.	qPCR, PCR-DGGE	Antibiotics
Ferraris L [80]	Prior to 2012 France	Presence of bacterial species	76 preterm infants aged 24-36 weeks GA. Samples were collected weekly until discharge and stored at -80°C.	Culture, PCR-TTGE	Antibiotics
Schwartz S [77]	Prior to 2012 USA	Counts of bacterial species	12 infants of undisclosed GA. Samples were collected on 3 rd month of life. No data on storage method was provided.	Sequencing	Diet: human milk, formula milk
Azad MB [60]	2008-09 Canada	Counts of bacterial species	24 full-term infants. One sample per infant was collected between 3 rd and 4 th month of life. No data on storage method was provided.	Sequencing	Diet: human milk, mixed Mode of delivery Antibiotics

1.2.iii.a Mode of delivery

Thirteen studies from this systematic review reported that mode of delivery has an impact on gut microbiota development; only one study did not find any correlation [39].

A vast number of studies found a negative correlation between Caesarean delivery and colonization by *Bacteroides* [37, 46-49, 57, 58], *Bifidobacterium* [27, 37, 46, 47, 49-55], *Escherichia coli* [44, 48, 49, 53, 55, 59], *Lactobacillus* and *Lactobacillus*-like bacteria [46, 47, 54, 56]. A number of studies found that vaginal delivery was negatively associated with the presence of *Clostridium* [37, 47-50]. No influence of mode of delivery on the *Streptococcus* population was reported [48, 50, 56-58] and findings for other bacteria were contradictory. The findings regarding the diversity and richness are contradictory as well. Three studies reported that Caesarean section leads to increased diversity of gut microbiota [48, 59, 83]. Rotimi *et al.* recorded that vaginally delivered infants had higher diversity of species [46] while Hallstrom *et al.* found that infants delivered by Caesarean section had higher number of species but equal number of genera [48]. At the same time Adlerberth *et al.* reported that infants delivered by Caesarean section had higher number of genera [59].

Additionally, research analysing gut microbiota during pregnancy and the correlation of maternal and infant's gut microbiota found no clear pattern. The results indicate that the infant after birth has microbiota of equal similarity with the microbiota of any mother during the beginning of pregnancy, but no similarity with the microbiota of any mother at the end of pregnancy [35]. That would suggest that perhaps maternal microbiota at the moment of birth does not affect infant gut colonisation. At the same time it is important to remember that the study analysed full-term infants and that the preterm gut may react to maternal microbiota differently.

1.2.iii.b Type of diet

Although formula milk aims to resemble the breast milk there are many differences in composition. One is higher content of proteins in formula milk than those in breast milk [34]. Higher protein content can potentially lead to a shift of microbiota towards species that specialise in protein fermentation. The process of the protein fermentation itself is known to produce toxic metabolites like

amines, ammonia, phenols and sulphides [84]. Untreated human milk (fresh, not pasteurised) has a protective role for the infant gut. Apart from having IgA antibodies it also contains a number of oligosaccharides of which some are homologous to cell surface pathogen receptors [18]. Thanks to this pathogens are less likely to bind to surface of the infant's gut as they will bind to free oligosaccharides coming from maternal milk. Additionally a study analysing the species of *Bifidobacterium* extracted from breast milk of healthy mothers found that some of these species were able to inhibit *in vitro* the growth of such pathogens as *Salmonella enterica*, *Shigella sonnei* and *Cronobacter sakazakii* while growth of *Clostridium difficile* or *Staphylococcus aureus* was not inhibited [85]. At the same time, it was noticed that some of the tested bifidobacterial strains were in fact facilitating adhesion of pathogens to intestinal mucus [85].

Most of the studies from systematic review found that presence of *Bifidobacterium* is associated with breast milk [54, 57, 58, 61-63, 66, 68-70, 72-76]. At the same time, the following species are more common among formula-fed infants: *Bacteroides* [37, 39, 57, 58, 62, 63, 66, 70, 76], *Clostridium* [37, 50, 57, 58, 66, 69, 72-76, 86], *Enterobacteriaceae* [37, 39, 44, 50, 57, 58, 65, 66, 71, 72, 76] and *Enterococcus* [63, 65, 66, 76]. Findings for the other genera and the diversity of gut microbiota are contradictory. One study found that formula-fed infants had higher number of species and genera in comparison to breast-fed ones [66]. Other study claims that total bacterial count was slightly higher for formula-fed infants within 1st week but after 1st month it was higher for breast-fed ones [63]. Other study found diversity (Shannon's Index) and richness (Chao Index) were much lower after administration of breast milk to preterm infants [87]. Other studies found that bacterial colonisation of the intestines was faster in breast-fed infants and that total bacterial count was higher during first 3 months of life [73, 76]. Finally, last study found that guts of infants fed with breast milk were colonised by more species than in case of formula fed infants and at the same time, the total bacterial count was similar in both groups [39].

According to Blakey *et al.* total parenteral feeding in first 4 days leads to delayed gut colonisation, lower bacterial diversity (lower number of species), half the number of *Bacteroides* and slightly higher number of *Clostridium* [41]. Another study found that both TPN and PN are associated with

presence of *Pseudomonas* and *Pseudomonacaceae* [71]. Parm *et al.* found that parenteral nutrition decreased chances of colonisation by *Klebsiella oxytoca*, *E. coli* and *Serratia* [44]. The decrease of microbial diversity as a result of parenteral nutrition was confirmed by Jacquot *et al.* [83].

1.2.iii.c Antibiotics

Exposure to antibiotics seems to affect many different genera. The population of beneficial bacteria like *Lactobacillus* [41, 78] and *Bifidobacterium* [37, 51, 79, 81] decreases in the presence of antibiotics. Two studies found that exposure to antibiotics negatively affect the population of *Enterococci* [78, 81]. Three studies report decreased number of infants colonized with *Clostridium* when they were given antibiotics [41, 49, 64]. A number of studies found that the total bacterial count and the diversity of species decrease when the antibiotics were given [39, 41, 42, 87]. However, one study found that early neonatal administration of drugs increased the diversity of gut microbiota [83].

Nine studies from this systematic review analysed the influence of antibiotics on gut microbiota. The findings in studies included in the systematic review are contradictory. This may be because of a number of factors. Firstly, different types of antibiotics are prescribed. Secondly, many species become resistant over time thus the older papers may have different finding than the modern one. Finally, infants that require antibiotics are usually sick and different types of diseases may lead to different microbiological profiles of the gut.

1.2.iii.d Gestational age

Gestational age affects the gut microbiota because of the immaturity of immune system. Additionally, some of the preterm births are linked to uterine infection thus gut microbiota may have an interesting link to gestational age [10, 88, 89]. Only three publications from this systematic review compared gut microbiota between preterm and full-term infants which was not enough to draw a conclusion [61, 64, 79]. The comparison of full-term babies with premature ones is not always informative. Preterm babies undergo different treatment than full-term ones (hospitalisation, prolonged exposure to antibiotics, parenteral nutrition) and as a result gestational age is not the only difference between preterm and full-term babies.

1.2.iii.e Other factors

There are number of other, less well reported factors that can potentially affect the gut microbiota. Those factors had not been included in the systematic review, but available findings are reported in this section. In adults, there had been found differences in the gut microbiota profiles between males and females. Men had been reported to have a higher relative abundance of species belonging to *Clostridia*, *Bacteroidetes* and *Proteobacteria* [38, 90]. The association of gut microbiota patterns with the geographical region is often mentioned in various reviews but can usually be explained by the difference in diets. Factors such as maternal education, use of probiotics or antibiotics by mother during pregnancy, prolonged rupture of membranes, sex, experiencing fever, living on a farm or having furry pets were found as possible contributors to gut microbiota profiles [37, 81].

1.2.iii.f Conclusion on systematic review

There had been only 11 studies that were conducted on preterm babies and met the criteria of this systematic review. Additionally, as seen in the Table 1 the sample collection times and methods used for bacterial profilings vary thus the comparison of the studies is difficult and results questionable. A large gap in the literature had been identified and this PhD project is contributing towards improving the knowledge regarding this issue.

1.2.iv Association of the gut microbiota with functional clinical outcomes

Another systematic review to assess what is known regarding the association of bacterial profiles with functional clinical outcomes had been conducted on the 1st Oct 2013. The systematic review was as well conducted in PubMed and the search term was: (gut OR gastrointestinal* OR enteric OR bowel OR colon OR intestine*) AND (biota OR microflora OR bacteria OR microbes OR microbiome OR microorganisms) AND (neonate* OR infant* OR baby OR babies) AND (disease OR allergy OR eczema OR asthma OR atopic OR obesity OR diabetes OR NEC OR necrotizing enterocolitis OR necrotising enterocolitis OR sepsis OR sirs). Studies were excluded if no sample was collected prior to the 6th month of life, or when pre/probiotics were administered to babies.

In total 62 studies had been included: five for sepsis, nine for NEC, four for diabetes, 10 for obesity, seven for autism, and 28 for various allergies and asthma.

1.2.iv.a Sepsis

Sepsis is a term that covers a number of infections including superficial skin infections, cellulitis, necrotizing fasciitis, wound infections, peritonitis, abscesses and septicaemia [7, 91]. Septicaemia occurs when bacteraemia is accompanied by the symptoms of infection [7]. Septicaemia is usually a result of spread of a localised infection and in 95% cases is caused by only one organism [7]. Sepsis syndrome is defined as septicaemia associated by an immunological reaction of the organism [7]. When the multiple organ failure occurs then the condition is called septic shock [7].

Sepsis is the most common problem in pediatric intensive care units and neonatal sepsis occurs in 0.8% of life births and in quarter of cases is associated by meningitis [92]. It is estimated that 25% of newborn infants with sepsis die. Early onset of neonatal sepsis occurs within 7 days of life and is usually caused by bacteria from the mother's vagina which can enter via the amnion or during the birth [92]. Late onset of sepsis occurs after 7th day and is believed to be caused by pathogens from the environment [92]. Risk factors for newborns to acquire sepsis include prematurity, low birth weight, prolonged rupture of membranes, maternal urinary tract infection and maternal colonization by group B *Streptococci* [92]. The major bacterial contributors of sepsis are group B *Streptococci*, *E. coli*, *Staphylococci*, *Enterococci*, *Klebsiella*, *Haemophilus* and *Pseudomonas* [92-96].

After analysing the findings from my systematic review investigating the correlation of gut microbiota profiles and sepsis, there were insufficient data to draw a conclusion for any of the genera or species. All of the studies found microbial differences between health and sepsis patients [97-101]. It was reported that microorganisms responsible for the onset of sepsis matched those obtained from the guts of patients. This included species such as *E. coli*, *Enterobacter cloacae* and *Acinetobacter baumani*. This would mean that gut microbiota can be a reservoir for the microorganism inducing sepsis. There was however no agreement on the influence of the high abundance of *Enterobacteriaceae* [98, 100].

Sepsis is a complex problem and it seems there is no common pattern of gut microbiota changes that could be linked with the higher incidence of sepsis.

1.2.iv.b *Necrotizing enterocolitis*

Necrotizing enterocolitis (NEC) is very important disease that can affect newborn infants. Up to 90% of all cases are seen in very low preterm infants. Additionally in more premature babies, NEC occurs later after birth [32]. There are studies analysing the contribution of gut microbiota to the development of NEC. Early detection of 'NEC microbial profile' could help prevent the onset of disease not only in premature but also in full term infants.

In this systematic review the presence of *Klebsiella* [102-105], *C. perfringens* [29, 103, 106, 107], group D *Streptococci* [102, 103], and *Enterobacter* [48, 103, 104] has been found to be associated with NEC. The correlation between the presence of *C. perfringens* and the NEC has also been observed in the studies on piglets [29]. The results for other bacteria are contradictory. There was also no conclusive answer regarding the incidence of NEC and diversity or richness of gut microbiota [83, 105, 108, 109].

Interesting finding was in relation to the relative abundance of *Proteobacteria*. Healthy infants had stable abundance of this phylum while patients that later developed NEC had lower abundance at birth and experienced a rapid increase in the abundance prior to onset of NEC [105, 109]. This was also observed on twins where only one of the infants developed NEC while the other remained healthy. Additionally, Smith *et al.* found in intestinal tissues surgically removed from NEC patients that members of the *Proteobacteria* phylum represented 49% of all bacteria [110]. It is possible that the presence of *Proteobacteria* at the early stage of life may lead to a proper maturation of the immune system resulting in lack of overgrowth of those opportunistic pathogens.

1.2.iv.c *Allergy, atopy and asthma*

Twenty-three studies reported findings for *Bifidobacterium* and allergy/atopy/eczema/asthma. Most of the studies reported a lack of correlation between prevalence [111-118] and total counts of the *Bifidobacterium* genus [113, 118-125] and the future development of allergy. On the other hand four

studies reported that the prevalence [119, 121, 126, 127] and the counts [123, 126-130] of *Bifidobacterium* were lower among allergic infants. There was not enough data about the abundance of this genus to draw any conclusions. No correlation of the development of the allergy and the prevalence of *Bacteroides* [111, 114, 117, 119, 121, 126, 131], *Lactobacillus* [111, 113, 114, 116, 117, 121, 126], *Streptococcus* [119, 121, 126], *Enterococcus* [121, 126], *Eubacterium* [119, 121, 126], *Enterobacteriaceae* [118, 121, 126] was found.

Additionally no correlation between the allergy and change of the bacterial counts of *Bacteroides* [114, 118, 119, 121, 122, 125, 126, 128, 130], *Clostridium* [113, 114, 118, 119, 121-126], *Lactobacillus* [113, 119, 121-128], *Enterococcus* [118, 119, 121, 122, 126-128], *Staphylococcus* [113, 118, 121, 126, 130], *Streptococcus* [119, 121, 125, 126], *Enterobacteriaceae* [118, 121-123, 125, 126], *Eubacterium* [121, 126] was observed.

A number of studies found no difference in the total bacterial counts [27, 123, 126, 128, 129]. Only Smehilova *et al.* reported lower total bacterial counts in the infants with allergic colitis [127]. Additionally two studies reported a lower diversity in the infants that developed eczema after sample collection [132, 133] while Bisgaard *et al.* found lower diversity prior to allergy but not prior to asthma [134]

The findings from this systematic review show that there are no known bacteria that are confirmed by few independent studies as strongly associated with the presence of NEC.

1.2.iv.d Obesity

The systematic review retrieved 10 studies that analysed the correlation of gut microbiota with obesity. Only two of those studies analysed bacterial profiles before infants got obese or overweight [135, 136]. Based on those two studies there seems to be a positive correlation between *Bacteroides* and obesity [135, 136]. However, the number of prospective studies is too low to draw any conclusions. There is an obvious gap in the literature regarding the colonisation of gut and future weight gain.

1.2.iv.e Diabetes

Only four studies were found to be eligible for the analysis of correlation of diabetes with changes in gut microbiota profiles. Diabetes was reported to be associated with higher counts of *Proteobacteria* and *Bacteroides* [137, 138] and also higher prevalence of the latter genus [139]. There was also a negative correlation with counts of *Bifidobacterium* [138, 139] and *Firmicutes* [137, 140]. Findings for *Clostridium*, *E. coli* and *Lactobacillus* were contradictory [137, 139, 140]. It is however important to remember that the number of included studies was low and only one study analysed microbiota prior to diagnosis and thus the conclusion may be biased. What is more important is that Giongio focused on type-1 diabetes while other studies analysed type-2. The systematic review had not retrieved enough data to draw any meaningful conclusions.

1.2.iv.f Neurodevelopmental disorders

The systematic review retrieved only six studies meeting inclusion criteria for the correlation of the gut microbiota with autism. None of the studies analysed microbiota prior to diagnosis thus the results cannot be treated as a proof of causality. All of the studies, that provided this data, had groups with a similar age of the individuals. Autism was reported to be negatively associated with the abundance of *Collinsella* [141, 142], *Verrucomicrobia* [142] and some species of *Clostridium* (*C. leptum* and *C. methylpentosum*) [141]. At the same time *Desulfovibrio* was found to be positively linked with autism [141].

1.2.iv.g Conclusion from the systematic review

The aspect of association of gut microbiota profiles with the functional clinical outcomes is not well investigated. There are very few studies that had looked into this and the results are usually inconclusive.

1.3 Preterm infants

1.3.i Definition and epidemiology

Human infants are usually born after 37 weeks of gestation. Such infants are called full-term [143-145]. Preterm infants, those born before 37th week of gestation are divided into early preterm and late preterm (Figure 1). Early preterm are those that were born before 32nd week of gestation while late preterm are those born between 32nd and 37th week [144-147]. Early preterm infant are classified as extremely preterm if born before 28th week while those born between 28th and 31st week of pregnancy are called very preterm [145]. Late preterm infants can be divided into mild and moderate if they are born between 32-33 week and 34-36 week respectively [145]. Late preterm birth occurs more often than early [89, 145, 147]. Approximately 1-1.5% of births in UK take place before the 31st week of gestation while worldwide up to 11% of births are preterm [143, 147, 148]. Due to the fact that it is difficult to know exact day of conception gestational age (GA) is usually estimated rather than exact value given. In some countries like USA gestational age is not used on daily basis. Instead a different terminology is used that is related to the weight of the baby. Full-term infants in the UK are usually born with a birth weight of around 3.5 kg [149]. If they are born with a weight lower than 2.5kg they are termed low birth weight infants. If their weight is lower than 1500g they are called very low birth weight and if less than 1000g, extremely low birth weight. Approximately 6% of live births in UK are low birth weight [150].

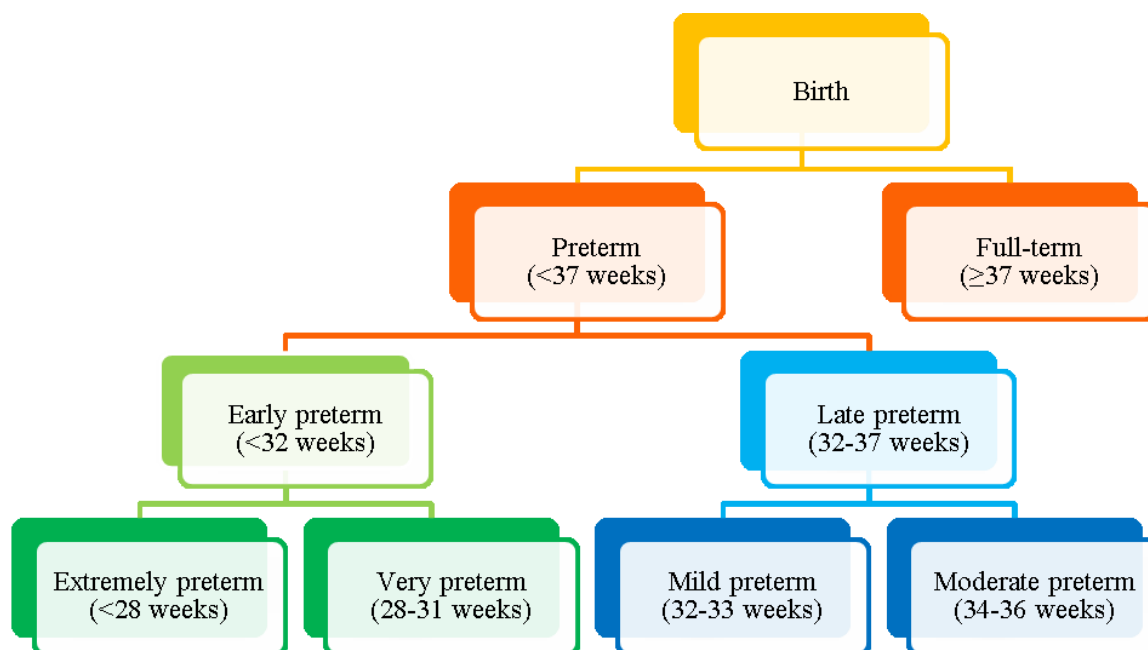


Figure 1: Flowchart presenting subgroups of premature birth. Based on the information from Moutquin 2003 [145]

1.3.ii Types of prematurity

The majority of preterm births, 21-68%, are idiopathic, which means that they are of an unknown aetiology. A common associate of preterm birth (7-58% cases) is preterm premature rupture of the membranes (PPROM). At the same time, on average 25% of preterm deliveries (range from 9% to 38%) are medically indicated [145-147, 151-153]. It is also believed that the early preterm birth (less than 32 weeks GA) is most frequently associated with the infection of the uterus [89, 154]. Prematurity is responsible for 75% of morbidity and 70% of mortality among neonates in developed countries and each year approximately 1 million infants die worldwide due to preterm complications [89, 144-147, 152, 153]. As a result of improvements in modern medicine the balance between mortality and morbidity has changed. Infants that survive increase the number of severe morbidity [151]. There are several factors that affect mortality and morbidity and they include gestational age (higher risk for early preterm infants than for late preterm infants), birth weight (smaller infants are at bigger risk) and plurality (higher rate of survival for singletons) [144, 147, 151].

1.3.iii Factors contributing to prematurity

There are number factors that contribute to the preterm delivery. This includes maternal age and health, occurrence of previous preterm birth, in-vitro fertilisation, low socio-economic status (at the same time there is an increase in preterm birth in wealthy communities), genetic factors, repeated 3rd trimester abortion, lifestyle and low BMI [89, 145, 147, 151, 152].

Medically induced preterm birth was found to be associated with intrauterine growth restriction, unstable fetal condition, maternal hypertension, in-vitro fertilisation, medical acute illness, obstetrical complications, multifoetal pregnancy, diabetes and maternal age above 35 years [145, 151, 152]. PPRM is more likely to occur in case of vaginal infection, uterine distension, cervical anomalies and in low socio-economic status [145]. Idiopathic preterm birth is related to previous preterm birth, low body mass and poor body mass gain, stress, uterine anomalies, smoking, drug abuse and maternal age below 18 years [145].

It is believed that intrauterine infections can be responsible for up to 40% of preterm labours [89, 147, 155]. However, it is more likely that the presence of bacteria in the uterus predisposes to the premature birth rather than is the main cause. Bacterial presence deep within the human foetal membranes had been observed in 70% elective caesarean sections at term [154]. Additionally, Steel observed that all of the intact fetal membranes obtained from preterm fetuses and some from term fetuses (included: not in labour, in labour, PPRM) had bacteria and/or inflammatory cells present. [154]. Additionally, bacterial presence in the amnion or chorio-amnion was found in up to 80% of preterm births and 30% of term births [156].

There are four main routes for microorganisms to enter the uterus (Figure 2). The first and the most common one is through the vagina and cervix, the second route is a transfer through placenta, the third is an accidental infection during an invasive procedure and the last possible route is a transfer of microorganisms from abdominal cavity through the fallopian tubes [89, 147, 155]. Additionally, bacterial vaginosis (BV) is believed to increase 2-5 times the risk of preterm birth although most women with BV deliver at term and most placental inflammation is not related to BV [88, 89, 156].

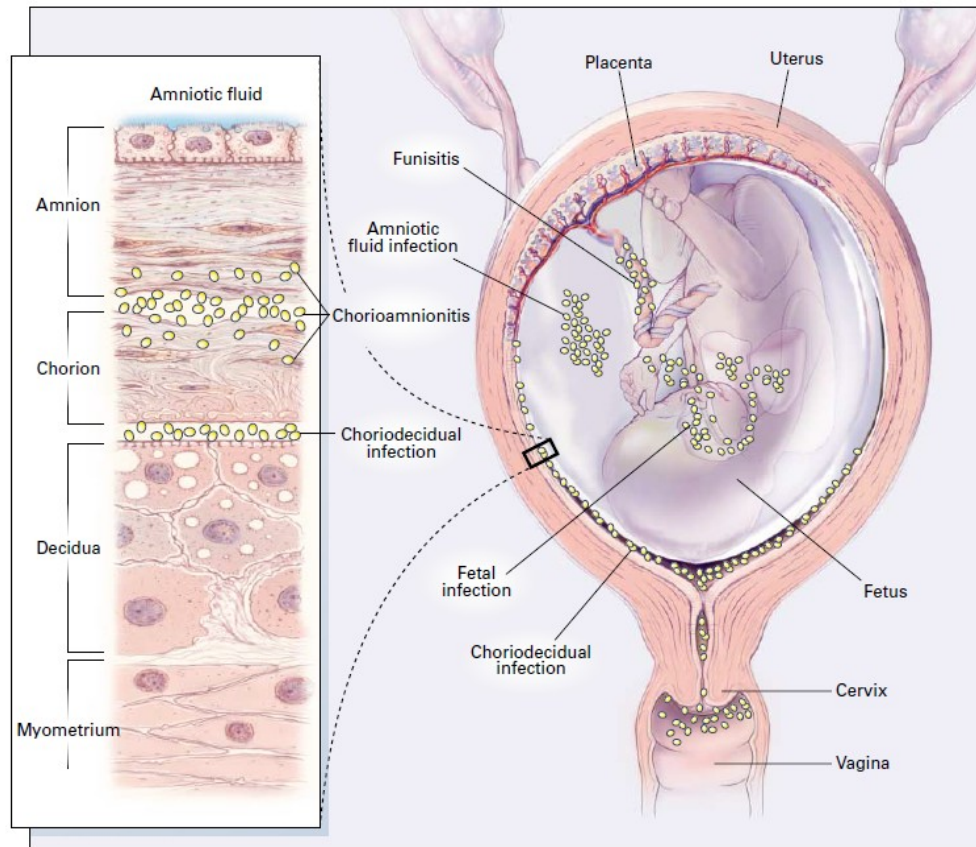


Figure 2: Potential sites of bacterial infections in the uterus. Reproduced with permission from Goldenberg et al. 2000 [89], Copyright Massachusetts Medical Society

Additionally, microorganism can colonise a number of sites in the uterus. They can colonise the placenta (villitis), the choriodecidual space (the space separating the fetal and maternal tissues), the fetal tissues, amnion and chorion (chorioamnionitis), the amniotic fluid (amnionitis), the umbilical cord (funisitis) or the fetus itself (Figure 2) [89]. One of the theories trying to find the link between the infection of the uterus and the preterm labour is that the fetal inflammatory response can lead to the onset of labour (Figure 3) [147, 154].

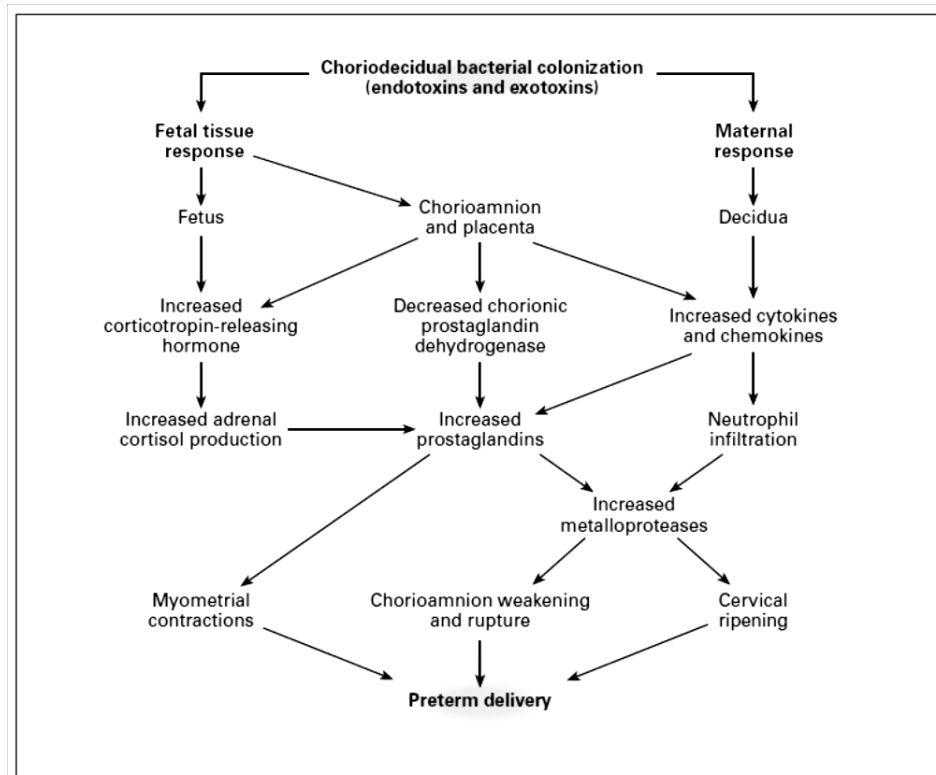


Figure 3. Potential Pathways from Choriodecidual Bacterial Colonization to Preterm Delivery.

Figure 3: Potential pathways from chorodecidual bacterial colonisation to occurrence of premature birth. Reproduced with permission from Goldenberg et al. 2000 [89], Copyright Massachusetts Medical Society

The most commonly found micro-organisms in the uterus of women delivering preterm without the rupture of membranes are *Gardnerella vaginalis*, *Mycoplasma hominis* and *Ureoplasma urealyticum* [89, 147, 156]. In the case of PPROM, the enteric gram-negative bacteria (*Enterobacteriaceae*) represent up to 40% of microorganisms recovered from the amniotic fluid [155]. Other common microorganisms that can be found in the womb after the rupture of membranes are *E. coli* and group B streptococci [89]. There is also a number of other species of either oral or vaginal origin that may be linked to preterm birth. They are usually members of one of the following genera: *Bacteroides*, *Fusobacterium*, *Gardnerella*, *Lactobacillus*, *Peptostreptococcus*, *Porphyromonas*, *Prevotella*, *Actinomyces*, *Campylobacter*, *Capnocytophaga*, *Mobiluncus*, *Leptotrichia* [89, 156].

1.3.iv Characteristics

Preterm infants are born immature. Different systems develop at different stages of their growth in the womb. For example suck swallow coordination is not fully developed until the 34th week [32]. This means that such infants are not able to be breast fed and require tube feeding. Additionally motility

and gastric emptying is delayed and quite often feeding intolerance is present [155]. Meconium is believed to suggest the maturity of the gut as it is rarely passed in utero before the 34th week of gestation [155]. It is hypothesised that early passage of meconium in the preterm infant may be caused by the swallowing of enteropathogenic bacteria from the womb [155]. In infants born before 30th week GA there is lower oesophageal tone [32]. The kidneys continue to develop in preterm infants after birth thus it is possible that urinary metabolites will show clusters relating to GA [157]. Nephrogenesis is usually finished by 34th to 36th week of gestation. Preterm infants at the 38th week of corrected age have smaller kidneys than the full term infants at the same age but when adjusted to the body size the kidneys of preterm infants were proportionally bigger [158]. Gut development is affected by the weeks post-partum which suggests that gut microbiota and faecal metabolites will be related mostly to weeks post-partum but still affected by GA [158]. Many preterm infants have problem with maintaining growth. They are usually lighter than a fetus remaining in the womb and once reaching term equivalent age, they are shorter than full-term infants [159]. A subnormal growth of head may indicate problems with neurodevelopment and may lead to lowering of IQ in future life [159]. Additionally, teenagers that were born extremely-low-birth-weight are on average shorter and lighter than those born at term [160].

1.3.v Mode of delivery

1.3.v.a Vaginal

Vaginal delivery occurs in the majority of mammals (with exceptions of monotremes). Vaginal delivery can be divided into three subgroups: spontaneous, assisted, and induced. Spontaneous vaginal delivery is defined as following labour that was not induced by drugs or any other techniques, and without the use of any instrumentation. Assisted vaginal delivery is when the use of special instrumentation is required to facilitate the birth. When labour does not start spontaneously and is induced manually or by drugs it is referred to as induced vaginal delivery. Spontaneous vaginal delivery is more likely to occur than Caesarean section in case of idiopathic preterm birth and PPROM (~ 65%) [152].

1.3.v.b *Caesarean section*

Caesarean section is a surgery that involves removal of the fetus/infant from the uterus through an incision in the abdomen. Historically Caesarean section has been performed when mother's or infant's life was put at risk during birth. Nowadays it may be done upon request. Caesarean section may take place pre-labour or in labour which often includes rupture of membranes. In medically indicated preterm births the most common mode of delivery is Caesarean section (60-90% of such births, depending on week of pregnancy) [152].

1.3.vi *Type of diet*

The average growth rate of the fetus is 15g/kg/day. At 24 weeks of gestation the fetus is composed of 90% water, 2% lipids and 8.8% protein [161]. At term the composition of the newborn has altered to 75% water with lipids and protein increased to 11% and 12% respectively [161]. Newborn babies have a rapid switch from a high carbohydrate and low fat diet to a high fat and low carbohydrate diet [162, 163]. The necessary glucose is obtained from the glycogen stored in the liver [162]. However, when preterm labour occurs, the storage of the glucagon in the liver is quite low and the enzymes participating in gluconeogenesis are not fully active yet [162]. Because of that there is a higher risk of developing hypoglycaemia in preterm infants than in full-term infants and intravenous administration of glucose is often required [162]. Until the third trimester, the fetus does not accumulating lipids in adipose tissue at levels high enough for survival on this source only [163]. The ability to oxidise long chain fatty acids and to produce ketone bodies is low but increases soon after birth [162]. The process of lipogenesis has been observed in preterm infants (28 weeks GA) in the first day of life [162].

1.3.vi.a *Enteral nutrition*

There are several types of milks that can be provided as an enteral nutrition (Figure 4). Fresh mother's milk is the milk of choice however it is not always possible to provide this [164]. In such situation either donor milk or formula is given. Donor milk is usually pasteurised which leads to elimination of beneficial bacteria and deactivation of the bile-salt-stimulated lipase which is responsible for increasing milk fat absorption [163]. Formula may or may not be supplemented with probiotic

bacteria. Levels of beneficial bacteria are different from those in the mother's milk even when probiotics are used. Preterm formula is usually composed in order to satisfy the nutrition requirements of a growing fetus. A blinded randomised trial of the influence of mother's milk vs. the preterm formula found that mother's milk reduces the number of days on parenteral nutrition and reduces the chances of necrotising enterocolitis (NEC) [164].

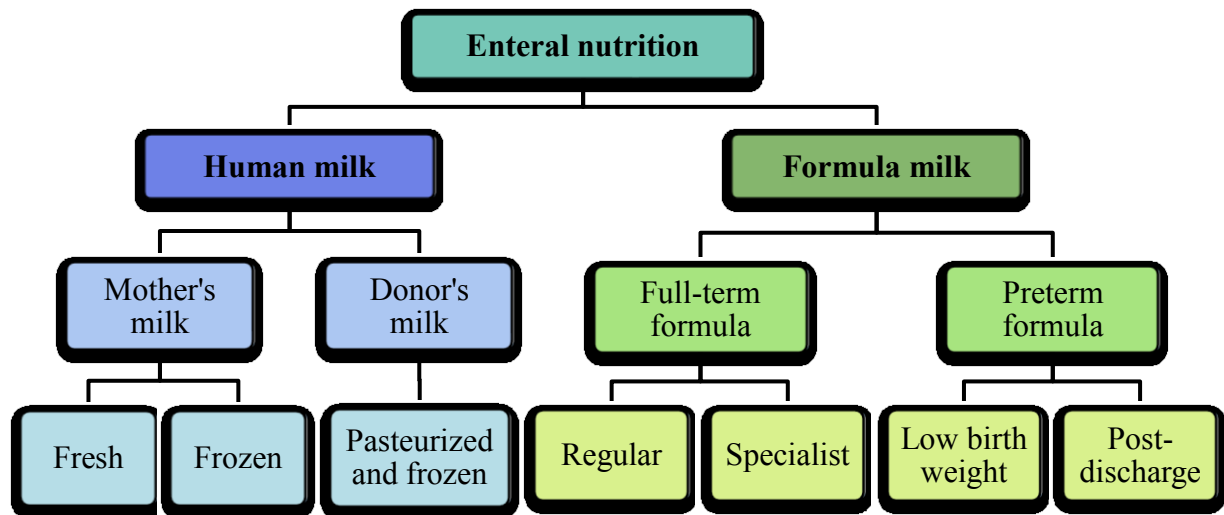


Figure 4: Types and subtypes of possible enteral nutrition

A proper composition of lipids, carbohydrates and amino acids is necessary for a healthy growth of preterm infant [157, 160]. The reason for administration of high-fat formulas is that the percentage of fat absorption remains stable regardless of the concentration of fat which means that high-fat formulas result in a bigger amount of absorbed and metabolised fat [165]. At the same time, it was considered desirable to reduce the concentration of fat and provide a caloric intake from carbohydrates and proteins [165]. Although preterm infants require proteins for rapid growth, the high concentration of those molecules can have an adverse effect on the infant's health [160]. High-protein formulas did often have high levels of sodium as well. Consumption of such formula results in a rapid weight gain due to the water retention which can lead to oedema [165]. Additionally prolonged high-protein diet may lead to metabolic acidosis as undeveloped kidneys of preterm infant are not able to cope with the metabolism of a non-volatile acid [165].

Newborn full-term infants are normally fed orally, however preterm infants are not always mature enough to be fed exclusively through mouth and need intravenous supplementation which is called a parenteral nutrition (PN) [157, 161, 162, 166]. Since the motor ability of suckling is not developed by the 34th week of gestation, babies born before are fed through a gastric tube.

1.3.vi.b Parenteral nutrition

Due to the immaturity of the gut (food intolerance, risk of necrotising enterocolitis) the majority of early preterm infants require some parenteral nutrition (PN) for 15.6 days on average [160, 167-169]. The time that an infant will spend on PN is affected by gestational age, body weight and morbidities [167]. A systematic review reported that there is a significant improvement in time to regain weight and in the reduction of maximum weight loss when parenteral nutrition is provided early after birth [170]. What is more, the same review mentioned that there was a significant reduction in chances of sepsis but no influence on the mortality when PN was provided straight after birth [170]. Energy requirements are higher for infants of small gestational age than for those born at term thus in order to maintain growth rate in preterm infants, a proper composition of nutrients must be provided [160, 161]. Resting energy is largely unaffected by the gestational age and is approximately 50 kcal/kg/day [160]. However, the energy needed for the development of new tissues is 4.5-4.9 kcal per gram of tissue [161]. The recommended daily caloric intake for preterm infants is 110-120 kcal/kg [161, 167] of which 20-25 kcal/kg/day should be from protein. In order to avoid growth faltering, enough energy of non-protein origin must be provided. In case of a disease more energy may be required [160].

When administering amino acid solutions, a daily intake of 1-1.5g/kg per day ($\text{g} \times \text{day} \times \text{kg}^{-1}$) is needed to prevent catabolism of proteins [166]. It is suggested that early administration of amino acids can have beneficial effect on head and brain growth and neurodevelopment and improves nitrogen balance, electrolyte balance and blood glucose control [167]. There are contradictory results whether rapid increase of amino acid concentration sustains optimum protein catabolism/anabolism balance [161, 166]. There is currently no gold standard of what should be the growth rate for preterm infants. It is assumed that the best option may be to maintain natural fetal growth [166] but this has not been established.

It was decided in the 1960s that parenteral administration of lipid emulsion may improve energy intake [171]. The presence of a lipid emulsion protects amino acids from being used as an energy fuel thus they can be used as protein building molecules [165, 172, 173]. Additionally, essential fatty acid deficiency may be prevented by the administration of lipids [173]. A common lipid emulsion provided intravenously is Intralipid which is composed of long chain triglycerides (LCT) from soybean oil and phospholipids from egg yolk [171]. The components of Intralipid are arranged into chylomicrons with phospholipids on the surface and triglycerides in the centre [172]. Soybean lipids are metabolised differently when they are given intravenously compared to oral nutrition [174]. Prolonged parental nutrition with Intralipid can lead to serious side effects like parental nutrition-associated liver disease (PNALD) [160, 175, 176]. It is suggested that the liver may not be able to metabolise artificial chylomicrons in the same manner as would it metabolise natural chylomicrons (created after digestion of fat) [174]. This is caused by the fact that vegetable oil-based lipid emulsions can lead to cholestasis and hepatostasis [160, 173, 175, 176]. Soybean oil is rich in omega-6-acids which can be proinflammatory and is often contaminated with phytosterols which are linked to liver damage [160, 171, 177]. Phytosterols are plant molecules that are analogous of cholesterol in animals [177]. Those molecules are not dangerous when administered orally, however when provided intravenously they are found to be toxic to the liver cells [177]. Additionally it was reported that administration of linoleic acid in high concentration can result in inhibition of macrophage and neutrophil activity [172]. An alternative source of lipids may be a solution to PNALD [160, 171, 175]. There is an alternative option to the Intralipid which is SMOFlipid, a lipid emulsion composed of soybean oil, medium-chained triglycerides, olive oil and fish oil.

Due to the fact that the major role of the introduction of the lipid emulsion is to preserve the amino acids from being used as an energy source, an interesting finding was that fish oil was associated with higher levels of branched-chained amino acids catabolites such as 2-methylbutyroylcarnitine, isovalerylcaritinine, hydroxyisovaleroyl carnitine, tiglyl carnitine, methylglutaroylcarnitine, compared to the soybean oil [173]. Fish oil is composed of long-chained fatty acids but is rich in omega-3-fatty acids which have more beneficial effect on the human body than omega-6-fatty acids that can be

found in soybean oil [173, 178]. Animal studies reported that fish oil based lipid emulsion does not impair bile flow, prevents the development of steatosis and can improve the health of PNALD patients [160, 173, 177, 178]. Metabolic analysis of the livers from mice fed either soybean oil or fish oil found that the latter led to a reduction of omega-6-fatty acids, arachidonic acid and hydroxyeicosatetraenoic acids (commonly produced during the inflammation) [173]. On the other hand eicosapentaenoic acid, which can be found in the fish oil, needs further testing on its safety in preterm babies [163].

Olive oil is believed to have a beneficial impact on health as some of its components like oleic acid, phytosterols, α -tocopherol, triterpenes, oleocanthal and phenolic compounds have been found to have anti-inflammatory and antioxidant abilities [171]. Additionally, *in vitro* and animal studies have found that olive oil based lipids, in contrast to soy-based lipids, did not lead to suppression of the T-lymphocyte activation and proliferation, delayed bacterial clearance, [157, 171, 179]. It was also reported that intravenous administration of olive oil lipids does not stimulate gluconeogenesis in contrast to administration of Intralipid [180].

Essential lipids for a newborn baby include linoleic, linolenic, arachidonic and docosahexaenoic acids [162]. All these lipids are included in the parenteral formula. Triglycerides are metabolised into fatty acids which then can enter ketogenesis. Ketones are considered to be a useful energy source in seriously ill patients thus administration of lipids that could be easily utilised into ketones is an important aspect of parenteral nutrition [181]. In the case of long chain fatty acids a molecule called carnitine is required for transportation into mitochondria. Seriously ill patients may have low levels of carnitine and thus a better energy source than long chain triglycerides was needed. Medium chain fatty acids can simply diffuse through mitochondrial wall without assistance of carnitine [172, 181]. Medium chain fatty acids are product of digestion of medium chain triglycerides (MCT). MCT are composed of three saturated fatty acids that have 6 to 12 carbon atoms. They are metabolised and absorbed two fold quicker than LCT due to the fact that they are 100 times more soluble in water than LCT and do not need attendance of pancreatic lipase or bile acids [172]. After parenteral administration MCT are neither accumulated in the liver nor stored in adipose tissue [165, 172, 182].

What is more, medium chain triglycerides do not participate in production of free radicals [172]. However, MCT are not advised for patients with ketosis problems [172]. The study performed by Denison *et al.* in the mid 80's found that administration of an emulsion composed of both long chain and medium chain triglycerides was metabolised more rapidly than an emulsion composed of only long chain triglycerides [181]. The finding was also confirmed by another study [183]. There were no side effects observed in adult men after 10 days of administration of emulsion composed of both long chain and medium chain triglycerides [181]. Long chain triglycerides cannot be totally removed from the lipid emulsion as they provide essential fatty acids which are crucial for proper visual and brain development [161, 172]. Fish oil is an interesting component of SMOF lipid which is not present in the Intralipid [175, 184]. Supplementation with fish oil lipid emulsion was found to decrease levels of pro-inflammatory cytokines, TNF- α and IFN- γ which are high in patients fed parenterally [168, 175, 185]. Furthermore, administration of fish oil lipid emulsion led to reversal of parental nutrition-associated liver disease symptoms [175]. A randomised double-blinded study analysing differences between lipid emulsions administered to preterm infants found that infusion of SMOF lipid led to decrease of total bilirubin and direct bilirubin when compared with infants fed with Intralipid [184]. At the same time there is not enough evidence that these hybrid lipids could improve the treatment of PNALD [178]. Additionally, animal studies have reported that supplementing PN with lipid emulsion reduces the levels of inflammatory markers Ccl2 and Apc5 (which are much higher in animals on PN without a lipid emulsion compared to control group) but does not affect the levels of inflammatory factors TNF α and IL1 β [186].

Carbohydrates are necessary for neurodevelopment as at early gestational ages the brain largely consumes glucose. The levels of glucose needed in preterm infant can be twice as high as in full-term infant [161]. The concentration of administered glucose must be balanced in order to avoid both hypo- and hyperglycaemia [168, 187]. Additionally, too high concentration of carbohydrates may lead to steatosis (abnormal retention of the fat in cells) [160]. Using exogenous insulin to obtain higher glucose delivery rates is not recommended as it can lead to reduction of protein synthesis [187].

Finally, electrolytes, vitamins and minerals are necessary for proper development of preterm infant. They should be provided at the same level as in utero. Phosphorus, magnesium and calcium are important to maintain function and structures of tissues [161]. When it comes to vitamins both fat-soluble and water-soluble must be provided with the latter one being administered on regular basis as humans are lacking abilities to effectively store them in their bodies [161].

Although the parenteral nutrition is required for survival of most of preterm infants, there may be some complications related to this mode of feeding. Firstly, there may be disorders related to the intravenous administration and this may include both thromboembolism and sepsis [93, 160, 161, 177, 188, 189]. Infection associated with catheters involves usually *Candida*, *Staphylococcus aureus* and coagulase-negative *Staphylococci* [93, 161]. Secondly, there may be an excess or deficit of administered nutrients. Preterm infants at discharge often have higher body fat composition than full-term infants [185]. This is why the proportion of molecules in emulsions is so important. Next, growth impairment or metabolic bone disease may be caused by parenteral nutrition as a result of improper calcium and phosphorus levels and aluminium contamination [161]. As previously mentioned PNALD is one of the biggest complications related to PN [160, 161, 177, 189] and it has a 100% mortality rate within the first year since the diagnosis if the transplantation is not available [160]. PN is only one of the risk factors for the development of PNALD. The other factors include young age, low birth weight, preterm birth, lack of EN, disruption of bile acids circulation, bile duct obstruction, nutrition excess or deficiency (especially taurine and choline) and sepsis [160, 177, 178]. Finally, there are reports suggesting that the origin of sepsis may not be the catheter but the gut microbiota that overgrows as a result of affected biliary flow and motility of gut [93, 160].

1.3.vi.c Total parenteral nutrition

Total parenteral nutrition (TPN) indicates that no oral food supplementation is given [185]. This form of diet is usually administered to very sick, very small or very preterm infants [185]. Oral consumption of food stimulates the secretion of gastrointestinal hormones like gastrin, peptide YY and cholecystokinin [177]. Lack of those enzymes may lead to reduced secretion of bile acids, intestinal stasis and consequently sepsis [177]. Prolonged administration of TPN to a number of

species (rats, pigs, humans) was found to increase body fat percentage [168, 185, 188, 190]. A study performed in rats found that animals on total parental nutrition for 21 days had altered organ weights when compared to their baseline characteristics. In particular they had lighter guts, and heavier kidneys, heart and lungs [188]. Additionally, TPN resulted in lower levels of haematocrit, haemoglobin and glucose [188]. A study in pigs found that administration of TPN results in insulin resistance and glucose intolerance [168]. It is believed that lack of EN plays a major role in the development of PN related complications [178].

1.4 Conclusions

Preterm babies differ in early life care practices from the full-term ones, it is therefore possible that the gut microbiota development will occur differently for full-term babies and preterm ones. The development and changes of the pattern of gut microbiota colonisation in premature babies had not been properly investigated and there is a large gap in the literature regarding the impact of early life interventions on the gut microbiota. This PhD has investigated how postnatal age, gestational age, mode of delivery, route of nutrition, type of milk, and sex of the baby is linked to gut microbial and metabolic profiles in preterm babies.

Chapter 2 – Materials and Methods

2.1 Study design

2.1.i NEON study

The NEON study stands for amino acid regimen and intravenous lipid composition in preterm parenteral nutrition: a randomised controlled trial of Nutritional Evaluation and Optimisation in Neonates. The NEON study was a randomised 2x2 factorial double blinded controlled trial which was funded by MRC/NIHR EME programme. The inclusion criteria for the NEON study were the prematurity (<31 weeks of gestation) and need of nutritional support in form of the PN and the exclusion criteria were life threatening anomalies. Infants were randomised within 12 h of life and assigned to one of 4 feeding mode groups. Factors variable in feeding mode are: concentration of amino acids (immediate recommended daily intake concentration vs. incremental concentration leading toward recommended daily intake) and lipid content (20% Intralipid vs. 20% SMOFlipid). The structure of the NEON study randomisation is presented in the Figure 5.

Additionally, infants differ in mode of delivery (vaginal delivery vs. elected caesarean section vs. emergency caesarean section) and antibiotic exposure. Full daily health record data were collected on the enrolled infants and this data is available for interrogation in the final analysis. This data included: gestation age, sex, mode of delivery (vaginal, Caesarean section), route of nutrition (parenteral, enteral), type of milk (formula, breast milk). Stool and urine samples were collected on weekly basis.

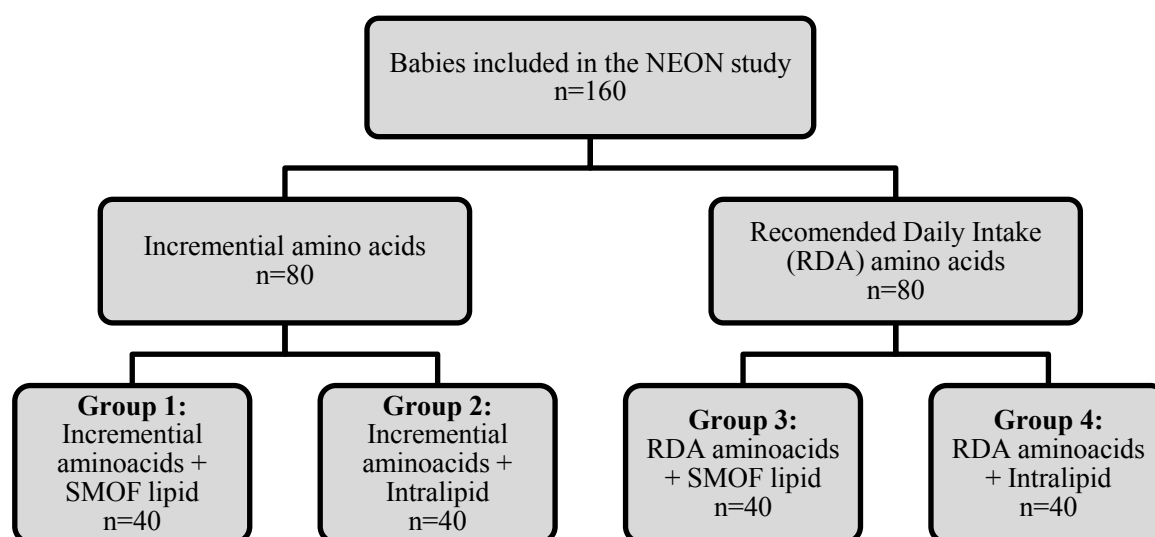


Figure 5: Study flowchart representing the four randomised groups.

The main aim of the NEON study is to demonstrate efficacy and confirm safety of introduction of SMOF lipid and immediate recommended daily intake concentration of amino acids to decrease non-adipose body mass deficit. Additionally the NEON study provides the opportunity to look at the effect of the three factors of interest (mode of delivery, parenteral feeding and type of milk) on development of gut microbiota in preterm infant's gut.

2.1.ii Sample collection

Stool and urine samples were collected on weekly basis since birth till discharge.

2.1.iii Sample analysis strategy

The analysis strategy of the biofluids obtained from the NEON study is presented in the Figure 6. The metabolic profiles of both stool and urine were designed to be obtained by the ^1H Nuclear Magnetic Resonance spectroscopy. The metataxonomic analyses were designed to be performed by a combination of bacterial identification by 16S rRNA gene sequencing.

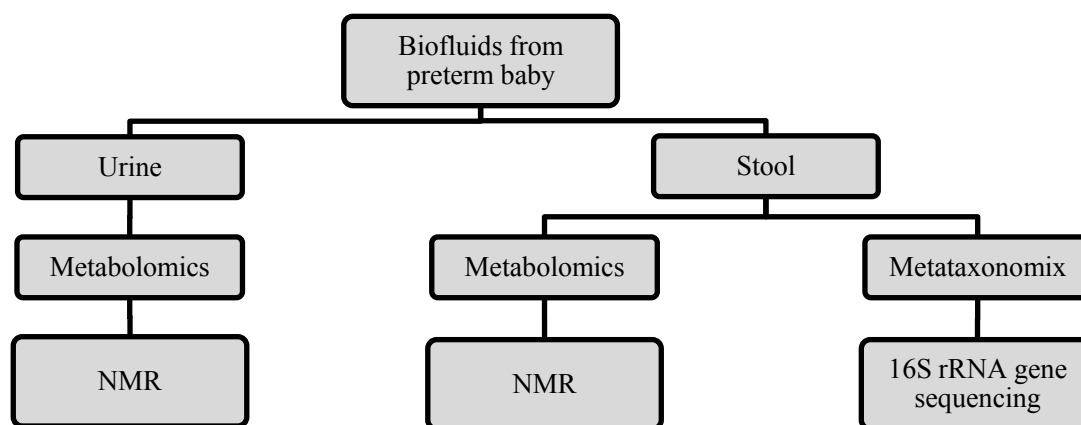


Figure 6: Flowchart presenting the sample analysis strategy

2.2 Metabonomics

Metabolomics is a branch of science that focuses on analysing the metabolic profile of biological samples. There are several techniques that can be used with the two most popular ones being Nuclear Magnetic Resonance (NMR) spectroscopy and Mass Spectrometry (MS). For the purposes of this study NMR was selected as the method of choice as it allows the global profiling, while MS is much more useful for the targeted analysis [191, 192].

2.2.i NMR spectroscopy

Nuclear Magnetic Resonance (NMR) spectroscopy is a powerful tool used for the identification of molecules and quantifying their relative concentration in the sample [193]. This method is especially good for untargeted global profiling of the samples as it is highly reproducible and does not destroy the sample [191, 192]. The high reproducibility of the NMR means that the comparison between the results obtained on different machines is reliable.

2.2.i.a Theory

Nuclear Magnetic Resonance spectroscopy is based on the principle that a nucleus has a magnetic momentum and an angular momentum (also known as spin) [194]. Many nuclei (those with an odd number of protons + neutrons) have a non-zero spin, which results in them creating a small magnetic field while spinning. When a magnetic field is applied, nuclei behave just like bar magnets in a magnetic field and they align themselves with or against the field, with a small energy difference existing between the several possible orientations (proton has two possible orientations) [192, 193]. In

order to excite nuclei in an external magnetic field, a 90° radio frequency pulse is applied. After the radio frequency is switched off, the nuclei relax and return to their preferred position in the applied magnetic field (Figure 7). This leads to a release of energy, which can be detected [191-194]. The exact resonance frequency produced when the nucleus returns to its natural position depends on the chemical and electromagnetic environment. For example, other nuclei and their electrons can shield nucleus from the effects of the magnetic field [191, 192]. These local differences in energy can be used to identify different chemical structures within a molecule. Typically the resonance frequency for each nucleus is transformed into parts per million (ppm) unit, with the use of the reference molecule – such as TSP (trimethylsilyl propionate) which is set at 0ppm [191, 192]. The NMR signals can provide further knowledge about the chemical structure and the absolute concentration of the molecule. The absolute concentration can be calculated from the area under the peak. When comparing the area to the control electromagnetic signal introduced in the run, it is possible to estimate the number of protons in the particular fragment of the molecule. If the chemical structure is known, then the total number of protons calculated from the area under the peaks belonging to the molecule is divided by the number of protons (^1H) in the chemical structure of the molecule. The structure is analysed based on the peak shape. Nuclei which are close to each other interact with what is called J-coupling. In the NMR spectroscopy, this interaction leads to J-splitting of the signal resulting in multiplet peaks. Apart from the shape of the peaks, an important information in the metabolite identification is the previously mentioned chemical shift given in parts per million (ppm). Each chemical compound has its own unique NMR fingerprint. For a detailed explanation of NMR properties refer to “Basic One- and Two-Dimensional NMR Spectroscopy” by Friebolin H. [191, 192, 194-197].

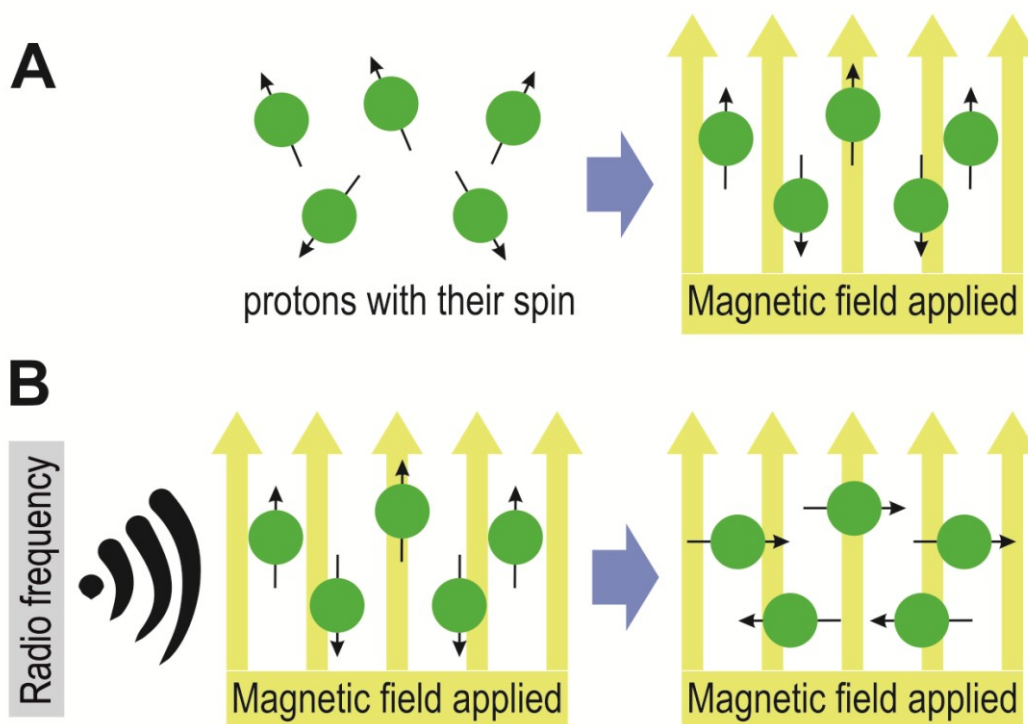


Figure 7: Principles of NMR. The segment A shows how the nuclei align themselves when the external magnetic field is applied. The segment B shows how the radio frequency ‘pushes’ the magnetic field of the nuclei.

In case of the biological samples the nuclei usually analysed are ^1H (protons) and ^{13}C as they are most common atoms in the organic molecules. However, occasionally other atoms are analysed – ^{15}N , ^{19}F , ^{23}Na and ^{31}P [191-193].

2.2.i.b One-dimensional NMR spectroscopy

The one dimensional (1D) ^1H -NMR spectra were acquired on a Bruker DRX-600 spectrometer (Bruker Biospin, Karlsruhe, Germany) using the water presaturation pulse sequence [relaxation delay- 90° - t_1 - 90° - t_m - 90° -acquire free induction decay (FID)]. The relaxation delay was 2 seconds; with 90° relating to the 90° radio frequency pulse applied, t_1 , referring to the interpulse delay, which was set to $3\mu\text{s}$, while t_m is the mixing time of 100ms. The probe was operating at 600.29 MHz at a temperature of 300 K and was matched and tuned automatically to the proton transmitter resonance frequency at the beginning of the run. The water signal has reduced its intensity by the presaturation technique. This is done with the use of long low power radio frequency pulse in order to selectively saturate the frequency at which the water is found during the relaxation delay. This is also referred to as the O1 parameter. Sixty-four scans were collected for each sample, composing of 64 thousand data points,

giving a total spectral width over 20 ppm (parts per million). The acquisition time was 2.73s. The signal to noise ratio was improved by applying an exponential weight function corresponding to a line broadening of 0.3Hz to each FID (free induction delay) prior to Fourier transformation. The parameters used for the standard pulse sequence of the faecal water sample were: P1=11.47 μ s, P11=12.50-dBW, P19=46.96dB and O1=2830Hz. For urine these were: P1=10.54 μ s, P11=11.10-dBW, P19=48.44dB and O1=2820.74Hz.

2.2.i.c Two-dimensional NMR spectroscopy

Two-dimensional NMR spectroscopy can help with identification of compounds in more complex samples [194, 195, 198]. The J-resolved (JRES) spectroscopy unravels the hidden peaks and releases information on peak structure for split peaks (e.g.: allows identification of whether a split peak is a doublet or two singlets) [195, 196]. Another type of two-dimensional pulse sequence is correlation spectroscopy (COSY), which gives information about the correlation/coupling of the neighbouring nuclei [197, 198]. TOSCY is total correlation spectroscopy and it shows couplings between the nuclei from the same molecule but which are further apart, typically up to 6 bonds distance [194, 199]. HSQC also known as heteronuclear single-quantum correlation spectroscopy, as the name suggests, shows the correlations between two neighbouring heterogenous atoms (e.g. ^1H and ^{13}C). In order to analyse the heterogenous atoms which are further apart there is a heteronuclear multiple-bond correlation spectroscopy (HMBC) [194, 197, 199].

2.2.ii Sample preparation for NMR analysis

The majority of samples were collected prior to the PhD as part of the NEON study and sample collection finished in the 19th month of the PhD. Samples were frozen straight after collection at -20°C and transferred within few days to a -80°C freezer. In order to avoid introducing batch effects in spectroscopic acquisition, samples were randomized by randomly assigning a number from 1 to 'n' where 'n' was the number of obtained containers with samples. This labeling was performed on dry ice to avoid any freeze-drying cycles.

2.2.ii.a Urine

Out of 1117 urine samples collected for the NEON study, 33 containers were empty or had volume of less than the required 540µl. The total working set therefore included 1086 samples, which were randomised and aliquoted for NMR analysis. An aliquot of 540µl of urine sample was mixed with 60µl of Phosphate buffer (containing 0.01% sodium 3-(trimethylsilyl)propionate-2,2,3,3-d₄ – used as a chemical shift reference) and transferred to a 5mm (outer diameter) NMR tube. After sealing the tubes with caps, they were inverted several times in order to mix the sample with buffer. Samples were left for few hours in 4°C to equilibrate. A quality control sample was prepared from pooled urine samples.

High concentration of antibiotics and other drugs, skewed baseline and line broadening across the spectrum were found in 44 samples, which were subsequently excluded from the analysis, leaving 1042 NEON urine samples and 110 quality controls (QCs). Mislabelling or a mismatch of available metadata led to exclusion of 129 samples and the remaining 913 samples were modelled using multivariate and univariate statistical analysis.

2.2.ii.b Stool

Faecal water was extracted following the protocol explained in the Method Development chapter. Briefly: 100mg of stool sample was mixed with 250µl of solvent (25% Acetonitrile in UPLC grade water) and homogenised by bead beating (1mm zirconium beads for 10 seconds). After centrifugation at >16,000×g for 20 min, 540µl of faecal water were transferred to a glass vial which was left for an hour for evaporation of solvent. The dehydrated powder was resuspended in 540 µl of D₂O (used as a field frequency lock for the spectrometer) and mixed with 60µl of Phosphate buffer (containing 0.01% sodium 3-(trimethylsilyl)propionate-2,2,3,3-d₄ – used as a chemical shift reference) and transferred to a 5mm (outer diameter) NMR tube. Tubes were closed with caps and inverted several times in order to mix the sample with buffer. Samples were left for few hours in 4°C to equilibrate.

Although 1187 stool samples were collected for the NEON study, 31 of the containers were empty. Sixty-nine duplicates were created to check the repeatability of the extraction method. A quality

control sample was prepared from pooled faecal water samples that were extracted from a random subset of leftover NEON stools. The working set of NEON stool samples equaled 1225. Sixty samples did not produce enough faecal water due to high concentration of lipids (which accounted for up to 90% of mass of the sample). Additionally 141 samples did not meet the criteria for shim and had to be removed from the modeling which left a total of 1024 samples for modeling.

2.2.iii Pre-processing of ^1H -NMR spectra

The spectra were pre-processed in the TopSpin 2.1 software where the baseline was corrected and the TSP peak was calibrated to 0.00ppm. Further pre-processing continued in the Matlab software with the use of in house created scripts. Spectra were aligned, normalised (probabilistic normalisation), scaled (uni-variance) and modelled using the quality controls (pooled sample of all samples in any given study). Due to the dynamic range of the molecules visualised within a typical biofluid spectrum, scaling to unit variance was used to allow the lower concentration metabolites to receive equal weighting to the high concentration metabolites such as creatinine, which would otherwise dominate the analysis without necessarily being biologically relevant.

2.2.iv Multivariate statistical analysis

Multivariate statistical analysis is a method of choice for the analysis of datasets rich in information, such as metabolic profiles, particularly where there are more variables than samples. This statistical approach helps to extract the latent variables and reduce the dimensionality of the dataset and facilitates the detection of which regions of spectra are discriminant, thus are significantly differ between the two analysed groups [200]. There are two broad approaches to multivariate statistical analysis: unsupervised and supervised. It is advisable to create both of the models as each of them carries different information which when combined gives a larger image of the metabolic changes between the analysed groups [200].

2.2.iv.a Unsupervised multivariate statistical analysis

Principal Component Analysis (PCA) is the unsupervised model used in this thesis for multivariate statistical analysis. The model has no knowledge regarding the classification of samples and performs

distribution of the samples only based on the NMR spectra and hence chemical similarities / dissimilarities between samples [200, 201]. Principal components are a set of values which were created by orthogonally transforming the observations of possibly correlated variables [200, 201]. There can be as many PCs as there are variables in the dataset but they contain progressively less information such that the last PC in a model will contain only noise and no useful information. Thus PCA is a dimension reduction technique that can show the information in the form of scores and loadings. Scores show the separation of samples based on their variability, while loadings indicates which characteristics causes the variability. Additionally, the model informs about the natural variation of the sample set by providing the R^2 for each PC (in a form of percentage of variation explained by this feature) [200, 201]. If PC1 has low R^2 this means that the samples present a very complex variation, a situation, which is very common in the human metabolic analysis. In this thesis, PCA models had been created in Matlab software using in house scripts.

2.2.iv.b Supervised multivariate statistical analysis

After performing the unsupervised analysis (the PCA models) the next natural step is to compare two or more groups of interest using a supervised modelling technique where information as to the class of each sample is used to maximize the distance between two or more classes in the data [200, 201]. Orthogonal Partial Least Squares Analysis (O-PLS) compares two (or more) groups of samples according to their class and reports any differences in the loadings plot. The regions of spectra are coloured according to the discriminance. The O-PLS model provides very important information in the form of R^2Y and Q^2Y . Quoting Triba *et al.* “ R^2Y measures the goodness of fit while Q^2Y measures the predictive ability of the model.” [202]. This means that if the R^2Y equals 1 it would indicate that model perfectly describes the data and if the Q^2Y equals 1 then it means the model is 100% predictive [200-202]. In order to test the robustness of the model, permutation analyses were performed. The model was repeated automatically 100 times against random permutations and the number of times the Q^2Y value for the permutations analysis was larger than the Q^2Y value for the true unpermuted model was considered the p -value. In order for the model to be considered robust the p -value had to be ≥ 0.05 which corresponds to 5%.

2.2.v Metabolite identification

There are two main methods to identify metabolites. First method is the visual inspection of the spectra and referencing them to any 2D datasets. Statistical methods can also be used to identify correlated information in the NMR spectra and hence to extract structural information. After identifying a discriminant region of the spectrum, a Matlab script called STORM was used to search for the correlation between the chosen region of the spectrum and the correlated parts. Because NMR is quantitative all the protons within a molecule should covary. STORM (subset optimization by reference matching) identifies which peaks are increasing in concentration with the peak of interest (Figure 8) [203]. Obtaining such information can provide the NMR fingerprint of a discriminant molecule. Metabolic databases are then searched to find a matching compound. However, it is not always possible to identify the molecule from the database. In case of uncertainty, the standards (pure solutions of a compound) are spiked into a sample to see how the metabolic profile changes. Additionally, the quality control samples had been given for the MS analysis in order to identify as many low concentration molecules as possible and try to match the findings from MS with the unknown metabolites.

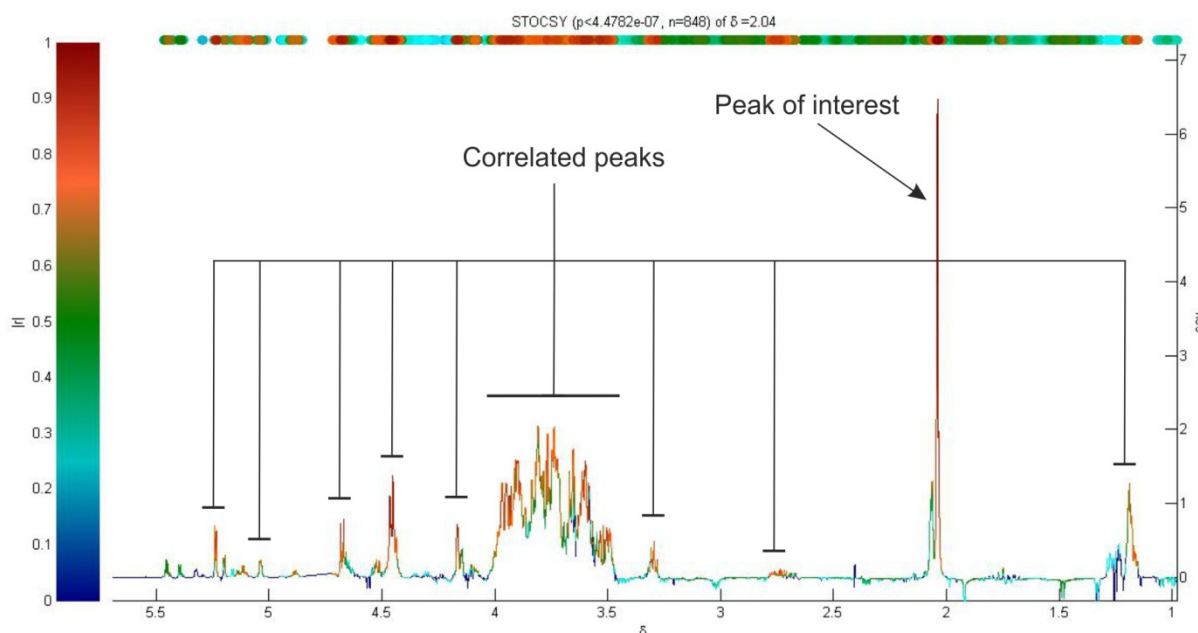


Figure 8: The result of the STORM analysis of the unknown peak of the chemical shift of 2.04ppm. The colour of the peaks indicates the level of correlation while the direction if the correlation is positive or negative. The above spectrum shows the ^1H NMR fingerprint of lactose.

2.3 Gut microbiota profiling

2.3.i Theory

Due to the fact that 99% of the known bacterial species are uncultivable it was necessary to find a new culture-independent technique to identify microbial profiles [204-209]. The potential biomarker used for the identification must have been present in all the bacteria yet present variability between the species [210]. The molecule perfect for the task seemed to be 16S rRNA gene.

2.3.i.a What is 16S rRNA

The 16S rRNA gene is a DNA molecule that codes the 16S rRNA, which is a part of small subunit (30S) of the prokaryotic ribosome [210, 211] (Figure 9). It is a molecule homologous to the eukaryotic 18S rRNA. Due to the invention of the polymerase chain reaction (PCR) and sequencing, it was possible to multiply and the sequence the 16S rRNA genes. The structure of the 16S rRNA is extremely useful as the conserved regions (identical among all the bacteria) are mixed with the variable regions carrying species-specific mutations (Figure 10).

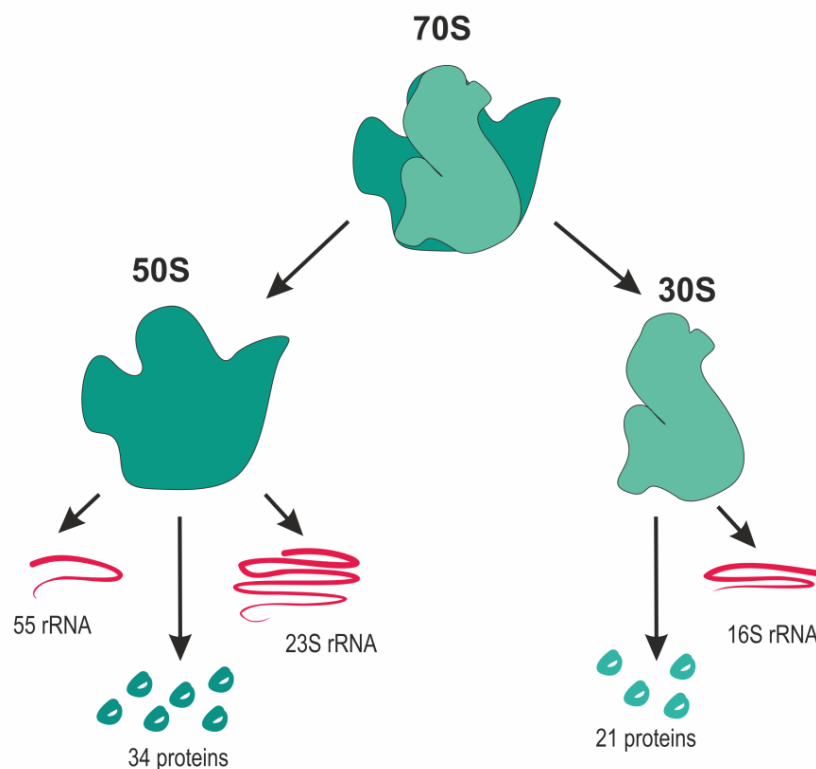


Figure 9: The simplified structure of the prokaryotic (70S) ribosome

Work done on the analysis of the mutations among the variable regions of the 16S rRNA gene led to the creation of the first phylogenetic tree [211]. This was later spread on other kingdoms like Archea and Eukaryota as analysis of the eukaryotic analogue of the 16S rRNA gene – the 18S rRNA gene, allowed the evolutionary assessment of the living organism [211]. The usage of the 16S rRNA gene in the microbial identification had found its application not only in the evolutionary research but also in environmental and clinical microbiology [210, 211].

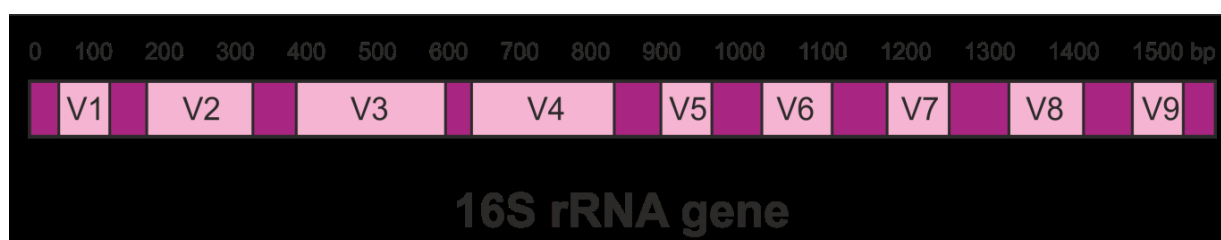


Figure 10: The simplified structure of the 16S rRNA gene with conserved regions dark shaded, and the variable regions light shaded

2.3.i.b Comparison of the analytical techniques

The discovery of the 16S rRNA gene created a shift in the microbial diagnostics from culture-based to culture-independent [210]. There are many techniques available, including: terminal restriction enzyme fragment polymorphism (T-RFLP), degradation gradient gel electrophoresis (DGGE), temporal temperature gradient gel electrophoresis (TGGE), Sanger's sequencing, pyrosequencing (a.k.a. 454 Roche sequencing platform), and MiSeq (Illumina platform). The detailed comparison of those techniques is given in the Table 3.

2.3.ii DNA extraction from stool samples

The stool samples had been chosen for the DNA extraction provided that there had been available:

1. NMR spectrum for the same sample
2. NMR spectrum for the matching urine sample (collected on the same day)
3. At least two more samples (meeting the above criteria) for the same baby
4. At least one sample (meeting the above criteria) for the same baby that was collected within the first 10 days of life

The DNA was extracted from the remaining stool pellets (after the extraction and removal of the faecal water) using the PowerSoil DNA extraction kit (MoBio Laboratories) according to standard protocol [212].

2.3.iii Quantitative Polymerase Chain Reaction (qPCR) analysis

The qPCR was performed to obtain the total bacterial count. Universal set of primers for the variable region V3 was chosen (515F, 806R). The primers and PCR grade water were purchased from MoBio Laboratories, while Maxima SYBR Green/ROX qPCR Master Mix was bought from Life Technologies. The reagent recipe for the master mix is given in the Table 2.

Table 2: The complete reagent recipe (master mix) for 1X PCR reaction

PCR Grade H2O	10.5 µL
Forward primer (10µM)	0.5 µL
Reverse primer (10µM)	0.5 µL
Template DNA	1 µL
Maxima SYBR green master mix	12.5 µL
Total	25 µL

The qPCR was performed in the 96 well plate thermocycler. The conditions were as follow:

1. 94°C 3 minutes
2. 94°C 45 seconds
3. 50°C 60 seconds
4. 72°C 90 seconds

Steps 2-4 repeated 35 times

5. 72°C 10 minutes
6. 4°C infinity

The calibration curve was done using five 10 fold dilutions of *E. coli* DNA at a concentration of 500µg/mL in triplicates. Additionally, each run had a positive and a negative control.

Development of gut microbiota and metabolic changes in preterm babies

Table 3: Comparison of the analytical tool for the bacterial identification

Analytical method	Culturing	Sequence-specific separation tools		Sequencing		
		T-RFLP	DGGE/TGGE	Sanger's sequencing	Pyrosequencing	MiSeq sequencing
Mechanism	Plating samples on various selective media, microscopic observations and biochemical tests	PCR products are digested with restriction enzymes to produce labelled terminal restriction enzyme fragments (TRFs). T-RFs are denatured, and the obtained single stranded DNA is separated by electrophoresis under denaturing conditions.	Molecules are separated in either: a) the gradient of the denaturant gel b) in constant concentration of denaturant in gel and in increasing temperature over the time	Single-stranded DNA is placed in a mixture. Fluorescently-labelled nucleotides are added (one by one), if they are incorporated into the daughter DNA strand – the fluorescence occurs.	Single-stranded DNA is attached to beads and then added to the matrix (each bead goes into a hole in the plate). Nucleotides are added separately and if they get incorporated into the daughter DNA strand – the luminescence occurs.	Single-stranded DNA molecule is attached to the matrix and amplified which results in a cluster of DNA molecules. Fluorescently-labelled nucleotides are added (one by one), if they are incorporated into the daughter DNA strand – the fluorescence occurs.
Identification made upon	Bacterial colonies	16S rRNA gene		16S rRNA gene		
Result	Colonies	Bands on gel electrophoresis = fingerprint		DNA sequence		
Strengths	<ul style="list-style-type: none"> Cheap Easy Little preparation 	<ul style="list-style-type: none"> Quick Can be automated Sensitive (LLD = 1%) Labelling with a different fluorochrome, allows accurate sizing of the fragments to within ± 1 nucleotide. 	<ul style="list-style-type: none"> Cheap Sensitive even to single mutation 64 samples per run Run lasts 16h Chemical denaturing gradient not needed 	<ul style="list-style-type: none"> The most sensitive Sequences can be classified according to taxonomy and function Results from different studies can be compared 		
				<ul style="list-style-type: none"> First ever sequencing technique Up to 800bp 	<ul style="list-style-type: none"> Cheaper than regular sequencing Up to 700bp Analysis time 10h No clone libraries required 	<ul style="list-style-type: none"> Cheaper than pyrosequencing Up to 600bp Analysis time – 4-16h per sample
Weaknesses	<ul style="list-style-type: none"> 50-90% of gut microbiota are not cultivable Growing bacteria takes days 	<ul style="list-style-type: none"> Microbes of different species often share one T-RF Digestion with up to three different restriction enzymes is usually necessary 	<ul style="list-style-type: none"> Difficulties with gradient creation Specially synthesized GC clamped primers needed 	<ul style="list-style-type: none"> Expensive One DNA strand at a time Requires clone libraries – risk for bias 	<ul style="list-style-type: none"> More expansive than separation methods Requires DNA libraries Need of using barcode reading to avoid multiplex results (specific primers or labelling the correct sentence) 	<ul style="list-style-type: none"> Need of using barcode reading to avoid multiplex results (specific primers or labelling the correct sentence)
References	[213-218]	[219-224]	[204-207, 221, 225-230]	[208, 221, 231-237]	[208, 209, 215, 221, 225, 232, 233, 235, 238-244]	[213, 221, 235, 238, 241, 243-253]

2.3.iv Pre-processing of the sequenced data

The sequencing was performed using the MiSeq (Illumina) platform by the Research and Testing Laboratory (USA) [245, 249]. Since the results from the sequencing were provided as two reads per each sequence (two halves with an overlapping region), they had to be combined into one structure. In order to do so, a program Mothur was used and a standard script was followed [254]. The cut-off of 2000 reads was used when processing the samples, which means that all the samples that had less than 2000 reads had been removed. Keeping those samples in would lead to big complications later on when the Mothur calculates the number of OTUs (operational taxonomic units a.k.a. bacterial species) as this is done by keeping only the OTUs which are in the 25th or above percentile. Keeping large numbers of samples with low number of sequences would reduce the percentile for all the OTUs and thus reduce how detailed the outcome could be [245, 249, 254]. In order to match the species from the company's database with the 16S rRNA gene sequences a command that calculates the uncorrected pairwise distances between the aligned DNA sequences was used, followed by the Bayesian classifier. The outcome files from the Mothur program were merged to create one final matrix of taxonomic information and number of reads in each sample.

2.3.v Analysing the microbial profiles

After obtaining the files from the Mothur and preparing the metadata file, the analysis of the microbial profiles was done using the STAMP software [255]. In order to avoid obtaining false results, a plan for statistical analysis had been prepared in advance (Figure 11). Since the data was not normally distributed the non-parametric tests were used. For comparison of two groups the two-sided (two-tailed) White's non-parametric t-test was used as this was the available non-parametric test [256]. There were two parametric tests available as T-test was designed for situation when both groups had similar variation (or standard deviation (SD) squared) while Welch's t-test was for analysis when the variations of groups were not similar [257]. Due to the fact that the dataset contains information about all the identified bacteria, the statistical tests in fact perform the multiple tests comparing all the bacteria types. In order to avoid the random false positive results, the Benjamini-Hochberg correction was applied [258].

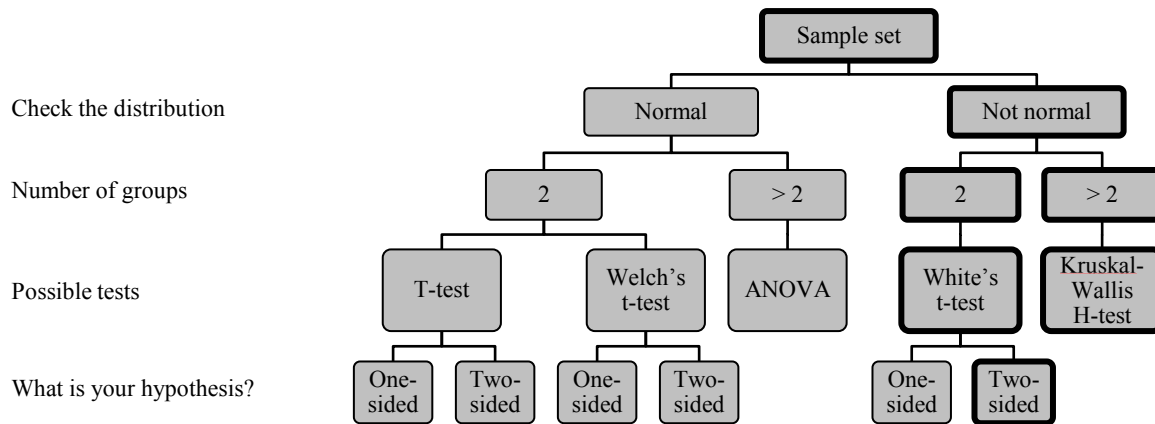


Figure 11: Flowchart presenting the statistical analysis options. The approach used for our analysis is presented in the thick-outlined boxes

For the analysis of > 2 groups the available non-parametric test was Kruskal-Wallis H-test [259]. Due to the fact that the Post-hoc analysis can produce type 1 error and give false positive results it was decided not to look at the Post-hoc results. Instead the statistical approach presented in the Figure 12 was followed.

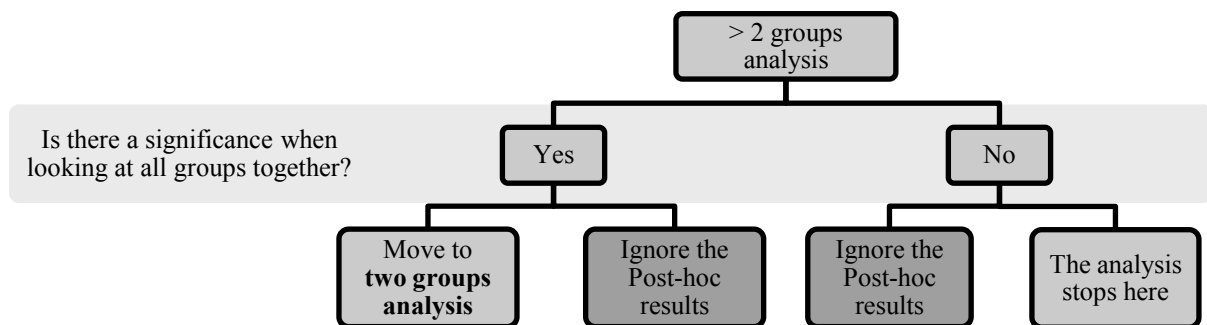


Figure 12: Flowchart presenting the statistical approach when working with comparing > 2 groups

When significant differences were found between the 3 groups it was investigated further which groups were significantly different from each other and this was pre planned to compare with the gold standard. For route of nutrition it was exclusive enteral nutrition (EN) and for type of milk it was exclusive breast milk (BM). The statistical approach to analyse the multiple groups that represent the continuous variable (like postnatal age or gestational age) is to do test for trend. Unfortunately this software did not offer such option therefore the best available option was chosen – description of any potential patterns of changes. The richness of gut microbiota had been calculated using the Chao

index while the diversity was calculated with the Shannon's index with the highest values corresponding to the highest richness/diversity [260].

2.4 Regression of metabolites against gut microbiota

In order to investigate the metabolic activity of gut microbiota, the taxa of interest were be regressed against the NMR spectra. Two approaches were taken. The first was the O-PLS regression of NMR spectra according to 'classes' which in fact will be the presence or lack of particular group of bacteria.

The second type of analysis was done using an algorithm in the R platform created by my colleague Daniel Homola. Features were selected using the Boruta feature selection method. Then spearman rank correlation and Benjamini-Hochberg correction for multiple testing were applied. The results were shown as heat maps of the regions of NMR spectrum (in ppm) and bacteria.

Chapter 3 - Method development

3.1 Introduction

Faecal water extraction protocols have been optimised for extraction from adult stool samples or full-term newborn's stools [261-264]. The preterm stool has a different composition than the one from full-term infant therefore the extraction method requires further development. Apart from the fact that the gut of preterm infants is not fully developed, these infants often receive parenteral nutrition instead of oral which can lead to lower bacterial yield in the intestine. Additionally, newborn stool often has higher levels of lipids than adult stool which results in accumulation of a lipid layer above the faecal water during the extraction protocol. A new protocol was developed to work with preterm infant's stool samples.

Other important consideration for method development is that preterm stool samples usually have small volume, which means that it is not possible to perform faecal water extraction and bacterial DNA extraction separately. Thus the new protocol was prepared in a manner that allowed using the same sample fraction for both NMR/MS analysis of faecal water and DNA extraction from stool microbiota. Each stage of faecal water extraction was tested for its influence on gut microbial DNA yield.

3.2 Faecal water extraction

In order to develop a method for faecal water extraction a number of factors had to be validated. This included: 1) solvent, 2) mass of the stool sample, 3) type of homogenisation, 4) solvent evaporation and 5) filtration. The final step is the lipid layer preparation for lipidomics.

3.2.i Solvents

A number of solvents were proposed for extraction of faecal water for NMR analysis. This included deionised water 18M Ω (dH₂O), 0.01M phosphate buffered saline (PBS), heavy water (D₂O) and phosphate buffer (2.84% w/v Na₂HPO₄, 0.48% w/v NaH₂PO₄, 0.02% w/v NaN₃, 0.017% w/v TSP (3-

(trimethyl-silyl)propionic acid-d₄), in D₂O). The first two solvents are commonly used for faecal water extraction while the latter two are commonly used for samples analysed by nuclear magnetic resonance (NMR) spectroscopy. Water was deionised (18M Ω) while D₂O, tablets for PBS preparation and all substrates for phosphate buffer preparation were sourced from Sigma.

A series of experiments comparing the previously mentioned solvents was performed. All of the solvents were not compared in the same experiment but in a series of experiments thus the results for each comparison will be presented separately with a final conclusion.

Stool sample ranging from 200mg to 500mg (depending on experiment) were mixed with twice the volume (ratio 1g of stool : 2ml of solvent) of either dH₂O (18M Ω), PBS (0.01M), D₂O or phosphate buffer (2.84% w/v Na₂HPO₄, 0.48% w/v NaH₂PO₄, 0.02% w/v NaN₃, 0.017% w/v TSP, in D₂O). Samples were homogenised either by vortexing for 2 min or by bead beating twice for 5 sec on Precellys 24-duo bead beater. Then samples were centrifuged for 20 min at >13,000 $\times$ g. Supernatant was transferred to new microcentrifuge tube, topped up with D₂O or phosphate buffer (depending on solvent) to a total volume of 550 μ l.

Standard one-dimensional ¹H NMR spectra were acquired on Bruker DRX 600 MHz spectrometer (Rheinstetten, Germany) with a 5mm TXI probe operating at 600.13 MHz. A relaxation delay-90°-t1-90°-tm-90°-acquire free induction decay (FID) sequence was used to obtain 1-dimensional NMR spectra. The 90° pulse length varied from 9.75 μ s to 11.25 μ s depending on sample type and condition. The total number of scans varied from 64 to 128 was saved as 32k data points with a spectral width of 20 ppm. The same settings (apart from the pulse length and the number of scans) were used for all other NMR analysis.

The first experiment comparing extraction of stool of full-term newborn found no significant difference between the PBS and D₂O solvents. Figure 13 shows that none of the peaks corresponding to recovered metabolites was lost and the general ratio of peaks was unaltered (with small differences for few peaks). An interesting observation was that when extracted in PBS, the lipid layer was much smaller and the faecal water was less clear than when extracted in D₂O.

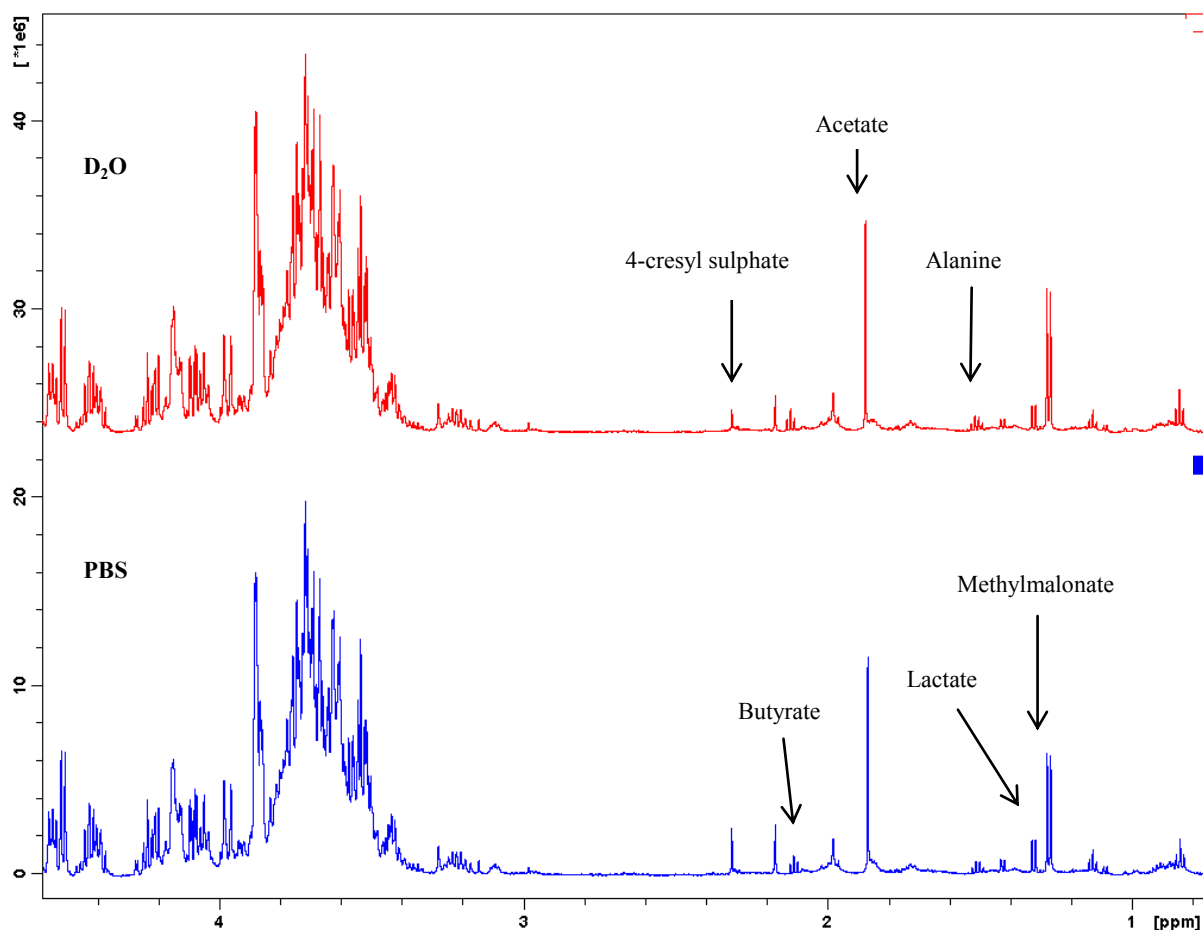


Figure 13: Comparison of CPMG one-dimensional ^1H NMR spectra of faecal water extracted in D_2O or PBS

The second experiment compared extraction of faecal water from fresh adult stool sample with D_2O , PBS or phosphate buffer. Results show that PBS was more effective in metabolite retrieval than D_2O (Figure 14). What is more, samples extracted in phosphate buffer were more difficult to shim than samples extracted in the other solvents indicating that there was a chemical interaction that occurred in phosphate buffer. Although the number of peaks or metabolites retrieved from each solvent was comparable, the ratio of peaks varied according to solvent.

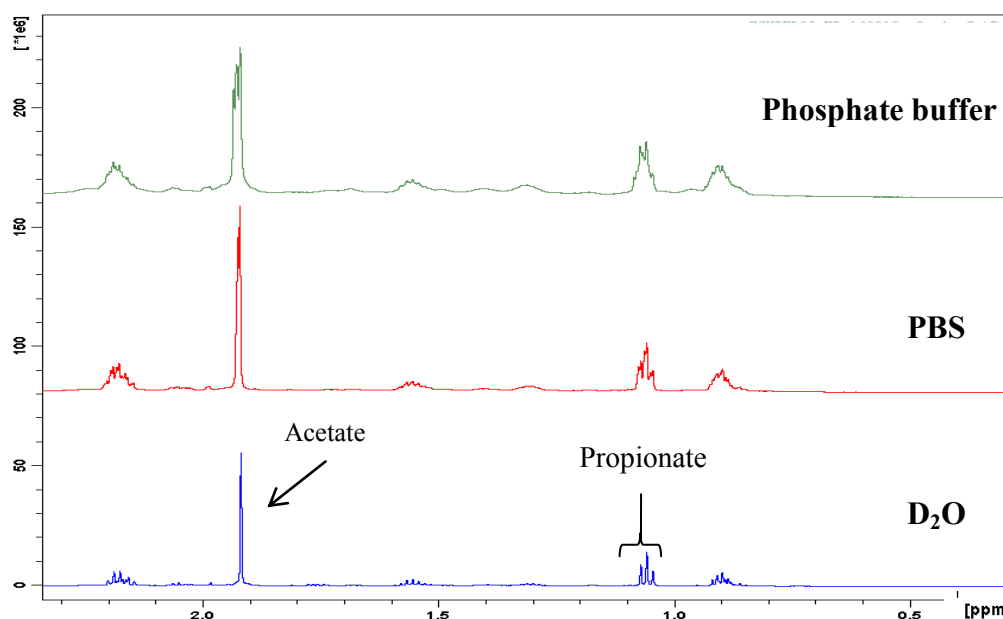


Figure 14: Comparison of standard one-dimensional ^1H NMR spectra of faecal water extracted in D_2O , PBS or phosphate buffer. Spectra expanded from 0.5 ppm to 2.3 ppm

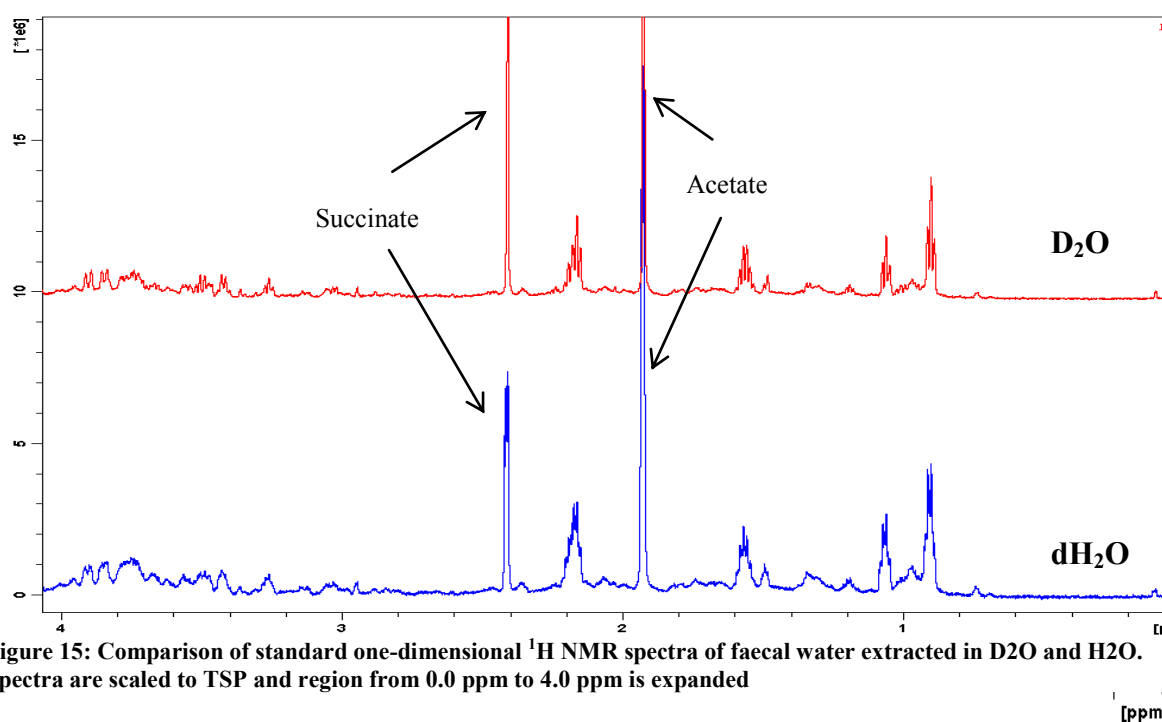


Figure 15: Comparison of standard one-dimensional ^1H NMR spectra of faecal water extracted in D_2O and H_2O . Spectra are scaled to TSP and region from 0.0 ppm to 4.0 ppm is expanded

The final experiment comparing solvents for faecal water extraction was performed on a stool sample of a one year-old infant. Comparison was made between dH_2O and D_2O . Results presented in Figure 15 show that the biggest difference was the shimming quality. Due to better shimming two peaks 2.41ppm (succinate) and 1.92 ppm (acetate) were higher when extracted in D_2O .

Due to the fact, that bile acid profiling by mass spectrometry (MS) was proposed as a part of this study, a new extraction method that would improve bile acid retrieval had to be created. It was proposed to use 25% acetonitrile solution in dH₂O (18MΩ) to extract faecal water. The faecal water was planned to be separated into two fractions of which one would be used for MS and the other one for NMR. The fraction for NMR would be speed vacuumed and the pellet would be resuspended in phosphate buffer. It was decided to check how extraction in 25% acetonitrile would affect NMR spectra. Extraction in dH₂O (18MΩ) was used as a control.

Three stool samples were divided into four fractions each. Two fractions were extracted in 25% acetonitrile while the other two in dH₂O (18MΩ). Samples were bead beaten with 1mm zirconium beads and centrifuged for 20 min at 16,000×g. The faecal water was transferred to new microcentrifuge tube and samples were speed vacuumed for 3h at 45°C. The powder was resuspended in 400µl of D₂O and 200µl of phosphate buffer was added. Samples were then vortexed for 1 min, sonicated for 5 min and vortexed again for 1 min. After that samples were centrifuged for 10 min at top speed at 4°C to precipitate any left proteins. Supernatant was transferred to NMR tubes and standard one-dimensional ¹H NMR spectra were acquired on Bruker DRX 600 MHz spectrometer (Rheinstetten, Germany) with a 5mm TXI probe operating at 600.13 MHz. A relaxation delay-90°-t1-90°-tm-90°-acquire free induction decay (FID) sequence was used to obtain 1-dimensional NMR spectra. The 90° pulse length varied from 15.00 µs to 20.00 µs depending on sample type. The total number of scans equalled 256 and was saved as 32k data points with a spectral width of 20 ppm.

The results showed that there was no difference between extraction in 25% acetonitrile and in dH₂O (18MΩ) (Figure 16). This means that 25% acetonitrile in dH₂O can be used to retrieve metabolites from stool for later NMR analysis (provided that acetonitrile will be evaporated and sample will be resuspended in phosphate buffer).

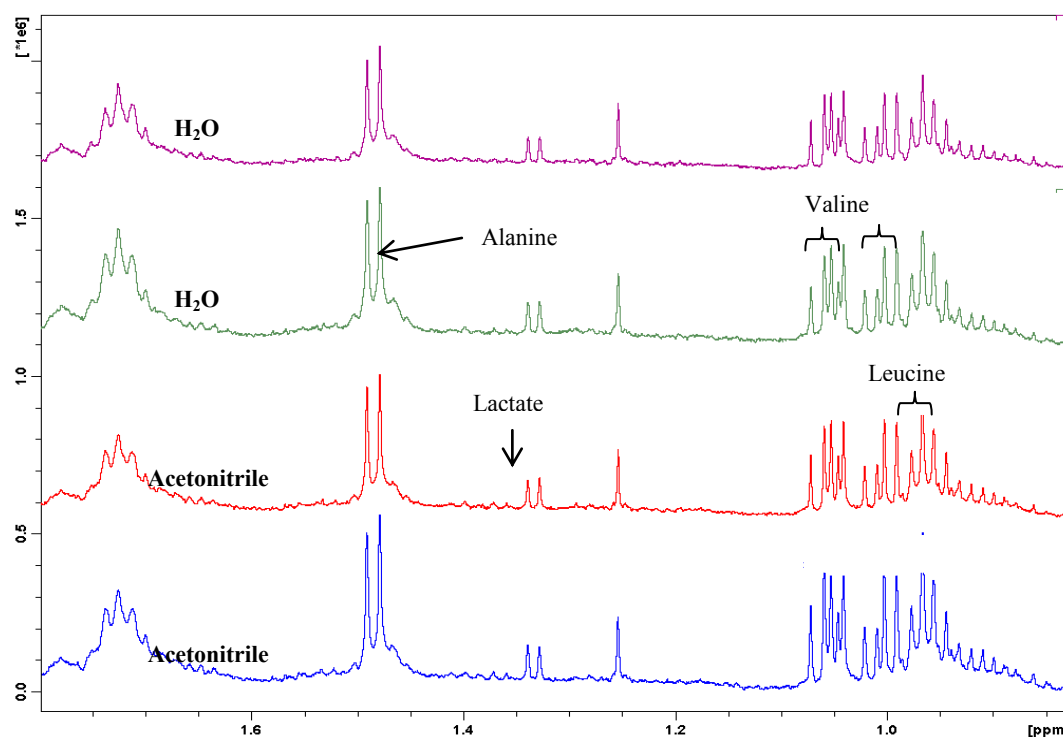


Figure 16: Comparison of standard one-dimensional ^1H NMR spectra of faecal water extracted by two different modes of bead beating (presented in duplicates). Figure showing an expanded fragment from 0.8ppm to 1.8ppm.

3.2.ii Stool sample mass

When using a centrifuge of speed below $16,000\times g$ the mass of stool sample used for faecal water extraction occurred to be an important issue. It was observed that in some samples there was ‘a wave effect’ (stool was shaped as wave) and often the top of the ‘wave’ collapsed into the sample trapping some fractions of faecal water (Figure 17)



Figure 17: Shape of stool pellet after extracting faecal water from 500mg of faeces

An experiment to compare optimal mass of stool was performed. Experiment was performed in duplicates and sample mass equalled 0.1g, 0.25g and 0.52g. The volume of added solvent was 200 μl , 500 μl and 1ml respectively. Samples were vortexed for 2 min and centrifuged for 20 min at speed $36,000\times g$. Supernatant was collected and samples were again centrifuged this time for 10 min. Supernatant was again collected and samples again centrifuged for 10 min. After that, samples were centrifuged again, this time for 5 min only. Table 4 presents results for this experiment.

Table 4: Results showing the influence of mass of stool sample and the shape of pellet and volume of retrieved faecal water when centrifuged at $< 16,000\times g$

Sample ID	Mass used	Volume of solvent	Volume of obtained faecal water and shape of pellet				
			After initial 20 min	After next 10 min	After next 10 min	After next 5 min	Total (% of solvent)
1 & 2	0.11g	200 μ l	150 μ l 	50 μ l 	0 μ l 	–	200 μ l (100%)
3 & 4	0.25g	500 μ l	300 μ l 	200 μ l 	20 μ l 	5 μ l 	525 μ l (104%)
5 & 6	0.52g	1000 μ l	800 μ l 	200 μ l 	50 μ l 	7 μ l 	1057 μ l (105%)

It was obvious that more than 20 min of centrifuging is needed if using a centrifuge slower than $16,000\times g$. Additionally, increasing the time of centrifugation leads to retrieval not only of 100% of used volume of solvent but also of additional water from the stool sample.

The next experiment was set to analyse how many sets of centrifugation should be performed. Three fractions of the same stool were weighted (0.26g) and mixed by vortexing with 500 μ l of D₂O. The first fraction was centrifuged once for 60 min, the second fraction – once for 40 min and the third fraction was centrifuged two times, each time for 20 min with supernatant being collected after each centrifugation. Table 5 presents results for this experiment.

Table 5: Results showing the correlation of volume of retrieved faecal water and the time of centrifuging at the speed $13,000\times g$

Sample ID	Mass used	Volume of solvent	Mode of centrifuging	Volume of obtained faecal water	Pellet shape
1	0.26g	500 μ l	1 x 60 min	340 μ l	
2	0.26g	500 μ l	1 x 40 min	300 μ l	
3	0.25g	500 μ l	2 x 20 min	450 μ l	 & 

Conclusion from those two experiments was that if a centrifuge of speed below $16,000\times g$ will be used, it is necessary to make sure stool sample is not heavier than 0.25g and that sample is centrifuged twice for 20 min. This will provide the highest retrieval of faecal water.

Due to these observations centrifuge with speed above $16,000\times g$ was obtained. This meant that time of centrifuging could be adjusted to 20 min.

3.2.iii Type of homogenisation

In order to extract the highest possible concentration of metabolites the most effective homogenisation method must be chosen. At the same time it was important to make sure that this step of faecal water extraction will not harm the bacterial cells as this would lead to a drop in bacterial DNA yield.

Two types of homogenisation have been compared – bead beating and vortexing. Firstly, the experiment was performed on stool obtained from full-term infant. Stool was divided into four fractions each weighting 250mg (experiment was performed in duplicates). 500 μ l of solvent was added and samples were homogenised by either 1 min of vortexing at top speed or by bead beating twice for 5 sec with 5sec break between beats at lowest possible speed which equalled 5000 rpm (Precellys 24-duo bead beater). The size of beads was used based on chances of accidental rupture of bacterial cells. It was decided that 250mg of 1mm zirconium beads should not cause a significant change in bacterial DNA yield (this was later confirmed and is explained in section 2.1.ii.b). Samples were centrifuged for 20 min at $16,000\times g$ and supernatant was transferred to NMR tube.

CPMG and standard one-dimensional ^1H NMR spectra were acquired on Bruker DRX 600 MHz spectrometer (Rheinstetten, Germany) with a 5mm TXI probe operating at 600.13 MHz. A relaxation delay- 90° - t_1 - 90° - t_m - 90° -acquire free induction decay (FID) sequence was used to obtain one-dimensional NMR spectra. The 90° pulse length varied from 9.75 μ s to 11.25 μ s depending on sample type and condition. The total number equalled 64 and was saved as 32k data points with a spectral width of 20 ppm. The same settings (apart from the pulse length and the number of scans) were used for all other NMR analysis. The results presented in Figure 18 show that both methods present similar information recovery. There has been visible difference in intensity of peaks.

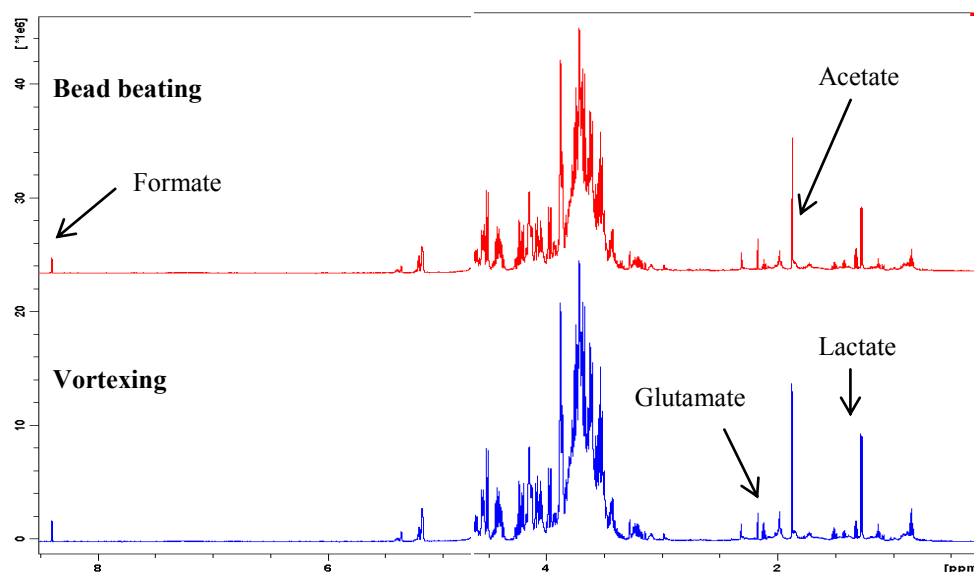


Figure 18: Comparison of CPMG one-dimensional ^1H NMR spectra of faecal water extracted by bead beating and vortexing

Due to the fact that bead beating is a high throughput technique it was selected for homogenisation. Another aspect of validation was to decide on time and speed of bead beating. The previously described bead beating was compared with the one from a different protocol (5500 rpm) bead beaten three times 40 sec with 5 sec breaks). Three preterm stool samples were divided into four fractions each (experiment run in duplicates). Approximately 100mg of stool was mixed with 200 μl of H_2O (0.02% NaN_3 w/v) and 250mg of 1mm zirconium beads and bead beaten. Samples were centrifuged for 20 min at 16,000 $\times g$, frozen and then they followed different protocol. After defrosting, 500 μl of H_2O was added and samples were centrifuged for 20 min at 16,000 $\times g$. 400 μl of supernatant was transferred to a new microcentrifuge tube with 200 μl of phosphate buffer (2.84% w/v Na_2HPO_4 , 0.48% w/v NaH_2PO_4 , 0.02% w/v NaN_3 , 0.017% w/v TSP, in D_2O) and vortexed for 10 sec. 550 μl were transferred to an NMR tube. The 90° pulse length varied from 12.00 μs to 13.25 μs depending on sample type and condition. The total number of scans equalled 128 and was saved as 32k data points with a spectral width of 20 ppm.

There was no difference between the spectra of faecal water samples obtained by the different types of bead beating (Figure 19). The influence of bead beating on DNA yield is described in Method Development chapter section 3.1.i.b. The choice of homogenisation method was made based on time and the shorter bead beating mode was chosen.

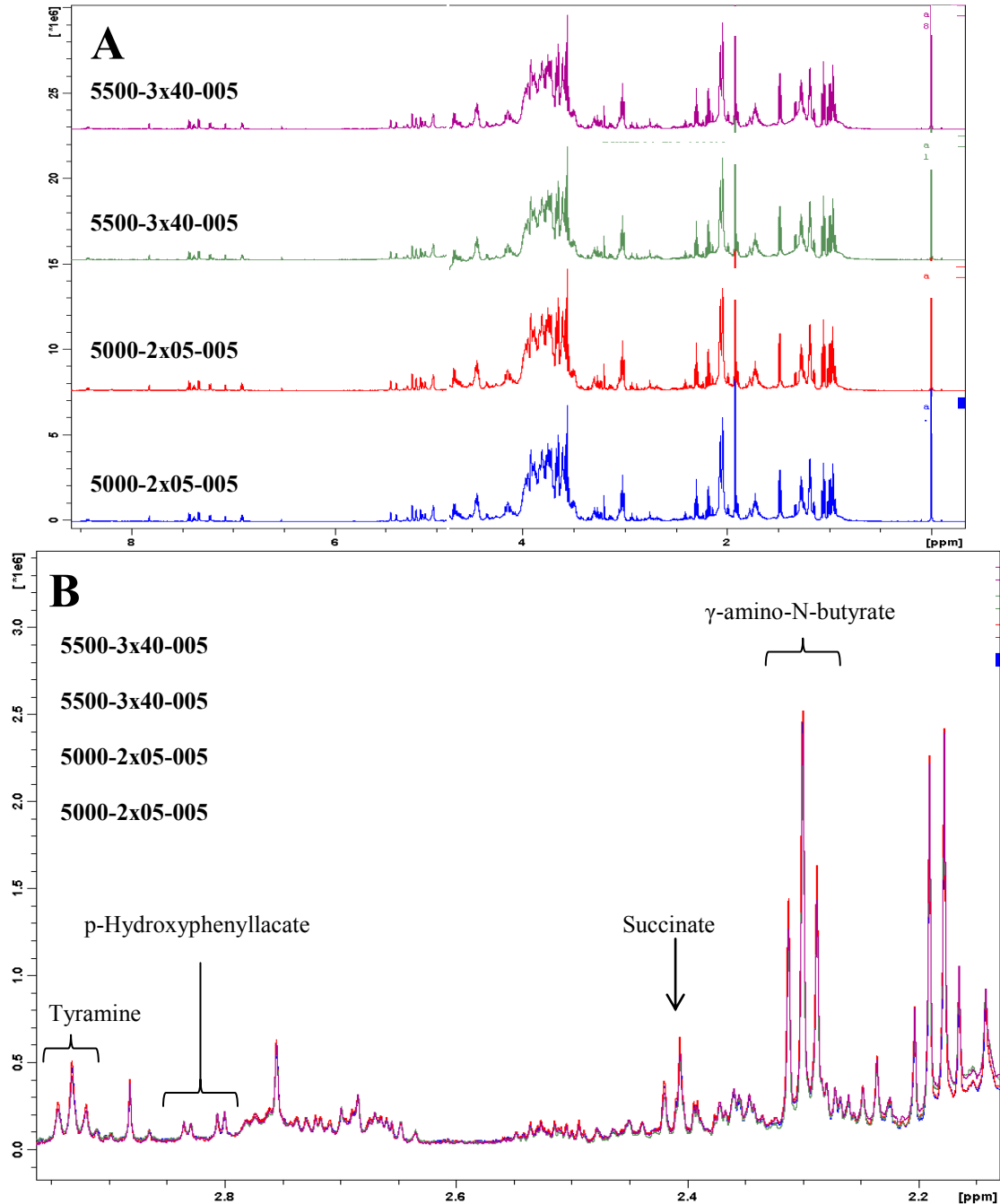


Figure 19: Comparison of standard one-dimensional ^1H NMR spectra of faecal water extracted by two different modes of bead beating (presented in duplicates). Segment A presents the whole spectra while segment B an expanded fragment from 2.1ppm to 2.9ppm.

Due to the fact that new, higher throughput bead beater had been purchased by the lab, another experiment comparing two bead beaters was performed. Three stool samples were divided into four fractions each (experiment in duplicates). Stool was mixed with H_2O containing 0.02% sodium azide (w/v) in a ratio of 1g to 2ml. 0.25g of 1mm zirconium beads were added to all fractions. Two fractions were bead beaten twice for 5 sec on Precellys 24 bead beater with speed of 5000 rpm while the other two fractions were set for 10 sec on Mini beadbeater with speed of 36 oscillations per 1 sec.

The supernatants of duplicates were pulled and phosphate buffer was added to final volume of 550 μ l. After vortexing the samples were transferred to NMR tubes and standard one-dimensional ^1H NMR spectra were acquired on Bruker DRX 600 MHz spectrometer with a 5mm TXI probe operating at 600.13 MHz. Due to the variable sample nature the value for the 90° pulse length which varied from 12.00 μ s to 30.00 μ s depending on sample type and condition. The total number of scans equalled 256 and was saved as 32k data points with a spectral width of 20 ppm. Only two stool samples were analysed by NMR as consistency of the third sample was emulsion like. Comparison of spectra (Figure 20) found no difference in information retrieval. This means that both beat beaters (Precellys 24 and Mini beadbeater) can be used for faecal water extraction.

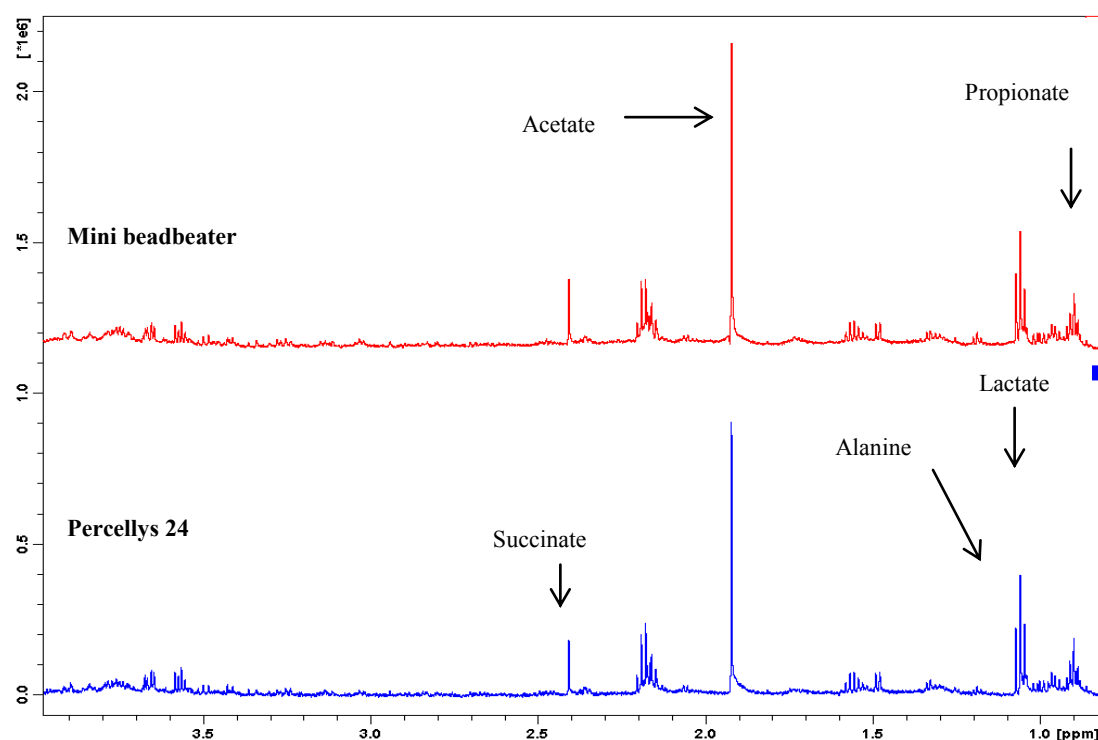


Figure 20: Comparison of standard one-dimensional ^1H NMR spectra of faecal water extracted by two different bead beaters (Precellys 24 and Mini beadbeater). Region from 0.8ppm to 4.0ppm was expanded.

3.2.iv Solvent evaporation

As previously mentioned, speed vacuuming step had to be incorporated into protocol in order to remove the acetonitrile and H_2O from sample before analysing on NMR. In order to assess if the process of speed vacuuming does not result in a loss of metabolites, three stool samples were divided into two fractions which followed regular protocol of faecal water extraction in H_2O . Then phosphate

buffer with TSP was added and standard one-dimensional, CPMG and diffusion edited spectra were obtained. Samples were then removed from the NMR tubes, transferred into microcentrifuge tubes, speed vacuumed for 3 hours and resuspended in either D₂O with TSP or phosphate buffer. Samples were then run on NMR to obtain spectra. There are flaws in the experiment which include removing samples from NMR tubes and higher final concentration of TSP in sample. Spectra were scaled to alanine (1.48ppm). The results presented in Figure 21 show that speed vacuuming led to reduction of acetate (1.92ppm), fumarate (5.63ppm) and a doublet at 8.43ppm shift which probably is formamide.

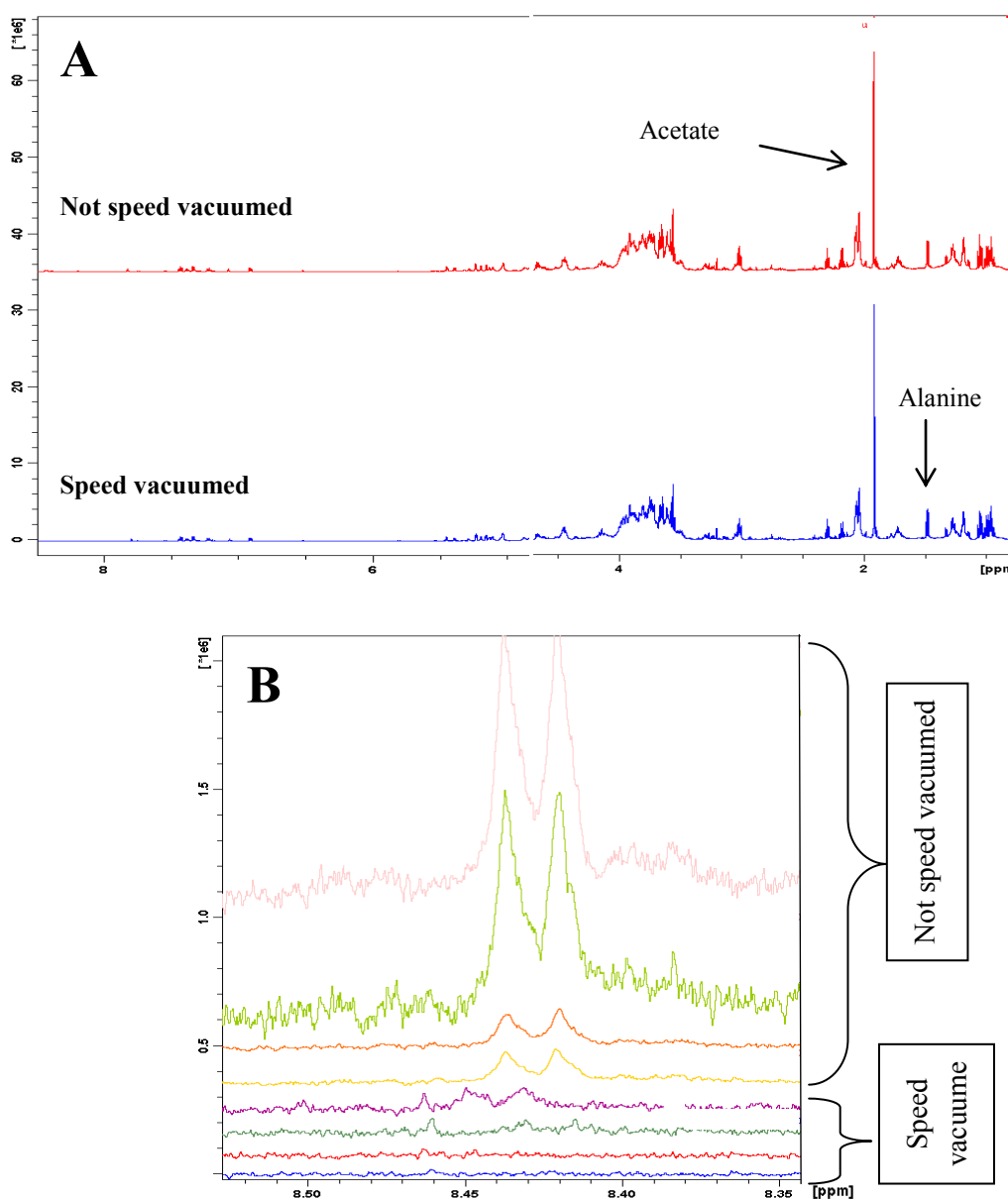


Figure 21: Comparison of standard one-dimensional ¹H NMR spectra of faecal water extracted before and after speed vacuuming. Segment A presents the whole spectra of selected sample while segment B is an expanded fragment of all spectra showing region from 8.25ppm to 8.50ppm.

To check the batch effect of speed vacuuming another experiment was performed. One stool sample was divided into two fractions which were extracted in dH₂O with use of bead beating. Faecal water for each fraction was then aliquoted (400µl) into two microcentrifuge tubes each. One set of the duplicates was speed vacuumed for 3h at 45°C while the other remained in the room temperature. All samples had sodium azide added during that time. After evaporation of dH₂O samples were resuspended in 400µl of dH₂O in order to maintain the same concentration of metabolites as in duplicates that were not speed vacuumed. After adding 200µl of phosphate buffer to every aliquot samples were vortexed and transferred into NMR tubes. Standard one-dimensional ¹H NMR spectra were acquired on Bruker DRX 600 MHz spectrometer with a 5mm TXI probe operating at 600.13 MHz. The 90° pulse length which varied from 15.00 µs to 20.00 µs depending on sample type and condition. The total number of scans equalled 256 and was saved as 32k data points with a spectral width of 20 ppm.

The results show that singlet at chemical shift 1.25ppm was introduced during speed vacuuming (Figure 22). The identification of this peak was not successful, however there are reports of “common impurity” presented as a singlet at the same chemical shift. At the same time singlets at chemical shifts 3.36ppm (methanol) and 4.42ppm (dihydroxyacetone) disappeared (Figure 23).

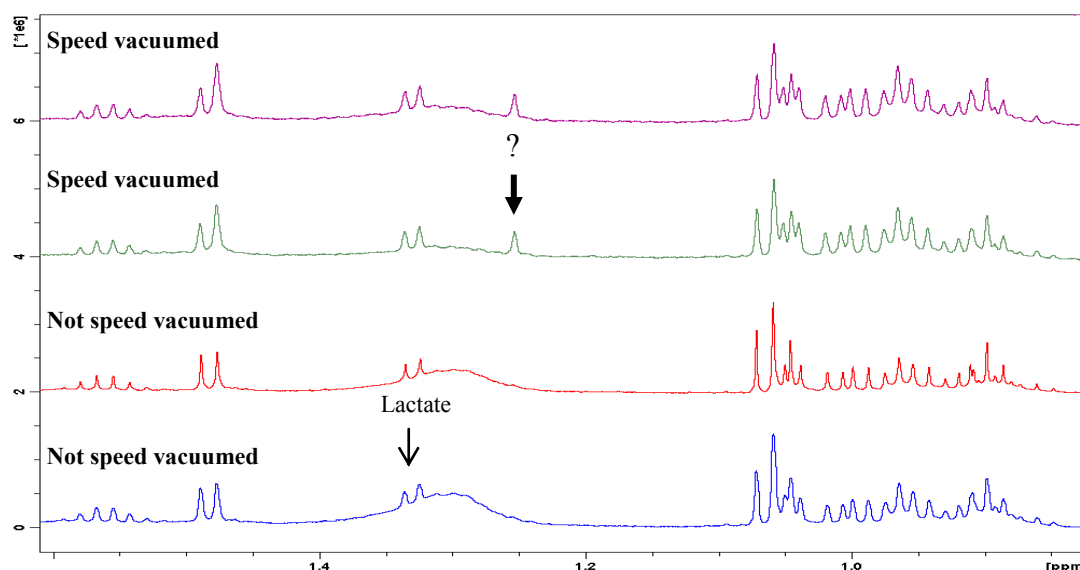


Figure 22: Comparison of standard one-dimensional ¹H NMR spectra of faecal water extracted before and after speed vacuuming

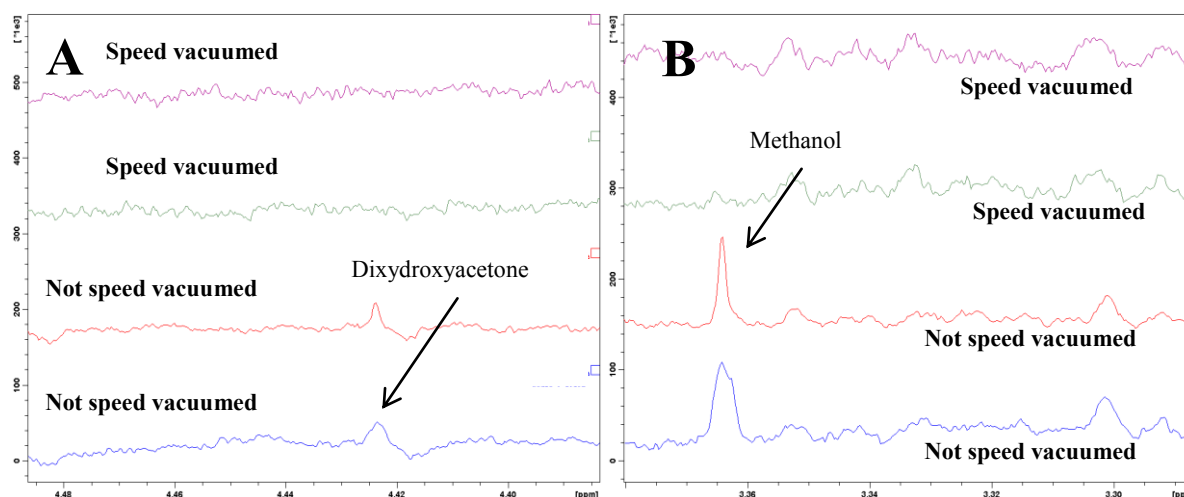


Figure 23: Comparison of standard one-dimensional ^1H NMR spectra presenting the influence of speed vacuuming of faecal water on the metabolic profile. Segment A presents fragment from 3.30ppm to 3.37ppm while segment B presents region from 4.41ppm to 4.48ppm.

Because the faecal water will be obtained with use of 25% acetonitrile it is necessary to evaporate this solvent and resuspend in D_2O prior to the NMR analysis. Speed vacuuming cannot be used for that purpose as it introduces the impurities. The other possibilities to remove the acetonitrile from faecal water are freeze-drying and Nitrogen flow.

Three sets of faecal waters that were extracted in D_2O or H_2O were divided into four fractions. Two of the fractions from each set were freeze-dried and later resuspended in D_2O while the other two fractions were used as a control. The results showed that freeze-drying did not introduce any impurities but at the same time caused decrease in the concentration of certain volatiles (Figure 24). The metabolites that were affected by freeze-drying included ethanol (triplet at 1.18ppm), acetate (singlet at 1.92ppm), citrate (doublets at 2.54ppm and 2.66ppm), methanol (singlet at 3.36ppm), α -hydroxyhippurate (singlet at 5.56ppm) and the unidentified singlet at chemical shift 2.23ppm (marked with a small black arrow on Figure 24).

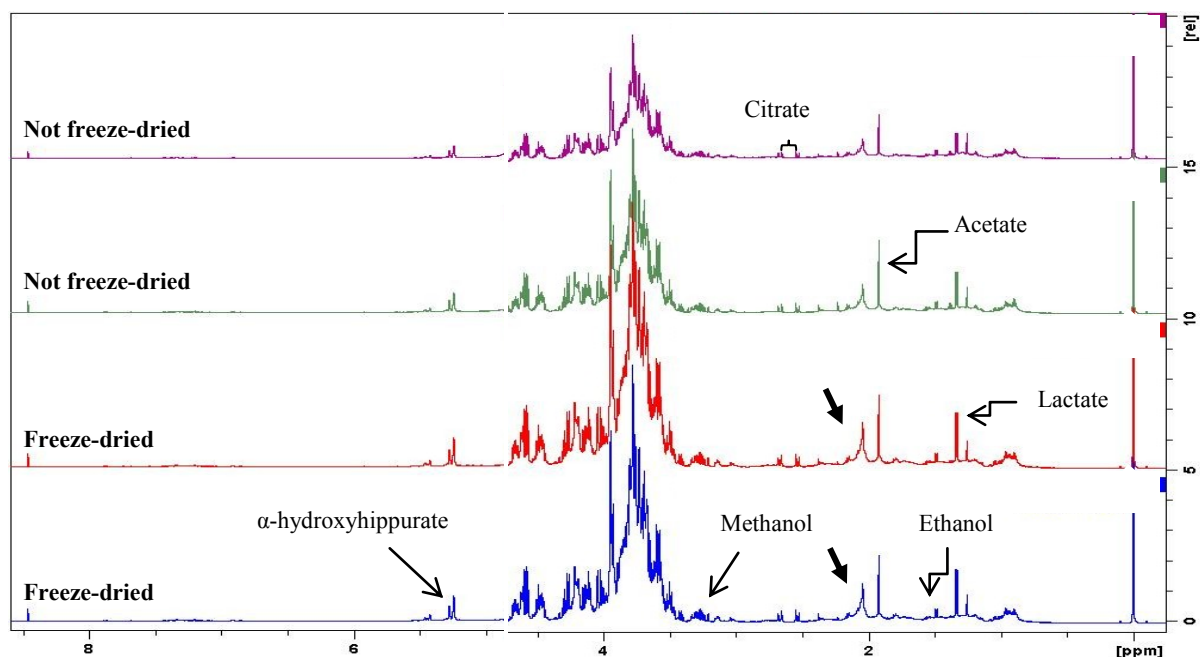


Figure 24: Comparison of standard one-dimensional ^1H NMR spectra presenting the influence of freeze-drying of faecal water on the metabolic profile

Unfortunately the available freeze-drier was operating only up to -50°C which meant that no samples containing acetonitrile could be used. The freezing point for acetonitrile is $\sim -50^\circ\text{C}$ which means that there was a risk that this solvent would escape the ice trap and damage the oil in the pump. In order to use freeze-drying as a method of acetonitrile removal, a new machine operating up to -100°C would have to be bought. Due to high costs of such machine, another potential method – the Nitrogen flow, was tested. The outline of the experiment to test the effect of Nitrogen flow on the quality of spectra was the same as for the previously described experiment testing the freeze-drying. The results showed, that similarly to the freeze-drying, the Nitrogen flow did not introduce any impurities but removed some of the volatile metabolites (Figure 25). The difference was observed for ethanol (triplet at 1.18ppm), a doublet at chemical shift 1.36 and a singlet at a chemical shift 2.23ppm.

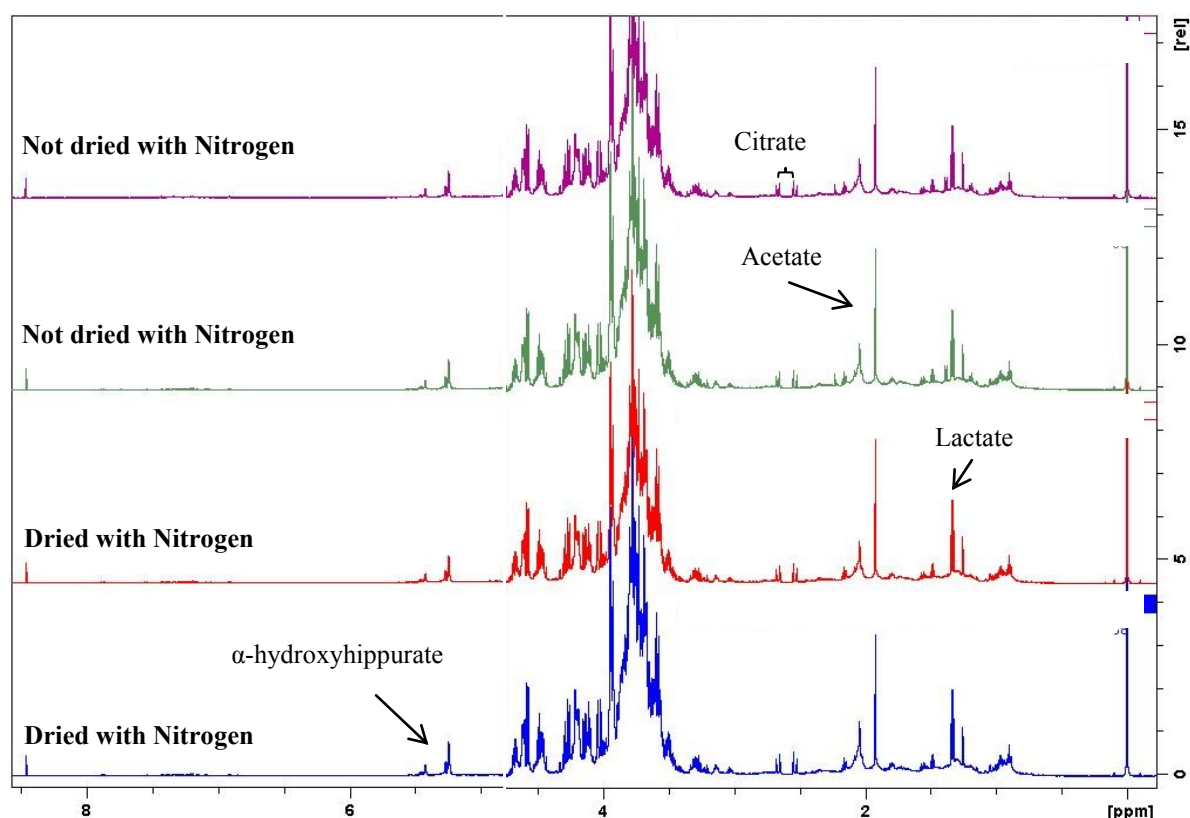


Figure 25: Comparison of standard one-dimensional ^1H NMR spectra presenting the influence of drying samples with Nitrogen flow on the metabolic profile

Using the Nitrogen flow was very fast (<1h needed for 50 μl) and did not require using water bath thus samples were kept at room temperature to avoid the degradation. Additionally in the Turbovap apparatus, up to 50 samples could be dried at the same time which made the method robust. The final decision was to use the Nitrogen flow (Turbovap) to remove the acetonitrile from the faecal water samples. In order to meet the Brukers criteria, samples were resuspended in 600 μl D_2O and 60 μl of phosphorus buffer (1.5M KH_2PO_4 , 2mM NaN_3 , 0.1% TSP (3-(trimethyl-silyl)propionic acid- d_4) was added.

3.2.v Filtration

One of the proposed steps in protocol was to filter faecal water through Costar's Spin-x centrifuge tube filters (Sigma CLS8160). At the very beginning of the method development filters were tested for their purity with use of both D_2O (Sigma, USA, 151882) and 0.01M PBS (Sigma, P4417) which was prepared from tablets according to manufacturer's instructions. 500 μl of solvent (the experiment was done for both 0.01M PBS and D_2O) was transferred to Costar's Spin-x centrifuge tube filter and

to Eppendorf 2ml minicentrifuge tube. The Eppendorf tube was used as a blank for solvent's contact with plastic. Both centrifuge tube filters and to Eppendorf minicentrifuge tubes were centrifuged at top speed for 10 min. Samples were transferred to NMR tubes and 50 μ l of D₂O was added in order to obtain final volume of 550 μ l. The 90° pulse length varied from 10 μ s to 10.5 μ s depending on sample type and condition. The total number of scans equalled 512 was saved as 32k data points with a spectral width of 20 ppm.

The results showed that both D₂O and 0.01M PBS absorbed elements of column which led to presence of impurities in NMR spectra (Figure 26). What is more important is that metabolites removed from column included acetate which is an important bacterial metabolite and is one our main focus while analysing spectra.

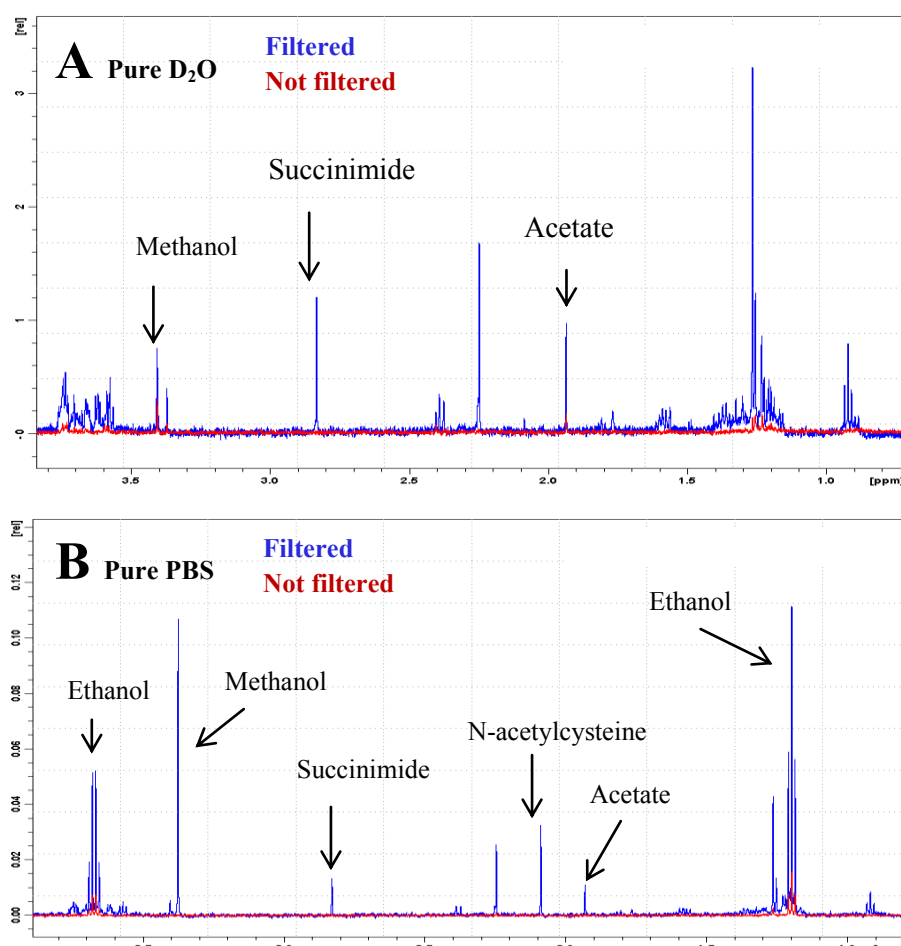


Figure 26: Standard one-dimensional ¹H NMR spectra presenting the influence of filtration of solvent through Spin-x centrifuge tube filters. Segment A compares filtered and not filtered D₂O while segment B compares filtered and not filtered PBS.

The experiment was later repeated for 25% acetonitrile solution. Costar's Spin-x centrifuge tube filter (Sigma CLS8160) was washed four times with new fractions of solvent. All fractions were collected and unfiltered 25% acetonitrile solution was treated as a control. Samples had been evaporated and stored in -20°C . Metabolites were resuspended in $550\mu\text{l}$ D_2O (Sigma, USA, 151882) with 0.017% TSP (Sigma). Samples were then vortexed for 1 min, sonicated for 5 min and again vortexed for 1 min. After that samples were centrifuged for 10 min at top speed at 4°C to precipitate any left proteins. Supernatant was transferred to NMR tubes and standard one-dimensional ^1H NMR spectra were acquired on Bruker DRX 400 MHz spectrometer (Rheinstetten, Germany) with a 5mm TXI probe operating at 400.13 MHz. A relaxation delay- 90° - t_1 - 90° - t_m - 90° -acquire free induction decay (FID) sequence was used to obtain 1-dimensional NMR spectra. The 90° pulse length varied from $12.00\mu\text{s}$ to $12.50\mu\text{s}$ depending on sample type. The total number of scans varied from 128 to 256 and was saved as 32k data points with a spectral width of 20 ppm. The results show that 25% acetonitrile solution does not retrieve impurities from Costar's Spin-x centrifuge tube filters. This means that filters can be used if faecal water was extracted with this solution (which may be a case if stool sample should be used also for mass spectrometry). Figure 27 illustrates that there are no impurities coming of the filters if 25% acetonitrile (in H_2O) is used as the solvent.

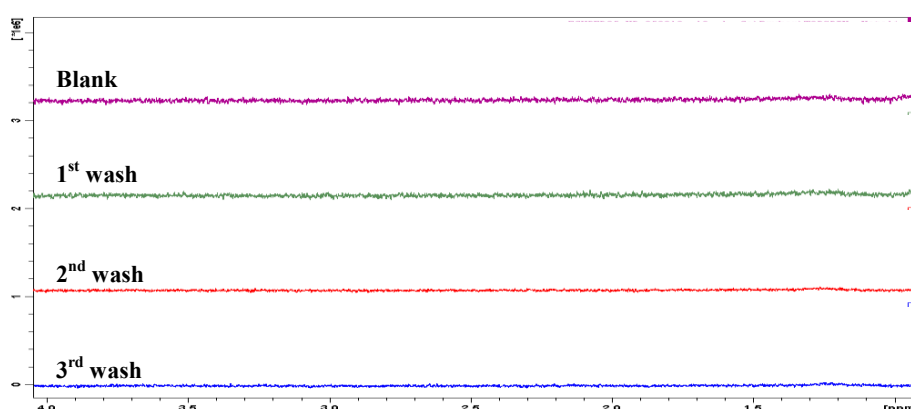


Figure 27: Standard one-dimensional ^1H NMR spectra presenting the influence of filtration of 25% acetonitrile solution through Spin-x centrifuge tube filter (256 scans on 400 MHz)

3.2.vi Preparation of lipid layer for lipidomics

The process of faecal water extraction was effective enough for the lipid layer separation from the faecal water that no organic solvent extraction was necessary. Faecal water was tested for the

presence of lipids in order to confirm the effectiveness of the method. There were no peaks visible on chromatogram. The whole lipid layer was collected and resuspended in 230µl of isopropanol/acetonitrile/water solution in a ratio 2:1:1. Three 1mm zirconium beads were added and sample was bead beaten for 10 seconds, centrifuged for 10 mins at $16,000 \times g$. 100µl of supernatant was transferred to glass vial and sealed.

3.2.vii Final protocol

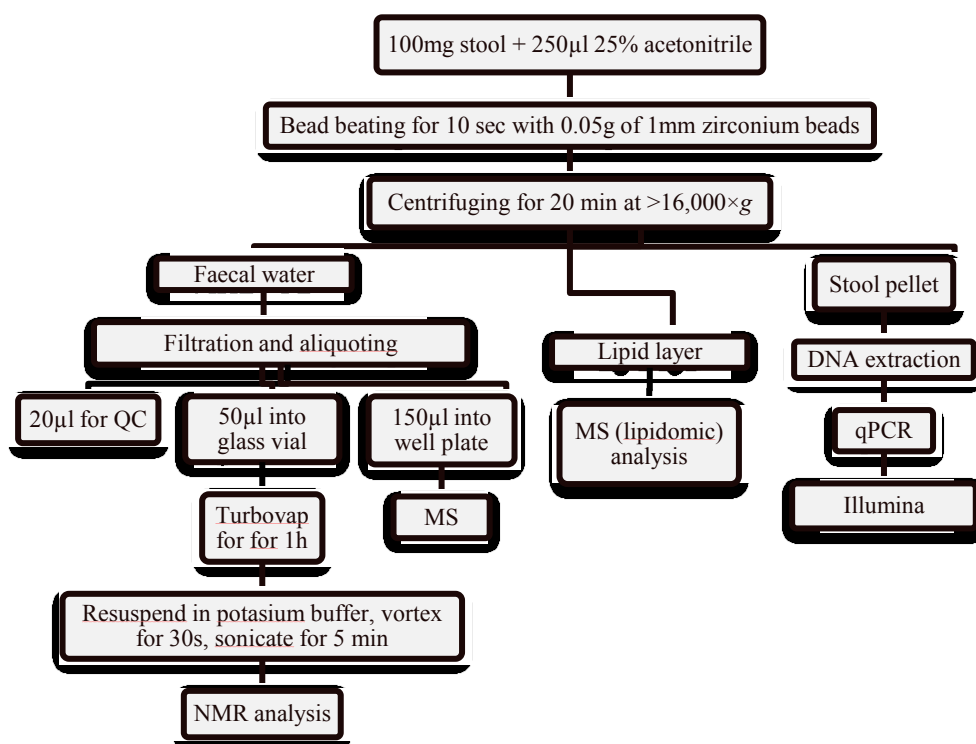


Figure 28: Flowchart presenting the final protocol for faecal water extraction

3.3 DNA extraction from stool pellet

Stool pellet from faecal water extraction was used for metataxonomic analysis of gut microbiota. In order to obtain accurate data it was important to assess if any of the faecal water extraction steps led to loss of DNA yield and/or quality. For the method development QIAamp Stool Tubes Pathogen Detection protocol and the QIAamp DNA Stool Mini Kit was used.

The whole pellet from faecal water extraction was placed in a sterile 2ml bead-beating tube that was prefilled with approximately 0.3g of 0.1mm glass beads. The sample was topped up with 1.5ml of

ASL buffer. There was left free space in the tube in order to obtain a more efficient bead beating. Sample was beaten 3 times for 30 sec with 5 sec intervals at 6500 rpm (Precellys 24 bead beater) and incubated for 5 min at 95°C in a heat block. After that the sample was centrifuged for 1 min 40 sec at 16,000×g (the protocol states that the sample should be centrifuged for 1 min at >20,000×g but it is advised that if the top speed of the centrifuge is lower, then the time must be proportionally elongated). After centrifuging 1.2ml of supernatant was transferred into a sterile 2ml Eppendorf microcentrifuge tube and the InhibitEX tablet was added. Immediately after that, the sample was vortexed for 1 min, incubated at room temperature for another 1 min and centrifuged at 16,000×g for 4 min. 200µl of supernatant was then transferred into a new sterile 1.5ml microcentrifuge tube prefilled with 15µl of proteinase K. 200µl of buffer AL were added into the sample making sure that proteinase K was not mixed with buffer AL before the supernatant was added. Sample was then vortexed for 15 sec and incubated for 10 min at 70°C in heat block. Next, 200µl of EtOH (96-100%) was added to the lysate and vortexed. The lysate was then transferred into QIAamp spin column, that was placed in 2ml microcentrifuge tube, and centrifuged at 16,000×g for 1 min 40 sec. QIAamp spin column was then transferred into a new 2ml microcentrifuge tube, 500µl of buffer AW1 was added and centrifuged at 16,000×g for 1 min 40 sec. QIAamp spin column was again transferred into a new 2ml microcentrifuge tube, 500µl of buffer AW2 was added and centrifuged at 16,000×g for 4 min. QIAamp spin column was finally transferred into a new 1.5ml microcentrifuge tube, 200µl of buffer AE was added, incubated for 1 min at room temperature and centrifuged at 16,000×g for 1 min 40 sec. Eluted DNA was stored in -20°C until further analysis.

Quantity of the DNA was measured with Nano-drop Spectrometer ND-1000. 2µl of the sample were applied and the absorbance at various wavelengths was measured. Buffer TE which was used for the elution of DNA from the column was used as blank. Apart from retrieving the value for DNA concentration per 1µl of solvent also the information about the purity of DNA were obtained. This included the ratio of absorbance 260/280nm and 260/230nm. The concentration of DNA is calculated with the use of the Lamberts-Beer law $A = e \cdot l \cdot c$, where 'A' stands for absorbance, 'e' for

absorption coefficient, 'l' for length that light has to travel and 'c' for concentration. The absorption coefficient 'e' is characteristic for a particular compound.

According to QIAamp protocol, the typical DNA concentration extracted from adult stool sample varies from 75 to 300 ng/μl. In order to make the validation process easier I used DNA concentration per 1g of stool. The concentration of DNA was recalculated using Equation 1 formulas.

Equation 1: DNA concentration recalculation

$$\text{Total extracted DNA ng} = \text{concentration of DNA ng/}\mu\text{l} \times \text{volume of eluent } [\mu\text{l}]$$

$$\text{Concentration of DNA [ng/1g]} = \text{Total extracted DNA ng} / \text{Stool mass [g]}$$

$$\text{Concentration of DNA } [\mu\text{g/1g}] = \text{Concentration of DNA [ng/1g]} \times 1000$$

3.3.i Untreated stool vs. faecal pellet

The first important experiment was to establish if it is possible to use stool pellet from faecal water extraction for effective DNA extraction. The efficiency was compared to this from untreated stool. Four stool samples (two from preterm infants 5 weeks post-partum, one from full-term infant 1 week post-partum and one from 1-year-old infant) were divided each into three fractions of equal masses. Two fractions were used for faecal water extraction and one was left untreated for the DNA extraction. The DNA extraction followed the protocol described in Method Development chapter section 3.2.ii. The results presented in Table 6 show that pellet from faecal water extraction can be effectively used for DNA extraction. Surprisingly both the DNA concentration per 1μl of elution buffer and the purity of samples are improved. In depth analysis of the sample purity are explained in the Method Development chapter section 3.2.ii.g.

Table 6: Comparison of DNA yields after extraction from untreated stool and stool pellet obtained from faecal water extraction. Results for pellets are presented as mean value of duplicates and range is provided in the brackets.

Sample ID	Type of sample	DNA concentration in ng/μl of elution buffer	Purity of DNA	
			260/280 nm	260/230 nm
A	Untreated	2.9	0.96	1.45
	Pellet	19.0 (16.2-21.8)	1.88 (1.85-1.90)	1.85 (1.56-2.13)
B	Untreated	1.9	1.03	0.46
	Pellet	3.1 (2.8-3.3)	1.47 (1.18-1.75)	0.48 (0.40-0.56)
C	Untreated	4.7	1.69	0.79
	Pellet	4.9 (4.5-5.3)	1.82 (1.50-2.13)	0.81 (0.78-0.83)
D	Untreated	63.8	2.22	1.96
	Pellet	83.4 (83.0-83.8)	2.15 (2.13-2.17)	2.11 (2.07-2.15)

3.3.ii Type of homogenisation

Type of homogenisation was one of the most important factors that could affect the DNA yield. There was a risk that bead beating with 1mm zirconium beads could rupture bacterial cells. What is worse, in such case it would not be clear if the percentage of loss was proportional (random loss) or if some species were more predisposed to damage. In order to assess the influence of bead beating on bacterial survival, pellets obtained by bead beating (twice for 5 sec at 5000 rpm on Percellys 24 bead beater) were compared with those that were obtained by vortexing. The DNA extraction followed protocol described in the Method Development chapter section 3.2.ii. As shown in the Table 7 vortexing of faecal water reduced the DNA yield of stool pellet. This means that bead beating with 1mm zirconium beads is a better method of stool sample homogenisation.

Table 7: Comparison of DNA yield from pellets that were bead beaten and those that were vortexed.

Sample ID	Homogenisation method	DNA concentration in $\mu\text{g/g}$ of stool	Purity of DNA	
			260/280 nm	260/230 nm
A	Bead beating 5000-2x05	4.6	2.13	0.78
	Vortexing	3.38	1.50	0.83
B	Bead beating 5000-2x05	21.8	1.90	1.56
	Vortexing	16.2	1.85	2.13
C	Bead beating 5000-2x05	2.7	1.18	0.56
	Vortexing	2.4	1.75	0.40

Next step was to decide which program of bead beating will give the best results for both metabolite profiles and DNA yield. Samples that have been bead beaten twice for 5 sec at 5000 rpm were compared with those that have been bead beaten three times for 40 sec at 5500 rpm (Percellys 24). The results shown in the Table 8, present that there is no clear pattern in influence of type of bead beating on DNA yield. It is more likely caused by the fact that stool itself is not homogenous and different fractions may contain more or less of bacterial cells. That is why it was decided to vortex stool sample for 1 min prior to separation of fraction for the faecal water extraction. What is more, sample “C” was meconium which explains low DNA concentration.

Table 8: Comparison of DNA yield from pellets that were bead beaten for different time.

Sample ID	Bead beating program	DNA concentration in $\mu\text{g/g}$ of stool	Purity of DNA	
			260/280 nm	260/230 nm
A	5500-3x40-005	8.2 (7.0-9.3)	2.67 (2.18-3.15)	0.27 (0.18-0.35)
	5000-2x05-005	15.7 (14.1-17.3)	3.18 (2.85-3.50)	0.39 (0.30-0.48)
B	5500-3x40-005	11.3 (10.5-12.0)	2.24 (1.98-2.49)	0.40 (0.39-0.40)
	5000-2x05-005	10.0 (8.7-11.2)	2.35 (2.32-2.38)	0.27 (0.26-0.27)
C	5500-3x40-005	1.1 (0.6-1.5)	1.98 (0.69-3.26)	0.15 (0.05-0.24)
	5000-2x05-005	1.3 (0.0-2.6)	0.59 (0.00-1.17)	0.10 (0.01-0.18)

As explained in the Method Development chapter section 3.2.i.b, it was necessary to perform another experiment comparing bead beating on Percellys 24 (twice for 5 sec at 5000 rpm) and Mini Bead beater (10 sec at 36 oscillations per 1 sec). Pellets from this experiment were extracted according to previously mentioned protocol (Method Development chapter section 3.2.ii). The results, presented in Table 9, show that there seems to be higher DNA yield when samples are homogenised at Mini Bead beater for 10 sec at 36 oscillations per 1 sec.

Table 9: Comparison of DNA yield from pellets that were bead beaten with different programs

Sample ID	Bead beating program	DNA [ug/1g]	Purity of DNA	
			260/280 nm	260/230 nm
A	10 sec at 36 osc per 1 sec	23.1 (21.8-24.4)	2.24 (2.18-2.29)	1.48 (1.37-1.58)
	twice for 5 sec at 5000 rpm	34.4 (32.8-36.0)	1.98 (1.96-1.99)	1.06 (1.02-1.10)
B	10 sec at 36 osc per 1 sec	29.6 (29.4-29.7)	2.13 (2.08-2.17)	1.87 (1.86-1.87)
	twice for 5 sec at 5000 rpm	26.9 (26.5-27.3)	2.04 (2.02-2.05)	1.65 (1.54-1.76)
C	10 sec at 36 osc per 1 sec	67.2 (53.0-81.4)	1.89 (1.81-1.97)	1.49 (1.14-1.83)
	twice for 5 sec at 5000 rpm	58.7 (51.0-66.4)	1.99 (1.98-2.00)	2.00 (1.99-2.00)

3.3.iii Solvents

As previously described in the Method Development chapter section 3.2.i.a, a number of solvents have been tested for faecal water extraction. The influence of those solvents on bacterial cell survival has been examined in additional experiments. The majority of the results (Table 10) suggest that there is no significant difference in DNA yield obtained from stool pellets that were extracted with use of different solvents. The only significant difference is visible for sample 'A' however due to lack of duplicates it is possible that those results are not representative and are caused by lack of homogeneity in the stool sample. When it comes to the issue of DNA purity all samples show that faecal water extraction in phosphate buffer leads to the best 260/280nm and 260/230nm ratios, followed by D₂O with PBS giving the lowest purity. The aspects of DNA purity are described in detail in the Method Development chapter section 3.2.ii.h. This finding suggests that faecal water extraction in phosphate buffer should provide good quality DNA sample.

Table 10: Comparison of DNA yield from pellets extracted either in D₂O, H₂O, PBS or phosphate buffer. Results are presented as mean value of duplicates and range is provided in the brackets.

Sample ID	Solvent	DNA concentration in ng/μl of elution buffer	Purity of DNA	
			260/280 nm	260/230 nm
A	D ₂ O	18.6	2.09	1.27
	PBS	27.3	1.69	0.71
	Phosphate buffer	68.3	1.84	1.42
B	D ₂ O	4.9 (4.5-5.3)	1.82 (1.50-2.13)	0.81 (0.78-0.83)
	PBS	5.0 (1.8-8.1)	1.52 (1.45-1.58)	0.43 (0.27-0.59)
C	D ₂ O	83.4 (83.0-83.8)	2.15 (2.13-2.17)	2.11 (2.07-2.15)
	H ₂ O	70.0 (61.0-79.0)	2.17 (2.14-2.19)	2.13 (2.11-2.14)
D	D ₂ O	57.0 (31.7-84.2)	1.86 (1.75-1.96)	1.09 (0.84-1.33)
	Phosphate buffer	52.1 (51.1-52.6)	1.81 (1.72-1.90)	1.23 (0.94-1.52)

As previously described in the Method Development chapter section 3.2.i.d, another experiment comparing faecal water extraction in 25% acetonitrile in dH₂O (18MΩ) with extraction in dH₂O (18MΩ) was performed. DNA extraction of the stool pellets was performed according to previously described DNA extraction protocol (Method Development chapter section 3.2.ii). The results of this experiment suggest that faecal water extraction in 25% acetonitrile in dH₂O (18MΩ) does not significantly affect the DNA yield (Table 11). Even though the mean value of DNA concentration may imply that faecal water extraction in water may produce better results, the range shows that there is a big variation among the sample and what is more, important the range values for acetonitrile and H₂O overlap. There is also no clear pattern of purity improvement or decrease.

Table 11: Comparison of DNA yield from pellets extracted either in 25% acetonitrile or H₂O. Results are presented as mean value of duplicates and range is provided in the brackets.

Sample ID	Solvent	DNA concentration in $\mu\text{g/g}$ of stool	Purity of DNA	
			260/280 nm	260/230 nm
A	H ₂ O	65.1 (36.6-93.6)	1.98 (1.82-2.13)	0.82 (0.77-0.87)
	Acetonitrile	117.6 (89.6-145.6)	2.08 (1.99-2.17)	0.95 (0.91-0.99)
B	H ₂ O	22.1 (15.7-28.5)	1.67 (1.57-1.76)	0.92 (0.63-1.11)
	Acetonitrile	16.9 (10.5-23.4)	1.94 (1.90-1.97)	1.12 (1.10-1.13)
C	H ₂ O	35.6 (16.7-54.6)	1.91 (1.79-2.02)	1.23 (1.15-1.30)
	Acetonitrile	31.7 (24.1-39.4)	1.81 (1.69-1.93)	1.02 (0.92-1.11)

3.3.iv Presence of sodium azide

In order to assure no bacterial metabolism in faecal water, 0.02% sodium azide (NaN₃) is commonly used as a preservative. The influence of sodium azide on bacterial cell survival is unknown and the aim of this experiment was to test if there is any significant drop in DNA yield after extracting in the presence of sodium azide. Four stool samples were divided into four fractions each and one stool sample was divided into two fractions. Two of the fractions were extracted in D₂O with no sodium azide and the other two fractions were extracted in phosphate buffer with 0.02% sodium azide. In the case of the fifth stool sample, which was divided only into two fractions, the experiment was not done in duplicates. In all the samples there was visible a slight drop in DNA yield when faecal water was extracted in phosphate buffer with 0.02% sodium azide, compared to D₂O (Table 12).

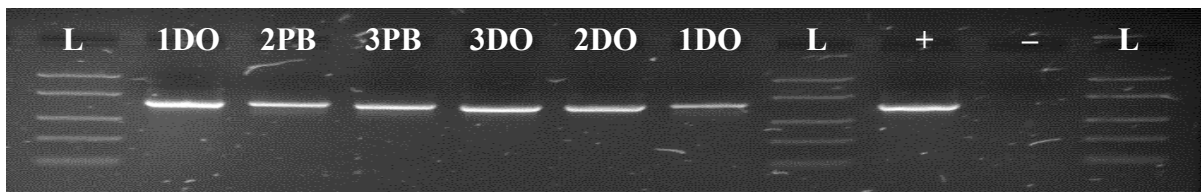


Figure 29: Gel electrophoresis of PCR products of 16S rRNA gene extracted from stool samples. Ladder is marked as L, positive control as + and negative control as -. Number refers to stool sample, PB refers to faecal water extraction in phosphate buffer (with 0.02% sodium azide) while D₂O refers to faecal water extraction in D₂O.

Table 12: Comparison of DNA yield from pellets extracted either in phosphate buffer or D₂O. Results are presented as mean value of duplicates and range is provided in the brackets.

Sample ID	Solvent (presence of sodium azide)	DNA concentration in ng/μl of elution buffer	Purity of DNA	
			260/280 nm	260/230 nm
A	Phosphate buffer with NaN ₃	35.1 (34.6-35.6)	1.76 (1.73-1.79)	1.08 (1.05-1.20)
	D ₂ O without NaN ₃	41.45 (34.6-48.3)	1.93 (1.92-1.93)	1.20 (1.04-1.43)
B	Phosphate buffer with NaN ₃	20.4 (20.3-20.5)	1.63 (1.59-1.66)	0.83 (0.80-0.86)
	D ₂ O without NaN ₃	21.3 (20.0-22.6)	1.76 (1.68-1.83)	0.81 (0.74-0.88)
C	Phosphate buffer with NaN ₃	17.6 (16.8-18.4)	1.70 (1.68-1.71)	0.87 (0.87-0.87)
	D ₂ O without NaN ₃	22.5 (22.3-22.6)	1.61 (1.61-1.61)	0.82 (0.74-0.91)
D	Phosphate buffer with NaN ₃	52.1 (51.1-52.6)	1.81 (1.72-1.90)	1.23 (0.94-1.52)
	D ₂ O without NaN ₃	57.0 (31.7-84.2)	1.86 (1.75-1.96)	1.09 (0.84-1.33)
E	Phosphate buffer with NaN ₃	2.2	1.70	0.49
	D ₂ O without NaN ₃	4.0	1.51	0.74

The quality of DNA was tested by Dr Julian Marchesi. PCR was done with use of the universal primers for 16S rRNA gene – 27F and 1492R. The PCR results (Figure 29) show that the quality of DNA was not affected by the presence of sodium azide during faecal water extraction. Due to the fact that the drop in DNA yield was not significant and may have been related to lack of homogeneousness in the stool itself, it was decided to use sodium azide. Presence of sodium azide in faecal water is very important as it stops any unwanted metabolism of the metabolites and provides representative spectra.

3.3.v Presence of the lipid layer

After analysing stool samples from different sources it was observed that newborn infants are more likely to have steatorrhoea which is an excess of fat in stool. This fat was forming a lipid layer which was floating above faecal water thus blocking the access (Figure 30). Due to the fact that the influence of presence of lipids on DNA extraction was unknown, it was decided to test it. As presented in Table 1Table 13, presence of lipid layer may affect the DNA extraction. It was decided to remove the lipid layer with metal spatula prior to removal of faecal water



Figure 30: Picture of faecal water presenting the lipid layer floating on top of faecal water.

Table 13: Comparison of DNA yields after extraction from stool pellet with and without lipid layer. Results for stool pellets without lipid layer are presented as mean value of duplicates and range is provided in the brackets.

Sample ID	Lipid layer	DNA concentration in ng/μl of elution buffer	Purity of DNA	
			260/280 nm	260/230 nm
A	Present	2.9	0.96	1.45
	Lack	19.0 (16.2-21.8)	1.88 (1.85-1.90)	1.85 (1.56-2.13)
B	Present	1.9	1.03	0.46
	Lack	3.1 (2.8-3.3)	1.47 (1.18-1.75)	0.48 (0.40-0.56)
C	Present	4.7	1.69	0.79
	Lack	4.9 (4.5-5.3)	1.82 (1.50-2.13)	0.81 (0.78-0.83)
D	Present	63.8	2.22	1.96
	Lack	83.4 (83.0-83.8)	2.15 (2.13-2.17)	2.11 (2.07-2.15)

3.3.vi DNA yield variation in preterm, full-term and adult stool samples

Results from all the experiments were pulled together in order to analyse the range of DNA concentration in each age group: preterm newborns, full-term newborns, 1-year-old babies and adults. Although the results suggest that preterm stool has higher concentration of DNA it is important to remember that stool from full-term infants available for analysis is usually from 1st week of life while samples from preterm infants were collected for few months after birth. In the method development, apart from 1st week stool samples from full-term infants also one set of stool from 7-week-old full-term infant was used (Table 14). The age postpartum range of preterm infants was 1-7 weeks. At the age of 1 year the adult microbiota starts to develop and by the 2nd year of life it is usually fully developed [26]. After few weeks of life newborn infants start to have higher yield of DNA that is within the adult range.

Table 14: Comparison of DNA yields from different types of stool samples including: preterm, full-term, 1-year-old and adult. Results are presented as mean value of duplicates and range is provided in the brackets.

Source of sample	DNA concentration in µg/1g of stool	Purity of DNA	
		260/280 nm	260/230 nm
Preterm baby (n=44)	17.2 (0.0-74.3)	1.79 (0.00-3.50)	0.69 (0.01-2.13)
Full-term baby (n=9)	14.6 (1.7-36.0)	1.89 (1.50-2.29)	0.97 (0.49-1.58)
1-year old baby (n=5)	71.1 (61-79-8)	2.17 (2.13-2.22)	2.09 (1.96-2.15)
Adult (n=12)	49.9 (26.5-81.4)	2.18 (1.72-2.29)	1.76 (0.94-2.00)

Due to the fact that samples collected in the first weeks of life are usually low in DNA yield, it was important to examine what DNA concentration is a threshold for PCR. Dr Julian Marchesi performed PCR with use of universal primers for 16S rRNA gene – 27F and 1492R. DNA concentration lower than 1.9ng/µl was not enough to give strong PCR product (Figure 31), concentration of 1.3ng/µl gave weak band after gel electrophoresis while <1ng/µl gave no PCR product. This means that at least 2ng/µl of DNA is required for metataxonomic analysis. It was decided to try concentrating DNA samples to avoid risk of losing some of the valuable samples.

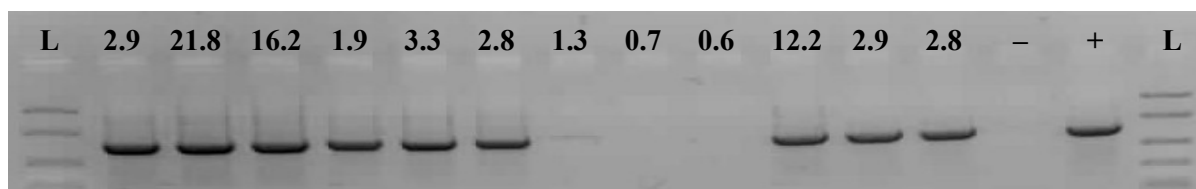


Figure 31: Gel electrophoresis of PCR products of 16S rRNA gene extracted from stool samples. Ladder is marked as L, positive control as + and negative control as -. Values represent the DNA concentration in ng per 1µl of elution buffer.

3.3.vii Concentration of DNA

In order to concentrate the DNA samples, the DNA extraction protocol got modified. In the original protocol the final step of DNA elution from the column (protocol described in detail in the Method Development chapter section 3.2.ii) consists of incubating 200µl of elution buffer for 1 min and then washing the column during centrifuging. The proposed change included using twice 100µl of elution buffer. 100µl of elution buffer was incubated on top of column for 1 min then centrifuged only to be collected and placed again on top of column for 1 min and then centrifuged. This not only introduced lower volume of solvent which should automatically concentrate the DNA two fold, but also elongated the incubation time which could result in washing of column higher amount of DNA. Those two variations of DNA elution from column were compared on four preterm stool samples collected 4-7 weeks post-partum. One of the stool samples had to be withdrawn from the experiment as the mass of the stool pellets after faecal water extraction was significantly higher in one of the duplicates. The results of this experiment (Table 15) show that proposed changes to the protocol result in higher concentration of DNA. This should lead to reduction in the number of stool samples with DNA concentration not meeting the threshold for the metataxonomic analysis.

Table 15: The association of DNA yield with the elution buffer volume and number of elutions

Sample ID	Time of elution and volume of elution buffer	DNA concentration in ng/µl of elution buffer	Purity of DNA	
			260/280 nm	260/230 nm
A	1x with 200µl	16.8	2.09	1.00
	2x with 100µl	51.8	1.66	0.91
B	1x with 200µl	12.8	2.09	0.87
	2x with 100µl	30.0	1.79	0.78
C	1x with 200µl	14.0	1.85	0.99
	2x with 100µl	80.0	1.71	0.86

The only worrying result was that the purity of samples seemed to have dropped. As a result, an experiment to improve the DNA purity was planned.

3.3.viii Purity improvement

The ratio 260/280nm (the ratio of absorbance) should be approximately 1.8 but values above 1.6 are acceptable. This ratio is mostly affected by presence of phenol (> 2.0) and proteins (< 1.8). The other reasons why the value may be different includes the composition of DNA and pH of the sample. The 260/280nm ratio for Guanine is 1.15, for Cytosine 1.51, for Thymine 1.47, for Adenine 4.50 and for Uracil 4.00. This means that DNA rich in A-U pairs would have higher 260/280nm ratio than DNA rich in G-C. Acidity plays an important role when comparing the 260/280nm ratio between regular spectrometer and nano-drop. If sample is in acidic pH environment the value obtained on nano-drop will be drop by 0.3 compared to results from standard spectrometer. In case of basic pH the value will rise by 0.3 [265]. The 260/230nm ratio should be between 2.0 and 2.2 however it is accepted if is above 1.5. It can be affected by EDTA, phenol, carbohydrates, guanidine hydrochloride which is a common compound in DNA extraction kits.

Two newborn stool samples from 1st and 3rd week of life were divided into four fractions each. Stool pellet was obtained after faecal water extraction with phosphate buffer. Two of the fractions were extracted accordingly to the DNA extraction protocol described in Method Development chapter section 3.2.ii and the other two had slightly modified protocol. Those modifications included additional 4 min centrifuging at 16,000 $\times$ g of supernatant obtained after adding InhibitEX tablet. The other modification was removing from column any leftover AW2 buffer by centrifuging for 1 min 40 sec at 16,000 $\times$ g. Although impurities and leftover AW2 buffer were removed from the samples, there was no clear improvement of total purity visible (Table 16). The mean values for 260/230nm ratio did seem to improve after modification of the DNA extraction protocol but the ranges of values overlapped thus those findings are not significant. On the other hand there is visible a drop in the 260/280nm ratio. The proposed modifications of the DNA extraction protocol will not be introduced to the final protocol.

Table 16: The association of DNA purity with the improvements of DNA extraction protocol

Sample ID	DNA extraction protocol	DNA concentration in ng/ μ l of elution buffer	Purity of DNA	
			260/280 nm	260/230 nm
A	Original protocol	12.0 (8.7-15.2)	1.72 (1.61-1.83)	0.44 (0.26-0.61)
	Additional centrifuging steps	10.4 (10.2-10.6)	1.37 (1.36-1.37)	0.56 (0.54-0.57)
B	Original protocol	6.8 (6.8-6.8)	1.53 (1.27-1.79)	0.34 (0.23-0.44)
	Additional centrifuging steps	5.8 (3.6-7.9)	1.28 (1.15-1.41)	0.43 (0.37-0.48)

The low ratio of 260/230nm was visible only in newborn stool samples which means that it is probably unavoidable obstacle. Dr Julian Marchesi performed an experiment to test if the impurities causing low 260/230nm ratio inhibit the PCR reaction. DNA extracts with low 260/230nm ratio were diluted before PCR reaction to see if there will be any improvement in the results. Dilution caused drop of DNA concentration below the threshold for PCR. However, the PCR reaction led to successful amplification of the DNA from samples that had the DNA concentration above the previously mentioned threshold and with the 260/230nm ratio as low as 0.35. The conclusion from this experiment was that low 260/230nm ratio does not inhibit the PCR reaction thus should not affect metataxonomic analysis. In order to check if low values for purity were not caused by low DNA concentration, four DNA extracts (two from preterm sample, one from 1-year-old infant and one from adult) were diluted two fold, fivefold and tenfold. Samples were analysed on nano-drop to see if there is any correlation between the DNA concentration and purity (Table 17).

The association of 260/280nm ratio value and DNA concentration is presented on the Figure 32, where the increasing DNA dilution corresponds to the decreasing DNA concentration. The graph clearly shows that there is underlying unknown factor that affects the 260/280nm ratio. Three out of four samples presented trend of 260/280nm ratio value decrease with decrease of DNA concentration (increase of sample dilution). However this is not visible for the DNA extracted from the adult stool sample.

Table 17: The association of DNA concentration with the DNA purity.

Sample ID	Dilution	DNA concentration in ng/μl of elution buffer	Purity of DNA	
			260/280 nm	260/230 nm
A	0	83.9	2.19	2.03
	2x	43.6	2.23	2.01
	5x	17.9	2.18	2.79
	10x	9.2	2.10	2.16
B	0	26.8	2.11	1.74
	2x	12.7	2.31	1.72
	5x	4.8	2.00	1.35
	10x	3.2	1.70	1.25
C	0	28.0	1.97	1.31
	2x	16.1	1.82	1.40
	5x	6.0	2.25	1.44
	10x	3.8	1.60	1.34
D	0	50.0	1.97	1.12
	2x	27.7	2.06	1.14
	5x	11.4	2.08	1.19
	10x	4.1	2.48	0.98

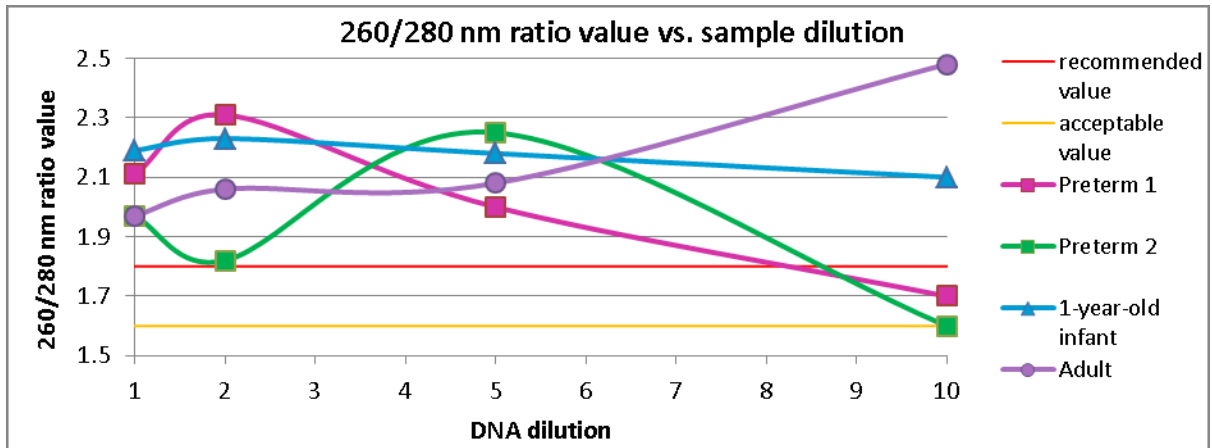


Figure 32: Graph showing correlation of 260/280nm ratio and DNA concentration. Red line represents the recommended value of 260/280nm ratio while yellow line represents the acceptable value of 260/280nm ratio. Squares represent results for preterm samples, triangles results for 1-year-old infant samples and circles – adults.

The 260/230nm ratio value seems to be more associated with the DNA concentration. As presented on Figure 33, there is a trend towards lower 260/230nm ratio in more diluted samples.

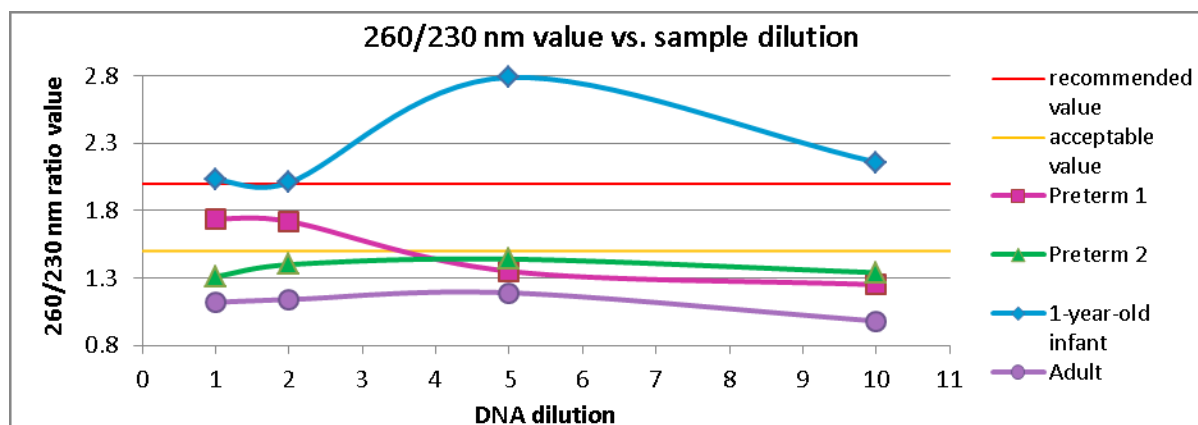


Figure 33: Graph showing correlation of 260/230nm ratio and DNA concentration (sample dilution). Red line represents the recommended value of 260/230nm ratio while yellow line represents the acceptable value of 260/230nm ratio. Squares represent results for preterm samples, triangles results for 1-year-old infant samples and circles – adults.

The purity of DNA extracted from preterm stool samples seems to be a complex issue and may be related to the nature of the sample. Classic methods of the purity improvement did not succeed. Due to the fact that metataxonomic analysis can be carried out on samples with low 260/230nm ratio, it was decided to accept the complex nature of preterm samples and continue with the experiments without improving the purity.

3.3.ix Final improvements to QIAamp DNA Stool Mini Kit

Whole pellet from faecal water extraction was weighted and placed in a sterile 2ml microcentrifuge tube with a screw cap that was prefilled with approximately 0.3g of 0.1mm glass beads. The sample was topped up with 1.5ml of ASL buffer (this should leave free space for a more efficient bead beating). Sample was beaten 3 times for 30sec with 5sec intervals at 6500 rpm (Precellys 24-duo bead beater) and incubated for 5 min at 95°C in a heat block. After that the sample was centrifuged for 1 min 40 sec at 16,000×g and 1.2ml of supernatant was transferred into a sterile 2ml Eppendorf microcentrifuge tube containing the InhibitEX tablet. Immediately after, the sample was vortexed for 1 min and incubated at room temperature for another 1 min. Sample was centrifuged at 16,000×g for 4 min, 200µl of supernatant was transferred into a new sterile 1.5ml microcentrifuge tube prefilled with 15µl of proteinase K. 200µl of buffer AL were added into the sample making sure that proteinase K is not mixed buffer AL before adding the supernatant. Sample was then vortexed for 15 sec and incubated for 10 min at 70°C in the heat block. Next, 200µl of EtOH (96-100%) was added to the lysate and vortexed. The lysate was then transferred into QIAamp spin column placed in 2ml

microcentrifuge tube and centrifuged at $16,000\times g$ for 1 min 40 sec. QIAamp spin column was then transferred into a new 2ml microcentrifuge tube, 500 μ l of buffer AW1 was added and centrifuged at $16,000\times g$ for 1 min 40 sec. QIAamp spin column was again transferred into a new 2ml microcentrifuge tube, 500 μ l of buffer AW2 was added and centrifuged at $16,000\times g$ for 4 min. QIAamp spin column was finally transferred into a new 1.5ml microcentrifuge tube, 100 μ l of buffer AE was added, incubated for 1 min at room temperature and centrifuged at $16,000\times g$ for 1 min 40 sec. Eluted buffer was transferred again on the column, incubated for 1 min at room temperature and centrifuged at $16,000\times g$ for 1 min 40 sec. Eluted DNA was stored in -80°C until further analysis.

3.3.x Final DNA extraction protocol

Due to the fact that Qiagen stopped the production of the kit used for the method development, a new kit was used after a careful analysis of the extraction protocols. The protocol which seemed most similar to the proposed by my improvements was for the MOBIO PowerFaeces DNA extraction kit. The only changes introduced to this protocol were centrifuging at $16,000\times g$ for 1 min 40 sec instead of centrifuging at $13,000\times g$ for 1 min.

Chapter 4 – The urinary metabonome of the preterm baby

4.1 Introduction

The human organism has a very complex system of metabolic pathways which are usually maintained in a state of equilibrium [266, 267]. This balance is partially achieved by excretion of waste products in the urine. The urine composition can reflect the reaction of an organism to various changes in environmental factors that take place (stress, change in health status or exposure to various chemical substances) [266-268]. Analysing the multi-component composition of the urine is a useful tool for quick assessment of an individual's health. Healthy adult kidneys have an efficient filtration system that typically does not allow certain metabolites to leave the system [267]. Therefore, any drastic metabolic profile alteration, which cannot be explained by diet, is an indicator of change in metabolism or filtration effectiveness. Urine is considered the sample specimen of choice for most biomarker studies, mostly because it is easily accessible and can be obtained in a non-invasive manner [267, 268]. Moreover, since it is composed of waste metabolites, there are few ethical issues in collecting it.

Various population studies have found that the variation of the urinary metabolic profiles is high not only between individuals but also changes within the same individual over time [266]. There are number of factors that can contribute to this variation. Since urine represents the waste products of the metabolism, the changes in the biochemical profile are most likely to be caused by diet and administration of drugs [269, 270]. It is also known that post puberty, males tend to have different metabolic profiles than females and at least 12 statistically significant differential metabolites had been identified including creatine, creatinine and citrate [266, 271]. Other factors affecting urinary metabolic profile are age, BMI, and even the time of day when the sample was collected [271]. There are a vast number of metabolic profiling studies of urine conducted in adults but far fewer in children, particularly in premature babies [268]. However, one study reported that gestational age and postnatal age can affect the changes in urinary metabolic profile (increase in the concentration of glycine and

threonine had been observed in relation to increasing postnatal age) [272]. Studies comparing urinary metabonome of preterm and full-term babies found that premature babies had lower concentration of formate, hippurate, betaine, tryptophan, phenylalanine, malate, tyrosine, hydroxybutyrate, *N*-acetylglutamate [267], proline and taurine [267, 268]. At the same time bile acids were found to decrease in concentration with increasing gestational age [273, 274].

Post-menstrual age (PMA) is a combination of gestational age (GA) with postnatal age (PNA) and can be explained by the formula: $PMA = GA + PNA - 1$. Knowing the post-menstrual age is essential for the urinary metabonome analysis as the development of kidneys is programmed genetically and they are mature at the 36th week of gestation [158, 275]. The development of kidneys has been shown to continue after premature birth resulting in larger volume of kidneys (after correcting for the body weight) than the term babies when compared at the same PMA [158]. In this chapter the dynamic metabolic changes in urine associated with development of the preterm infant are characterised taking PMA, GA, PNA and feeding strategy into account.

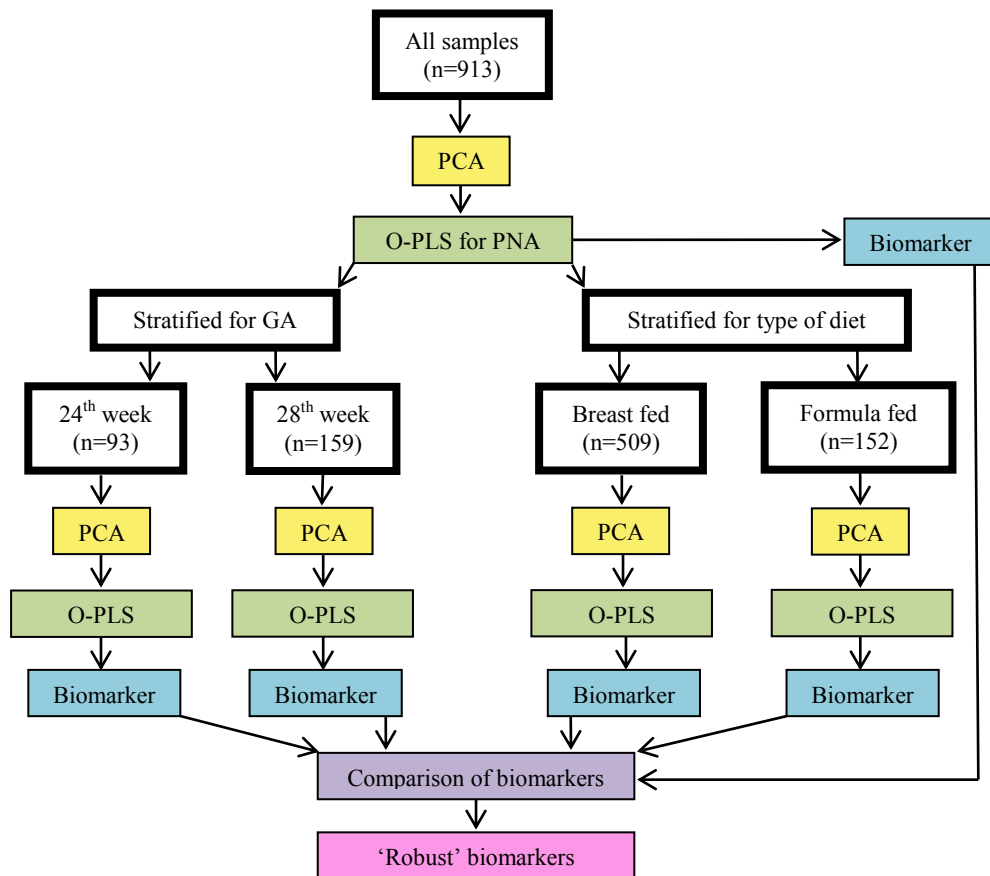


Figure 34: Flowchart presenting the analysis pipeline for urine samples obtained from premature babies characterising changes in the urinary metabonome associated with postnatal age (PNA).

4.2 Experimental design

4.2.i Urine sample preparation

In total 913 urine samples were eligible for the modeling. Please see the Materials and Methods chapter for the detailed sample preparation.

4.2.ii Nuclear Magnetic Resonance spectra acquisition

The acquisition of all the urine samples took approximately a month due to a large sample set and several issues with the NMR machine. For the detailed information on the acquisition of NMR spectra please see the Materials and Methods chapter. The parameters used for the standard pulse sequence of the urine sample were: P1=10.54 μ s, P11=11.10-dBW, P19=48.44dB and O1=2820.74Hz.

4.2.iii Sample pre-processing

Spectra for the NEON urine samples were preprocessed and modeled in Matlab. Initially a model was build using all spectra and the quality controls (QCs) in order to ascertain if there was any instrumental batch effect. The Principal Component Analysis (PCA) scores plot for the model was calculated using 3 principal components returning a cumulative R^2X of 18.9%. The quality control samples were clustering in the centre of the PCA scores plot suggesting that appropriate method of normalization had been chosen and that the method was reproducible (Figure 35). It was important to confirm that the prolonged acquisition time (over a month) did not lead to a batch effect in the sample set. Subsequent stages of analysis were undertaken after removal of the QCs from the model. The global model for all samples (including multiple time points) that is presented in the results section (section 4.3) was calculated using 3 principal components (PC) giving a cumulative R^2 of 19.1%. Since the cumulative variance for three components only reached 18.9%, this indicated that the dataset contained a high degree of chemical variance. The PCA scores plots presented in this chapter for the whole data set were created based on this model.

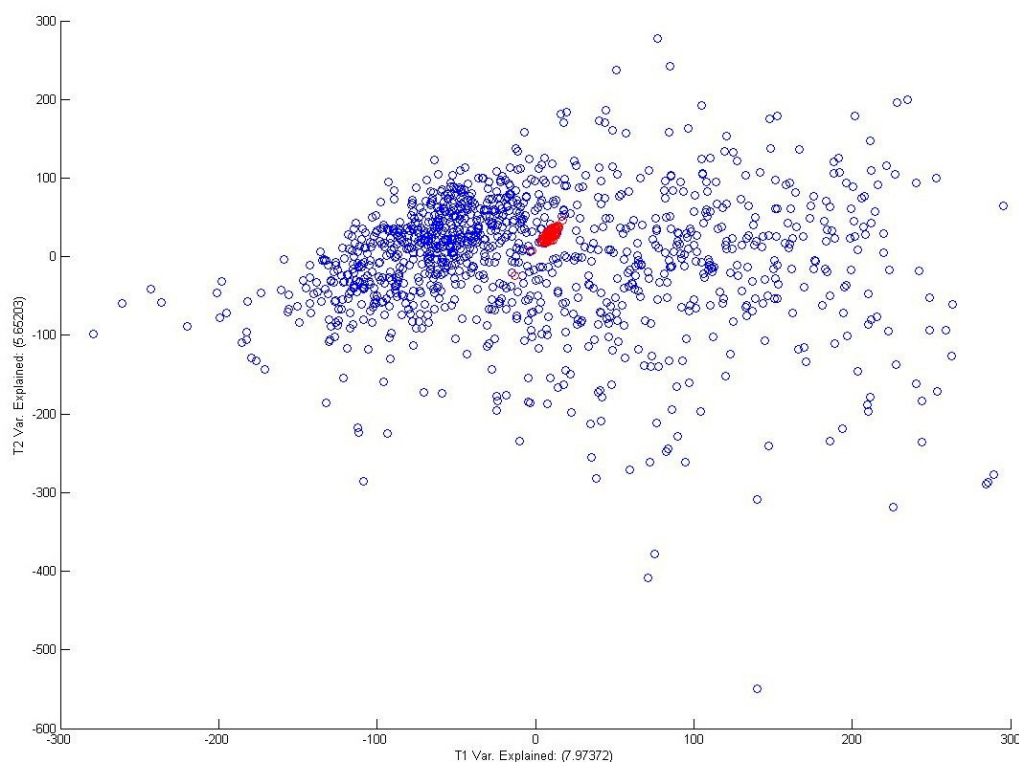


Figure 35: Scores scatter plot of the principal component analysis (PCA) model of urine NMR spectral data generated from samples collected at multiple time points (whole data set). Quality control samples are coloured red.

The first step of the analysis was to check if the randomized groups had different metabolic profiles, if not then the samples could be merged into further exploratory analysis. Initially, in order to get an overview of the dataset and to identify any biological outliers, for the exploratory analysis all spectra corresponding to multiple time points for 160 babies were modeled together. The unsupervised models for multiple time points was coloured according to classes obtained from the NEON metadata to identify the factors contributing to the main source of the variation in the scores plot. Those variables include: a) postnatal age, b) gestational age at birth, c) mode of delivery, d) presence of labour, e) sex, f) route of nutrition (parenteral or enteral), and g) type of enteral nutrition (formula milk or human milk). Another step in the preliminary analysis was to carry out a supervised analysis – Orthogonal Partial Least Squares model (O-PLS). This approach allowed finding discriminant metabolites associated with chosen classes of samples. The preliminary analysis helped to understand which of the known factors need to be controlled or accounted for while assessing the influence of the other factors. Several models exploring these class variables were undertaken in smaller subsets (Figure 34).

4.3 Results

4.3.i Metabolic profiles associated with randomised PN feeding group

The analysis on randomised groups was done after subtracting samples from the whole data set. Three time points were chosen: (a) 2nd to 9th day of life (babies exposed to parenteral nutrition), (b) 15th to 22nd day of life (babies off the trial parenteral nutrition), and (c) last sample collected provided it was after the 32nd week of life. If there was more than one sample available in particular time frame for a baby then the one collected earlier was included. All three models were calculated using 3 principal components and the cumulative R^2 equalled: (a) 24.2% (b) 20.8% and (c) 20.6%.

4.3.i.a Intralipid vs. SMOF lipid

The PCA scores plots showed no separation according to the type of lipid solution that the babies had been given intravenously (Appendix: Figure 78, Figure 79 and Figure 80). The O-PLS models for the three previously selected time points confirmed the lack of difference between the groups as the Q^2Y values were negative (Figure 36).

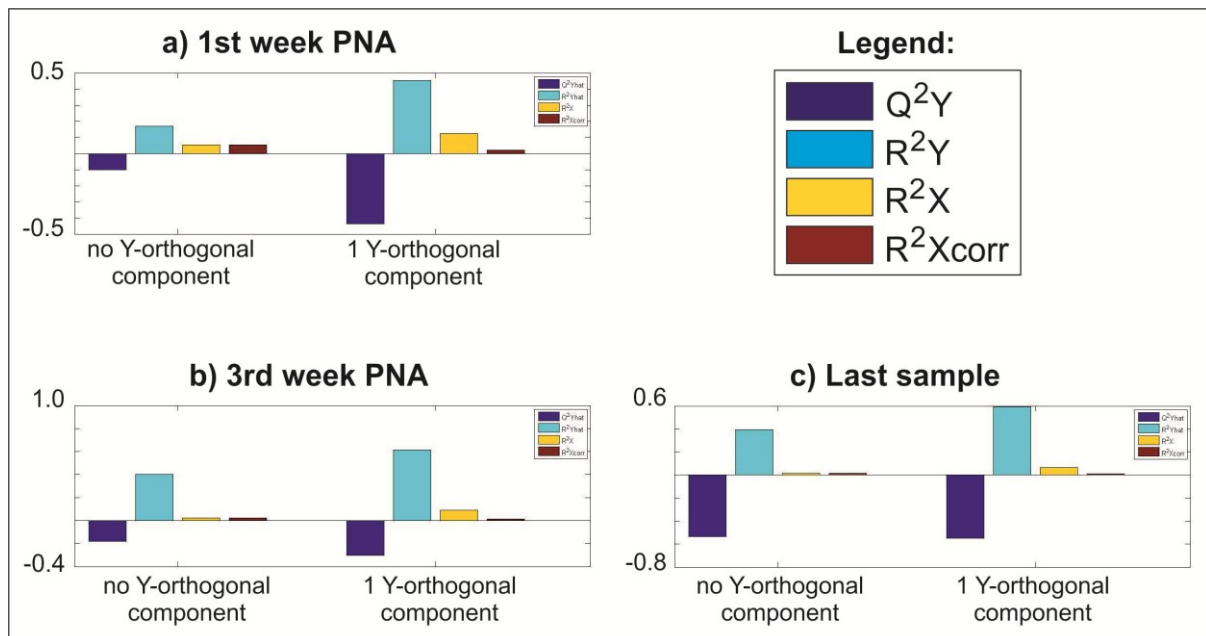


Figure 36: The O-PLS models for urine samples comparing the types of lipids. The R^2Y and Q^2Y values for the O-PLS of urinary metabolic profiles representing samples collected (A) between 2nd-9th day of life, (B) between 15th-21st day of life, and (C) the last sample prior to discharge. The dark blue bar presents Q^2Y while the light blue – R^2Y . The bars on the right of each figure show the change after introduction of the orthogonal component.

4.3.i.b Recommended daily intake (RDI) vs. increasing concentration of amino acid

When analysing the concentration of amino acids, the PCA model showed no separation (Appendix: Figure 81, Figure 82 and Figure 83). Additionally the O-PLS models created for each of the preselected time points returned a negative Q^2Y values. This means that there is no significant difference between the groups randomised according to the concentration of amino acids (Figure 37).

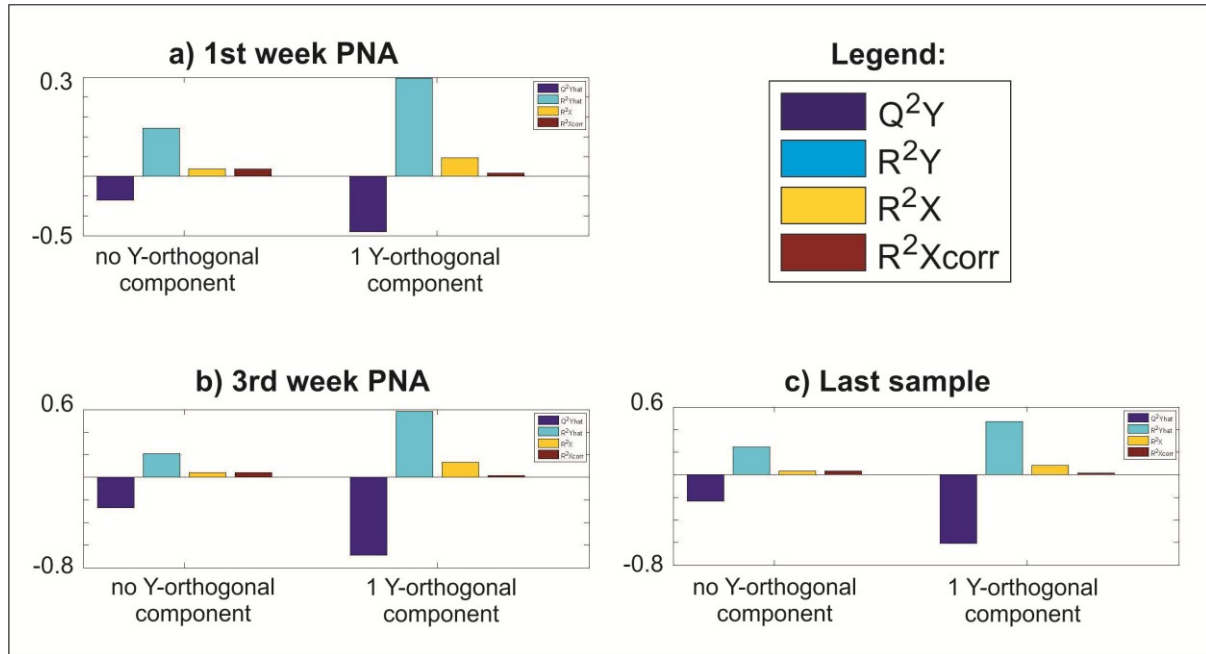


Figure 37: The O-PLS models for urine samples comparing the concentrations of amino acids. The R^2Y and Q^2Y values for (A) whole sample set; samples collected (B) at 1st week of life, (C) 3rd week of life, (D) from babies born at 24th week of GA, (E) from babies born at 28th week of GA. The dark blue bar presents Q^2Y while the light blue - R^2Y . The bars on the right of each figure show the change after introduction of the orthogonal component.

4.3.ii Metabolic profiles associated with gestational age (GA)

The Principal Component Analysis scores plot for components PC1 vs. PC2 represented the most evidence of clustering according to GA (Figure 38) with earlier GA showing a tendency to cluster to the right hand side of the plot and older GA groups clustering on the left.

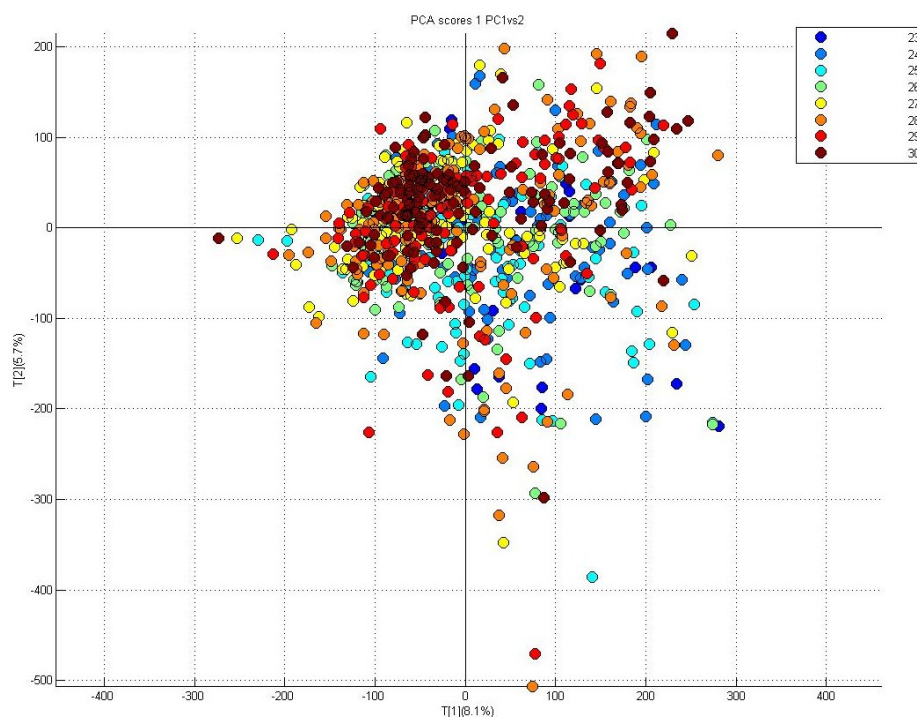
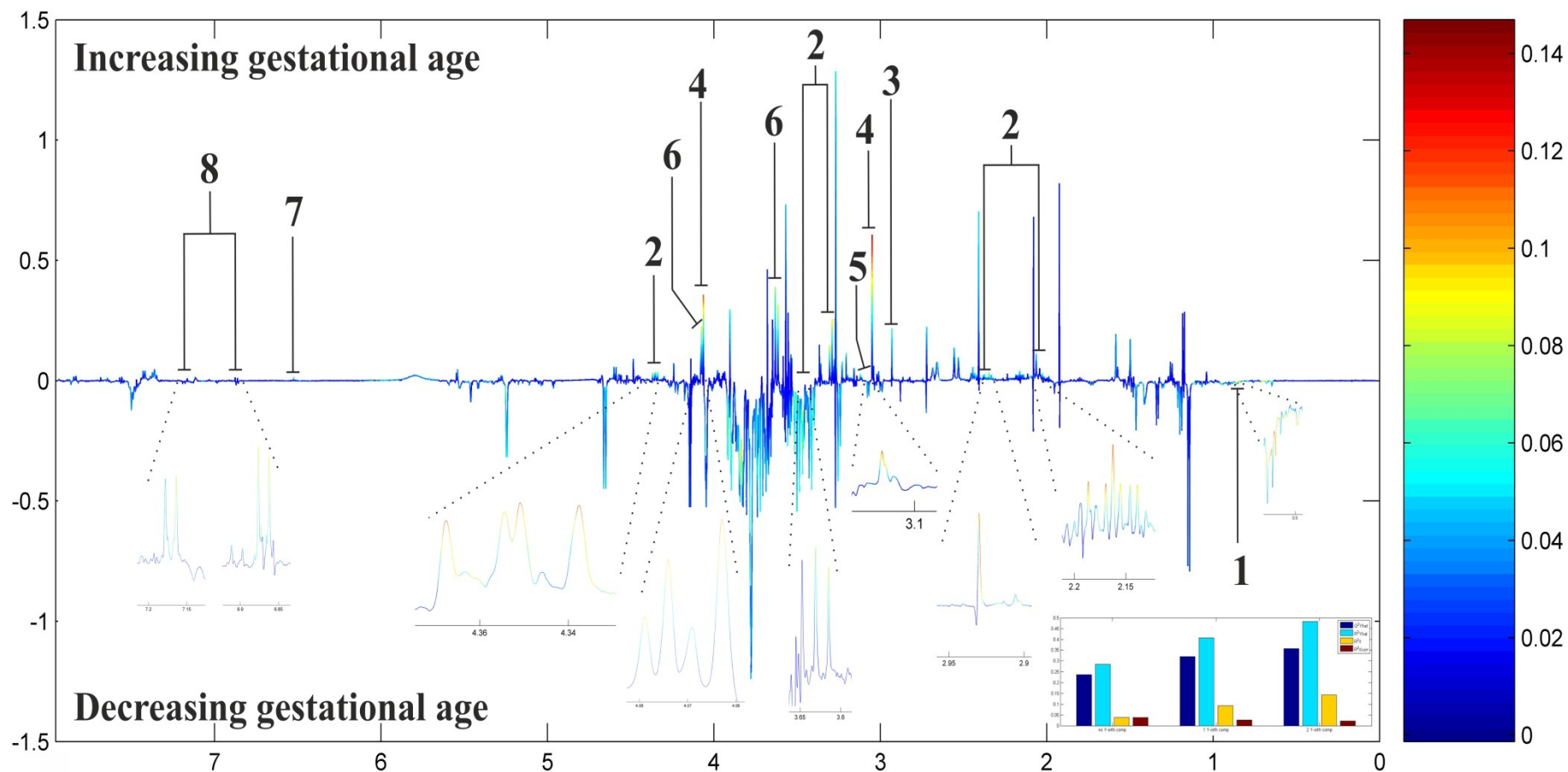


Figure 38: PCA Scores Scatter Plot of whole data set of NMR urine spectra. Samples coloured according to gestational age (in weeks)

A supervised model (O-PLS) was built with two orthogonal components resulting in a R^2Y value of 47% and a Q^2Y value of 36%. The O-PLS score plot shows the results for the increasing GA from the bottom to the top (Figure 39). Metabolites associated with higher GA included creatinine (3.05ppm s, 4.06ppm s), myo-inositol (3.29ppm t, 3.52ppm dd, 3.61ppm t, 4.05ppm t), unidentified molecule (3.62ppm d, 3.64ppm t, 3.74ppm s), hydroxyproline (2.16ppm multi, 2.43ppm multi, 3.37ppm dt, 3.47ppm dd 4.35ppm dd) and fumarate (6.52ppm s). Bile acids (0.84ppm s) were associated with lower GA (Figure 39). The permutation tests to validate the significance of the Q^2Y showed that the model was not robust as the p -value = 0.91.

The influence of postnatal age was removed by creating models on subsets of samples that were collected at the same time points. The choice of the postnatal age time point was made based on the number of available samples. The first analysed time point included samples collected between the 2nd and the 9th day of life (126 samples). The second time point comprised the samples collected between the 15th and the 22nd day of life (104 samples). Those time points will be referred to as 1st and 3rd week of postnatal age, respectively.



The PCA scores plot for the samples collected in the 1st week of life was calculated using three principal components giving a cumulative R^2X of 24.2%. The PCA scores plot showed no clustering based on the GA (Appendix: Figure 84). The supervised O-PLS model conducted for the continuous variable (gestational age in weeks) showed that the reliability of the model (Q^2Y) increased from 28% to 41% after the introduction of the orthogonal component. The p -value for the Q^2Y was 0.49 which means that the model was not robust. The O-PLS loadings plot showed that methylhydantoin (2.93ppm s) and creatinine (3.05ppm s, 4.06ppm s) were associated with the higher gestational age. At the same time the oligosaccharide region (3.70ppm-3.90ppm s) was found to be associated with the lower GA.

The PCA scores plot for the samples collected on the 3rd week of life was also calculated from 3 principal components and the cumulative R^2X equaled 25.8%. There was no clustering visible in the PCA scores plot (Appendix: Figure 85) and the Q^2Y (37%) for the O-PLS dropped after introduction of the orthogonal component, indicating that the model was not stable. The PLS model was robust as the p -value from the permutation tests was 0.01. The loadings plot showed that creatinine (3.05ppm s, 4.06ppm s), malonate (3.12ppm s) and 4-hydroxyphenylacetate (3.45ppm s, 6.87ppm d, 7.17ppm d) were associated with the higher GA. No metabolites were found to be associated with the lower GA.

4.3.iii Metabolic profiles associated with postnatal age (PNA)

Both PCA score plots (for PC1 vs. PC2, and for PC1 vs. PC3) showed clustering according to postnatal age (PNA) presented in weeks. The Figure 40 shows the PCA scores plot for PC1 vs. PC2.

The O-PLS model was calculated using four orthogonal components which indicate a very high variability between the samples as the orthogonal components remove the systematic differences caused by the natural variability, but not related to the class of interest (Figure 41). The final R^2Y and Q^2Y values equaled 82% and 72% respectively. The model was robust as the p -value from the permutation tests was 0.01. The O-PLS loadings plot showed that metabolites associated with increasing PNA were: pantothenate – vit B₅ (0.90ppm s, 0.94 ppm s). There were no metabolites associated with lower PNA (Figure 41).

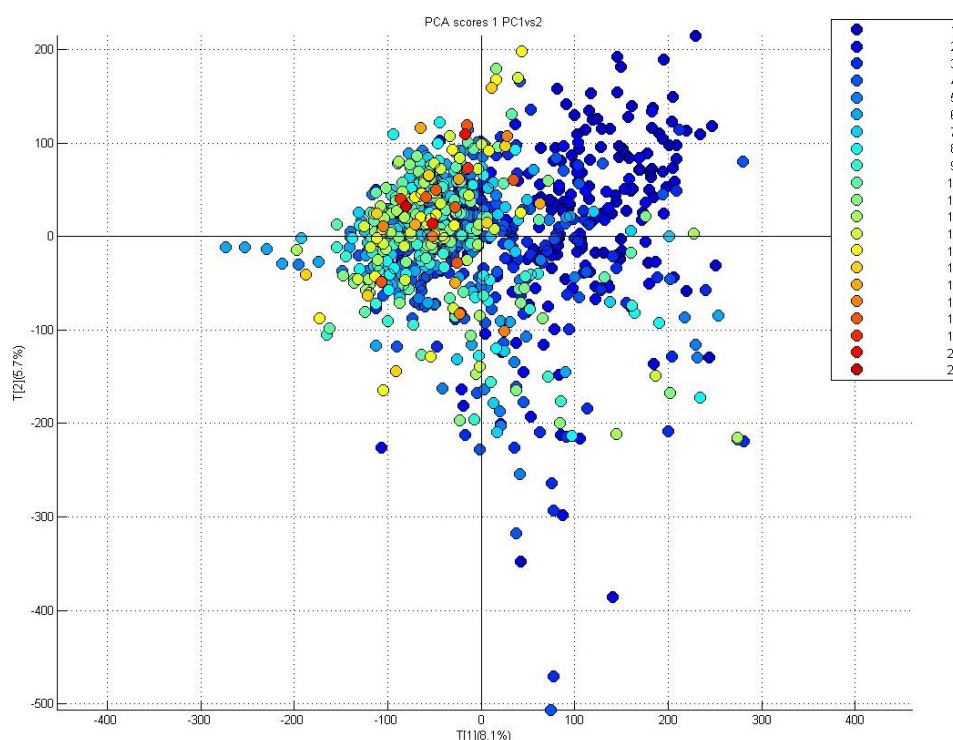


Figure 40: PCA Scores Scatter Plot of whole data set of NMR urine spectra. Samples coloured according to postnatal age (in weeks)

In order to remove the potential bias of the variable GA, the analysis was continued in stratified subgroups as presented in Figure 34. Both PCA scores plot for the 24th GA and for 28th GA were calculated using 3 principal components giving cumulative R^2X values of 24% and 26.9% respectively. A clear separation of samples (related to PNA) was visible on PC1 in both cases (Appendix: Figure 86 and Figure 87). The O-PLS model for the samples collected from babies born on the 24th week of GA was built with 2 orthogonal components and final R^2Y and Q^2Y values equaled 95% and 82% respectively. The O-PLS model for the samples collected from babies born on the 28th week of GA was built with only one orthogonal component and the R^2Y equaled 77% while Q^2Y 63%.

After removing the GA variable, the O-PLS scores plots confirmed the findings for pantothenate – vit B₅ (0.90ppm s, 0.94 ppm s). Additionally acetyl carnitine (2.13ppm s, 2.50ppm dd, 3.18ppm s, 3.62ppm d, 3.83ppm dd), unidentifiable chemical compound (2.33ppm dd), unidentifiable chemical compound (0.91ppm t, 2.78ppm multi), trimethylamine (2.90ppm s) and malonate (3.12ppm s) were associated with increasing PNA. At the same time mixture of sugars (3.33ppm multi, 3.38ppm s,

3.44ppm s, 3.49ppm s, 3.67ppm d, 3.69ppm t, 3.75ppm multi, 3.83ppm d, 3.85ppm d, 4.04ppm t, 4.14ppm d, 4.20ppm s, 4.24ppm s, 4.51ppm s, 4.96ppm t, 5.03ppm d, 5.11ppm s, 5.24ppm multi, 5.37ppm s, 5.40ppm s, 6.49ppm d) had been identified as associated with lower PNA.

Due to the fact that nutrition is known to have a potential impact on the metabolic profile of urine, samples were stratified based on the type of enteral nutrition (exclusively fed by human milk, and exclusively formula fed) (Figure 34). This gave two models, both of which were calculated with 3 principal components. The first model (samples collected only from exclusively human milk fed babies) gave cumulative R^2X of 20.1%.

The PCA scores plots showed clear separation based on PNA and the O-PLS was build using 5 orthogonal components and gave final R^2Y of 90% and final Q^2Y of 71%. The second model was built for samples collected from exclusively formula-fed infants and gave cumulative R^2X values of 21.2%. The PCA scores plots did not show a strong separation based on PNA in formula-fed infants. The O-PLS model was build using 4 orthogonal components and the final R^2Y and Q^2Y values equaled 95% and 62% respectively.

After removing the variability of diet the O-PLS scores plots showed that higher PNA was associated with pantothenate – vit B₅ (0.90ppm s, 0.94 ppm s) and acetyl carnitine (2.13ppm s, 2.50ppm dd, 3.18ppm s, 3.62ppm d, 3.83ppm dd). Findings for the time mixture of sugars (3.33ppm multi, 3.38ppm s, 3.44ppm s, 3.49ppm s, 3.67ppm d, 3.69ppm t, 3.75ppm multi, 3.83ppm d, 3.85ppm d, 4.04ppm t, 4.14ppm d, 4.20ppm s, 4.24ppm s, 4.51ppm s, 4.96ppm t, 5.03ppm d, 5.11ppm s, 5.24ppm multi, 5.37ppm s, 5.40ppm s, 6.49ppm d) had been confirmed. Additionally unidentifiable chemical compound (2.04ppm s, 3.60ppm t, 3.92ppm s, 3.95ppm s, 3.96ppm s, 3.98ppm t, 7.91ppm s) was correlated with lower PNA.

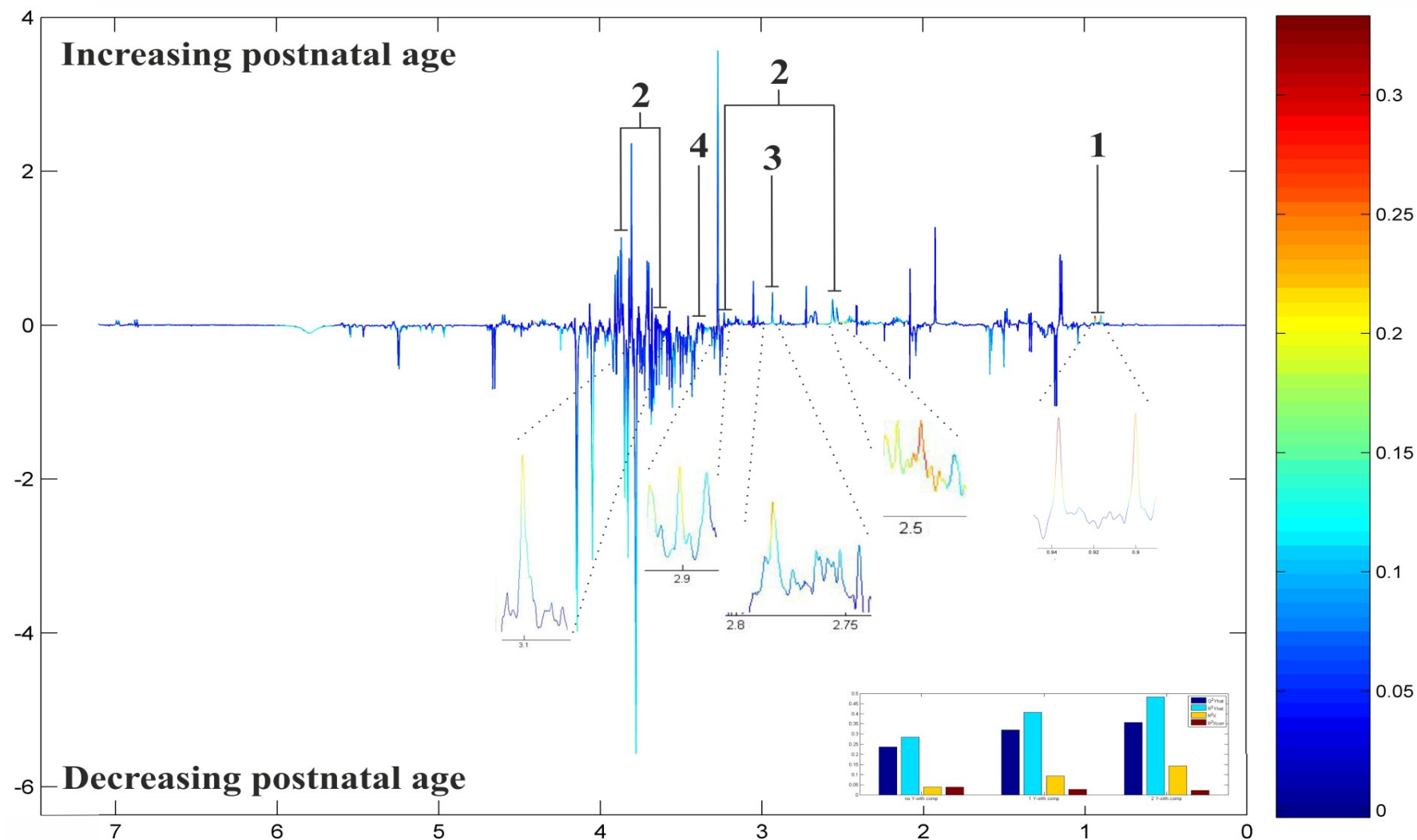


Figure 41: Orthogonal Partial Least Squares (O-PLS) loadings plot of urine metabolic profiles coming from samples collected at multiple time points and created in accordance to the postnatal age. The X axis shows chemical shift in ppm, while the Y axis represents metabolites associated with (increasing – up or decreasing – down) postnatal age. Annotated metabolites: 1. pantothenate, 2. acetyl carnitine, 3 trimethylamine, 4 malonate. The insert bottom right: R^2Y (light blue bar) and Q^2Y (dark blue bar) values showing 82% and 72% respectively. The permutation tests for Q^2Y retrieved p -value = 0.01.

4.3.iv Metabolic profiles associated with postmenstrual age (PMA)

The PCA score plots showed clustering according to PMA presented in weeks. This was especially visible in the plot of PC1 vs. PC2 (Figure 42). The O-PLS model was built using 4 orthogonal components and final R^2Y and Q^2Y values equaled 80% and 67% respectively (Figure 43). The model was robust as the p -value from the permutation tests was 0.01. The O-PLS scores plot showed that metabolites associated with increasing PMA were the same as for increasing PNA: pantothenate – vit B₅ (0.90 ppm s, 0.94 ppm s) (Figure 43). There were no metabolites associated with decreasing PMA.

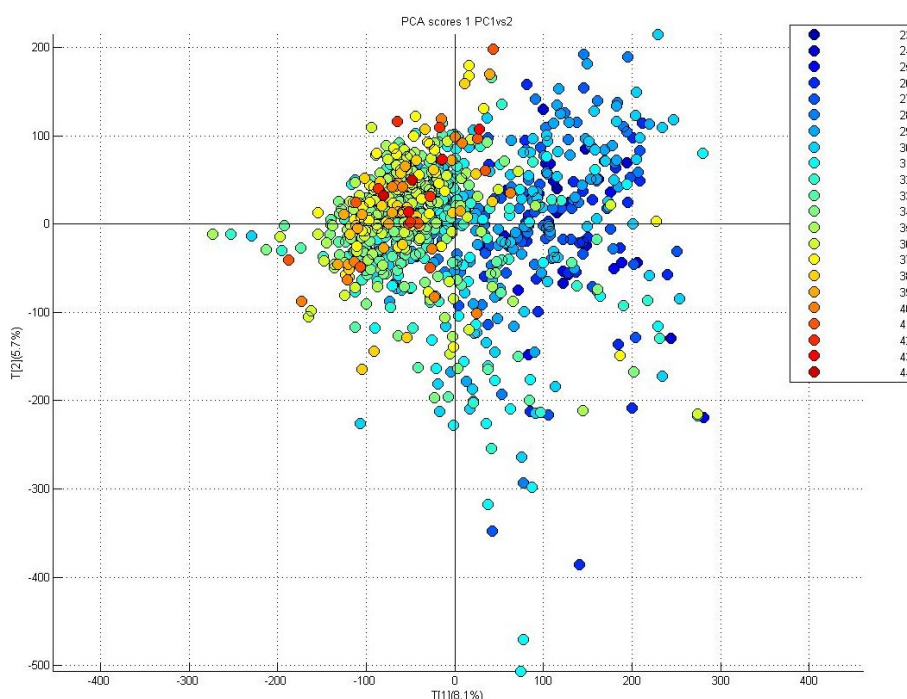


Figure 42: PCA Scores Scatter Plot of whole data set of NMR urine spectra. Samples coloured according to postmenstrual age (in weeks).

Figure 43: Orthogonal Partial Least Squares (O-PLS) loadings plot of urine metabolic profiles coming from samples collected at multiple time points and created in accordance to the postmenstrual age. The X axis shows chemical shift in ppm, while the Y axis represents metabolites associated with (increasing – up or decreasing – down) postmenstrual age. Annotated metabolites: 1. pantothenate. The insert bottom right: R²Y (light blue bar) and Q²Y (dark blue bar) values showing 80% and 67% respectively. The permutation tests for Q²Y retrieved *p*-value = 0.01.

4.3.v Metabolic profiles associated with sex

The PCA scores plot showed no clustering according to the sex of the baby (Appendix: Figure 88). Additionally, the Q^2Y in the O-PLS analysis was below the threshold of 5% (Figure 46). After working on each postnatal age separately, no difference between no difference about male and female urinary metabonomes was observed as the Q^2Y value was negative. The same was observed when analysing the samples collected only from babies born in the 24th week of GA, and from babies born only on the 28th week of GA (Figure 46).

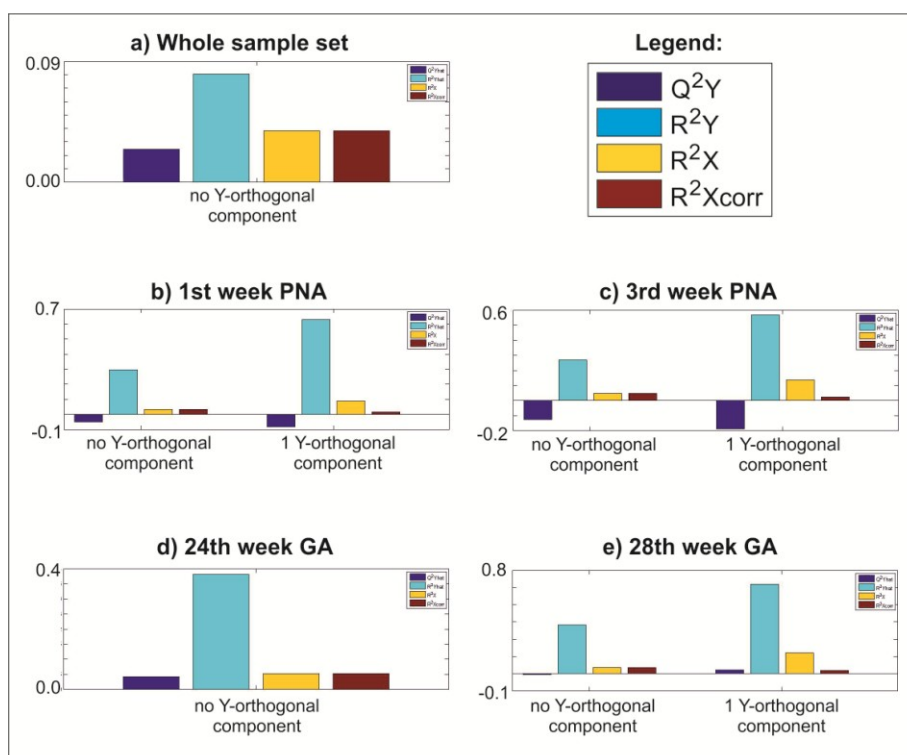


Figure 44: The O-PLS models for urine samples comparing sex of baby. The R^2Y and Q^2Y values for the O-PLS of urinary metabolic profiles for (A) whole sample set; samples collected (B) at 1st week of life, (C) 3rd week of life, (D) from babies born at 24th week of GA, (E) from babies born at 28th week of GA. The dark blue bar presents Q^2Y while the light blue - R^2Y . The bars on the right of each figure show the change after introduction of the orthogonal component.

4.3.vi Metabolic profiles associated with mode of delivery

The NEON dataset provided a detailed information about mode of delivery which was divided into: a) vaginal - natural, b) vaginal – with use of forceps, c) Caesarean elected – in labour, d) Caesarean elected – not in labour, e) Caesarean emergency – in labour and f) Caesarean emergency – not in labour. However in order to simplify analysis and have a reasonable number of babies in each

analysed groups, the analysis were done for vaginal delivery vs. Caesarean section, and in labour vs. not in labour.

The PCA model for the whole sample set showed no separation based on either mode of delivery or occurrence of labour (Appendix: Figure 89 and Figure 90). The supervised O-PLS model analysing the differences between the modes of delivery was created using one orthogonal component and the final R^2Y value was 23% while Q^2Y value was 13% (Figure 47). Although the Q^2Y value was above the threshold of 5%, the predictability of the model was very low indicating that the variance attributable to mode of delivery was much less, for example, than that attributed to GA.

Before identifying the discriminant metabolites, further OPLS analysis were performed on the subsets of samples collected: a) in the 1st week of life, b) in the 3rd week of life, c) from babies born in the 24th week of GA, and d) from babies born at the 28th week of GA. The results for all of the four subgroup analysis showed that there is no difference in metabolic composition between the vaginal delivery and Caesarean section as the Q^2Y values were either negative or below the threshold of 5% (Figure 47).

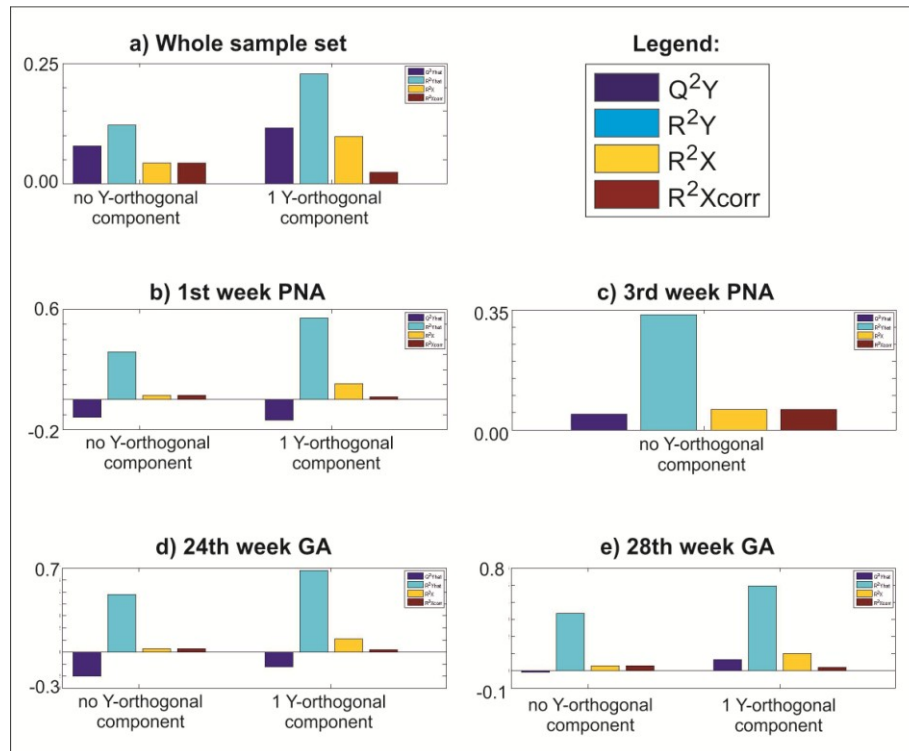


Figure 45: The O-PLS models for urine samples comparing mode of delivery. The R^2Y and Q^2Y values for the O-PLS of urinary metabolic profiles for (A) whole sample set; samples collected (B) at 1st week of life, (C) 3rd week of life, (D) from babies born at 24th week of GA, (E) from babies born at 28th week of GA. The dark blue bar presents Q^2Y while the light blue - R^2Y . The bars on the right of each figure show the change after introduction of the orthogonal component.

Similar results were observed for the O-PLS models comparing the samples collected from babies born while the mother was in labour with those born while the mother was not in labour. The Q^2Y values for the whole sample set and four subgroups were negative or below 5% (Figure 48).

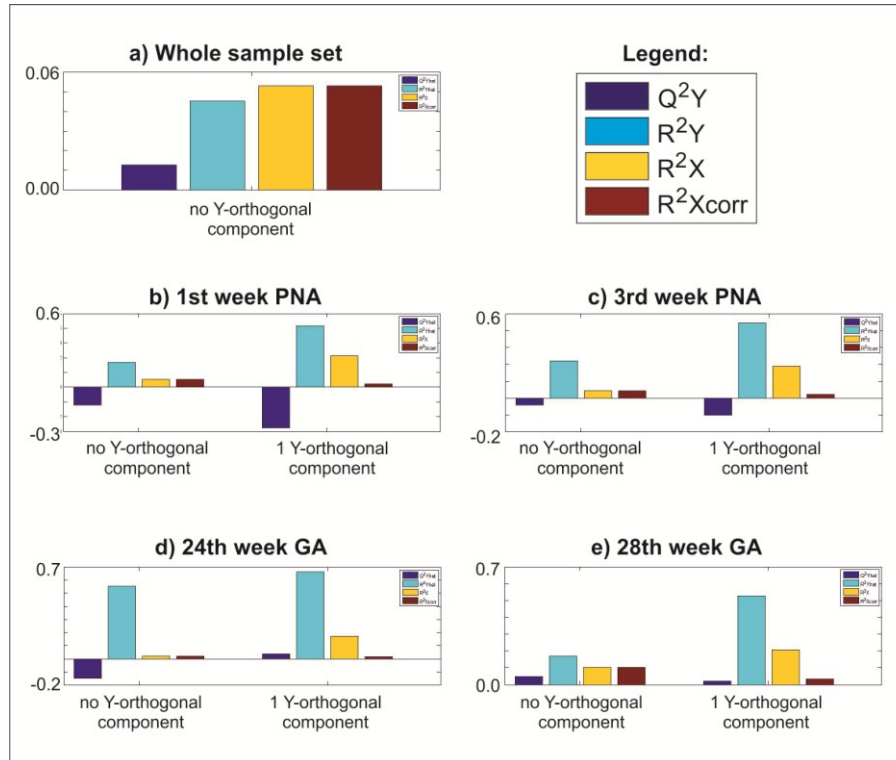


Figure 46: The O-PLS models of urine samples created for presence of labour. The R^2Y and Q^2Y values for the O-PLS of urinary metabolic profiles for (A) whole sample set; samples collected (B) at 1st week of life, (C) 3rd week of life, (D) from babies born at 24th week of GA, (E) from babies born at 28th week of GA. The dark blue bar presents Q^2Y while the light blue - R^2Y . The bars on the right of each figure show the change after introduction of the orthogonal component.

These findings suggest that the mode of delivery or presence of labour does not affect the urinary metabolic profile, although when the whole data set was used then some evidence of effect on mode of delivery was seen. Since this effect is very subtle, it would require validation with a much larger test set of samples.

4.3.vii Metabolic profiles associated with diet

Due to the complex nature of the feeding regimes for the premature babies, there are number of factors that need to be accounted for. The first step undertaken was the comparison of samples collected while the baby was on parenteral nutrition (PN) with those collected while the baby was on enteral nutrition (EN).

As previously mentioned parenteral nutrition is an intravenous injection of nutrients. This route leads to the distribution of the macroelements to the tissues without reaching the lumen of the gut. At the same time, the liver of the baby on parenteral nutrition is affected differently than when baby is fed orally. Enteral nutrition on the other hand, is the food given orally therefore the nutrients must be absorbed through the digestive tract.

Second diet analysis included comparison between two types of enteral nutrition – human milk and formula milk. The composition of the human milk and the formula milk which is based on cow's milk differs. One of such examples is the presence of human milk oligosaccharides in the breast milk. Those molecules are known to support the growth of beneficial bacteria and reduce the chances of the adhesion of pathogenic bacteria to the lumen of the gut. Additionally, human milk, when not pasteurised, is a source of important beneficial bacteria. Those factors have a potential to contribute to the metabolic profile of the faecal water

4.3.vii.a Parenteral nutrition vs. enteral nutrition

The PCA scores plot showed clustering of samples according to the route of nutrition (Figure 47). TPN (1) corresponds to total parenteral nutrition which is simply a parenteral nutrition without any enteral supplementation. PN + EN (2) are a mixture of parenteral nutrition and enteral nutrition while EN (3) is an exclusive enteral nutrition. Clear metabolic differences were apparent relating to the route of nutrition and were still retained after removing the potential bias from such variables as PNA or GA.

Due to the fact that there was no information on the ratio of EN to PN, the analysis were simplified by classifying samples as collected either when no PN was given (thus purely enteral nutrition) or when baby was on support in form of PN (includes TPN and PN+EN).

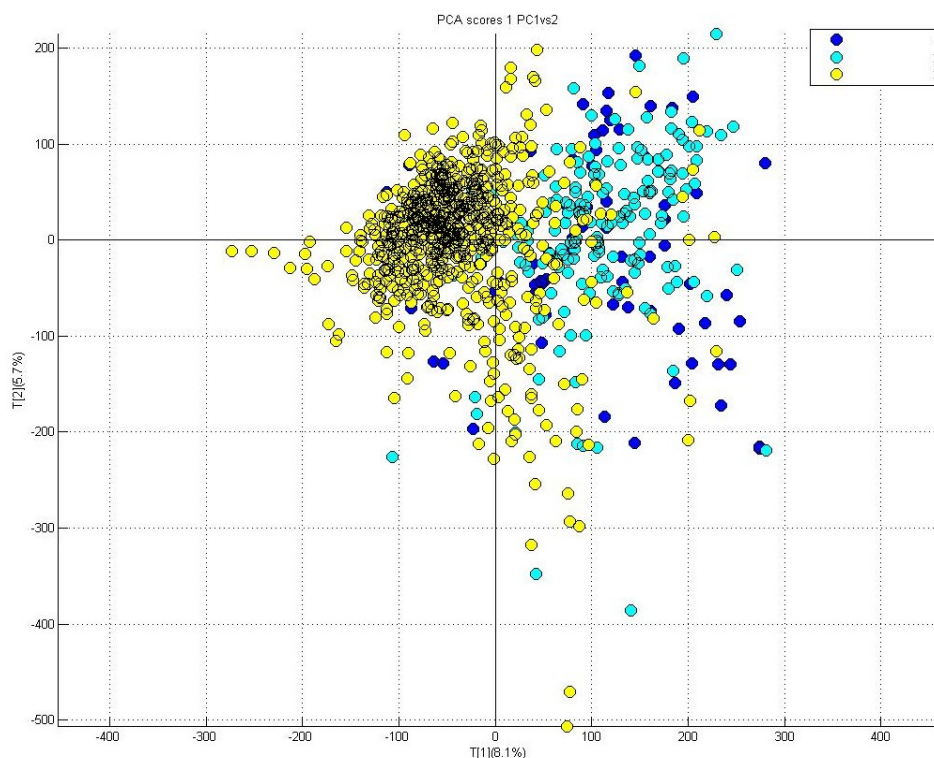


Figure 47: PCA Scores Scatter Plot of whole sample set of NMR urine spectra. Samples are coloured according to feeding route. Dark blue (1) represents total parenteral nutrition (TPN), light blue (2) is a mixture of parenteral nutrition (PN) with enteral nutrition (EN) and yellow (3) is enteral nutrition (EN).

The O-PLS model was build using 2 orthogonal components with the final R^2Y and Q^2Y values being 68% and 60% respectively. The model was robust as the p -value from the permutation tests was 0.01. There were no metabolites associated with exclusive enteral nutrition. However, the presence of parenteral nutrition was associated with: myo-inositol (3.27ppm s, 3.29ppm t, 3.54ppm dd, 3.62ppm t, 4.06ppm t) and a mixture of sugars (3.22ppm s, 3.33ppm multi, 3.38ppm s, 3.42ppm s 3.44ppm s, 3.49ppm s, 3.51ppm s, 3.67ppm d, 3.69ppm t, 3.72ppm s, 3.75ppm multi, 3.83ppm d, 3.85ppm d, 3.89ppm s, 3.91ppm s, 4.04ppm t, 4.14ppm d, 4.20ppm s, 4.24ppm s, 4.51ppm s, 4.65ppm d, 4.96ppm t, 5.03ppm d, 5.11ppm s, 5.24ppm multi, 5.37ppm s, 5.40ppm s, 6.49ppm d) (Figure 58).

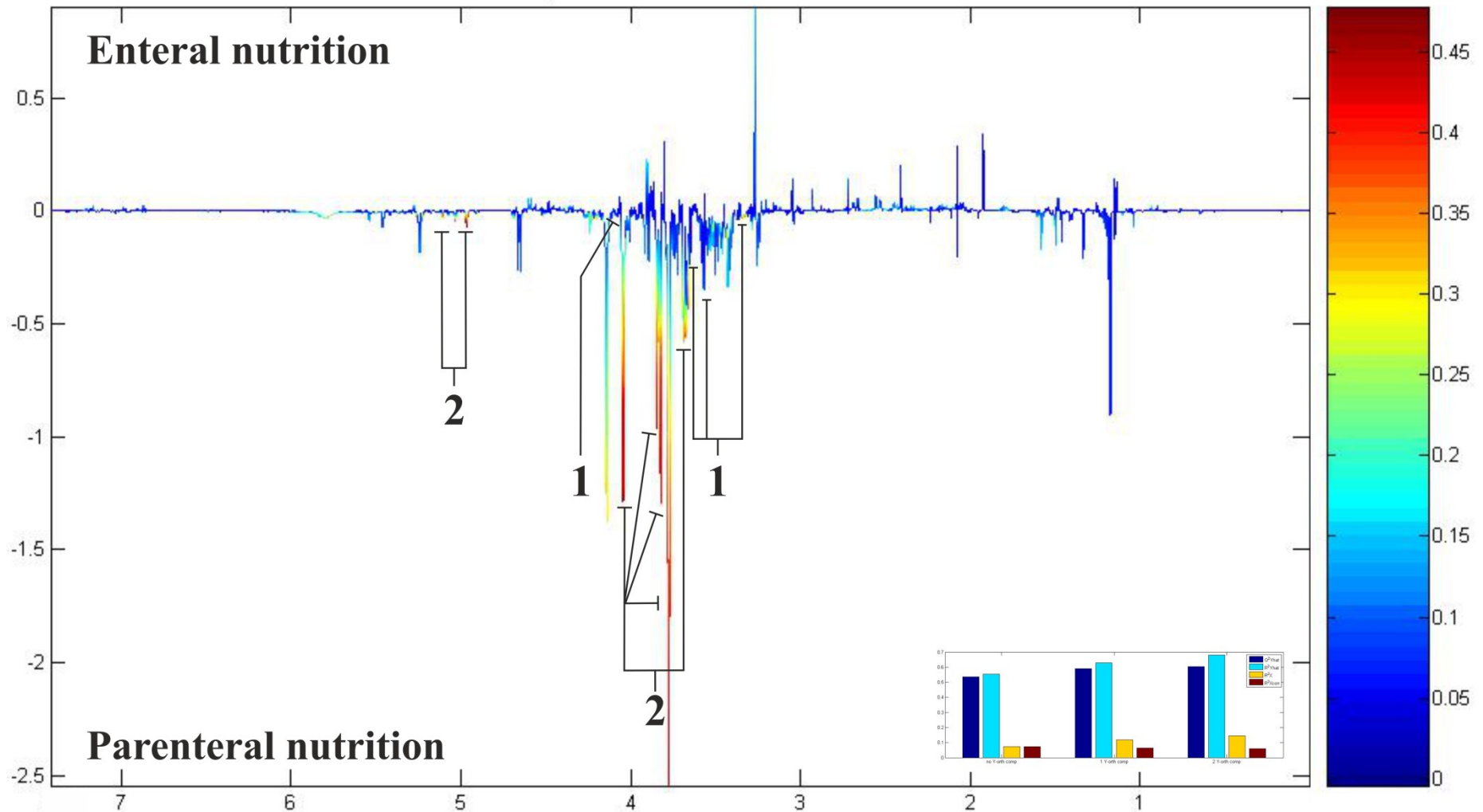


Figure 48: Orthogonal Partial Least Squares (O-PLS) loadings plot of urine metabolic profiles coming from samples collected at multiple time points and created in accordance to the route of nutrition. The X axis shows chemical shift in ppm, while the Y axis represents metabolites associated with enteral nutrition (up) or parenteral nutrition (down). Annotated metabolites: 1. myo-inositol, 2. sugars. The insert bottom right: R^2Y (light blue bar) and Q^2Y (dark blue bar) values showing 68% and 60% respectively. The permutation tests for Q^2Y retrieved p -value = 0.01.

4.3.vii.b Human milk vs. formula milk

The analysis focused on the influence of human milk thus samples were classified as either a) human milk given (includes exclusive and mixed feeding) or b) formula milk given. The PCA model showed a slight separation of samples based on the type of milk (Figure 49).

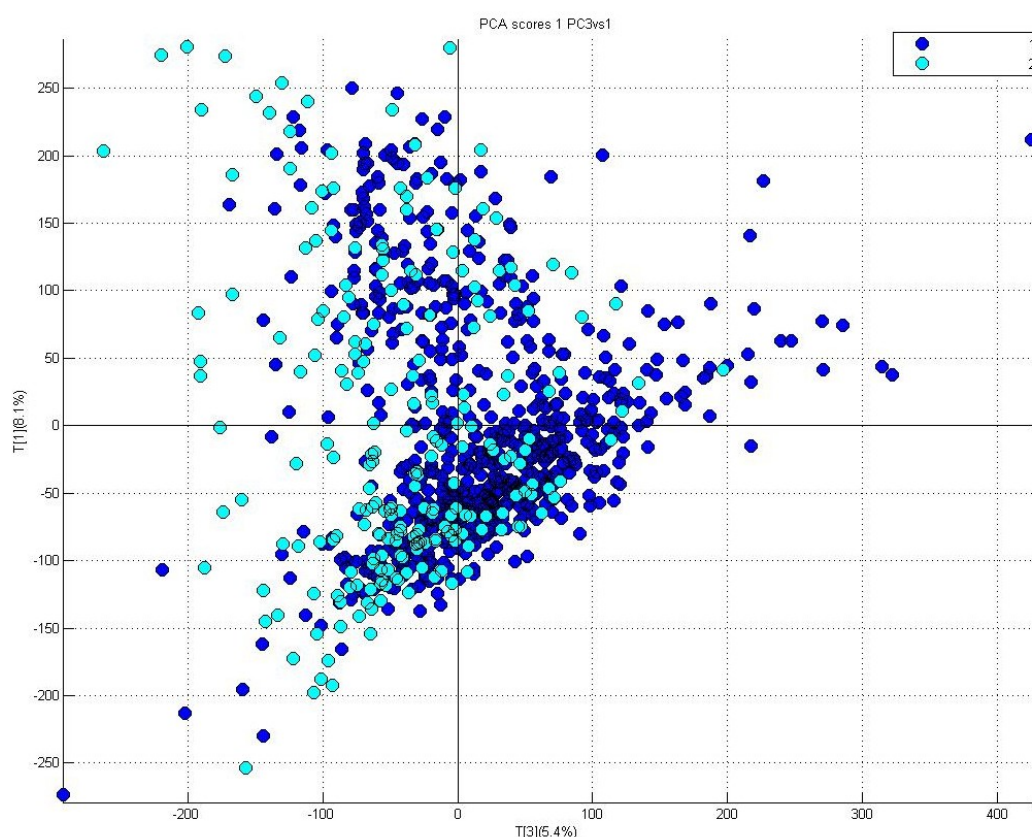


Figure 49: PCA Scores Scatter Plots of whole sample set of NMR urine spectra (principal component 3 vs. principal component 1) coloured according to whether human milk was given (dark blue) or formula milk was given (light blue).

The O-PLS model was done using one orthogonal component and the R^2Y and Q^2Y values equalled 43% and 33% respectively. The model was robust as the p -value from the permutation tests was 0.01. Introduction of formula milk is associated with an unidentifiable chemical compound 1 (0.90ppm t, 1.70ppm multi, 1.78ppm s, 2.33ppm dd, 4.00ppm dd), and chemical compound 2 (1.87ppm multiplet, 2.33ppm dd, 2.38ppm d, 3.10ppm t, 6.51ppm d, 6.54ppm d). There were no metabolites associated exclusively with human milk (Figure 50).

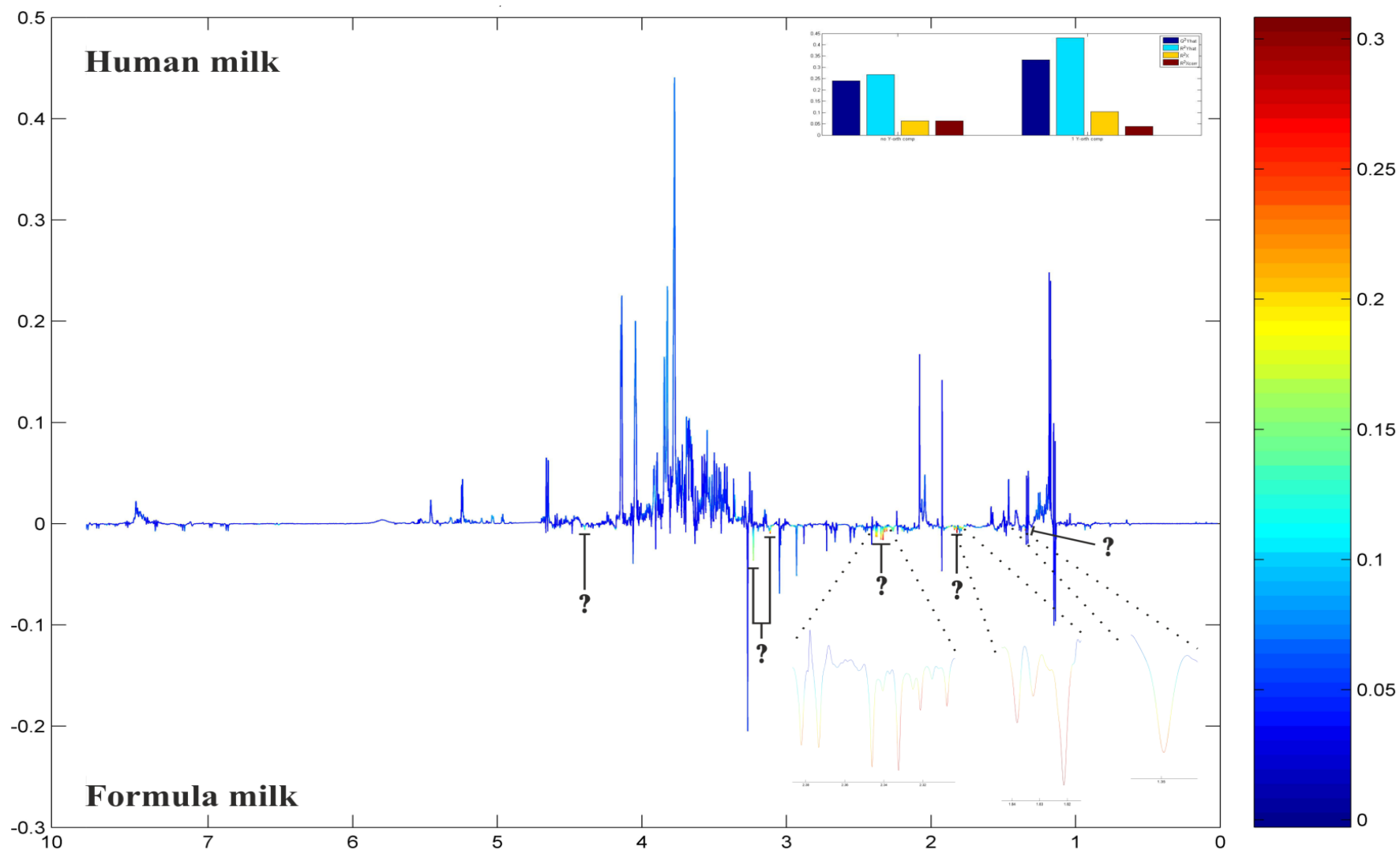


Figure 50: Orthogonal Partial Least Squares (O-PLS) loadings plot of urine metabolic profiles coming from samples collected at multiple time points and created in accordance to the type of milk. The X axis shows chemical shift in ppm, while the Y axis represents metabolites associated with human milk (up) or formula milk (down). The insert bottom right: R^2Y (light blue bar) and Q^2Y (dark blue bar) values showing 43% and 33% respectively. The permutation tests for Q^2Y retrieved p -value = 0.01.

Table 18: The summary of the models created for the analysis of the changes between the urinary metabolic profiles

Corrected for	Samples	PNA	GA	Sex	Mode of delivery	Labour	Human milk vs. formula	PN vs. EN
N/A	Whole data set	Y (4)	Y (2)	Y (0)	Y (1)	Y (1)	Y (1)	Y (1)
Gestational age	24 th week GA	Y (2)	X	Y (0)	N	N	X	X
	28 th week GA	Y (1)	X	N	N	N	X	X
Postnatal age	2 nd -9 th day of life	Y (1)	Y (1)	N	N	N	X	X
	15 th -22 nd day of life	N	Y (0)	N	Y (0)	N	X	X
	Last day of sample collection	N	N	N	N	N	X	X
Postmenstrual age	37 th week PMA	N	X	N	N	N	X	X
Type of enteral nutrition	Formula milk	Y (4)	Y (4)	N	N	N	X	Y (0)
	Human milk	Y (5)	Y (3)	Y (0)	Y (1)	N	X	Y (1)

Key:Y – difference between the classes, positive Q^2Y above the threshold value of 0.05N – no difference between the classes, negative Q^2Y or below the threshold value of 0.05

(2) – the value in the brackets indicates the number of orthogonal components used in the model

X – no data as the analysis was not performed

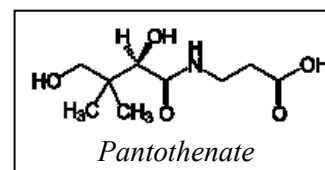
Table 19: The list of discriminant urinary metabolites identified to be associated with each analysed factor.**Key:** GA – gestational age, PMA – postmenstrual age, PNA – postnatal age.

Factor	Association	Metabolite	Chemical shift
PNA	Increasing	pantothenate	0.90ppm s, 0.94 ppm s
		acetyl carnitine	2.13ppm s, 2.50ppm dd, 3.18ppm s, 3.62ppm d, 3.83ppm dd
		unknown	2.33ppm dd
		unknown	0.91ppm t, 2.78ppm m
		trimethylamine	2.90ppm s
		malonate	3.12ppm s
PMA	Decreasing	sugars	3.33ppm m, 3.38ppm s, 3.44ppm s, 3.49ppm s, 3.67ppm d, 3.69ppm t, 3.75ppm m, 3.83ppm d, 3.85ppm d, 4.04ppm t, 4.14ppm d, 4.20ppm s, 4.24ppm s, 4.51ppm s, 4.96ppm t, 5.03ppm d, 5.11ppm s, 5.24ppm m, 5.37ppm s, 5.40ppm s, 6.49ppm d
		pantothenate	0.90ppm s, 0.94 ppm s
		creatinine	3.05ppm s, 4.06ppm s
		myo-inositol	3.29ppm t, 3.52ppm dd, 3.61ppm t, 4.05ppm t
		unknown	3.62ppm d, 3.64ppm t, 3.74ppm s
		hydroxyproline	2.16ppm m, 2.43ppm m, 3.37ppm dt, 3.47ppm dd 4.35ppm dd
GA	Increasing	fumarate	6.52ppm s
		malonate	3.12ppm s
		4-hydroxyphenylacetate	3.45ppm s, 6.87ppm d, 7.17ppm d
		methylhydantoin	2.93ppm s
		bile acids	0.84ppm s
		oligosaccharides	3.70ppm-3.90ppm s
Parenteral nutrition	Decreasing	myo-inositol	3.27ppm s, 3.29ppm t, 3.54ppm dd, 3.62ppm t, 4.06ppm t
		sugars	3.22ppm s, 3.33ppm m, 3.31ppm s, 3.38ppm s, 3.42ppm s 3.44ppm s, 3.49ppm s, 3.51ppm s, 3.67ppm d, 3.69ppm t, 3.72ppm s, 3.75ppm m, 3.83ppm d, 3.85ppm d, 3.89ppm s, 3.91ppm s, 4.04ppm t, 4.14ppm d, 4.20ppm s, 4.24ppm s, 4.51ppm s, 4.65ppm d, 4.96ppm t, 5.03ppm d, 5.11ppm s, 5.24ppm m, 5.37ppm s, 5.40ppm s, 6.49ppm d
Formula milk		unknown	0.90ppm t, 1.70ppm m, 1.78ppm s, 2.33ppm dd, 4.00ppm dd
		unknown	1.87ppm m, 2.33ppm dd, 2.38ppm d, 3.10ppm t, 6.51ppm d, 6.54ppm d

4.4 Discussion

The randomised feeding groups (SMOF lipid vs. Intralipid, and incremental vs. RDI concentration of amino acids) presented no difference when it came to the urinary metabolic profile suggesting that there is no difference in the production of the waste metabolites. The urinary metabonome of preterm babies is strongly affected by the postnatal age and gestational age. In case of diet, not only the route of nutrition but also the type of milk caused metabolic differences. There were no associations visible between the mode of delivery, presence of labour during delivery, sex of the baby and the urine metabonome.

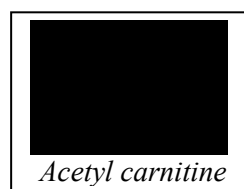
The association of increasing postnatal age with increasing concentration of pantothenate (vitamin B₅) is not surprising. Pantothenate is known to be synthesised by various bacteria including those colonising human gut



[3, 276-280]. Additionally, animals cannot synthesise vitamin B₅ and rely solely on dietary consumption or from gut microbial biosynthesis [277, 279, 281]. It seems logical that the older the

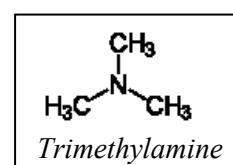
baby the more likely it is to be colonised by bacteria capable of producing this vitamin. Furthermore, it is also known that the concentration of the vitamin B₅ increases in the breast milk with duration of the lactation and this was reported for both mothers of full-term and preterm babies [282, 283]. Therefore it is possible that the increase in the concentration of the vitamin B₅ is caused by two factors simultaneously – the increase of bacterial biosynthesis, and the increase in concentration in mother's milk. Pantothenate is an essential molecule for all living organisms as it is a precursor for Coenzyme A (CoA) [276-278, 281, 284, 285] and it was reported before that, since pantothenate is a water soluble vitamin, it can be excreted with urine and the excreted concentration correlates to the consumption levels [286].

Another metabolite associated with increasing postnatal age was acetyl carnitine. Presence of this chemical compound in urine is not surprising as it has been reported before [264, 287-289]. It is created in the human body from



carnitine and acetyl-CoA [285]. Carnitine is responsible for the transfer of free fatty acids across the membranes of mitochondria where they can be utilised [264, 285, 288, 289]. In newborn babies carnitine is removed from the organism in the form of acetyl carnitine via the urine [264, 288-290]. It had been reported before that the concentration of the acetyl carnitine in blood significantly increases after first day of postnatal age [290]. Additionally, carnitine can be found in breast milk and since the volume of breast milk given to preterm babies increases over time this could support the observed increase in the concentration of acetyl carnitine with age [291]. An interesting observation is that the concentration of carnitine in the amniotic fluid drops significantly after the 33rd week of gestation which means that babies born before 33rd week of GA are exposed prior to birth to a much higher concentration of this molecule than the full-term babies [290]. How this affects the preterm babies is yet unknown [290].

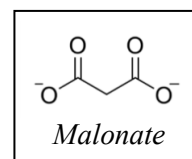
Trimethylamine, which I found to be positively associated with postnatal age, is mostly known as a product of decomposition that possesses an unpleasant smell [288, 292, 293]. The molecule can be excreted in urine if there was a high



concentration of carnitine due to the fact that it is a product of bacterial carnitine degradation [264,

288, 292, 294, 295]. Since it was already reported that concentration of carnitine increases with postnatal age it is possible that the products of its degradation increase in concentration as well. Trimethylamine can be also produced as a product of the degradation of choline – it is therefore unsure which of the metabolites has contributed as the source for the trimethylamine production [296-298]. Additionally, women that went into premature labour had lower levels of urinary trimethylamine in their first trimester [299].

Another compound associated with postnatal age is malonate. It reacts with acetyl-CoA to form a malonyl-CoA which is an important molecule in biosynthesis of fatty acids. It is normal to see a low concentration of malonate excreted in urine

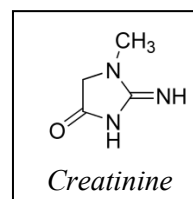


[300]. It has been reported that diet rich in medium chain fatty acids can result in increased urinary excretion of malonate [301]. One possible explanation could be that medium chain fatty acids from milk lead to increased concentration of urinary malonate when compared to babies that are mostly on parenteral nutrition. Since the shift from fully parenteral to exclusively enteral nutrition is related to age and occurs quicker in more mature babies, this could explain why malonate is more concentrated in older (higher PNA) and more mature (higher GA) babies. The higher the postnatal age the more likely the baby is to be on enteral nutrition. Also the higher the gestational age at birth, the more mature the baby is thus again more likely to be fed orally on earlier stage than less mature babies. However, malonate was not observed to be associated with any particular route of nutrition.

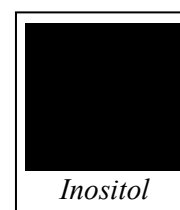
It was not possible within the timeframe of this PhD to properly identify the sugars that were associated with lower postnatal age since targeted spectroscopic characterisation of sugars is challenging. The presence of sugar in urine can be indicative of diabetes. However, premature babies have underdeveloped kidneys which do not function properly. The kidneys continue to mature after the premature birth and it is logical that the older the baby gets, the more mature the kidneys will be. The presence of sugars in the urine may however reflect malfunctioning sugar reabsorption.

The association of creatinine with an increasing gestational age is a predictable result. The levels of creatinine are related to the muscle mass and babies born later in gestation are usually larger than

those more premature [302-305]. However, there may be at least two reasons for the creatinine levels being associated with gestational age. It is known that in healthy well-functioning kidneys there is no reabsorption of creatinine back to blood [302-306]. However, premature babies do not have properly functioning kidneys and it has been reported that higher gestational age is associated with lower creatinine levels in serum which indicates that urinary excretion was higher [302, 305, 307, 308]. It is however important to mention that newborn babies have urinary creatinine levels that are initially the same as their mothers due to creatinine transfer through placenta [302, 307, 309]. Subsequently the creatinine level drastically drops to an infant's own level and rises again in correspondence with the infant's growth and gain in muscle mass. Therefore, if not enough timed samples have been collected or the time-points fall in the time-frame too close to birth, it is possible to observe the opposite result – with creatinine levels dropping with postnatal age.

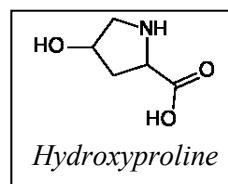


Myo-inositol, a compound found to be associated with gestational age, is a most important stereoisomer of inositol [285, 310]. Inositol is produced in the human body from glucose and is used as a building block of many cellular secondary messengers [285, 310, 311] and a number of clinical trials had found that inositol helps to cure various psychiatric conditions [310-312]. There are reports of increased levels of inositol in the urine of babies with inappropriate size for gestational age (both for too small and too large babies) [313-315]. The most likely explanation why myo-inositol was found to be associated with higher gestational age is that breast milk contains this compound and, as previously described, more mature babies are more likely to be put on exclusive enteral nutrition earlier [316, 317]. In contrast to above hypothesis, it was observed that myo-inositol in urine was associated with parenteral nutrition which is consistent with the fact that myo-inositol should be fully reabsorbed to blood and only diabetics or people with renal disorders generally have this compound present in their urine [318]. Babies on parenteral nutrition are usually younger (postnatal age) than those on full enteral nutrition thus it would be logical for their kidneys to be unable to reabsorb the myo-inositol. The above hypothesis is also supported by our finding that other sugars are also found at higher levels in the urine of babies on



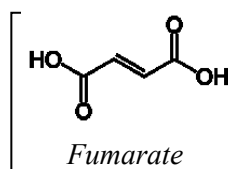
parenteral nutrition (compared to those fed orally). To my knowledge, there is no evidence in the literature as to why babies born at later gestational age have more myo-inositol in their urine thus this issue requires further investigation. The model which reported this association, even though being predictable and representative for the sample set – was not robust.

As previously reported by Atzori *et al.* proline has been found to have a decreased concentration in premature babies compared with full term ones [267]. Therefore, it is not surprising that it was observed an association of



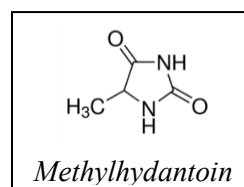
hydroxyproline (a hydroxylated proline) with gestational age. Proline and hydroxyproline are important components of collagen and are necessary for maintenance of a healthy skin. The reported association came from a not robust model.

Fumarate is a product of both citric acid and urea cycles thus it is possible to find it in urine [319-321]. The excretion of citric acid cycle metabolites is said to be a form of acid-base equilibrium management [321]. It was previously



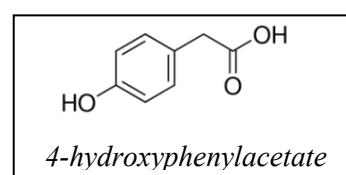
reported that babies fed orally have higher concentration of fumaric acid in their urine compared to parenterally fed ones [320]. However, this study did not account for potential bias from postnatal age. However, if the findings from this study are considered to be correct, the results can be interpreted as an association between higher concentration of fumarate and higher gestational age and could be explained by the fact that more mature babies are more likely to switch from parenteral nutrition to enteral nutrition earlier than those less mature. Another explanation for positive association of fumarate with higher gestational age may be due to better functioning kidneys in the more mature babies thus more efficient filtration and management of the acid-base balance. Another hypothesis is that perhaps the urea cycle, which takes place in the liver, is more effective in more mature babies whose liver is less overloaded with effects of parenteral nutrition. Yet again, there may be multiple reasons that contribute to the observed difference in the fumarate concentration.

A singlet at $\delta 2.93$ consistent with the chemical shift of methylhydantoin has been found as associated with the gestational age. The other singlet expected



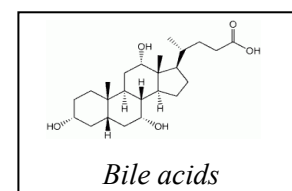
from this molecule ($\delta 4.07$) had not been observed possibly because the urinary concentration of the compound is very low (barely detectable by the ^1H NMR spectroscopy). Methylhydantoin is a product of bacterial degradation of creatinine [322, 323] which was observed in *Pseudomonas*, *Bacillus*, *Corynebacterium*, *Flavobacterium*, *Escherichia*, *Proteus* and *Clostridium* species [322-327]. It would be a very interesting finding as it would support the findings from the studies that reported presence of creatinine degrading bacteria in human stool and urine [324-329]. Additionally, since levels of creatinine in urine are associated with gestational age, it would make sense that the product of its degradation would be associated as well. The reported association came from a not robust model.

Finding for 4-hydroxyphenylacetate being associated with the gestational age is interesting as this molecule can be found in olive oil which is one of the components of the SMOF lipid that is given



intravenously to babies [330, 331]. However, this source of the metabolite is probably not important as no difference between the randomized feeding groups was observed and if the 4-hydroxyphenylacetate was coming solely from the olive oil from the SMOF lipid the separation of samples based on the type of lipid the babies were receiving intravenously would be observed. Therefore, it is far more likely that the source of 4-hydroxyphenylacetate was the gut microbial catabolism of amino acids [320, 332-336]. Following the more possible hypothesis, if the source of 4-hydroxyphenylacetate is from the catabolism of amino acids, then perhaps the reason why this occurs more often in more mature babies is probably due to the fact that they are less deprived of amino acids than the more preterm babies. This issue would require further investigation.

Bile acids were the only group of metabolites found to be associated with lower gestational age. Such an association has been reported previously



[273, 274]. Cholic acid and chenodeoxycholic acid are primary bile acids which are produced in the liver from cholesterol and are excreted to facilitate the lipid absorption [5, 33, 337]. In humans bile acids are conjugated to glycine and are transferred from ileum to the liver [33]. Some bacterial species can deconjugate those bile acids which results in shortage of their reabsorption to the liver and leads

to production of secondary bile acids [33]. Additionally, if the liver is dysfunctional the bile acids can be excreted in the urine [274, 337-339]. Premature babies do not have fully functional livers and additionally they can be affected by the hepatotoxic effects of parenteral nutrition [340]. One possible explanation as to the reason that urinary excretion of bile acids is associated with lower gestational age, therefore, is probably due to not fully functioning liver. Altering the bile acid balance can lead to obesity and insulin resistance [33].

The sex of the baby did not appear to have any influence on the metabolic profile of urine. Perhaps at this stage of life there are no metabolic differences between boys and girls that would be observable in the urine. Additionally, neither mode of delivery nor presence of labour had an effect on the urinary metabolism of the babies. It is possible that it is too early to observe any metabolic changes between babies born via two different delivery routes. When it comes to diet, the route of nutrition (parenteral vs. enteral) has a strong influence on the urinary metabolic profile of the preterm babies. It is however very difficult to discuss the results as the switch from parenteral to enteral nutrition is strongly linked with increasing postnatal age. Therefore, it is not possible to determine which of those factors (or if both simultaneously) affect those metabolic changes. There was visibly a clear separation between samples based on the type of milk they were given but the identification of chemical compounds responsible for this difference was not possible.

4.5 Conclusion

Metabolites found to be associated with increasing postnatal age were either related to bacterial activity or to the introduction of enteral nutrition. The presence of sugars in the earlier postnatal age indicates underdeveloped kidneys, which with time improve their functionality and reabsorptive capacity for sugars. It seems that the most important factors contributing to metabolic changes presented in the urine are postnatal age and diet (both the route of nutrition and the type of milk). Not all of the discriminant metabolites were identified, thus further research should be undertaken to fully answer the question how exactly diet affect the metabolic changes in premature babies.

Chapter 5 – The stool metabonome of the preterm baby

5.1 Introduction

Stool is mostly composed of water, undigested food, bacteria and their metabolites and its composition can hold information relating to the nutritional, pathologic and physiological environment of an individual [341]. A number of studies have been performed to assess the metabolic profile (metabonome) of adult stool [261-264, 342-347]. It is known that diet and bacterial composition are the major factors affecting the metabonome [261, 346, 348, 349]. However, little is known about the metabolic changes in premature babies [261]. Premature babies not only have underdeveloped guts but also undergo a different nutrition regime compared to term born babies, are exposed to prolonged hospitalisation and antibiotic therapies [350, 351]. When analysing the results of the faecal water metabolic profiles it was important to take into account factors which are already known to affect metabolic profile in stool or that are linked to changes in gut microbiota and thus may lead to changes in metabonome. Those factors include age [38, 86, 352], diet [9, 34, 58, 60, 65, 70, 71, 77, 85, 93, 353-365], mode of delivery [47, 53, 55, 56, 60, 73-75, 366-372], exposure to antibiotics [78, 373-375], and sex of the baby [20, 38, 86, 376]. The detailed explanation of the impact of each of those factors on the microbiota profile had been given in the Background chapter (section 1.2.iii). Ability to understand the influence of early life interventions on the metabolic profile could help to improve the care at the hospital and later in life. Prematurity is already known to be associated with number of adverse health outcomes like obesity and high blood pressure. This analysis could contribute to further understanding of the changes that occur after premature birth [143, 144, 351, 377, 378].

In order to answer this question faecal water was extracted and ^1H NMR acquisition of all the stool samples was performed. After obtaining the spectra an unsupervised Principal Component Analysis (PCA) model for the whole data set (including multiple time points for each baby) was created. The

samples were coloured according to their class a) gestation age, b) postnatal age), c) sex, d) mode of delivery and e) diet. Next, a supervised Orthogonal Partial Least Squares analysis (O-PLS) for those classes was performed in order to find discriminant metabolites – biomarkers for each factor. The analyses were continued on stratified subgroups of samples in order to control for some of the confounding factors.

5.2 Experimental design

5.2.i Stool sample preparation

Faecal water had been extracted from 100mg of stool sample by mixing with 250µl of 25% Acetonitrile (in the molecular grade water). Mixture was homogenized by bead beating for 10s with 1mm Zirconium beads. Faecal water was collected and aliquoted. After evaporating the acetonitrile from the fraction for the NMR analysis, the powder was resuspended in the starting volume (50µl) of D₂O and topped-up with the buffer to a final volume of 600µl. Please see the Materials and Methods chapter (section 2.2.ii.b) for the detailed sample preparation and the Method Development chapter for the details of the protocol validation.

5.2.ii Nuclear Magnetic Resonance spectra acquisition

Please see the Materials and Methods chapter (section 2.2.i.b) for the detailed information on the acquisition of NMR spectra. The parameters used for the standard pulse sequence of the faecal water sample were: P1=11.47µs, P11=12.50-dBW, P19=46.96dB and O1=2830Hz.

5.2.iii Sample pre-processing

Spectra for the NEON faecal water samples were preprocessed and modeled in Matlab. The first step was to model all spectra and the quality controls (QCs) in order to check if there are any instrumental batch effects. The Principal Component Analysis (PCA) scores plot for the model was calculated for 3 principal components giving a cumulative R² of 25.7%. The QC samples clustered together in the centre of the plot indicating that the analytical data and the chosen normalization method were appropriate (Figure 51). The second stage of analysis included modeling only the NEON faecal water

samples, without the QCs. The model for the whole data set (also referred to as multiple time points) that is presented in the results section (section 5.3) was calculated using 3 principal components (PC) giving a cumulative R^2 of 21.7%. This shows that there is a very high variability between the faecal water samples which makes the analysis more difficult.

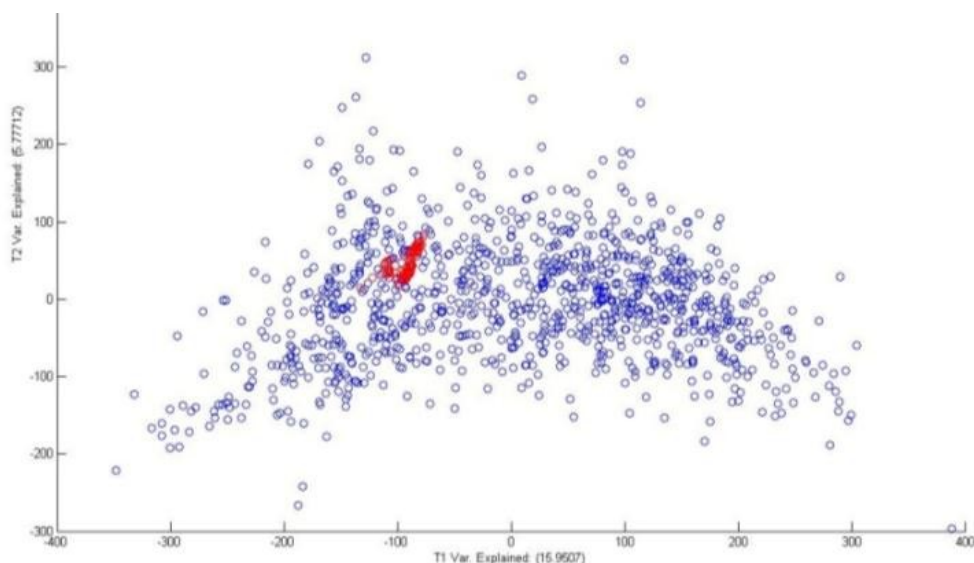


Figure 51: Scores scatter plot of the principal component analysis (PCA) model of faecal water NMR spectral data generated from samples collected at multiple time points (whole data set). Quality control samples coloured red.

Initially, all spectra corresponding to multiple time points for 160 babies were modeled to get an overview of the dataset. Due to the fact that the samples came from the randomized clinical trial it was crucial to firstly assess if there is any difference between the randomised groups. Then, in order to identify the factors contributing to the main source of the variation on scores plot, the unsupervised models for multiple time points was coloured according to classes obtained from the NEON metadata. Those variables included: a) postnatal age, b) gestation age at birth, c) mode of delivery, d) presence of labour, e) sex, f) route of nutrition (parenteral or enteral), and g) type of enteral nutrition (formula milk or human milk). The second step of preliminary analysis was to perform a supervised analysis using the Orthogonal Partial Least Squares model (O-PLS) in order to find discriminant metabolites associated with particular classes of samples. The results from those steps help to understand which of the known factors have to be controlled for while analysing the influence of the others. In order to do so, analysis is continued on several smaller subsets.

5.3 Results

5.3.i Metabolic profiles associated with randomised PN feeding group

In order to avoid a possible bias from PNA, the analysis on randomised groups was done after subtracting samples from the whole data set. Three time points have been chosen: (a) 1st-14th day of life when babies were still exposed to parenteral nutrition and high number of samples was available, (b) 22nd-35th day of life when babies were off parenteral nutrition and large number of samples was available, and (c) last sample collected provided it was after 32nd week of life. If there was more than one sample available in particular time frame for a baby then the one collected earlier was included. All three PCA models were calculated using 3 principal components and the cumulative R^2 equalled: (a) 34.3% (b) 30.6% and (c) 27.9%.

5.3.i.a Intralipid vs. SMOF lipid

The scores plots for all of the unsupervised Principal Component Analysis models showed no separation based on the type of lipid emulsion (Appendix: Figure 91, Figure 92 and Figure 93). The supervised models (O-PLS) confirmed the lack of difference between the randomised groups (Figure 53). The Q^2Y value was negative for the PLS model (no orthogonal component) in all three tested time points.

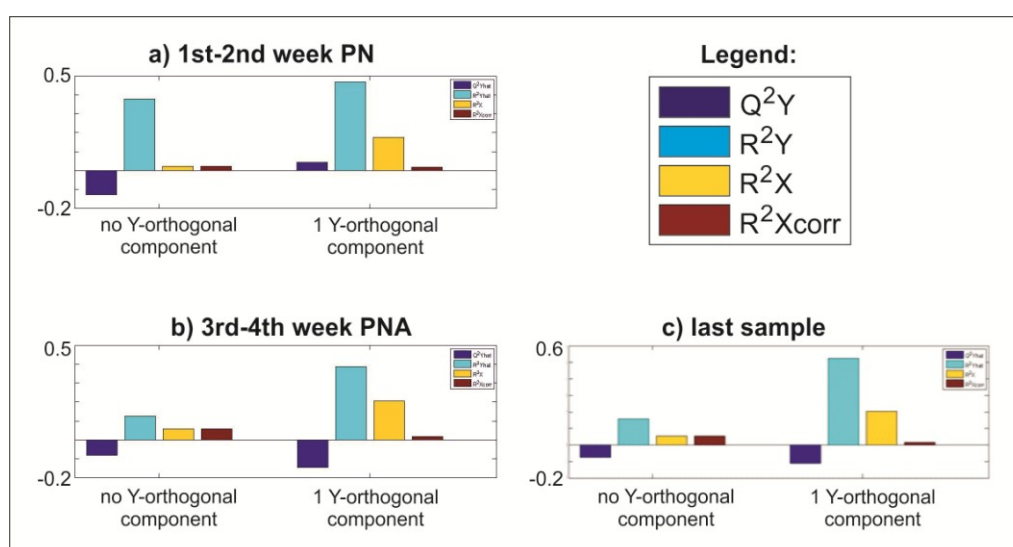


Figure 52: The O-PLS models for stool samples comparing the types of lipids. The R^2Y and Q^2Y values for the O-PLS of faecal water metabolic profiles representing samples collected (A) between 1st-14th day of life, (B) between 22nd-35th day of life, (C) the last sample prior to discharge. The dark blue bar presents Q^2Y while the light blue – R^2Y . The bars on the right of each figure show the change after introduction of the orthogonal component.

The introduction of the orthogonal component led to the increase of the Q^2Y value for the first tested time point, however it was still below the threshold of 5%. The two remaining time points had a negative Q^2Y value. Those findings suggest that the type of lipid emulsion provided intravenously does not influence the metabolic profile of faecal water. It is important to mention that faecal water was tested for the presence of lipids with a negative result.

5.3.i.b Recommended daily intake vs. increasing concentration of amino acid

The scores plots for the unsupervised models (PCA) showed no separation based on the concentration of amino acids (Appendix: Figure 94, Figure 95 and Figure 96). The O-PLS models gave negative Q^2Y values for all tested time points (Figure 53). Such result confirmed the lack of difference between the groups randomised according to the concentration of amino acids.

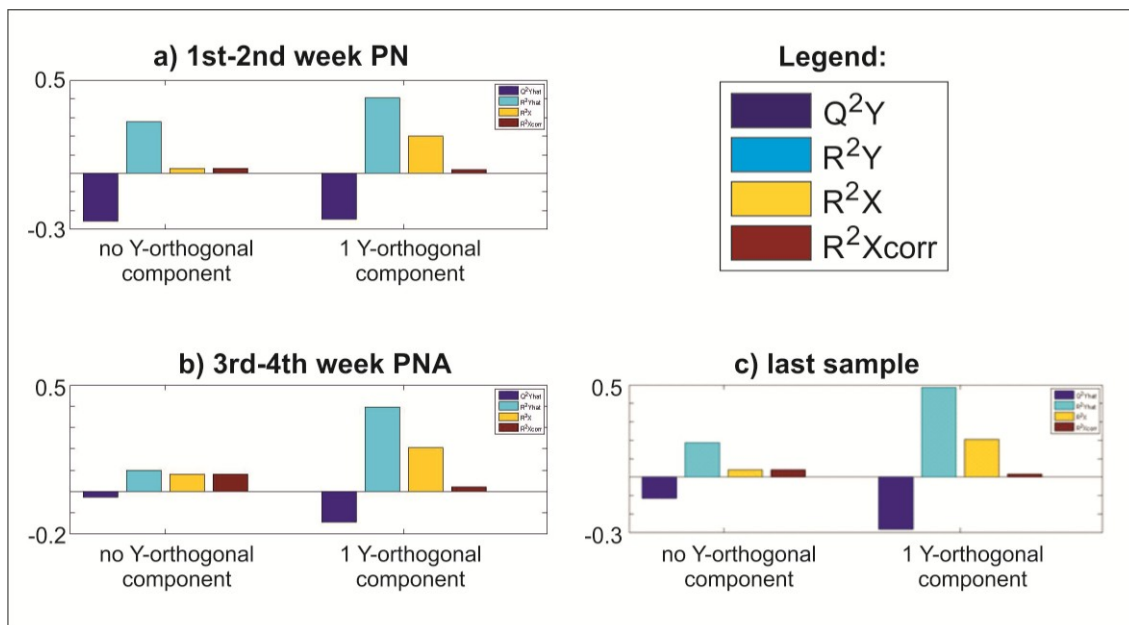


Figure 53: The O-PLS models for stool samples concentration of amino acids. The R^2Y and Q^2Y values for the O-PLS of faecal water metabolic profiles representing samples collected (A) between 1st-14th day of life, (B) between 22nd-35th day of life, (C) the last sample prior to discharge. The dark blue bar presents Q^2Y while the light blue – R^2Y . The bars on the right of each figure show the change after introduction of the orthogonal component.

Due to lack of differences between the randomised groups it was acceptable to merge the groups for further exploratory analysis.

5.3.ii Metabolic profiles associated with gestational age (GA)

The PCA scores plot for components PC1 vs. PC2 coloured according to gestational age (GA) showed no clustering of samples according to class. Components PC3 vs. PC1, which are shown in Figure 55, represent the most evidence of clustering according to GA.

Using the supervised modeling approach to maximize the effect of GA on sample composition it was observed that there is a metabolic difference related to the GA. The O-PLS model was built with one orthogonal component leading to the increase of the R^2Y value to 23% and in the Q^2Y value to 16%. The O-PLS score plot, shown in Figure 56, presents the results for the increasing GA from the bottom to the top. Metabolites associated with higher GA include two unknown molecules (1.21ppm s, and 1.84ppm dd, 2.21ppm d, 2.23ppm d), acetate (1.92ppm s), N-acetylated glycoprotein (2.05ppm s), glutamate (2.12ppm d, 2.14ppm s, 2.36ppm multi), γ -aminobutyrate (1.90ppm multi, 2.30ppm t, 3.01ppm t), and aspartate (2.68ppm q, 2.81ppm dd) (Figure 55). There were no metabolites associated with lower GA.

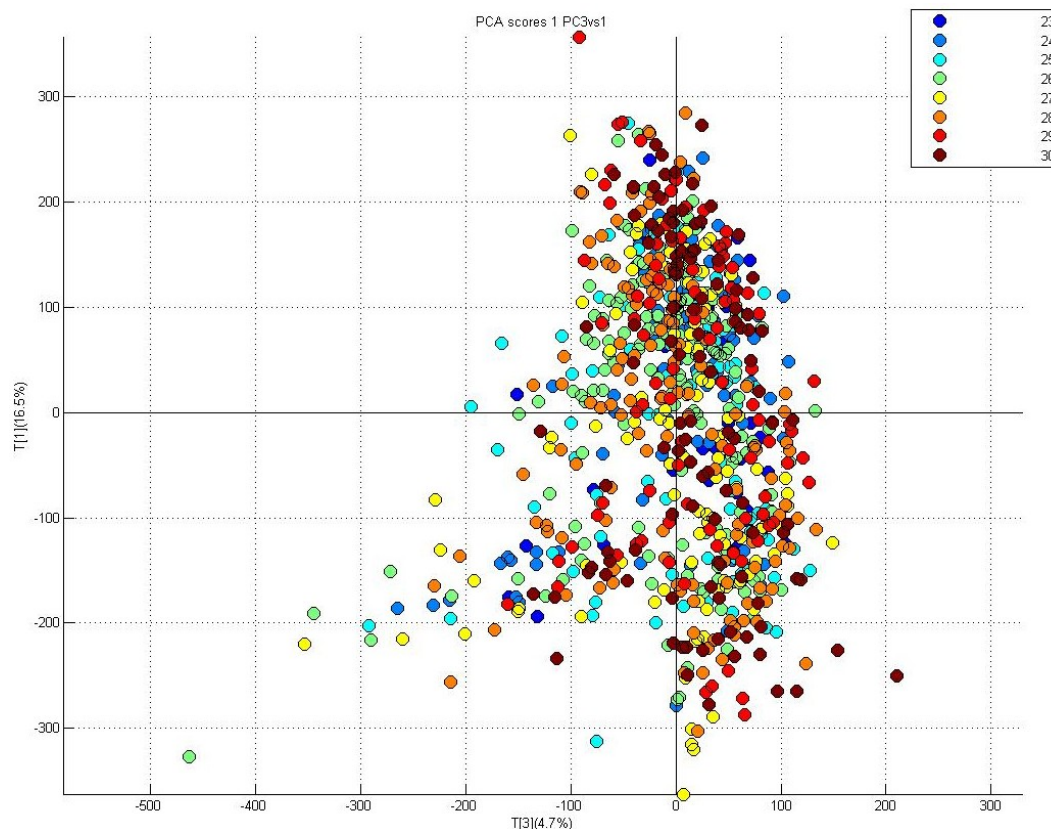


Figure 54: PCA Scores Scatter Plot of whole data set of NMR stool spectra. Samples coloured according to gestation age (in weeks).

In order to confirm the findings, samples were grouped in order to have the same variable for postnatal age and the analysis was continued on two subsets of samples: a) collected in the 2nd week of life (113 samples) and b) collected in the 3rd week of life (127 samples). Those time points were chosen due to the availability of the highest numbers of samples. Samples from 1st week of life, although well represented in numbers, were not used due to possible bias from combining normal stool samples with meconium. Meconium is a first stool sample passed by the newborn baby. It is predominately made of mucus and amniotic fluid therefore its composition varies drastically from the normal stool.

The PCA scores plot for the samples collected on the 2nd week of life was calculated from 3 principal components giving a cumulative R^2 of 30.8%. The PCA scores plot showed no clustering based on GA (Appendix: Figure 97) and the Q^2Y for the O-PLS dropped from 11% for no orthogonal component to 9% for the orthogonal component. The orthogonal component was removed and supervised PLS analysis was performed. The PLS loadings plot showed that bile acids (1.09ppm s), are associated with lower gestation age. At the same time betaine (3.28ppm s, 3.90ppm s) was found to be associated with increasing GA.

The PCA scores plot for the samples collected on the 3rd week of life was also calculated from 3 principal components and the cumulative R^2 equaled 29.3%. There was no clustering visible on the PCA scores plot (Appendix: Figure 98) and the Q^2Y for the O-PLS dropped again, this time from 32% to 29% after introduction of the orthogonal component. The PLS model was used instead and the loadings plot showed that unidentified molecule (1.37ppm s) and N-acetyllysine (1.48ppm multi, 1.71ppm multi, 1.99ppm s, 3.01ppm t, 3.18ppm t) were associated with lower GA. No metabolites were found to be associated with increasing GA.

Figure 55: Orthogonal Partial Least Squares (O-PLS) loadings plot of stool metabolic profiles coming from samples collected at multiple time points and created in accordance to the gestational age. The X axis shows chemical shift in ppm, while the Y axis represents metabolites associated with (increasing – up or decreasing – down) gestational age. Annotated metabolites: 1. bile acids, 2. N-acetyllysine, 3. γ -aminobutyrate, 4. acetate, 5. N-acetylated glycoprotein, 6. glutamate, 7. aspartate, and 8. betaine. The insert bottom right: R^2Y (light blue bar) and Q^2Y (dark blue bar) values showing 23% and 16% respectively. *P*-value from the permutation analysis was 0.01.

5.3.iii Metabolic profiles associated with postnatal age (PNA)

The PCA score plots showed clustering according to postnatal age (PNA) presented in weeks. This was especially visible for the components PC3 vs. PC1 (Figure 56).

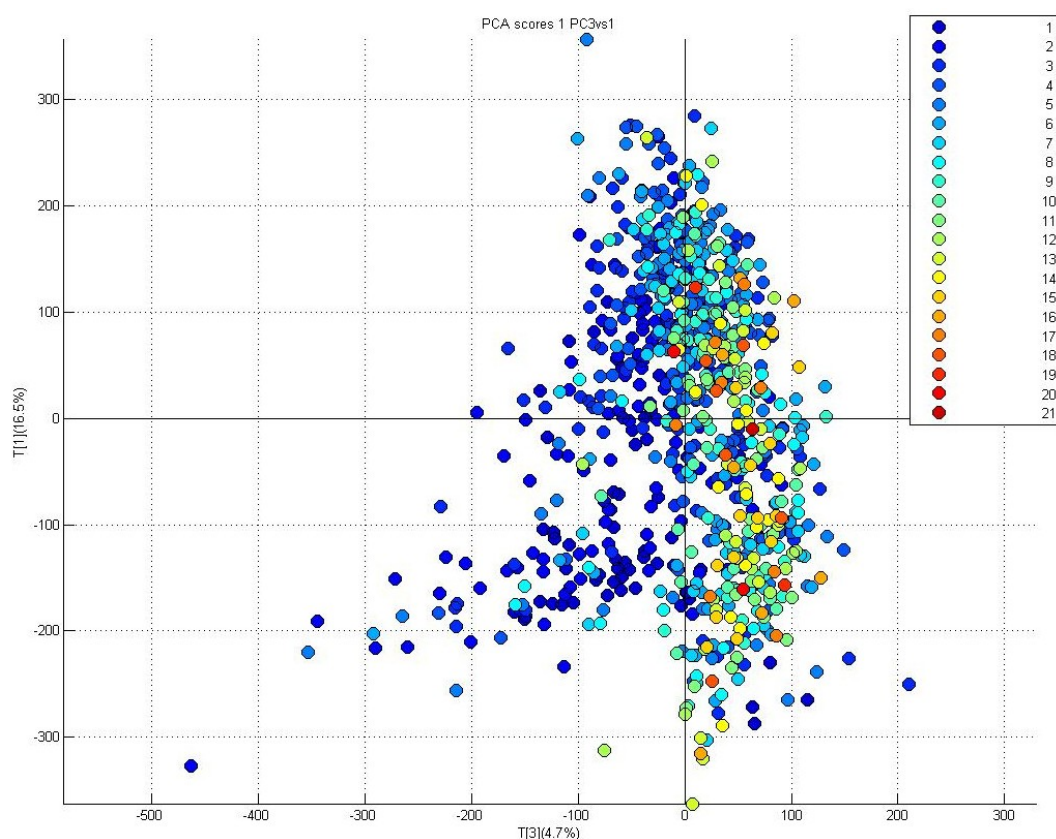


Figure 56: PCA Scores Scatter Plot of whole data set of NMR stool spectra. Samples coloured according to postnatal age (in weeks)

The O-PLS was calculated with three orthogonal components which suggest that there was a very high variability between the samples as orthogonal component are meant to remove the differences caused by the natural variability (Figure 57). The final R^2Y and Q^2Y values were 66% and 52% respectively. The O-PLS scores plot showed that bile acids (0.64ppm s, 0.78 ppm s, 1.09 ppm s, 5.54 ppm multi) were associated with decreasing PNA. There were no metabolites associated with increasing PNA (Figure 57). In order to avoid the possible bias of the gestational age (GA) I have chosen two subsets of samples: a) samples from babies born at the 24th week of gestation (94 samples) and b) samples from babies born at the 28th week of gestation (142 samples).

The PCA scores plot for the 24th GA was calculated using 3 principal components giving a cumulative R^2 of 30.9%. No clear separation was visible (Appendix: Figure 99). However, R^2Y and Q^2Y for the

O-PLS increased after the introduction of the orthogonal component and equaled 85% and 38% respectively. The PCA scores plot for the 28th GA was calculated using 3 principal components giving cumulative R^2 of 30.5% (Appendix: Figure 100). The introduction of the orthogonal component in the O-PLS analysis increased the R^2Y and Q^2Y to 57% and 45% respectively. The O-PLS scores plots showed that increasing PNA is associated with alanine (1.48ppm d), acetate (1.92ppm s), trimethylamine (2.88ppm s), tyramine (2.93ppm t, 6.91ppm d, 7.23ppm d). Bile acids (0.78ppm s, 1.09ppm s, 5.54ppm multi) were found to be associated with decreasing PNA.

In order to avoid a potential bias from the type of enteral nutrition, samples were stratified based on the type of milk given. This gave two models, both of which were calculated with 3 principal components. The first model (built on samples collected only while exclusively on breast milk) gave cumulative R^2 of 26.7%. The PCA scores plots showed separation according to PNA and the O-PLS gave R^2Y of 46% and Q^2Y of 38%. The second model built on samples collected only while exclusively formula fed gave cumulative R^2 of 29.1% and. The PCA scores plots showed separation according to PNA and the O-PLS gave R^2Y of 62% and Q^2Y of 40%. After removing the variability of diet the O-PLS scores plots showed that increasing PNA is associated with pantothenate also known as vitamin B₅ (0.95ppm s), unknown metabolite (0.92ppm multi, 1.39ppm s, 1.98ppm s, 3.01ppm t, 3.18ppm t), alanine (1.48ppm d), cadaverine (1.45ppm multi, 1.72ppm multi, 3.01ppm t), unknown metabolite (2.13ppm d, 2.36ppm multi, 2.80ppm d, 2.83ppm d), trimethylamine (2.88ppm s) thiamine (3.19ppm t), scyllo inositol (3.35ppm s), unknown metabolite (2.87ppm s, 3.35ppm s, 3.39ppm s, 3.42ppm s), and an unknown metabolite (2.60ppm s, 2.87ppm s, 5.07ppm , 7.32ppm s, 7.61ppm multi). Findings for bile acids (0.64ppm s, 0.78ppm s, 1.09ppm s, 5.54ppm multi) being associated with decreasing PNA were confirmed. Since postnatal age has a strong influence on the model, the analysis of the other factors was additionally done working on smaller subgroups of samples collected on certain weeks of life.

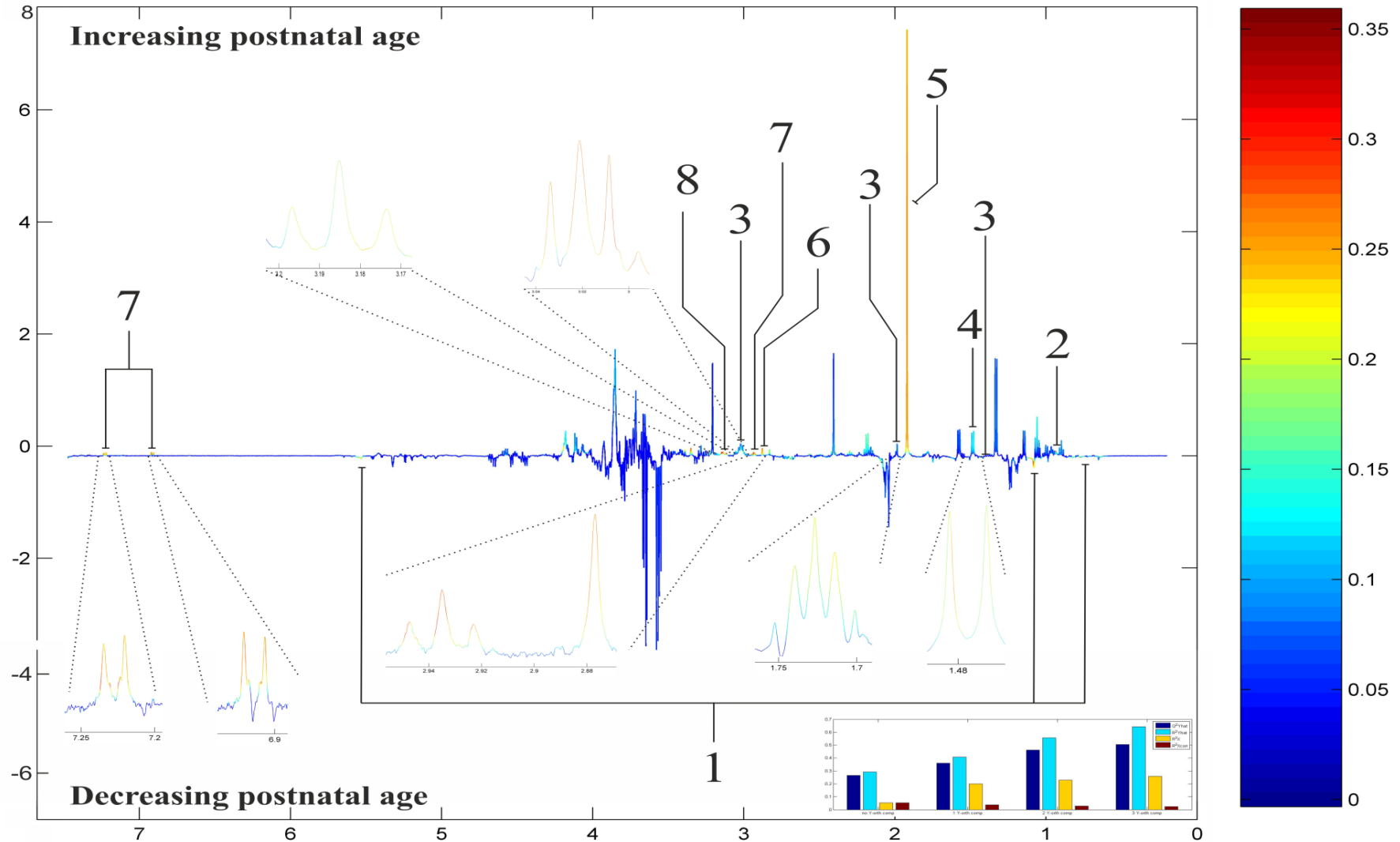


Figure 57: Orthogonal Partial Least Squares (O-PLS) loadings plot of stool metabolic profiles coming from samples collected at multiple time points and created in relation to postnatal age. The X axis shows chemical shift in ppm, while the Y axis represents metabolites associated with (increasing – up or decreasing – down) postnatal age. Annotated metabolites: 1. bile acids, 2. pantothenate, 3. cadaverine, 4. alanine, 5. acetate, 6. trimethylamine, 7. tyramine, 8. thiamine. The insert bottom right: R^2Y (light blue bar) and Q^2Y (dark blue bar) values showing 66% and 52% respectively. P -value from the permutation analysis was 0.01.

5.3.iv Metabolic profiles associated with sex

The PCA scores plots modeled for the whole data set showed no clustering according to sex of the baby (Appendix: Figure 101). Additionally the Q^2Y value in the O-PLS analysis was below 5% which indicates a very weak prediction (Figure 58). Controlling for postnatal age found no difference between the sexes of the babies while the findings after controlling for the gestational age were inconsistent (Figure 58).

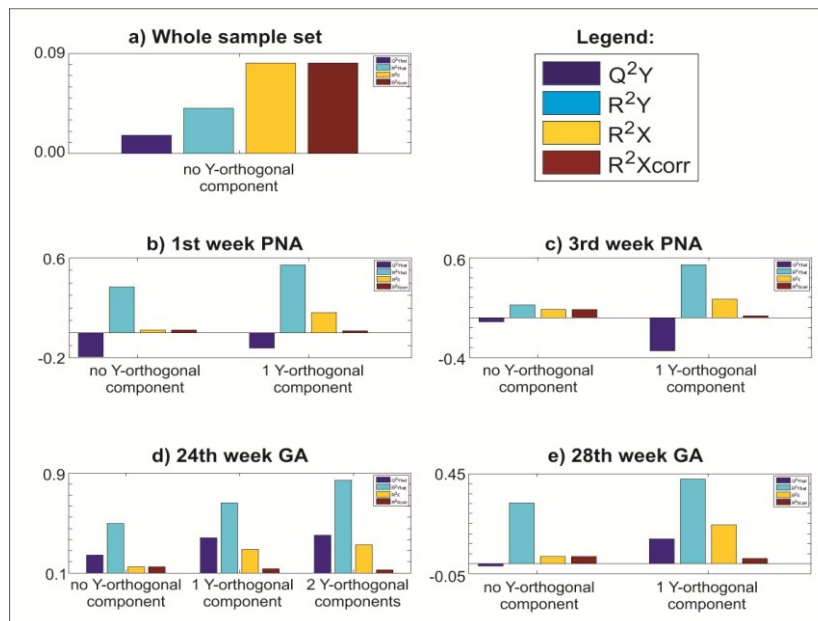


Figure 58: O-PLS models for stool samples comparing sex of baby. The R^2Y and Q^2Y values for the O-PLS of faecal water metabolic profiles representing (A) whole sample set; samples collected (B) from babies born at 24th week of GA, (C) from babies born at 28th week of GA, (D) at 2nd week of life, (E) 3rd week of life. The dark blue bar presents Q^2Y while the light blue – R^2Y . The bars on the right of each figure show the change after introduction of the orthogonal component.

This strong variability between results suggests that the underlying factors affecting the model were not normally distributed between the sexes of the baby. The multiple time points analysis is however of no use for assessing the influence of sex on the metabolic profile as prematurity itself is associated with sex and males are more likely to be born preterm. There were not enough samples to create a subgroup with both GA and PNA variables removed simultaneously. Therefore, obtained results suggest that there is no influence of sex of the baby on the metabolic profile of the stool. Yet there seems to be a possible association between the sex of the baby and the length of the stay at the unit which would explain the bias coming from the unevenly distributed PNA between the samples. Therefore controlling for the sex of the baby was not necessary in further analysis of the faecal water metabolome.

5.3.v Metabolic profiles associated with mode of delivery

The NEON dataset provided a detailed information about mode of delivery which was divided into: a) vaginal - natural, b) vaginal – with use of forceps, c) Caesarean elected – in labour, d) Caesarean elected – not in labour, e) Caesarean emergency – in labour and f) Caesarean emergency – not in labour. For the analysis the classification was simplified as either Caesarean section or vaginal delivery.

The PCA scores plots showed no separation based on the mode of delivery (Appendix: Figure 102) and the Q^2Y value in the O-PLS was below 5% which indicated a very weak prediction of the model (Figure 59). After controlling for the variables like PNA and GA, the results showed that there is no difference between vaginal delivery and Caesarean section when it comes to stool metabolic profile as O-PLS retrieved a negative Q^2Y value (Figure 59). Therefore it suggests that mode of delivery has no impact on the metabolic profile of the stool.

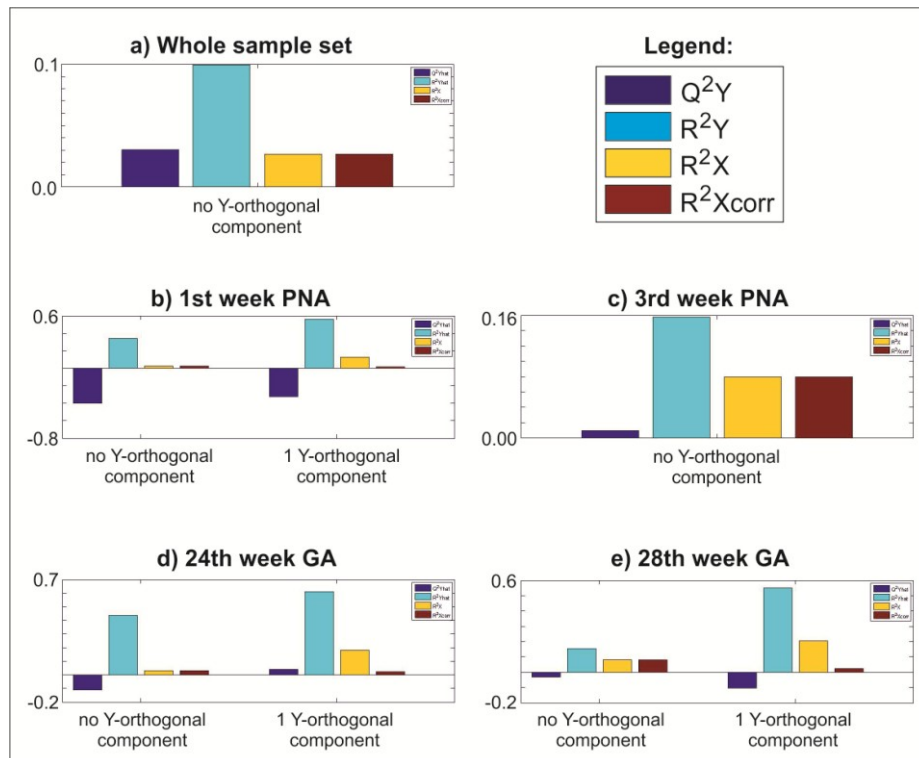


Figure 59: The O-PLS models for stool samples comparing mode of delivery. The R^2Y and Q^2Y values for the O-PLS of faecal water metabolic profiles representing (A) whole sample set; samples collected (B) from babies born at 24th week of GA, (C) from babies born at 28th week of GA, (D) at 2nd week of life, (E) 3rd week of life. The dark blue bar presents Q^2Y while the light blue – R^2Y . The bars on the right of each figure show the change after introduction of the orthogonal component.

The same PCA scores plots were coloured according to labour (all vaginal deliveries were classified as in labour). There was no separation visible on the PCA scores plots (Appendix: Figure 103) and the O-PLS analysis showed a Q^2Y value below 5% (Figure 60). Similarly as to vaginal delivery vs. Caesarean section, the sample subtracting for PNA and GA resulted in negative Q^2Y in the O-PLS analysis (Figure 60).

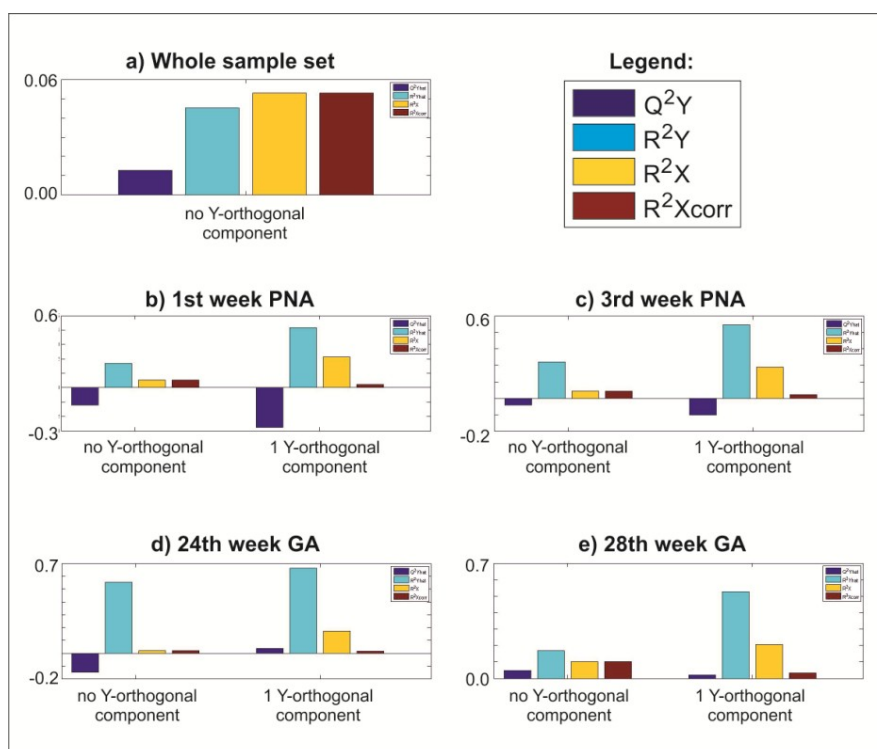


Figure 60: The O-PLS models for stool samples comparing presence of labour. The R^2Y and Q^2Y values for the O-PLS of faecal water metabolic profiles representing (A) whole sample set; samples collected (B) from babies born at 24th week of GA, (C) from babies born at 28th week of GA, (D) at 2nd week of life, (E) 3rd week of life. The dark blue bar presents Q^2Y while the light blue – R^2Y . The bars on the right of each figure show the change after introduction of the orthogonal component.

This suggest that being born while mother was in labour does not significantly affect the metabolic profile of stool, compared to being born while mother was not in labour. Those results show that there is no need to worry about the bias from the variable mode of delivery or labour while analysing the faecal water metabolic profiles.

5.3.vi Metabolic profiles associated with diet

Similarly to the urinary metabolome (Urine Metabolome chapter, section 4.3.vi), there are number of dietary factors that need assessment. Those include comparison of: a) parenteral nutrition (PN) with enteral nutrition (EN), and b) type of milk (human milk vs. formula milk).

5.3.vi.a Parenteral nutrition vs. enteral nutrition

The PCA scores plot for components PC3 vs. PC1 is shown in Figure 61 which presents clustering of samples according to route of nutrition. TPN (1) corresponds to total parenteral nutrition which is simply a parenteral nutrition without any enteral supplementation. PN + EN (2) are a mixture of parenteral nutrition and enteral nutrition while EN (3) is an exclusive enteral nutrition. After removing the variables for PNA, and GA, the distinctive clustering based on route of nutrition remained.

Due to lack of the detailed information on the ratio of EN over PN it was not possible to do the supervised analysis for the continuous variable (increasing ratio of enteral nutrition). Exclusion of the samples that were collected while the baby was on the mixed PN+EN diet would result in strongly unbalanced groups since only 4% of samples were collected during being exclusively on TPN while 74% of samples while exclusive enteral nutrition. Therefore the analysis was performed comparing two groups: a) samples from babies fed only orally, and b) samples from babies on parenteral nutrition regardless if were or were not supplemented orally (Figure 62).

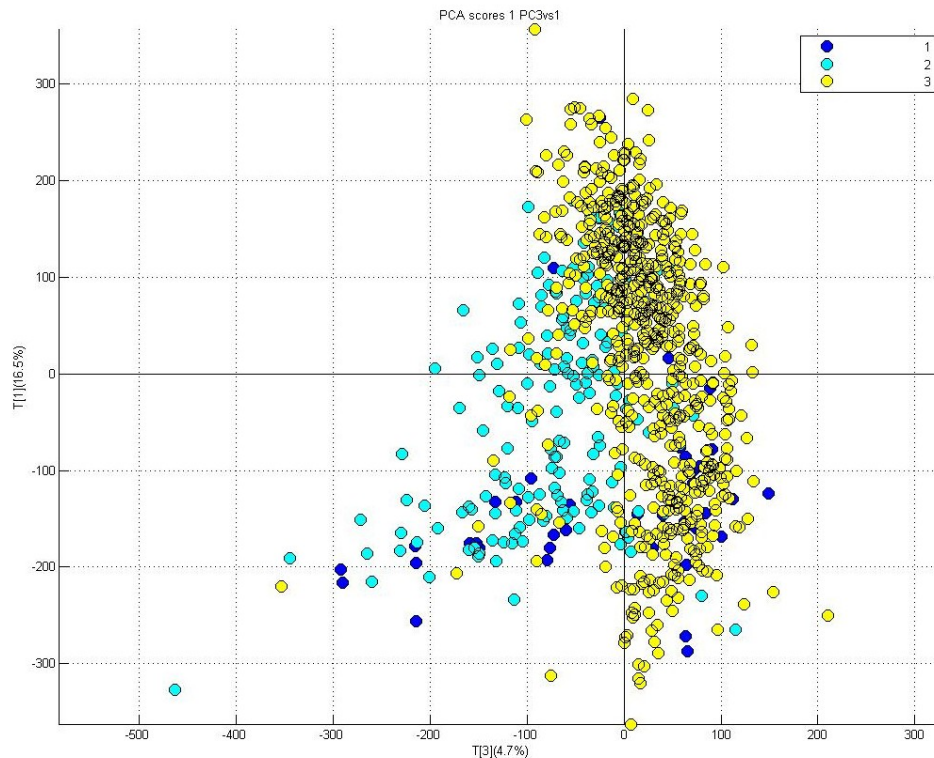


Figure 61: PCA Scores Scatter Plot of whole sample set of NMR stool spectra. Samples are coloured according to feeding route. Dark blue (1) represents total parenteral nutrition (TPN), light blue (2) is a mixture of parenteral nutrition (PN) with enteral nutrition (EN) and yellow (3) is enteral nutrition (EN).

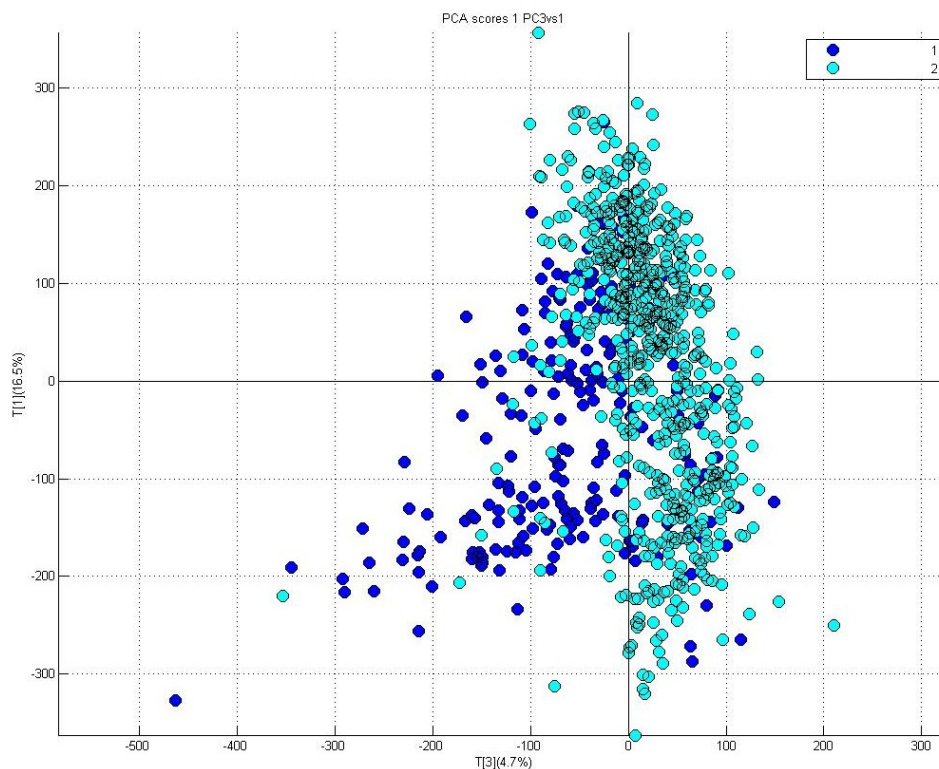
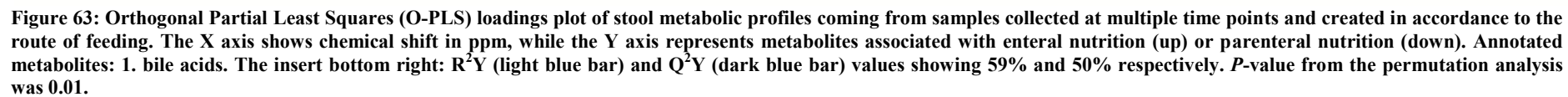


Figure 62: PCA Scores Scatter Plot of whole sample set of NMR stool spectra. Samples are coloured according to feeding route. Dark blue (1) represents samples collected while the baby was on the support of parenteral nutrition, while the light blue (2) are the samples collected while the baby was exclusively on enteral nutrition (EN).

The O-PLS model was built with two orthogonal components giving the final R^2Y and Q^2Y values 59% and 50% respectively. The results showed that the introduction of parenteral nutrition is associated with bile acids (0.63-0.83 ppm, 1.09 ppm s, 5.54 ppm s) (Figure 63).



5.3.vi.b Human milk vs. formula milk

The PCA scores plots of the whole data set showed a distinctive separation based on the type of enteral nutrition. The R^2 for PC1, PC2 and PC3 was 16.5%, 5.5% and 4.7% (Figure 64).

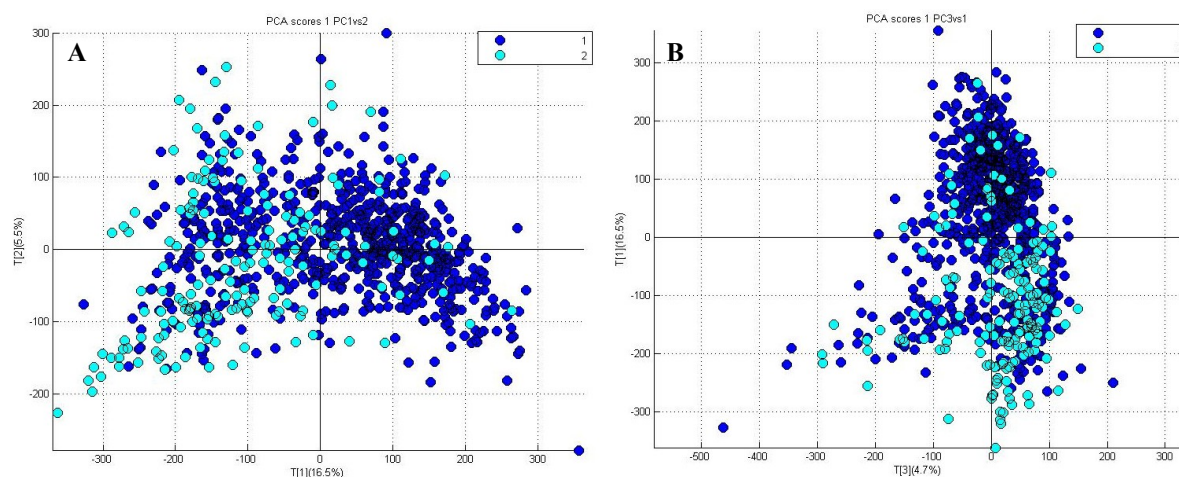


Figure 64: PCA Scores Scatter Plot of whole sample set of NMR stool spectra. Samples are coloured according to type of enteral nutrition. Panel A shows principal components 1 and 2 while panel B principal components 1 and 3. The dark blue (1) dots represent the samples collected while the baby was on human milk while the light blue (2) represents the samples collected while the baby was fed formula milk.

The supervised O-PLS model was calculated using two orthogonal components and the R^2Y and Q^2Y values were 46% and 34% respectively (Figure 65). The discriminant metabolites associated with human milk include: an N-acetylated glycoprotein (2.04ppm s), and lactose (3.29ppm t, 3.50-4.00ppm multi, 4.45ppm d, 4.67ppm d, 5.23ppm d). The formula milk are associated with: pantothenate – vit B₅ (0.91ppm s and 0.94ppm s), unknown chemical compound (0.92ppm d), N-acetyllysine (1.42ppm multi, 1.72ppm multi, 2.02ppm s, 2.98ppm t, 3.00ppm t) and unknown chemical compound (1.52ppm multi, 1.59ppm multi, 1.84ppm multi, 1.88ppm multi, 3.12ppm multi, 3.14ppm multi, 3.74ppm multi) (Figure 65).

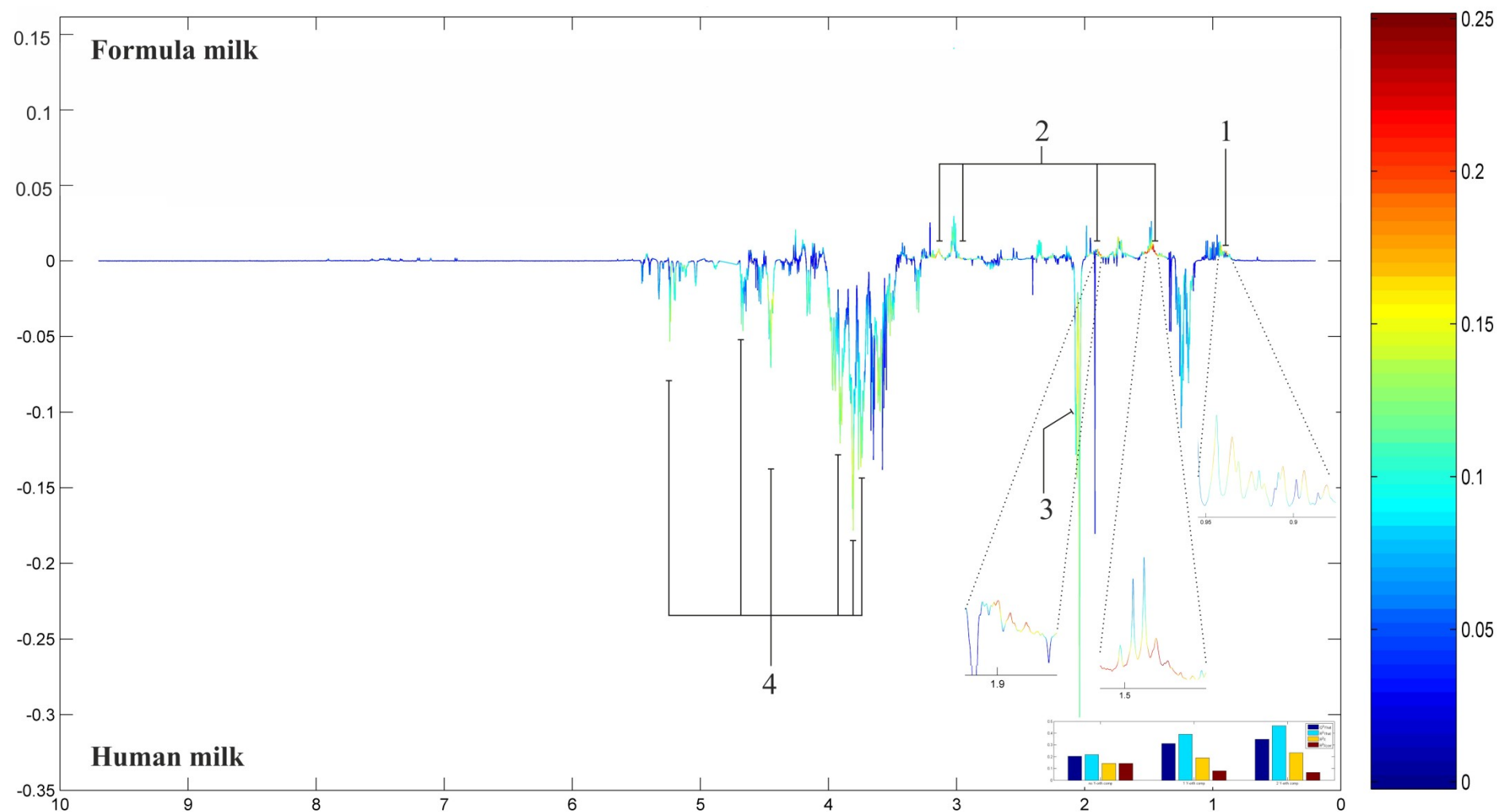


Figure 65: Orthogonal Partial Least Squares (O-PLS) loadings plot of stool metabolic profiles coming from samples collected at multiple time points and created in accordance to type of milk. The X axis shows chemical shift in ppm, while the Y axis represents metabolites associated with human milk (up) or formula milk (down). 1. pantothenate, 2. N-acetyllysine, 3. N-acetylated glycoprotein, 4. lactose. The insert bottom right: R^2Y (light blue bar) and Q^2Y (dark blue bar) values showing 46% and 34% respectively. P -value from the permutation analysis was 0.01.

Table 20: The summary of produced models testing the differences in metabolic profile of faecal water

Accounted for	Samples	PNA	GA	Sex	Mode of delivery	Labour	Human milk vs. formula	PN vs. EN
N/A	Whole data set	Y (3)	Y (1)	N	N	N	Y (2)	Y (2)
Gestational age	24 th week GA	Y (1)	X	Y (1)	N	N	X	X
	28 th week GA	Y (1)	X	N	N	N	X	X
Postnatal age	2 nd week of life	X	Y (0)	N	N	N	X	X
	3 rd week of life	X	Y (0)	N	N	N	X	X
Type of enteral nutrition	Formula milk	Y (1)	X	N	N	N	X	Y (1)
	Human milk	Y (1)	X	N	N	N	X	Y (1)

Key:Y – difference between the classes, positive Q^2Y above the threshold value of 0.05N – no difference between the classes, negative Q^2Y or below the threshold value of 0.05

(2) – the value in the brackets indicates the number of orthogonal components used in the model

X – no data as the analysis was not performed

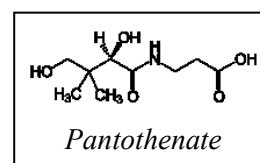
Table 21: The list of discriminant faecal metabolites identified to be associated with each analysed factor.**Key: GA – gestational age, PNA – postnatal age.**

Factor	Association	Metabolite	Chemical shift	
PNA	Increasing	Unknown	0.92ppm multi, 1.39ppm s, 1.98ppm s, 3.01ppm t, 3.18ppm t	
		pantothenate	0.95ppm s	
		cadaverine	1.45ppm multi, 1.72ppm multi, 3.01ppm t	
		Alanine	1.48ppm d	
		Acetate	1.92ppm s	
		unknown	2.13ppm d, 2.36ppm multi, 2.80pm d, 2.83ppm d	
		unknown	2.60ppm s, 2.87ppm s, 5.07ppm , 7.32ppm s, 7.61ppm multi	
		unknown	2.87ppm s, 3.35ppm s, 3.39ppm s, 3.42ppm s	
		trimethylamine	2.88ppm s	
		tyramine	2.93ppm t, 6.91ppm d, 7.23ppm d	
	thiamine	3.19ppm t		
Decreasing	Bile acids	0.78ppm s, 1.09ppm s, 5.54ppm multi		
GA	Increasing	unknown	1.21ppm s	
		unknown	1.84ppm dd, 2.21ppm d, 2.23ppm d	
		γ -aminobutyrate	1.90ppm multi, 2.30ppm t, 3.01ppm t	
		Acetate	1.92ppm s	
		N-acetylated glycoprotein	2.05ppm s	
		glutamate	2.12ppm d, 2.14ppm s, 2.36ppm multi	
		aspartate	2.68ppm q, 2.81ppm dd	
	Decreasing	betaine	3.28ppm s, 3.90ppm s	
		bile acids	1.09ppm s	
		unknown	1.37ppm s	
Parenteral nutrition		N-acetyllysine	1.48ppm multi, 1.71ppm multi, 1.99ppm s, 3.01ppm t, 3.18ppm t	
Human milk		bile acids	0.63-0.83 ppm, 1.09 ppm s, 5.54 ppm s	
		N-acetylated glycoprotein	2.04ppm s	
		lactose	3.29ppm t, 3.50-4.00ppm multi, 4.45ppm d, 4.67ppm d, 5.23ppm d	
Formula milk			pantothenate	0.91ppm s and 0.94ppm s
			unknown	0.92ppm d
			N-acetyllysine	1.42ppm multi, 1.72ppm multi, 2.02ppm s, 2.98ppm t, 3.00ppm t
			unknown	1.52ppm multi, 1.59ppm multi, 1.84ppm multi, 1.88ppm multi, 3.12ppm multi, 3.14ppm multi, 3.74ppm multi

5.4 Discussion

The metabonomic analysis of the faecal water from preterm babies showed that the diet and postnatal age were the strongest contributors to the metabolic changes. In case of diet, not only the route of nutrition but also the type of milk caused metabolic differences. Gestational age also contributed to the metabolic profile. There were no associations visible between the mode of delivery, presence of labour during delivery, sex of the baby and the faecal metabonome.

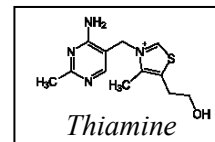
The findings for the association of the pantothenate (vitamin B₅) in stool with increasing postnatal age (PNA) are very interesting as I had reported the same findings for association of urinary pantothenate with the PNA (Chapter 4,



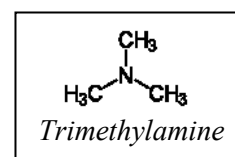
section 4.4). As previously mentioned the presence of vitamin B₅ in the human organism can originate from dietary sources like breast milk or from bacterial production [277, 279, 281-283, 285, 379, 380].

Our study also found the association of pantothenate with formula milk. To my knowledge, there is no literature explaining the possible reason for this. Although the preterm formula milk is supplemented with many vitamins, the B₅ is not one of them [381].

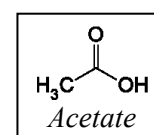
The finding that thiamine (also known as vitamin B₁) was associated with increasing postnatal age is not surprising as humans (and other animals) are not capable of producing it and rely on bacterial biosynthesis or plant consumption [379, 382]. *Bifidobacteria* and *Lactobacilli* are known to produce thiamine [380, 383]. Therefore, the older the baby is the higher the likelihood is of having colonisation of the gut by thiamine-producing bacteria. Thiamine has a role in neural function and its deficiency is very dangerous for the health of the brain [285, 382].



Another finding confirming our previous reports (from the urinary metabonome) is the association of trimethylamine with increasing postnatal age (Chapter 4, section 4.4). As previously mentioned, this compound is a product of bacterial decomposition of either choline or carnitine [264, 285, 288, 292, 294-298, 384]. The presence of trimethylamine in the stool of premature baby has not been reported before [385]. In animal models the presence of this compound has been associated with diarrhoea (in calves) and prematurity (in pigs) [386, 387].

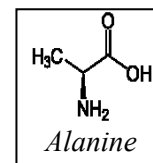


Acetate was found to be associated with both increasing postnatal age and higher gestational age. Acetate is the most common short-chain fatty acid (SCFA) and is a well-known end product of bacterial metabolism [388-393]. Its main function in the body is to provide fuel for both the liver and muscle metabolism but is also linked to reduced risk of: obesity, insulin resistance, and colon cancer [388-394]. The possible explanation for the observed association of acetate with increasing postnatal age may be the fact that the older the baby gets, the more bacteria colonises the gut thus the higher concentration of the end metabolite – acetate. Additionally, acetate is

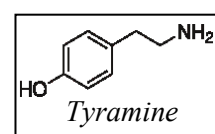


produced in higher concentration in infants than in adults [394]. No other SCFAs (such as butyrate or propionate) were observed associated with increasing postnatal age.

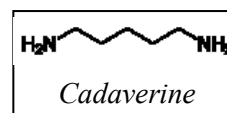
Alanine, which is associated with increasing postnatal age, is one of the most common amino acids [388]. As a non-essential amino acid, alanine can be produced by humans and is mainly used in the glucose-alanine cycle which produces energy from the protein degradation [388]. One of the possibilities why there is a higher concentration of alanine in the stool when the baby gets older can be the fact that older babies have higher muscle mass.



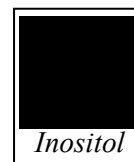
Tyramine is a derivative of a common amino acid – tyrosine [395-400]. Such conversion can be also performed by lactic acid bacteria [395-398, 401]. The association of tyramine with increasing postnatal age is probably due to the fact that with increasing age, the baby is more likely to be colonised by lactic acid bacteria that can perform the decarboxylation of tyrosine to tyramine. Tyramine is biologically relevant as it is a vasoactive amino acid and can trigger the nonpsychoactive peripheral sympathomimetic affects [395, 396, 398-400].



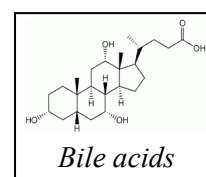
Cadaverine, which was reported associated with increasing postnatal age, is a product of protein hydrolysis [402-404]. This molecule originates from the amino acid lysine and is usually associated with the putrefaction of dead meat but is also produced in small quantities in living organisms and is a common metabolite produced during vaginal bacteriosis [405]. A very small concentration of this molecule was observed in the stool and the reason why it increases with postnatal age may be related to the increasing muscle mass of the babies, thus increased natural metabolism of proteins. Additionally, certain bacterial species that can be found in the human gut are capable of facilitating the degradation of lysine to cadaverine [404, 406-409]. Therefore it is possible that increased concentration of cadaverine is in fact related to changes of bacterial composition that occur over the time.



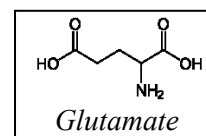
The last molecule associated with increasing postnatal age is scyllo-inositol – a common stereoisomer of inositol [310]. As previously mentioned in the Chapter 4 (section 4.4) inositol is produced from the glucose and is used as a building block cellular secondary messengers [310, 311]. It had also been found to improve the condition of various psychiatric disorders [310-312]. Since the human milk contains inositol, and the increasing postnatal age is linked with dietary shift from parenteral nutrition to exclusive enteral nutrition, it seems possible that increasing levels of scyllo-inositol are caused by the increased volume of enteral nutrition [316, 317].



Bile acids have been found to be associated with lower postnatal age, lower gestational age and parenteral nutrition. Exclusive parenteral nutrition in premature babies occurs mostly in the first days of life thus the common metabolic finding for both of those factors is not surprising. The association of bile acids with lower gestational age had been observed in the urinary metabonome (Chapter 4, section 4.4.). Additionally, it had been reported by other studies [273, 274]. As previously described, bile acids are produced either by the liver (primary bile acids) or by bacterial metabolism of the primary bile acids (secondary bile acids) [5, 33, 337]. The reason for higher concentration of the bile acids in the stool of younger (postnatal age-wise) babies may be due to malfunctioning reabsorption of those molecules. The maturing of the preterm gut continues after birth and the older the baby, the more functional the gastrointestinal tract becomes [351]. Therefore, it is possible that with the increasing postnatal age, the gut improves the absorption of bile acids.

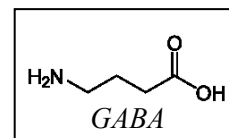


Glutamate, which is associated with increasing gestational age, is non-essential amino acid that has an important role as a neurotransmitter [410, 411]. Reduced levels of glutamate in premature babies had been found to be associated with the lesions in the white matter of the brain [412] and had been found in preterm babies with restricted growth rate [387]. Additionally, glutamate is a precursor of an important neurotransmitter – γ -aminobutyric acid (GABA) [410, 411, 413, 414]. The transformation of glutamate into GABA can be performed by



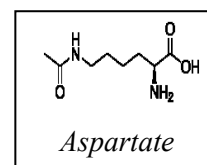
Lactobacilli and *Bifidobacteria* which are members of gut microbiota and as previously mentioned are commonly found in breast milk and breast fed babies [414-417].

GABA is another compound found to be associated with increasing gestational age. It supports the findings for its precursor – glutamate [410, 413, 414].



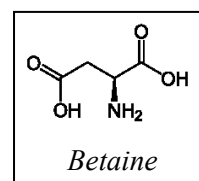
GABA, in a mature brain is an inhibitory neurotransmitter which reduces the excitability of neurons [413]. However, in the developing brain this molecule is responsible for excitement of the developing neurons [418-424]. Additionally, until the 32nd week of gestation, 90% of oligodendrocytes are still immature [425]. The findings of association of GABA with higher gestational age are extremely fascinating. This could potentially be one of the explanations why the more premature babies have higher chances of neurodevelopmental disorders such as autism [426-430].

Aspartate, is associated with increasing gestational age, is a non-essential amino acid that is produced in ureic cycle from ornithine and citrulline [431, 432].



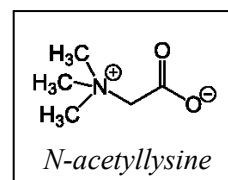
Aspartate can activate the same receptor as glutamate, however with a much weaker strength [432]. A study on premature infants found that all the dietary aspartate is metabolised and none is lost in the stool [431].

Betaine, which is associated with gestational age, is a name describing a group of molecules which lack hydrogen and have a positively charged cationic functional group [285, 433]. The first discovered betaine was trimethylglycine and in this

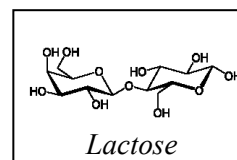


study this is the compound that was identified [285, 433]. Betaine is a product of choline oxidation whose main function in the organism is to support the methylation [285, 433, 434]. The most important metabolites that require methylation are coenzyme Q10, dopamine, serotonin and melanine [285, 433]. Betaine can enter the human organism from diet or be produced by gut microbiota [285]. A possible explanation why betaine is associated with higher gestational age, may be the fact that more premature babies are usually more sick thus are exposed to antibiotics for prolonged time.

N-acetyllysine, which is associated with lower gestational age and milk other than human, is an acetylated form of an amino acid – lysine [435]. The function of this molecule is binding to histones which in result regulates the availability of DNA for the gene expression [435].



Lactose is a disaccharide derived from galactose and glucose and is a well-known component of the human milk [283, 381, 436-450]. Association of lactose with breast milk is not a surprising discovery as it had been reported



before that formula milk has a significantly lower concentration of this sugar than breast milk collected from mothers who delivered prematurely. Although the metabonomic analysis were performed on the milk samples, not the stool samples of babies that consumed them, it seems logical that if breast milk has higher concentration of lactose, then there are higher chances that more of his molecule will be disposed undigested. Lactose is an important compound as its role is not only to provide nutrition for the baby, but also to promote the growth of healthy microbiota and model the immune system [443-446, 448-454].

We observed no difference in the stool metabonome in relation to the sex of the baby. Additionally, the mode of delivery or the incidence of labour had no influence on the faecal metabolic profile either. Perhaps the time frame I analysed was not enough to observe any permanent changes in the metabolism.

5.5 Conclusion

The findings from this study show that the metabolic profile of faecal water is strongly affected by the increasing postnatal age, gestational age at birth, and diet. A number of important molecules such as vitamins and neurotransmitters had been positively associated with the increasing maturity (both PNA and GA) in newborn babies. The metabolic profile in relation to the types of milk did not surprise as lactose is the most prominent sugar in the human milk.

Chapter 6 – The development of gut microbiota in the preterm baby

6.1 Introduction

Gut microbiota is an essential element in the development and maturation of the human gut. As previously described in the Background chapter (section 1.2.iii), there are number of factors affecting the development of microbiota. Those include: a) gestational age [61, 64, 79, 455], mode of delivery [27, 37, 46-58, 67], route of nutrition [41, 44, 71, 83], type of milk [37, 39, 50, 54, 57, 58, 61-63, 65, 66, 68-71, 73-76], and exposure to antibiotics [15, 37, 41, 44, 49, 51, 64, 71, 78, 79, 81].

A detailed report, conducted as a part of systematic review, on what is known for the previously mentioned factors and bacterial profiles can be found in the Background chapter (section 1.2.iii).

Postnatal age is related to the changes in bacterial population due to the fact that at this stage of life the gut microbiota is still in the process of development and in the first days of life there is a known sequence of bacterial succession, therefore perhaps there are some universal patterns of colonisation over the longer period of time. Gestational age may contribute to different patterns of colonisation for several reasons. Firstly, premature babies are more likely to be sicker and be on prolonged exposure to antibiotics. Additionally, their gut and immune systems are more immature thus they might react differently to first colonisers.

Mode of delivery contributes to the first exposure of the baby to the bacteria outside the womb, and since Caesarean section is performed in sterile environment and baby is firstly exposed to the skin and environmental bacteria. The current believe is that the gut microbiota of those babies differ from the babies born vaginally thus having contact with maternal vaginal and gut microbiota.

Route of nutrition in this study basically means that baby is either fed intravenously or orally. This factor is believed to have a strong impact on the gut bacteria as they require the nutrients in the gut to flourish. However, in case of premature babies, the shift of the route of nutrition from total parenteral nutrition (TPN) through mixed parenteral (PN) with enteral nutrition to finally fully enteral nutrition

(EN), is strongly linked with the postnatal age. Babies in first days of life are on TPN, and then with increasing postnatal age the ratio of EN over PN is increasing.

The effect of the type of milk on the gut microbiota is an interesting subject. There can be two types of human milk: mother's fresh milk and donor's pasteurized milk. Mother's fresh breast milk is a rich source of bacteria, while the donor's milk is considered bacteria-free due to the pasteurization process. Every type of human milk has oligosaccharides which act as prebiotics not only serving as a source of nutrition for the beneficial bacteria but also blocking the receptors on the surface of the pathogenic bacteria thus reducing the chance of pathogens colonizing the gut. Additionally, there are many different types of preterm formulas. Nutriprem1 – a formula given in Chelsea and Westminster Hospital, is a preterm formula rich in prebiotics, however not all the preterm formulas are enriched in prebiotics.

6.2 Experimental design

6.2.i DNA extraction

Samples for metataxonomic analysis had been selected if they met both of the criteria:

1. The matching stool and urine metabolic profile (collected on the same day) must have been available
2. The sample must have been collected from a baby for whom at least 3 samples had met the requirements of the 1st criterion and one of those must have come from the first 10 days of life.

Such strategy allowed us to ensure there would be multiple time points for each included baby and the starting point would be in the first 10 days of life. Additionally, the matching spectra would allow us to search for potential biomarkers for significant bacteria and bacterial processes. Please see the Materials and Methods chapter (section 2.3.ii) for the detailed description of the DNA extraction protocol. 214 samples were selected for the DNA extraction.

6.2.ii Sequencing

The sequencing was performed by the Research and Testing Laboratory (USA) on the MiSeq platform. The MiSeq platform is a next generation sequencing-by-synthesis method. The first step is the attachment of single strands of DNA and multiplication in order to create clusters. Then, with use of fluorescently labelled nucleotides, the synthesis of the complementary DNA strand takes place. The camera records fluorescent signals which allows the reconstruction of the DNA sequence. More details about the MiSeq platform can be found in the Materials and Methods chapter (section 2.3.i.b).

6.2.iii Sample pre-processing

The sequencing results were pre-processed and analysed using the Mothur program [254]. After assembling the two reads from the paired-end reads (reverse and forward) of the DNA sequence, the read depth per sample was analysed and 10 samples with less than 2000 reads were excluded. This reduced the number of samples for modeling from 214 to 204. The characteristics of the removed samples are presented in the Table 22. Please see the Materials and Methods chapter (section 2.3.iv) for the more detailed description of the preprocessing of the sequencing results.

Table 22: The comparison of the characteristics of samples that were included in the modeling and those that were excluded from the modeling.

Sample set	Postnatal age ¹	Gestational age at birth ¹	Sex ²	Mode of delivery ²	On parenteral nutrition ²	Breast milk given ^{2,3}	Formula milk given ^{2,3}
Included (n=204)	5 (3.56)	28 (1.67)	Male (56%)	Vaginal (42%)	Yes (79%)	Yes (81%)	Yes (31%)
Excluded (n=10)	5 (2.23)	27.5 (1.16)	Male (50%)	Vaginal (20%)	Yes (10%)	Yes (80%)	Yes (10%)

Key:

¹ median value and standard deviation given in brackets

² Percentage of total population given in brackets

³ Samples could have been collected when baby was on breast milk, formula milk, both types of milks or none, thus the values for the milk do not add up to 100%

6.2.iv Data analysis

6.2.iv.a *Analysed factors*

The factors that were analysed include: a) randomised groups b) postnatal age, c) gestational age, d) mode of delivery, e) labour f) sex, g) route of nutrition, h) type of milk. The impact of antibiotics on the gut microbiota was not assessed as all the babies in this study had been on the broad spectrum of antibiotics.

The first step of the analysis was to check if there are any differences between the randomised groups. Postnatal age (in weeks) was regressed against the microbial profiles in order to observe if there are any visible patterns of bacterial succession. The time frame of the samples collected from the 1st to the 15th week of life with the median of samples being collected at 5th week. Gestational age (in weeks) varied from 25th to 31st week with the median for samples collected being from the babies born in the 28th week of gestation (Table 22). Even though the detailed information on the type of delivery (vaginal delivery natural or with forceps, Caesarean section elected or emergency) was available, it was decided to keep the analysis simple and group babies according to the mode of delivery as in vaginal or Caesarean. For the occurrence of labour babies were divided into those that were not in labour (this included emergency caesarean section and elected caesarean section) and those that were delivered during the labour (this included spontaneous vaginal delivery, vaginal delivery with forceps, emergency caesarean section and elected caesarean section). There have not been enough babies in order to look at the influence of the labour for each type of Caesarean section separately. The route of nutrition was classified as TPN if there was no enteral nutrition given before, PN if the baby was on both enteral nutrition and parenteral nutrition for at least 2 days prior to sample collection, and EN if the baby was on the exclusive enteral nutrition for at least 2 days prior to sample collection. Due to a high variability in types of milks that the babies were receiving, the primary analysis for this factor was to answer the question what impact does the fresh mother's milk have on the gut microbiota. In order to do so mother's fresh milk (given exclusively or together with any other type of milk) was compared with other types of milk.

6.2.iv.b Reporting gut microbiota

The results for the gut microbiota was analysed for all the taxonomic levels: phylum, class, order, family and genus. The Figure 67, Figure 68 and Figure 68 show the taxonomic levels for the most significant (for this study) bacteria.

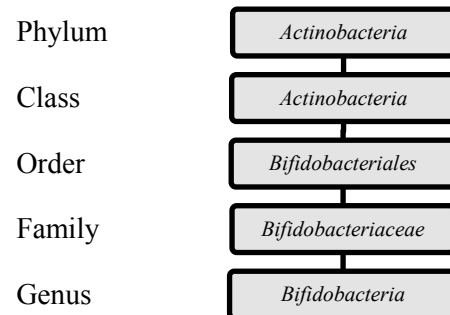


Figure 66: The most significant for this study members of the *Actinobacteria* phylum, organized according to their taxonomic level.

Development of gut microbiota and metabolic changes in preterm babies

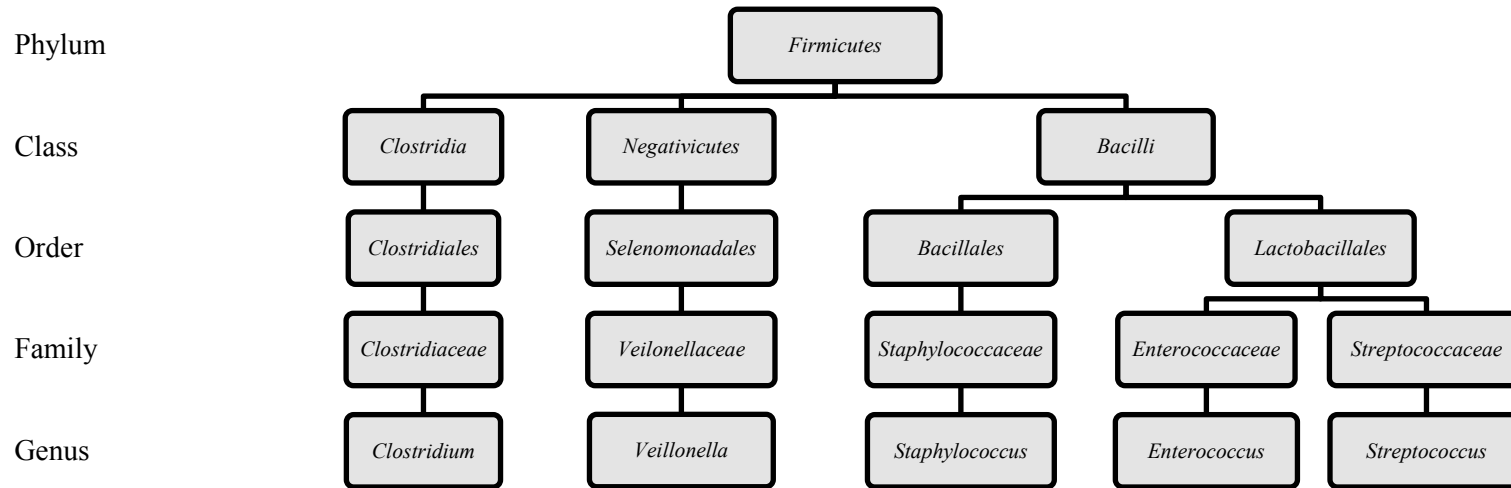


Figure 67: The most significant for this study members of the *Firmicutes* phylum, organized according to their taxonomic level.

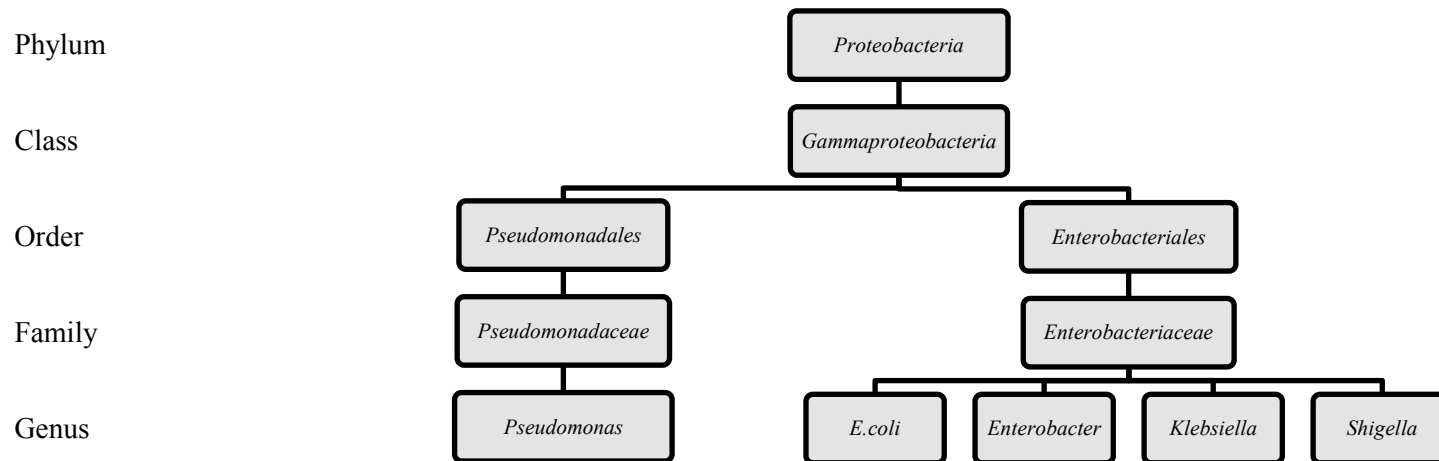


Figure 68: The most significant for this study members of the *Proteobacteria* phylum, organized according to their taxonomic level.

6.3 Results

The Principal Coordinates Analysis (PCoA) plots showed that there are 3 most abundant phyla, and 3 most abundant genera (Figure 69). The phyla were identified as *Proteobacteria*, *Firmicutes* and *Actinobacteria*. While the genera were *Enterobacteria* (belonging to the *Proteobacteria* phylum), *Staphylococci* (belonging to the *Firmicutes* phylum) and *Bifidobacteria* (belonging to the *Actinobacteria* phylum). The 79.2% of variability between the samples was due to the difference in the relative abundance of *Proteobacteria* and *Firmicutes*. While 18.5% was due to the changes of the relative abundance of *Actinobacteria* phylum. On the genus level the 61.8% of variability was explained by the relative abundance of *Enterobacteria* while the 20.4% of variability was caused by the difference in the relative abundance of *Bifidobacteria* and *Staphylococci*.

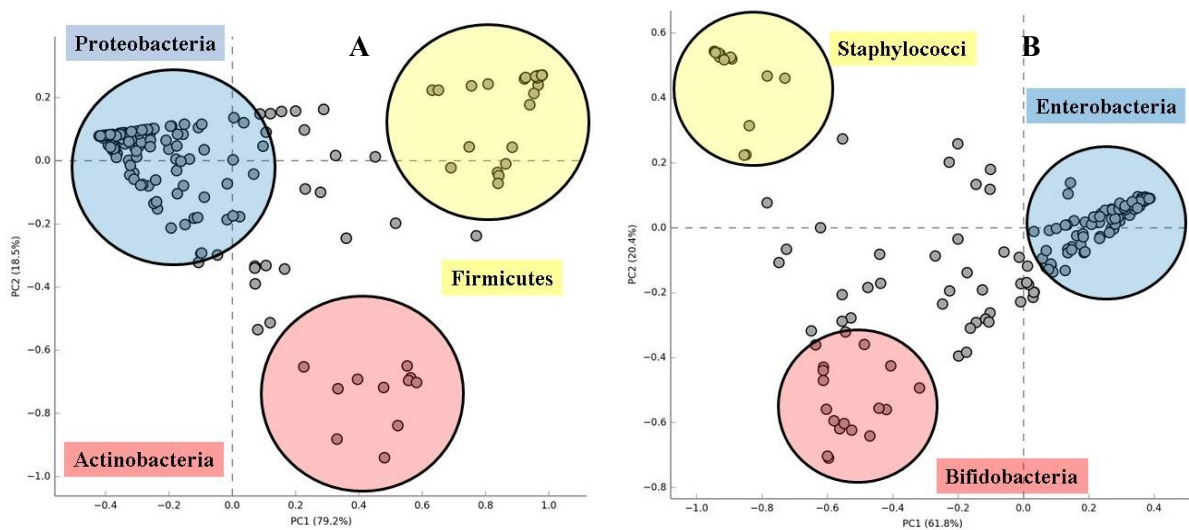


Figure 69: The Principal Coordinates Analysis plots presenting the separation based on the relative abundance of (A) phyla and (B) genera.

Proteobacteria was the most dominant of the all phyla with the average relative abundance of 65% (Figure 70). Second most abundant phylum was *Firmicutes* with the average relative abundance of 24%. *Actinobacteria* had average relative abundance of 10%. The remaining 1% of relative abundance was distributed between the remaining phyla: *Acidobacteria*, *Bacteroidetes*, *Fusobacteria*, *Terenicutes* and *Verrucomicrobia*. The most abundant families were: *Enterobacteriaceae* (64% of the average relative abundance), *Staphylococcaceae* (11%), *Bifidobacteriaceae* (10%), *Enterococceae*

(7%), *Streptococcaceae* (2%), and *Veillonellaceae* (2%). All the remaining families had the average relative abundance below 1%.

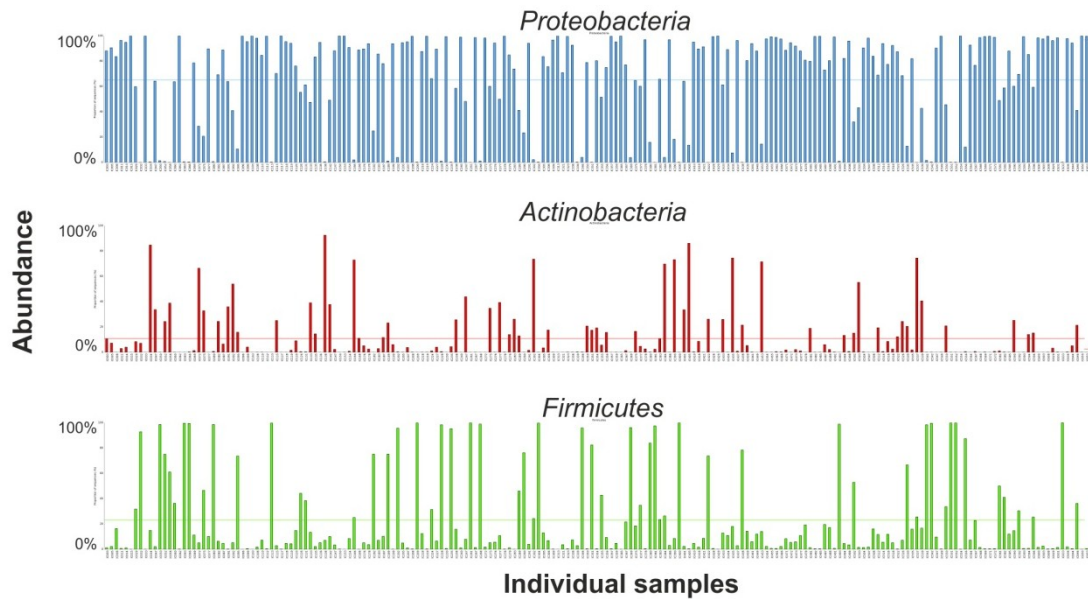


Figure 70: The relative abundance (Y axis) of the *Proteobacteria* (blue), *Actinobacteria* (red), and *Firmicutes* (green) phyla among all the samples collected from premature babies (X axis).

The results from qPCR were difficult to interpret without daily information on the antibiotic exposure. There were visible patterns suggesting that the bacterial loads were affected by the antibiotics as the concentration of bacteria was low (often below the detection threshold) at the birth for every baby. In the consecutive weeks the concentration of bacteria was increasing. In some babies there were occasions of sudden drop of bacterial load followed by the slow regrowth of the community. Because of that, the results of the qPCR will not be mentioned in the further subsections of this chapter.

6.3.i Randomised groups

No microbial differences have been observed for any of the randomised groups. The richness (Chao index) analysis showed that there was no difference between the types of lipids (11.5 for Intralipid and 11 for SMOF lipid) or the two concentrations of amino acids (10.87 for incremental and 12 for recommended daily intake). The diversity (Shannon's index) of the gut microbiota did not differ for amino acids (0.66 for incremental and 0.61 for RDI). However, the samples collected from babies on SMOF lipid had more diverse gut microbiota (Shannon's index = 0.73) than samples from babies on Intralipid (Shannon's index = 0.40).

6.3.ii Gut microbiota associated with gestational age (GA)

In order to analyse the association of gut microbial profiles with gestational age, samples were grouped according to the GA of the baby at birth. This meant that there had been samples collected at different time points. The sample size was not allowing creating classes with the same distribution of factors that would cause possible bias for the analysis. There were 20 samples for the 25th week of GA, 52 samples for the 26th week, 15 samples for the 27th week, 35 samples for the 28th week, 48 samples for the 29th week, and 34 samples for the 30th week. After analysing the groups using the Kruskal-Wallis H-test few interesting features were observed. At the phylum level, babies born before the 28th week of gestation had significantly lower relative abundance of the Actinobacteria. After investigating all the taxonomic levels, it was discovered that this trend was due to the *Bifidobacteria* genus (Figure 71). At the same time the relative abundance of the *Proteobacteria* phylum was much more constantly higher in babies born below the 26th week of gestation. The *Enterobacteriaceae* family contributed to this trend (Figure 71). No other microbial differences were observed.

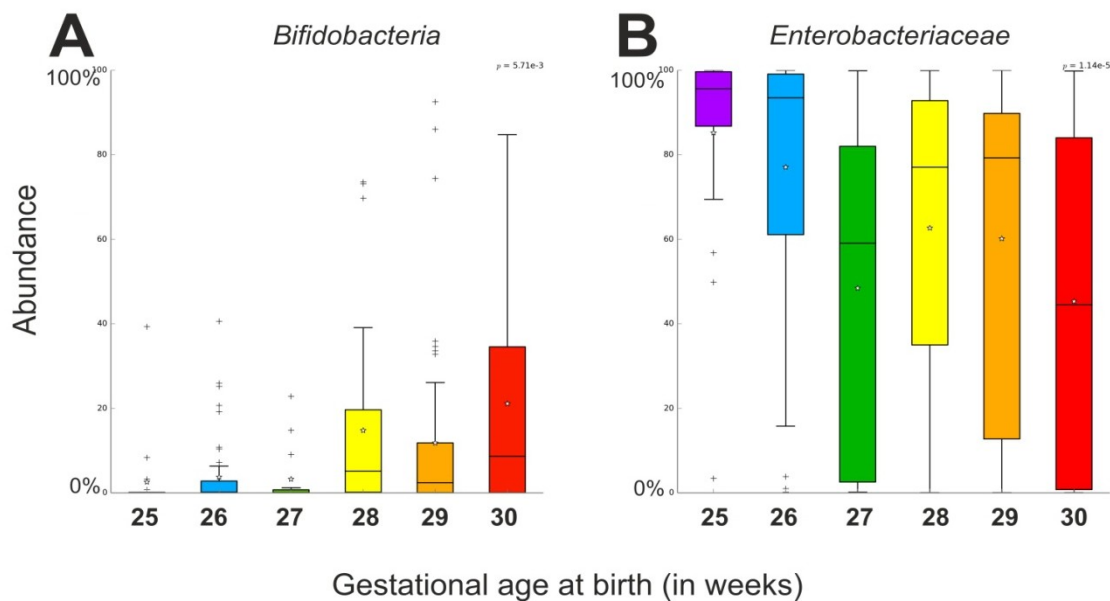


Figure 71: The relative abundance of the (A) *Bifidobacteria* genus and (B) *Enterobacteriaceae* family among samples collected from babies born at different gestational age.

6.3.iii Gut microbiota associated with postnatal age (PNA)

Comparing the gut microbiota association with postnatal age at the sample collection was also done using the Kruskal-Wallis H-test. Unfortunately some of the groups had less than required by the test 5

samples. The number of samples collected in each week of life were: 22 (week 1), 28 (week 2), 28 (week 3), 22 (week 4), 23 (week 5), 14 (week 6), 18 (week 7), 6 (week 8), 9 (week 9), 10 (week 10), 7 (week 11), 7 (week 12), 5 (week 13), 4 (week 14), 1 (week 15). The differences in PNA at five different levels: a) phylum, b) class, c) order, d) family, and e) genus were looked at. This gave a much deeper understanding of the development of gut microbiota over time.

There was no significant difference between the relative abundance at any of the taxonomic levels (phyla/classes/orders/families/genera). It was however possible to observe, that at the phylum level the relative abundance of *Firmicutes* had increased in the 1st week of life reaching its peak at the 2nd week after which it started to drop until the 10th week of life when the relative abundance started to increase again (Figure 72).

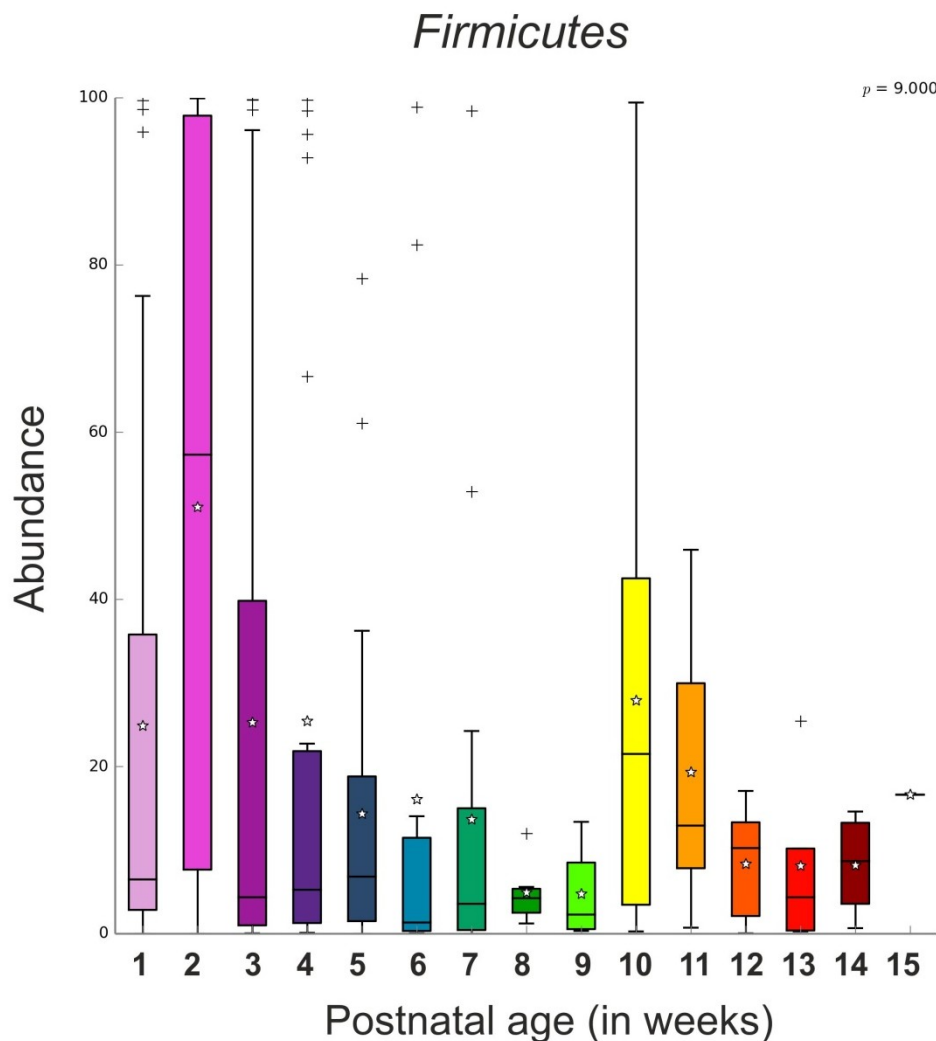


Figure 72: The relative abundance of *Firmicutes*.

After looking at the other taxonomic levels it was observed that the same was observed for the *Bacilli* class. The interesting shape of the relative abundance of this class was caused by two different genera *Staphylococcus* and *Clostridium* (Figure 73). There was no other difference visible.

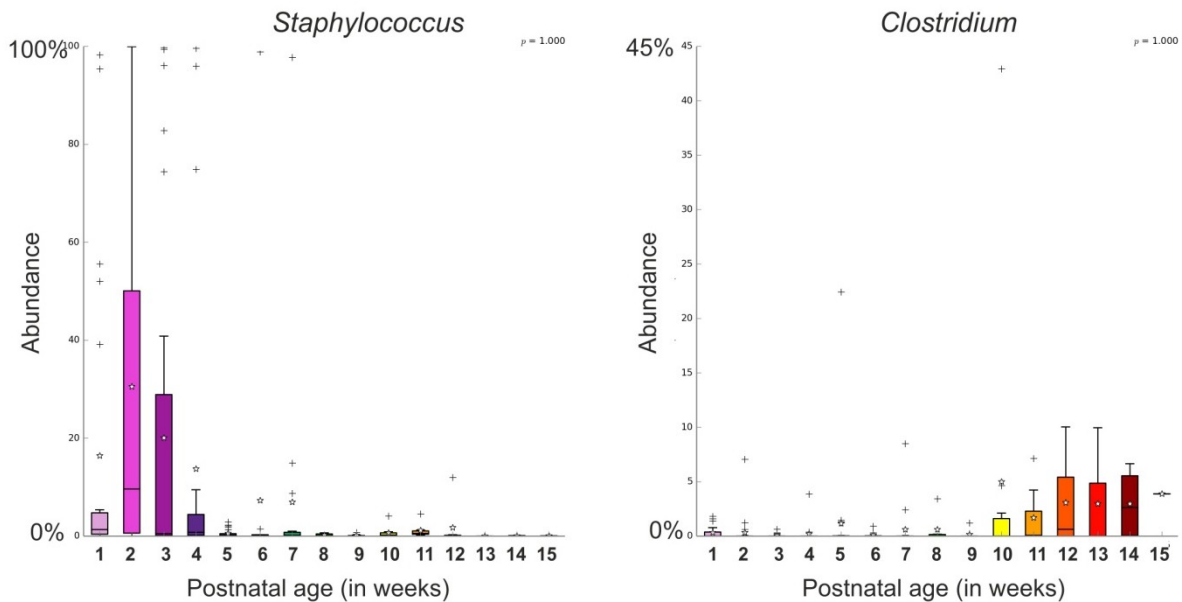


Figure 73: The relative abundance of the (A) *Staphylococcus*, and (B) *Clostridia* grouped according to the postnatal age of the baby at the time of sample collection. Please note that the scale for the relative abundance of *Staphylococcus* (segment A) is up to 100% while the scale for *Clostridium* (segment B) is up to 45%.

6.3.iv Gut microbiota associated with sex

There had been no significant difference at any of the taxonomic levels. The richness of the gut microbiota was slightly higher in females than in males with Chao alpha diversity being 12.5 and 11 respectively. The Shannon's diversity index was however similar in females (0.67) and in males (0.63).

6.3.v Gut microbiota associated with mode of delivery

At the phylum level, *Proteobacteria* was significantly associated with vaginal delivery while *Firmicutes* with Caesarean section (Panel A of Figure 74). At the class level, *Gammaproteobacteria* were identified as associated with vaginal delivery, while at the genus level *Clostridium* and *Staphylococcus* were found to be associated with Caesarean section (Panel B and C of Figure 74).

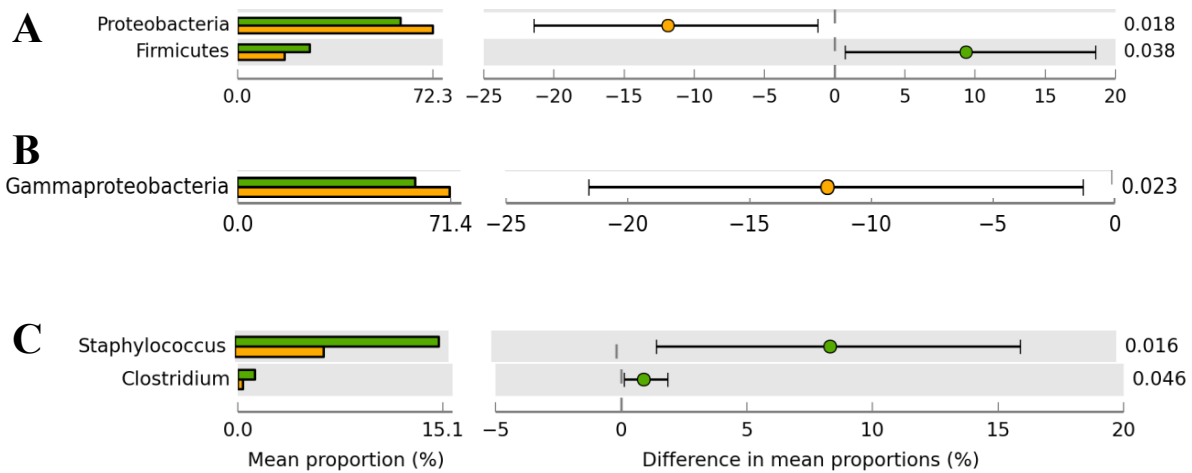


Figure 74: The (A) phyla, (B) classes, and (C) genera found to be significantly associated with the mode of delivery of the baby. Green represents babies born by caesarean section while orange babies born by vaginal delivery.

The richness and the diversity of the gut microbiota was similar in babies born by vaginal delivery (Chao index = 11.25; Shannon's index = 0.63) and in those born by Caesarean section (Chao index = 11; Shannon's index = 0.65).

6.3.vi Gut microbiota associated with occurrence of labour

There was no difference between the relative abundance of any of the groups of bacteria (tested at all the taxonomic levels). The richness of the microbiota was similar between samples collected from babies that were born during the labour (Chao index 0.62) and samples from babies that were not born during the labour (Chao index 0.68). The diversity was also similar with the Shannon's index equalling 11 and 12 for labour and non-labour respectively.

6.3.vii Gut microbiota associated with diet

The results from this study show that gut microbiota had been found to be associated with both the route of nutrition and the type of milk. The detailed nutrition metadata collected for the NEON study allowed proper classification of the samples according to the feeding.

6.3.vii.a Parenteral nutrition vs. enteral nutrition

The analysis of association of gut microbiota with route of nutrition was done after classifying the samples as either collected during: a) exclusive parenteral nutrition (TPN), b) mix of parenteral and

enteral nutrition (PN), or c) exclusive enteral nutrition (EN). The classification process is explained in detail in the 6.2.iv.a section of this chapter.

At the phylum level, the results showed that the relative abundance of the *Firmicutes* was increasing after the introduction of enteral nutrition, while the relative abundance of *Proteobacteria* started to decrease after the babies were orally fed (Figure 75).

In order to fully investigate the difference between the enteral nutrition and parenteral nutrition the analysis was continued comparing only two groups: the TPN group with the EN one.

The results confirmed the findings for the *Proteobacteria* and *Firmicutes* phyla (Figure 76). After investigating the other taxonomic levels, this difference between *Proteobacteria* and *Firmicutes* was explained by the relative abundance of two families: *Staphylococcaceae* (belonging to *Firmicutes* and associated with EN) and *Enterobacteriaceae* (belonging to the *Proteobacteria* and associated with TPN) families (Figure 76). There were no other differences in association to the route of nutrition at any of the taxonomic levels.

Both the richness and diversity were higher in samples collected while baby was on total parenteral nutrition (Chao index = 0.96; Shannon's index = 18.75) compared to the samples collected while babies were on exclusive enteral nutrition (Chao index = 0.63; Shannon's index = 13).

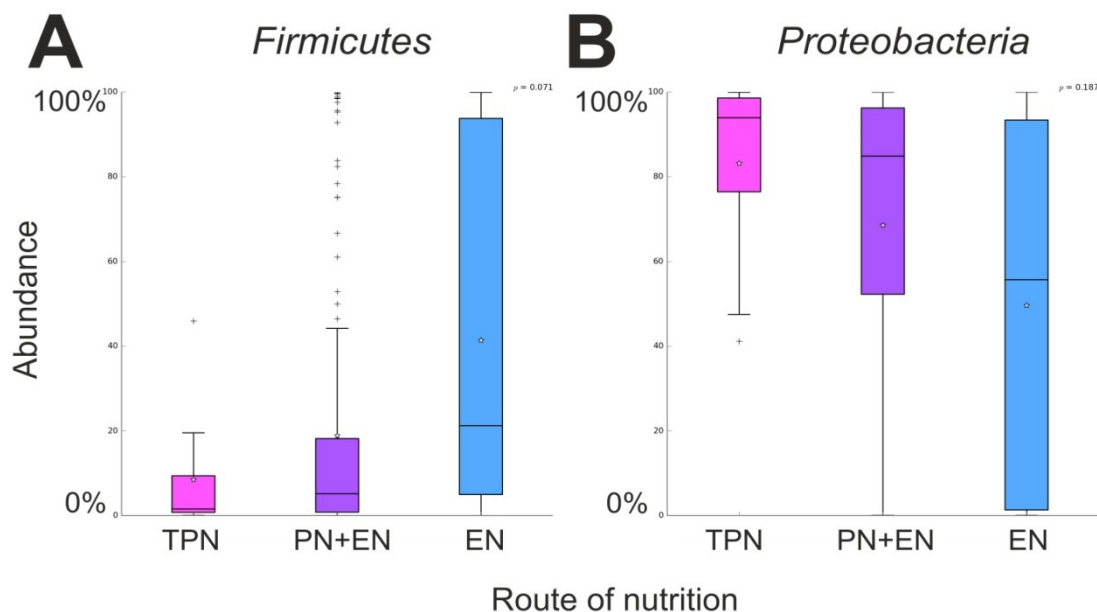


Figure 75: The relative abundance of (A) *Proteobacteria* phylum, and (B) *Firmicutes* phylum grouped according to the route of nutrition that the baby was on during the sample collection (pink = exclusive parenteral nutrition (TPN), purple = mix of parenteral and enteral nutrition (PN), blue = exclusive enteral nutrition (EN)).

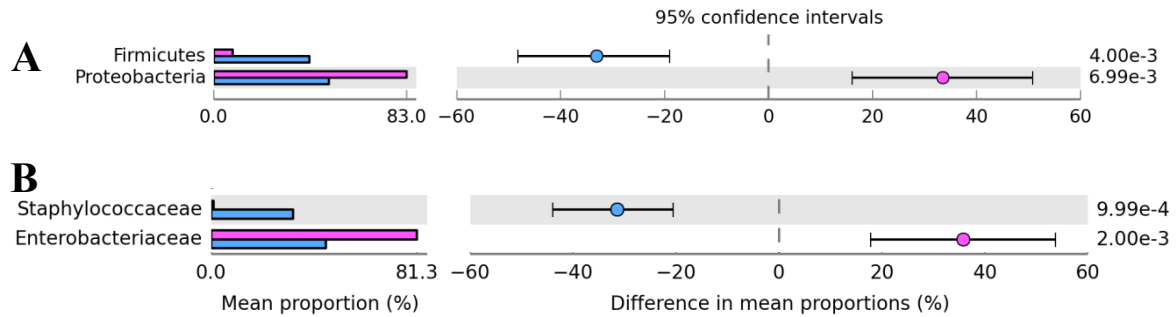


Figure 76: The (A) phyla and (B) families found to be significantly associated with the route of nutrition. Pink represents exclusive parenteral nutrition, while blue – exclusive enteral nutrition.

6.3.vii.b Breast milk vs. other milk

Samples had been classified as breast milk, or other milk based on the feeding for the seven days prior to the sample collection. Seven days are known to be enough to fully reprogram gut microbiota after the shift of diet. This gave 165 samples collected while on breast milk and 39 samples while the baby was not breast fed.

At the phylum level only *Proteobacteria* significantly differed in relation to presence or lack of breast milk (Figure 77). The investigation of the lower taxa found that this association was due to the higher relative abundance of *Enterococcaceae* in the samples obtained from babies that had been breast fed (Figure 77).

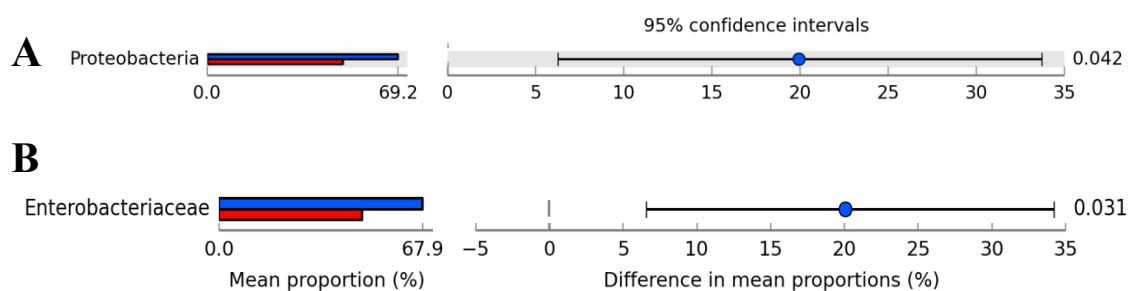


Figure 77: The (A) phylum and (B) family found to be significantly associated with the breast milk (dark blue). The samples collected while baby had been on any other milk for at least 7 days are presented in red.

The richness of the gut microbiota was lower in samples collected while baby was on breast milk than in those from babies on other milk (Chao index: 10 and 16.5 respectively). The diversity of the gut

microbiota was also lower in babies on breast milk (Shannon's index = 0.60) than those on other milk (Shannon's index = 1.09).

6.3.viii Regression of the stool metabolites against gut microbiota

The first analytical approach – using the O-PLS regression showed no discriminant metabolites for any of the bacteria of interest. However, the analysis performed by Daniel Homola found that *Enterococcus* was positively correlated with acetate, and *Bifidobacterium* was positively correlated with GABA.

6.4 Discussion

In this study it was observed that *Proteobacteria* are the most abundant phylum in the guts of the premature newborn babies. There are reports in the literature supporting our findings [371]. A recent study by Dogra *et al.* observed high levels of *Enterobacteriaceae* in the first week of life, which started to decline with the increasing postnatal age [456]. Additionally, Parm *et al.* reported that hospitalisation in the Neonatal Intensive Care Unit led to higher chances of being colonised by *Enterobacteriaceae*, a family that belongs to the *Proteobacteria* phylum [44]. In adults the most prominent phyla are *Firmicutes* and *Bacteroides* and their ratio is known to be related to the eubiosis/dysbiosis balance which is linked to the health outcomes [6, 457]. *Bacteroides* do not seem to play any role in the gut of premature babies, at least in this sample set, since they were almost absent or below the limits of detection.

The randomised groups did not have significantly different microbiologic profiles. However, a higher diversity of gut microbiota in samples of babies that had been randomised to receive the 20% SMOF lipid solution was observed when compared to babies randomised to receive the 20% Intralipid solution. To our knowledge, there is no biological explanation of this and a decision was made that the groups can be merged for further analysis.

Looking at the results it seems that the abundance of *Firmicutes* is more or less stable over the first 14 weeks of life; however, the members of this phylum change. In the first weeks *Staphylococci* are

responsible for the high representation of *Firmicutes*, but as their relative abundance start to decline, the *Clostridia* start to take over, which results in stable levels of *Firmicutes*. As shown in the above example, it is very important to analyse bacterial populations on all taxonomic levels as the crucial changes may go unnoticed when looking only at the higher taxa. The high abundance of the *Staphylococcus* genus in the first weeks of life can be explained by the fact that this genus had been previously reported as one of the first colonisers of the gut [36]. The reason for this may be in the localisation of *Staphylococcus*, which is mostly found on the human skin [458]. The contact with mother, especially during breast feeding, would expose the baby to this bacterium early on in life [39, 72, 73, 76]. Since the gut at this stage does not have a stable microbiota, it creates a perfect niche for *Staphylococci* to grow. After some time, when more oxygen is removed and the environment becomes anoxic, the facultative anaerobes which are *Staphylococci* may lose their influence in the presence of the anaerobic bacteria [36]. Alternatively, the gut may be colonised with bacteria which exclude and prevent colonisation of the gut by *Staphylococci*.

There was no significant microbial change observed in relation to gestational age at birth. However, there had been two trends observed. A higher relative abundance of the *Bifidobacteria* was visible after the 28th week of gestation. Dogra *et al.* had recently reported a similar observation [456]. Additionally, it had been reported before that full-term babies have in general a much higher relative abundance of *Bifidobacteria* than the preterm neonates [87]. One of the possible explanations may be the fact that mothers who went into the early premature labour may experience the problems with lactation more often thus their babies will not receive the fresh breast milk, which is a source of bacteria [451, 459]. Instead the early premature babies are often given the donor's milk which is pasteurised thus has no living bacteria in them. Such milk would not support the growth of *Bifidobacteria* at the same level as fresh breast milk or formula milk with prebiotics. The other possibility is that since more premature babies are usually sicker, they require receiving antibiotics more frequently thus their microbiota is more at risk.

As suggested by our hypothesis, mode of delivery is associated with changes in the gut microbiota. The association of *Gammaproteobacteria* with vaginal delivery have been reported before, especially

for *Enterobacter* [48, 50], *E. coli* [44, 48, 49, 53, 55, 59] and *Klebsiella* [46, 53, 55]. All of those genera are considered normal members of the maternal gut microbiota and when a baby is born by vaginal delivery the first exposure to bacteria from the outside world is during the passage through the birth canal thus the baby is exposed to maternal perianal bacteria [44, 46, 48-50, 53, 55, 59]. Babies born by caesarean delivery are exposed to environmental and skin microbiota which explains the higher abundance of the *Staphylococcus* genus [48, 50, 56]. There have also been previous reports supporting the association of *Clostridium* with caesarean section [37, 47-50]. However, some researchers found that specific species of *Clostridia* (*C. perfringens*, *C. leptum* and *C. difficile*) are not affected by the mode of delivery [47, 57, 58]. In accordance to the findings from the literature, in our study the population of *Streptococci* [48, 50, 56-58] and *Lactobacilli* [57, 58, 67] did not differ between the babies in regard to mode of delivery. However, contrary to other studies, this study did not find any difference in the relative abundance of *Bacteroides* [37, 46-49, 57, 58], *Bifidobacterium* [27, 37, 46, 47, 49-55, 456], *Bacillus* or *Atopobium* [48, 50, 56-58]. The reports from the literature regarding the diversity and richness of the gut bacterial population in newborn babies are contradictory. I had not observed any difference in relation to mode of delivery, however three studies had reported higher diversity in samples from babies born by Caesarean section [48, 59, 83] and one study found stool of vaginally delivered babies to be more diverse [46].

The introduction of enteral nutrition, as predicted, is associated with a change in microbial profile of the premature baby's gut. The genera found to contribute to the difference between the total parenteral nutrition (TPN) and enteral nutrition (EN) were *Enterobacteriaceae* and *Staphylococci* respectively. One of the reasons why the relative abundance of *Staphylococci* is higher when babies are fed orally could be due to at least two reasons. Since the majority of the babies in this analysis were breast fed, one of the reasons for a higher relative abundance of *Staphylococci* could be due to the contact with maternal skin during feeding [364]. Additionally, human milk oligosaccharides are known to promote the development of *Staphylococci* [363]. *Enterobacteriaceae*, as previously mentioned, are the bacteria colonising the baby's gut straight after the birth. And since there is limited consumption of the bacteria, the primary community may not be affected until the enteral nutrition takes place. The

results for *Staphylococci* being associated with EN, and the non-significant trend was observed of this family's relative abundance increasing drastically after the first week of postnatal life may be linked to it. Total parenteral nutrition occurs usually only in the first few days of life (1st week of PNA). Therefore it is possible that the interesting trend for the *Staphylococci* in accordance to the postnatal age is in fact related to the introduction of enteral nutrition. It is however surprising that no significant difference between the abundance of *Lactobacilli* or *Bifidobacteria* was observed as those are the species known to be correlated with both breast and formula milk [42, 52, 68, 81, 437, 438, 460]. Unfortunately, there are very few studies comparing gut microbiota of preterm babies on TPN vs. those on EN. In the systematic review that I had conducted in 2012 only 4 studies, which looked at the route of nutrition in premature babies were identified [41, 44, 71, 83]. All of those studies reported a lower bacterial diversity when baby was on TPN [41, 44, 71, 83]. Our study reported the opposite findings. The explanation why the gut microbiota becomes less diverse after the introduction of enteral nutrition could be that milk (regardless if human or formula) provides nutrition for selected bacteria which leads to a stable but not diverse microbiota. This argument is supported by the metabonomic analysis as I had observed the reduction in diversity of metabolic profiles in both urine and stool samples after the introduction of enteral nutrition.

When analysing the influence of breast milk on the gut microbiota, a significant difference at the phylum level was observed. *Proteobacteria* were more abundant among the breast fed babes than among those fed with other types of milk. Ward *et al.* had reported a similar finding before [451]. In our study, the higher relative abundance of *Proteobacteria* was due to the *Enterobacteriaceae* family. A higher relative abundance of *Enterobacteriaceae* in breast fed babies does not comply with the other research, which had reported this family to be associated with formula feeding [39, 50, 57, 58, 65, 71, 72, 76]. However, Neut *et al.* observed the same association as reported in this study [73]. In contrary to other studies, no difference between the type of milk and the abundance of *Clostridia* [37, 66, 69, 72-76, 86], *Lactobacilli* [63, 66, 67, 70, 76] (though three studies found no difference [57, 58, 73]), *Streptococcus* [50, 66, 72, 73] and *Bifidobacteria* [42, 51, 54, 57, 58, 61-63, 66, 68-70, 72-76] was found. It is important to remember that the majority of the studies comparing breast milk with

formula milk have been done on full term babies and often using different to our method of bacterial detection. A recent study done by Khodayar-Pardo *et al.* found that mothers who went into a premature labour had significantly lower concentration of *Bifidobacteria* in their breast milk than women who gave birth in term [461]. Additionally, it was observed that adding oligosaccharides to the preterm formula milk increases the counts of *Bifidobacteria* in the stool samples of babies fed with this milk [460]. What is more, the relative abundance of *Bifidobacteria* in the whole sample set was 10%. Similar finding has been observed in preterm babies that participated in a probiotic study done by Underwood *et al.* [462]. The babies selected for our analysis had not been supplemented with any probiotics. Our findings regarding the lower diversity and richness of gut microbiota in babies on breast milk confirm the reports from the literature [66, 68, 87, 463]. Breast milk introduces a very limited number of bacterial species which may contribute to the lower diversity of the microbiota in the baby. Additionally, as mentioned before, breast fed babies were predominately colonised by *Enterobacteriaceae* thus if one family of bacteria contributed to the highest relative abundance, there was not much diversity in the gut.

The sex of the baby had no association with changes of gut microbiota profiles. Additionally neither the richness nor diversity differed between the males and females. This suggests that at this stage of life, there are no differences between male and female babies when it comes to the colonisation and development of gut microbiota. The reasons why adult females and males have differences in the gut microbial profiles are probably due to the hormonal changes that are triggered during puberty.

We also did not observe any microbiological differences in accordance to the presence or lack of labour during the delivery. It is however important to mention that the labour group consisted of vaginal delivery, emergency caesarean section and elective caesarean section, while the non-labour deliveries included only the latter two options. The numbers of the babies were too small to look at the labour vs. non-labour issue in each of the groups separately.

We also observed that both the baby that developed sepsis (positive blood culture – coagulase-negative *Staphylococci*) and the baby that was tested for sepsis but had negative result showed the

same pattern of microbial changes – increasing relative abundance of *Enterobacteriaceae* that started one week prior to the blood culture taken, and resulted in almost 100% of bacteria belonging to this family.

The regression of discriminant metabolites against the relevant genera showed that *Enterococcus* was positively associated with the presence of acetate [464, 465]. Similar finding has been observed by my colleagues in their sample sets (data not published). The fact that GABA was positively associated with the presence of *Bifidobacteria* is an interesting finding as there had been reports suggesting that those bacteria are able to produce GABA [466, 467].

6.5 Conclusion

Our hypothesis that mode of delivery and type of diet affect the microbial community has been confirmed. The visible pattern of gut microbial changes over the time is probably associated with the transition from the parenteral to enteral nutrition. The introduction of enteral nutrition leads to the reduction in the diversity of gut microbiota which is in accordance with the metabonomic results. The differences in the gut microbial profiles in relation to mode of delivery are probably caused by the different exposure to the environmental/maternal bacteria straight after birth. The metataxonomic analysis on the samples collected as a part of the NEON trial had proved that early life interventions contribute to the changes in the gut microbiota of the premature babies.

Chapter 7 – General discussion

The aim of this PhD project was to investigate how microbial and metabolic profiles develop in preterm babies. Factors such as type of parenteral nutrition (the randomised feeding groups of the NEON study), route of nutrition (parenteral nutrition vs. enteral nutrition), type of milk, mode of delivery, mother being in labour during delivery, gestational age and postnatal age and their effect on the gut microbiota and metabolic profiles were investigated. I did not investigate the influence of antibiotics as all the babies were given broad spectrum antibiotics in the neonatal intensive care unit (NICU).

The results from this study show that the biggest contributors to the changes in metabolic profiles of both stool and urine were postnatal age (PNA), gestational age (GA) and route of nutrition. However, it is important to remember that the route of nutrition in premature babies is linked to the postnatal age, so the shift from total parenteral nutrition (TPN) through mixed parenteral nutrition (PN) with enteral nutrition to fully enteral nutrition (EN) is linked with postnatal age. Preterm babies are on TPN for only the few first days of life and for the majority of their stay in NICU (at least in this study) they are on both PN and EN. The ratio of EN over PN increases with postnatal age until the baby is ready to be fully on EN. The start of enteral nutrition not only introduces chemical compounds in to the gut, but also stimulates the development of the gut microbiota [41, 503, 510-512]. Metabolites such as sugars and bile acids were observed to be associated with both early parenteral age and TPN, and as described in previous chapters both of those factors (lack of enteral nutrition, and early postnatal age) can on their own have a biological explanation for such metabolic changes. Therefore, it is very difficult to decide which of the factors – postnatal age or route of nutrition, is in fact responsible for observable metabolic changes. On the other hand, it is not really a limitation as such situation is a representation of what is happening to the premature baby while in NICU.

In general, the major metabolic and microbiota changes over time represented themselves as metabolites which can be linked either to bacterial activity or composition of human milk. Since babies are more likely to go on full enteral nutrition when they are older (higher postnatal age) it makes biological sense why there are metabolites known to be present in human milk (such as vitamin

B₅, inositol, carnitine) to be associated with increasing postnatal age. The bacterial changes over time, for example the sudden increase in the relative abundance of staphylococci from the second week of life, can also be explained by the changes in the nutrition pattern. In the 2nd week of life babies were supplemented orally with either breast or formula milk (81% of samples were collected while the baby was breast fed). Since breast feeding is associated with colonisation by staphylococci (present on skin and in the milk, and having their growth supported by the human milk oligosaccharides [363]) the changes of microbial profiles over time closely mirror the timeline of feeding. The influence of the type of milk on the metabolic profile gave expected results of lactose being positively associated with breast milk. Marincola *et al.* compared the metabolic profiles of breast milk expressed from women that delivered babies preterm with formula milk. The results showed that breast milk has a significantly higher concentration of lactose than formula milk [513]. The same study also found differences in the composition of breast milk depending on the gestational age when the preterm birth occurred. This difference in the chemical composition of milk linked with the level of prematurity of the baby may in fact be a key to understanding why there are observable metabolic and microbiota differences related to the gestational age. Results reported in this study show that babies born below the 28th week of gestation had significantly lower relative abundance of the beneficial bifidobacteria. This low abundance could be caused by one (or all) of the following factors: different composition of milk depending on the gestational age, the numbers of bifidobacteria in the breast milk are positively associated with the increasing gestational age, and last but not least higher exposure to antibiotics of more premature babies due to their more advanced health problems. Additionally, the average relative abundance of *Bifidobacteriaceae* in stool samples collected from the babies in this study was on the level of 10%. The studies conducted on full term babies report the relative abundance of bifidobacteria at the levels from 18% to 75% [514, 515]. Moreover, the results showing that the assumption that premature babies exclusively breast fed will have high abundance of bifidobacteria is wrong. The high exposure to antibiotics is probably the leading cause of lower relative abundance of bifidobacteria, when compared to full term babies. Though, it is important to remember that this study was not designed for the analysis of the influence of gestational age at birth with the abundance of

Bifidobacteria and in order to confirm this hypothesis a properly designed study on both preterm and full term babies should be conducted.

Another interesting finding is that neither being in labour during delivery nor the mode of delivery affected the metabolic profile of the babies, even though the latter factor had been found to be associated with changes in the gut microbiota. One possible reason for this is that the changes in gut microbiota in relation to the mode of delivery do not have any metabolic effect on the babies, at least in this time frame. It is possible that the bacteria that had been found to differ – *Staphylococcaceae* for Caesarean section and *Enterobacteriaceae* for vaginal delivery, have similar metabolic activity. Perhaps the general function of those bacteria is stable regardless of the species that comprise it. This similarity would mean that even though there are different bacteria present in the gut, their metabolism and the metabolites produced are similar. Another option is that the discriminant bacteria produce different metabolites which are used by other species common in both groups of the babies, and what we observe as a metabolic profile are simply the end products of the long bacterial metabolic chain. Our metataxonomic results did, however, support what had been previously reported in full-term babies and thus supports the theory that the gut microbiota is affected by the first environment with which the baby has contact after birth [37, 47-50, 56]. *Enterobacteriaceae* are common bacteria found in maternal perianal area and vaginally delivered babies have a profound contact with this area. At the same time, babies born by caesarean section are not exposed to maternal gut microbiota and have in fact higher relative abundance of staphylococci which are known to be a member of the skin microbiota.

It is difficult to compare the findings from this present study with the literature as the majority of observations were done in full-term babies. The development of gut microbiota and the factors affecting it have been investigated since the 80's [37, 39, 41, 43, 46, 61, 62, 501-505], but very few have been conducted in premature babies [220, 506-509]. Additionally, even if a study was conducted on preterm babies, factors such as time of sample collection (postnatal age), analytical technique for microbiota analysis or even choice of bacterial DNA extraction kit, and the primers used for the amplification of a one of many variable regions of the 16S rRNA gene could lead to different results.

Therefore it is difficult to directly compare results found in the literature with this study as there are many factors that can affect the result.

The limitations of this study were that it was not designed to look into the factors that have ultimately been analysed. Since there was no microbial or metabolic difference between the randomised groups exploratory analysis was conducted to look at all the other early life interventions that could have interesting results with full awareness that the study was not created for it. The study was randomised for type of parenteral nutrition and it had an equal distribution of all other factors between the four randomised groups. However, because principal component analysis was used, which is an unsupervised model; the interference of the other factors may be observed (which would be corrected for in the standard statistical approach). Therefore in order to remove any possible interference the subgroups must be individually compared (e.g. analyse mode of delivery in samples collected from the same week of postnatal age from the babies born at the same gestational age). This quite often resulted in a sample size too small to apply any meaningful statistics on therefore certain compromises had to occur. It is important to remember that metabonomic analysis of human samples is extremely challenging. Humans are by nature diverse and trying to compare metabolic differences between two groups as an effect of an intervention/event is difficult. The analytical approach, although perhaps was not perfect, was the best that could have been done in those circumstances.

The major strength of this study is that samples were collected as part of a clinical trial therefore providing very detailed high quality data. The babies were assessed on a daily basis and access to the feeding diaries for each day of their lives was available up until the point of discharge. This allowed for efficient classification of the samples, i.e. not purely based on the feeding that the baby had received on the day of sample collection but also based on the type of feeding from the previous days. The gut microbiota requires a certain amount of time to shift – 24h to shift as a result of change from TPN to EN and 7 days in adults to shift as a result of dietary change. Over 1000 urine samples and another 1000 stool samples were collected from 160 babies giving the opportunity to build strong models and even work on preselected features. The sample collection was standardised and efficient and gave a good representation of samples coming from the same baby from consecutive weeks which

gave an interesting opportunity to look into the development of gut microbiota over time (postnatal age). Additionally, the gold standard methods for the microbial analysis, (16S rRNA gene sequencing) was used which means that the results should be easily comparable with future studies conducted on preterm babies. Last but not least, in contrary to many other studies in the field of gut microbiota, the bacterial populations were assessed alone but also in context of their ecosystem. The metabolic profiling combined with bacterial identification builds a whole story of the changes occurring in the gut of the baby. It is far more informative when we know not only who the gut inhabitants are but also what they are doing. Microbial profiling on its own is not as reliable as many of the identified species (especially when culture independent identification methods are used) may have been dead for a while or simply inactive therefore not having any impact on the gut environment.

In conclusion, the results of this study contribute strongly to the scientific field. Our results are in accordance with what is considered the most accurate techniques available on the market and the samples came from a clinical trial. Evaluation of the association between the early life interventions and both microbial and metabolic changes were investigated. Previous findings have been re-confirmed on the microbiota and association with the mode of delivery, and commenced a new discussion on why the microbiota changes are not accompanied by the metabolic changes. Additionally, it has been shown that there exists a strong link between the first several weeks of life post-partum (in preterm babies) and the shift in the route of nutrition. The fact that introduction of fully enteral nutrition leads to a more unified microbial and metabolic profiles (lower diversity of bacteria and higher clustering of metabolic profiles on principal component analysis) is another argument in support of the approach of introducing enteral nutrition as soon as it is safe and possible.

The future studies should conduct a further investigation into the link between the factors associated with altered microbiota and long term functional clinical outcomes. So far it is known that the gut microbiota is an important component of the mammalian organism and that it can influence the maturation of the gut and the development of the immune system [3, 10-13, 23, 25, 468-477]. There are a number of health complications that are linked to the alteration in gut microbiota, however, it is

not yet certain if this is merely an association or a correlation with a direct causality [1, 6, 28, 137-140, 146, 318, 457, 473, 478-500]. There are three possible hypotheses regarding this aspect, and they should be investigated in greater detail by future research.

The first hypothesis suggests that the dysbiosis (alteration in the gut microbiota) affects the host organism causing metabolic changes which result in functional clinical outcomes. This hypothesis is usually supported by the results obtained from the animal studies, especially those performed on germ free subjects. However, it is important to remember that animal species differ from each other and the mechanism observed in mice may not reflect that in humans. The other approach to investigate the potential effect of alteration of gut microbiota on the health is to design a prospective study. Such studies require a large number of participants and regular gut microbiota examination over a prolonged period of time with hope that an appropriate number of participants will develop a certain disorder. Such studies are expensive and due to prolonged time, they are at risk of high drop-out rate which as a result decreases the chances of observing the outcome of interest.

The second hypothesis is that the correlation between the alteration of microbiota and functional clinical outcomes is reversed and therefore suggests that gut microbiota changes are as a result of a functional clinical outcome. Again, the best way to investigate this hypothesis is by a prospective large scale study. So far the majority of the studies in the field have either a very small sample size or are retrospective therefore the causality cannot be determined.

The third popular hypothesis is that there is a common factor causing both the shift in microbiota and the functional clinical outcome. This hypothesis is the most difficult to investigate due to the fact that if any of the above hypotheses are supported, it would still not answer the question whether or not there is an unknown underlying factor or factors. It is possible that all of the above hypotheses are correct and that different functional clinical outcomes have a different type of correlation/association with altered gut microbiota.

Since we already know how the gut microbiota is affected by various early life interventions it is possible to design a study that would account for the factors causing potential bias, thus the models

can be adjusted leading to obtaining more reliable and accurate results. Since the animal studies have reported a strong impact that the first colonisers of the gut have on the immune system and the gut itself, more and more interest has been put on the development of the microbiota in humans. Researchers believe that understanding the mechanism of gut colonisation and programming of the immune system to differentiate between the beneficial bacteria and pathogens may hold the answer to the causality of allergies and potentially lead to the elimination of this disorder [10, 23, 477].

References

- Greiner, T. and F. Backhed, *Effects of the gut microbiota on obesity and glucose homeostasis*. Trends Endocrinol Metab, 2011. **22**(4): p. 117-23.
- Traskalova-Hogenova, H., et al., *The role of gut microbiota (commensal bacteria) and the mucosal barrier in the pathogenesis of inflammatory and autoimmune diseases and cancer: contribution of germ-free and gnotobiotic animal models of human diseases*. Cell Mol Immunol, 2011. **8**(2): p. 110-20.
- Hooper, L.V., T. Midtvedt, and J.I. Gordon, *How host-microbial interactions shape the nutrient environment of the mammalian intestine*. Annu Rev Nutr, 2002. **22**: p. 283-307.
- Gill, S.R., et al., *Metagenomic analysis of the human distal gut microbiome*. Science, 2006. **312**(5778): p. 1355-9.
- Blaut, M. and T. Clavel, *Metabolic Diversity of the Intestinal Microbiota: Implications for Health and Disease*. The Journal of Nutrition, 2007. **137**(3): p. 751S-755S.
- Guarner, F. and J.R. Malagelada, *Gut flora in health and disease*. Lancet, 2003. **361**(9356): p. 512-9.
- Sleigh, J.D. and M.C. Timbury, *Notes on Medical Bacteriology*. 1998, Edinburgh: Churchill Livingstone.
- Nicholson, J.K., E. Holmes, and I.D. Wilson, *Gut microorganisms, mammalian metabolism and personalized health care*. Nat Rev Microbiol, 2005. **3**(5): p. 431-8.
- Marques, T.M., et al., *Programming infant gut microbiota: influence of dietary and environmental factors*. Current Opinion in Biotechnology, 2010. **21**(2): p. 149-156.
- Bengmark, S., *Gut microbiota, immune development and function*. Pharmacol Res, 2013. **69**(1): p. 87-113.
- Hooper, L.V., *Bacterial contributions to mammalian gut development*. Trends in Microbiology, 2004. **12**(3): p. 129-134.
- Biol, M.C., A. Martin, and P. Louisot, *Nutritional and developmental regulation of glycosylation processes in digestive organs*. Biochimie, 1992. **74**(1): p. 13-24.
- Bry, L., et al., *A model of host-microbial interactions in an open mammalian ecosystem*. Science, 1996. **273**(5280): p. 1380-3.
- Umesaki, Y., et al., *Segmented filamentous bacteria are indigenous intestinal bacteria that activate intraepithelial lymphocytes and induce MHC class II molecules and fucosyl asialo GM1 glycolipids on the small intestinal epithelial cells in the ex-germ-free mouse*. Microbiol Immunol, 1995. **39**(8): p. 555-62.
- Fak, F., et al., *Microbial manipulation of the rat dam changes bacterial colonization and alters properties of the gut in her offspring*. Am J Physiol Gastrointest Liver Physiol, 2008. **294**(1): p. G148-54.
- Bleich, A. and A.K. Hansen, *Time to include the gut microbiota in the hygienic standardisation of laboratory rodents*. Comp Immunol Microbiol Infect Dis, 2012. **35**(2): p. 81-92.
- Stappenbeck, T.S., L.V. Hooper, and J.I. Gordon, *Developmental regulation of intestinal angiogenesis by indigenous microbes via Paneth cells*. Proc Natl Acad Sci U S A, 2002. **99**(24): p. 15451-5.
- Kelly, D. and A.G. Coutts, *Early nutrition and the development of immune function in the neonate*. Proc Nutr Soc, 2000. **59**(2): p. 177-85.
- Hooper, L.V., et al., *Angiogenins: a new class of microbicidal proteins involved in innate immunity*. Nat Immunol, 2003. **4**(3): p. 269-73.
- Shroff, K.E., K. Meslin, and J.J. Cebra, *Commensal enteric bacteria engender a self-limiting humoral mucosal immune response while permanently colonizing the gut*. Infect Immun, 1995. **63**(10): p. 3904-13.
- Umesaki, Y., et al., *Expansion of alpha beta T-cell receptor-bearing intestinal intraepithelial lymphocytes after microbial colonization in germ-free mice and its independence from thymus*. Immunology, 1993. **79**(1): p. 32-7.
- Umesaki, Y., et al., *Differential roles of segmented filamentous bacteria and clostridia in development of the intestinal immune system*. Infect Immun, 1999. **67**(7): p. 3504-11.
- Moreau, M.C. and G. Corthier, *Effect of the gastrointestinal microflora on induction and maintenance of oral tolerance to ovalbumin in C3H/HeJ mice*. Infect Immun, 1988. **56**(10): p. 2766-2768.
- Hanson, L.A. and S.A. Silfverdal, *The mother's immune system is a balanced threat to the foetus, turning to protection of the neonate*. Acta Paediatr, 2009. **98**(2): p. 221-8.
- Calder, P.C., et al., *Early nutrition and immunity - progress and perspectives*. Br J Nutr, 2006. **96**(4): p. 774-90.
- Dore, J. and G. Corthier, *[The human intestinal microbiota]*. Gastroenterol Clin Biol, 2010. **34** Suppl 1: p. S7-15.
- Abrahamsson, T.R., et al., *Low diversity of the gut microbiota in infants with atopic eczema*. J Allergy Clin Immunol, 2012. **129**(2): p. 434-40, 440 e1-2.
- Ly, N.P., et al., *Gut microbiota, probiotics, and vitamin D: Interrelated exposures influencing allergy, asthma, and obesity?* Journal of Allergy and Clinical Immunology, 2011. **127**(5): p. 1087-1094.
- Cilieborg, M.S., et al., *Preterm birth and necrotizing enterocolitis alter gut colonization in pigs*. Pediatr Res, 2011. **69**(1): p. 10-6.
- Madan, J.C., et al., *Gut microbial colonisation in premature neonates predicts neonatal sepsis*. Archives of Disease in Childhood - Fetal and Neonatal Edition, 2012.
- Smith, B., et al., *Investigation of the early intestinal microflora in premature infants with/without necrotizing enterocolitis using two different methods*. Pediatr Res, 2012. **71**(1): p. 115-120.
- Neu, J., *Gastrointestinal maturation and feeding*. Semin Perinatol, 2006. **30**(2): p. 77-80.
- Valentina, T. and B. Fredrik, *Functional interactions between the gut microbiota and host metabolism*. Nature, 2012. **489**(7415): p. 242-249.
- Mischke, M. and T. Plosch, *More than just a gut instinct-the potential interplay between a baby's nutrition, its gut microbiome, and the epigenome*. Am J Physiol Regul Integr Comp Physiol, 2013. **304**(12): p. R1065-9.
- Koren, O., et al., *Host remodeling of the gut microbiome and metabolic changes during pregnancy*. Cell, 2012. **150**(3): p. 470-80.

36. Solis, G., et al., *Establishment and development of lactic acid bacteria and bifidobacteria microbiota in breast-milk and the infant gut*. *Anaerobe*, 2010. **16**(3): p. 307-10.
37. Penders, J., et al., *Factors Influencing the Composition of the Intestinal Microbiota in Early Infancy*. *Pediatrics*, 2006. **118**(2): p. 511-521.
38. Mueller, S., et al., *Differences in fecal microbiota in different European study populations in relation to age, gender, and country: a cross-sectional study*. *Appl Environ Microbiol*, 2006. **72**(2): p. 1027-33.
39. Gewolb, I.H., et al., *Stool microflora in extremely low birthweight infants*. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, 1999. **80**(3): p. F167-F173.
40. Magne, F., et al., *Low species diversity and high interindividual variability in faeces of preterm infants as revealed by sequences of 16S rRNA genes and PCR-temporal temperature gradient gel electrophoresis profiles*. *FEMS Microbiology Ecology*, 2006. **57**(1): p. 128-138.
41. Blakey, J.L., et al., *Development of Gut Colonisation in Pre-term Neonates*. *Journal of Medical Microbiology*, 1982. **15**(4): p. 519-529.
42. Butel, M.-J., et al., *Conditions of Bifidobacterial Colonization in Preterm Infants: A Prospective Analysis*. *Journal of Pediatric Gastroenterology and Nutrition*, 2007. **44**(5): p. 577-582 10.1097/MPG.0b013e3180406b20.
43. Bezirtzoglou, E., *The intestinal microflora during the first weeks of life*. *Anaerobe*, 1997. **3**(2-3): p. 173-7.
44. Parm, U., et al., *Risk factors associated with gut and nasopharyngeal colonization by common Gram-negative species and yeasts in neonatal intensive care units patients*. *CORD Conference Proceedings*, 2011. **87**(6): p. 391-399.
45. Schiffrin, E.J. and S. Blum, *Interactions between the microbiota and the intestinal mucosa*. *Eur J Clin Nutr*, 2002. **56 Suppl 3**: p. S60-4.
46. Rotimi, V.O., S.A. Olowe, and I. Ahmed, *The development of bacterial flora of premature neonates*. *J Hyg (Lond)*, 1985. **94**(3): p. 309-18.
47. Gronlund, M.M., et al., *Fecal microflora in healthy infants born by different methods of delivery: permanent changes in intestinal flora after cesarean delivery*. *J Pediatr Gastroenterol Nutr*, 1999. **28**(1): p. 19-25.
48. Hallstrom, M., et al., *Effects of mode of delivery and necrotising enterocolitis on the intestinal microflora in preterm infants*. *Eur J Clin Microbiol Infect Dis*, 2004. **23**(6): p. 463-70.
49. Adlerberth, I., et al., *Reduced enterobacterial and increased staphylococcal colonization of the infantile bowel: an effect of hygienic lifestyle?* *Pediatr Res*, 2006. **59**(1): p. 96-101.
50. Magne, F., et al., *Low species diversity and high interindividual variability in faeces of preterm infants as revealed by sequences of 16S rRNA genes and PCR-temporal temperature gradient gel electrophoresis profiles*. *FEMS Microbiol Ecol*, 2006. **57**(1): p. 128-38.
51. Butel, M.J., et al., *Conditions of bifidobacterial colonization in preterm infants: a prospective analysis*. *J Pediatr Gastroenterol Nutr*, 2007. **44**(5): p. 577-82.
52. Chen, J., W. Cai, and Y. Feng, *Development of intestinal bifidobacteria and lactobacilli in breast-fed neonates*. *Clin Nutr*, 2007. **26**(5): p. 559-66.
53. Biasucci, G., et al., *Cesarean delivery may affect the early biodiversity of intestinal bacteria*. *J Nutr*, 2008. **138**(9): p. 1796S-1800S.
54. Mitsou, E.K., et al., *Fecal microflora of Greek healthy neonates*. *Anaerobe*, 2008. **14**(2): p. 94-101.
55. Biasucci, G., et al., *Mode of delivery affects the bacterial community in the newborn gut*. *Early Human Development*, 2010. **86**(1, Supplement): p. 13-15.
56. Dominguez-Bello, M.G., et al., *Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns*. *Proc Natl Acad Sci U S A*, 2010. **107**(26): p. 11971-5.
57. Fallani, M., et al., *Intestinal microbiota of 6-week-old infants across Europe: geographic influence beyond delivery mode, breast-feeding, and antibiotics*. *J Pediatr Gastroenterol Nutr*, 2010. **51**(1): p. 77-84.
58. Fallani, M., et al., *Determinants of the human infant intestinal microbiota after the introduction of first complementary foods in infant samples from five European centres*. *Microbiology*, 2011. **157**(Pt 5): p. 1385-92.
59. Adlerberth, I., et al., *Intestinal colonization with Enterobacteriaceae in Pakistani and Swedish hospital-delivered infants*. *Acta Paediatr Scand*, 1991. **80**(6-7): p. 602-10.
60. Azad, M.B., et al., *Gut microbiota of healthy Canadian infants: profiles by mode of delivery and infant diet at 4 months*. *CMAJ*, 2013. **185**(5): p. 385-94.
61. Stark, P.L. and A. Lee, *The bacterial colonization of the large bowel of pre-term low birth weight neonates*. *J Hyg (Lond)*, 1982. **89**(1): p. 59-67.
62. Stark, P.L. and A. Lee, *The microbial ecology of the large bowel of breast-fed and formula-fed infants during the first year of life*. *J Med Microbiol*, 1982. **15**(2): p. 189-203.
63. Yoshioka, H., K. Iseki, and K. Fujita, *Development and differences of intestinal flora in the neonatal period in breast-fed and bottle-fed infants*. *Pediatrics*, 1983. **72**(3): p. 317-21.
64. Merida, V., et al., *Significance of Clostridium difficile and its cytotoxin in children*. *Eur J Pediatr*, 1986. **144**(5): p. 494-6.
65. Bonang, G., et al., *Influence of breastmilk on the development of resistance to intestinal colonization in infants born at the Atma Jaya Hospital, Jakarta*. *Scand J Infect Dis*, 2000. **32**(2): p. 189-96.
66. Harmsen, H.J., et al., *Analysis of intestinal flora development in breast-fed and formula-fed infants by using molecular identification and detection methods*. *J Pediatr Gastroenterol Nutr*, 2000. **30**(1): p. 61-7.
67. Ahrne, S., et al., *Lactobacilli in the intestinal microbiota of Swedish infants*. *Microbes Infect*, 2005. **7**(11-12): p. 1256-62.
68. Bezirtzoglou, E., et al., *Occurrence of Bifidobacterium in the intestine of newborns by fluorescence in situ hybridization*. *Comp Immunol Microbiol Infect Dis*, 2006. **29**(5-6): p. 345-52.

69. Roger, L.C. and A.L. McCartney, *Longitudinal investigation of the faecal microbiota of healthy full-term infants using fluorescence in situ hybridization and denaturing gradient gel electrophoresis*. Microbiology, 2010. **156**(Pt 11): p. 3317-28.
70. Bezirtzoglou, E., A. Tsitsis, and G.W. Welling, *Microbiota profile in feces of breast- and formula-fed newborns by using fluorescence in situ hybridization (FISH)*. Anaerobe, 2011. **17**(6): p. 478-82.
71. Morowitz, M.J., et al., *Strain-resolved community genomic analysis of gut microbial colonization in a premature infant*. Proc Natl Acad Sci U S A, 2011. **108**(3): p. 1128-33.
72. Lundequist, B., C.E. Nord, and J. Winberg, *The composition of the faecal microflora in breastfed and bottle fed infants from birth to eight weeks*. Acta Paediatr Scand, 1985. **74**(1): p. 45-51.
73. Neut, C., et al., *Bacterial colonization of the large intestine in newborns delivered by cesarean section*. Zentralbl Bakteriell Mikrobiol Hyg A, 1987. **266**(3-4): p. 330-7.
74. Bezirtzoglou, E. and C. Romond, *Effect of the feeding practices on the establishment of bacterial interactions in the intestine of the newborn delivered by cesarean section*. J Perinat Med, 1990. **18**(2): p. 139-43.
75. Bezirtzoglou, E. and C. Romond, *Occurrence of Enterobacteriaceae and E. coli in the intestine of the newborn by caesarean section*. Acta Microbiol Bulg, 1990. **25**: p. 76-81.
76. Kleessen, B., et al., *Influence of two infant formulas and human milk on the development of the faecal flora in newborn infants*. Acta Paediatr, 1995. **84**(12): p. 1347-56.
77. Schwartz, S., et al., *A metagenomic study of diet-dependent interaction between gut microbiota and host in infants reveals differences in immune response*. Genome Biol, 2012. **13**(4): p. r32.
78. Bennet, R., et al., *Intestinal bacteria of newborn Ethiopian infants in relation to antibiotic treatment and colonisation by potentially pathogenic gram-negative bacteria*. Scand J Infect Dis, 1991. **23**(1): p. 63-9.
79. Arbolea, S., et al., *Establishment and development of intestinal microbiota in preterm neonates*. FEMS Microbiol Ecol, 2012. **79**(3): p. 763-72.
80. Ferraris, L., et al., *Clostridia in premature neonates' gut: incidence, antibiotic susceptibility, and perinatal determinants influencing colonization*. PLoS One, 2012. **7**(1): p. e30594.
81. Favier, C.F., W.M. de Vos, and A.D. Akkermans, *Development of bacterial and bifidobacterial communities in feces of newborn babies*. Anaerobe, 2003. **9**(5): p. 219-29.
82. O'Toole, P.W. and M.J. Claesson, *Gut microbiota: Changes throughout the lifespan from infancy to elderly*. International Dairy Journal, 2010. **20**(4): p. 281-291.
83. Jacquot, A., et al., *Dynamics and Clinical Evolution of Bacterial Gut Microflora in Extremely Premature Patients*. The Journal of Pediatrics, 2011. **158**(3): p. 390-396.
84. Windey, K., V. De Preter, and K. Verbeke, *Relevance of protein fermentation to gut health*. Mol Nutr Food Res, 2012. **56**(1): p. 184-96.
85. Arbolea, S., et al., *Characterization and in vitro properties of potentially probiotic Bifidobacterium strains isolated from breast-milk*. Int J Food Microbiol, 2011. **149**(1): p. 28-36.
86. Cooperstock, M., et al., *Influence of age, sex, and diet on asymptomatic colonization of infants with Clostridium difficile*. J Clin Microbiol, 1983. **17**(5): p. 830-3.
87. LaTuga, M.S., et al., *Beyond bacteria: a study of the enteric microbial consortium in extremely low birth weight infants*. PLoS One, 2011. **6**(12): p. e27858.
88. Watts, D.H., et al., *The association of occult amniotic fluid infection with gestational age and neonatal outcome among women in preterm labor*. Obstet Gynecol, 1992. **79**(3): p. 351-7.
89. Goldenberg, R.L., J.C. Hauth, and W.W. Andrews, *Intrauterine infection and preterm delivery*. N Engl J Med, 2000. **342**(20): p. 1500-7.
90. Li, M., et al., *Symbiotic gut microbes modulate human metabolic phenotypes*. Proceedings of the National Academy of Sciences, 2008. **105**(6): p. 2117-2122.
91. Shimizu, K., et al., *Altered gut flora are associated with septic complications and death in critically ill patients with systemic inflammatory response syndrome*. Dig Dis Sci, 2011. **56**(4): p. 1171-7.
92. Steele, R.W., *The Clinical Handbook of Pediatric Infectious Disease*. 1999, New York: The Parthenon Publishing Group.
93. Donnell, S.C., et al., *Nutritional implications of gut overgrowth and selective decontamination of the digestive tract*. Proc Nutr Soc, 1998. **57**(3): p. 381-7.
94. Soeorg, H., et al., *Genetic Relatedness of Coagulase-Negative Staphylococci from Gastrointestinal Tract and Blood of Preterm Neonates with Late-Onset Sepsis*. Pediatr Infect Dis J, 2012.
95. van Saene, H.K., et al., *Gut overgrowth with abnormal flora: the missing link in parenteral nutrition-related sepsis in surgical neonates*. Eur J Clin Nutr, 2003. **57**(4): p. 548-53.
96. Smith, A., et al., *Concordance of Gastrointestinal Tract Colonization and Subsequent Bloodstream Infections With Gram-negative Bacilli in Very Low Birth Weight Infants in the Neonatal Intensive Care Unit*. Pediatr Infect Dis J, 2010. **29**(9): p. 831-5.
97. Das, P., et al., *Colonization of the gut with Gram-negative bacilli, its association with neonatal sepsis and its clinical relevance in a developing country*. J Med Microbiol, 2011. **60**(Pt 11): p. 1651-60.
98. Madan, J.C., et al., *Gut microbial colonisation in premature neonates predicts neonatal sepsis*. Arch Dis Child Fetal Neonatal Ed, 2012. **97**(6): p. F456-62.
99. Mai, V., et al., *Distortions in development of intestinal microbiota associated with late onset sepsis in preterm infants*. PLoS One, 2013. **8**(1): p. e52876.
100. Parm, U., et al., *Mucosal surveillance cultures in predicting Gram-negative late-onset sepsis in neonatal intensive care units*. J Hosp Infect, 2011. **78**(4): p. 327-32.

101. Stewart, C.J., et al., *Bacterial and fungal viability in the preterm gut: NEC and sepsis*. Arch Dis Child Fetal Neonatal Ed, 2013. **98**(4): p. F298-303.
102. Bell, M.J., et al., *Epidemiologic and bacteriologic evaluation of neonatal necrotizing enterocolitis*. J Pediatr Surg, 1979. **14**(1): p. 1-4.
103. Blakey, J.L., et al., *Enteric colonization in sporadic neonatal necrotizing enterocolitis*. J Pediatr Gastroenterol Nutr, 1985. **4**(4): p. 591-5.
104. Peter, C.S., et al., *Necrotising enterocolitis: is there a relationship to specific pathogens?* Eur J Pediatr, 1999. **158**(1): p. 67-70.
105. Wang, Y., et al., *16S rRNA gene-based analysis of fecal microbiota from preterm infants with and without necrotizing enterocolitis*. ISME J, 2009. **3**(8): p. 944-54.
106. de la Cochetiere, M.F., et al., *Early intestinal bacterial colonization and necrotizing enterocolitis in premature infants: the putative role of Clostridium*. Pediatr Res, 2004. **56**(3): p. 366-70.
107. Carlisle, E.M. and M.J. Morowitz, *The intestinal microbiome and necrotizing enterocolitis*. Curr Opin Pediatr, 2013. **25**(3): p. 382-7.
108. Smith, B., et al., *Investigation of the early intestinal microflora in premature infants with/without necrotizing enterocolitis using two different methods*. Pediatr Res, 2012. **71**(1): p. 115-20.
109. Mai, V., et al., *Fecal microbiota in premature infants prior to necrotizing enterocolitis*. PLoS One, 2011. **6**(6): p. e20647.
110. Smith, B., et al., *Community analysis of bacteria colonizing intestinal tissue of neonates with necrotizing enterocolitis*. BMC Microbiol, 2011. **11**: p. 73.
111. Penders, J., et al., *Gut microbiota composition and development of atopic manifestations in infancy: the KOALA Birth Cohort Study*. Gut, 2007. **56**(5): p. 661-7.
112. Stsepetova, J., et al., *Molecularly assessed shifts of Bifidobacterium ssp. and less diverse microbial communities are characteristic of 5-year-old allergic children*. FEMS Immunol Med Microbiol, 2007. **51**(2): p. 260-9.
113. Vael, C., et al., *Early intestinal Bacteroides fragilis colonisation and development of asthma*. BMC Pulm Med, 2008. **8**: p. 19.
114. Sjogren, Y.M., et al., *Altered early infant gut microbiota in children developing allergy up to 5 years of age*. Clin Exp Allergy, 2009. **39**(4): p. 518-26.
115. Suzuki, S., et al., *Differences in the composition of intestinal Bifidobacterium species and the development of allergic diseases in infants in rural Japan*. Clin Exp Allergy, 2007. **37**(4): p. 506-11.
116. Vael, C., et al., *Denaturing gradient gel electrophoresis of neonatal intestinal microbiota in relation to the development of asthma*. BMC Microbiol, 2011. **11**: p. 68.
117. van Nimwegen, F.A., et al., *Mode and place of delivery, gastrointestinal microbiota, and their influence on asthma and atopy*. J Allergy Clin Immunol, 2011. **128**(5): p. 948-55 e1-3.
118. Waligora-Dupriet, A.J., et al., *Diversity of gut Bifidobacterium species is not altered between allergic and non-allergic French infants*. Anaerobe, 2011. **17**(3): p. 91-6.
119. Bjorksten, B., et al., *The intestinal microflora in allergic Estonian and Swedish 2-year-old children*. Clin Exp Allergy, 1999. **29**(3): p. 342-6.
120. Bjorksten, B., et al., *Allergy development and the intestinal microflora during the first year of life*. J Allergy Clin Immunol, 2001. **108**(4): p. 516-20.
121. Sepp, E., et al., *Intestinal microbiota and immunoglobulin E responses in 5-year-old Estonian children*. Clin Exp Allergy, 2005. **35**(9): p. 1141-6.
122. Songjinda, P., et al., *Differences in developing intestinal microbiota between allergic and non-allergic infants: a pilot study in Japan*. Biosci Biotechnol Biochem, 2007. **71**(9): p. 2338-42.
123. Thompson-Chagoyan, O.C., et al., *Changes in faecal microbiota of infants with cow's milk protein allergy--a Spanish prospective case-control 6-month follow-up study*. Pediatr Allergy Immunol, 2010. **21**(2 Pt 2): p. e394-400.
124. Storro, O., et al., *Temporal variations in early gut microbial colonization are associated with allergen-specific immunoglobulin E but not atopic eczema at 2 years of age*. Clin Exp Allergy, 2011. **41**(11): p. 1545-54.
125. Thompson-Chagoyan, O.C., et al., *Faecal microbiota and short-chain fatty acid levels in faeces from infants with cow's milk protein allergy*. Int Arch Allergy Immunol, 2011. **156**(3): p. 325-32.
126. Watanabe, S., et al., *Differences in fecal microflora between patients with atopic dermatitis and healthy control subjects*. J Allergy Clin Immunol, 2003. **111**(3): p. 587-91.
127. Smehilova, M., et al., *Comparison of intestinal microflora in healthy infants and infants with allergic colitis*. Folia Microbiol (Praha), 2008. **53**(3): p. 255-8.
128. Kalliomaki, M., et al., *Distinct patterns of neonatal gut microflora in infants in whom atopy was and was not developing*. J Allergy Clin Immunol, 2001. **107**(1): p. 129-34.
129. Kirjavainen, P.V., et al., *Characterizing the composition of intestinal microflora as a prospective treatment target in infant allergic disease*. FEMS Immunol Med Microbiol, 2001. **32**(1): p. 1-7.
130. Mah, K.W., et al., *Distinct pattern of commensal gut microbiota in toddlers with eczema*. Int Arch Allergy Immunol, 2006. **140**(2): p. 157-63.
131. Verhulst, S.L., et al., *A longitudinal analysis on the association between antibiotic use, intestinal microflora, and wheezing during the first year of life*. J Asthma, 2008. **45**(9): p. 828-32.
132. Forno, E., et al., *Diversity of the gut microbiota and eczema in early life*. Clin Mol Allergy, 2008. **6**: p. 11.
133. Wang, M., et al., *Reduced diversity in the early fecal microbiota of infants with atopic eczema*. J Allergy Clin Immunol, 2008. **121**(1): p. 129-34.

134. Bisgaard, H., et al., *Reduced diversity of the intestinal microbiota during infancy is associated with increased risk of allergic disease at school age*. J Allergy Clin Immunol, 2011. **128**(3): p. 646-52 e1-5.
135. Kalliomaki, M., et al., *Early differences in fecal microbiota composition in children may predict overweight*. Am J Clin Nutr, 2008. **87**(3): p. 534-8.
136. Vael, C., et al., *Intestinal microflora and body mass index during the first three years of life: an observational study*. Gut Pathog, 2011. **3**(1): p. 8.
137. Larsen, N., et al., *Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults*. PLoS One, 2010. **5**(2): p. e9085.
138. Xu, X., H. Hui, and D. Cai, *[Differences in fecal Bifidobacterium species between patients with type 2 diabetes and healthy individuals]*. Nan Fang Yi Ke Da Xue Xue Bao, 2012. **32**(4): p. 531-3, 564.
139. Wu, X., et al., *Molecular characterisation of the faecal microbiota in patients with type II diabetes*. Curr Microbiol, 2010. **61**(1): p. 69-78.
140. Giongo, A., et al., *Toward defining the autoimmune microbiome for type 1 diabetes*. ISME J, 2011. **5**(1): p. 82-91.
141. Finegold, S.M., et al., *Pyrosequencing study of fecal microflora of autistic and control children*. Anaerobe, 2010. **16**(4): p. 444-53.
142. Kang, D.-W., et al., *Reduced Incidence of <italic>Prevotella</italic> and Other Fermenters in Intestinal Microflora of Autistic Children*. PLoS One, 2013. **8**(7): p. e68322.
143. Wen, S.W., et al., *Epidemiology of preterm birth and neonatal outcome*. Semin Fetal Neonatal Med, 2004. **9**(6): p. 429-35.
144. WHO, *Preterm birth*. 2012. **Fact sheet Nr 363**.
145. Moutquin, J.M., *Classification and heterogeneity of preterm birth*. BJOG, 2003. **110 Suppl 20**: p. 30-3.
146. Wen, L., et al., *Innate immunity and intestinal microbiota in the development of Type 1 diabetes*. Nature, 2008. **455**(7216): p. 1109-1113.
147. Goldenberg, R.L., et al., *Epidemiology and causes of preterm birth*. Lancet, 2008. **371**(9606): p. 75-84.
148. Dusick, A.M., et al., *Growth failure in the preterm infant: can we catch up?* Semin Perinatol, 2003. **27**(4): p. 302-10.
149. WHO, *UK Growth Charts. Pre-school charts 2nd Edition* WHO, Editor. 2013.
150. Bearfield, C., et al., *Possible association between amniotic fluid micro-organism infection and microflora in the mouth*. BJOG, 2002. **109**(5): p. 527-33.
151. Tucker, J.M., et al., *Etiologies of preterm birth in an indigent population: is prevention a logical expectation?* Obstet Gynecol, 1991. **77**(3): p. 343-7.
152. Henderson, J.J., et al., *Preterm birth aetiology 2004-2008. Maternal factors associated with three phenotypes: spontaneous preterm labour, preterm pre-labour rupture of membranes and medically indicated preterm birth*. J Matern Fetal Neonatal Med, 2012. **25**(6): p. 642-7.
153. Ananth, C.V. and A.M. Vintzileos, *Epidemiology of preterm birth and its clinical subtypes*. J Matern Fetal Neonatal Med, 2006. **19**(12): p. 773-82.
154. Steel, J.H., et al., *Bacteria and inflammatory cells in fetal membranes do not always cause preterm labor*. Pediatr Res, 2005. **57**(3): p. 404-11.
155. Rebarber, A., et al., *Shigellosis complicating preterm premature rupture of membranes resulting in congenital infection and preterm delivery*. Obstet Gynecol, 2002. **100**(5 Pt 2): p. 1063-5.
156. Srinivasan, U., et al., *Vaginal and oral microbes, host genotype and preterm birth*. Med Hypotheses, 2009. **73**(6): p. 963-75.
157. Simmer, K., A. Rakshasbhuvankar, and G. Deshpande, *Standardised parenteral nutrition*. Nutrients, 2013. **5**(4): p. 1058-70.
158. Kandasamy, Y., et al., *Extra-uterine renal growth in preterm infants: oligonephropathy and prematurity*. Pediatr Nephrol, 2013. **28**(9): p. 1791-6.
159. Vasu, V. and N. Modi, *Assessing the impact of preterm nutrition*. Early Hum Dev, 2007. **83**(12): p. 813-8.
160. Gianni, M.L., et al., *Nutrition and growth in infants born preterm from birth to adulthood*. Early Hum Dev, 2013. **89 Suppl 2**: p. S41-4.
161. Velaphi, S., *Nutritional requirements and parenteral nutrition in preterm infants*. South African Journal of Clinical Nutrition, 2011. **3**(24): p. S27-S31.
162. Beek, R.v., *Metabolism in preterm infants on the first days of life. The effect of corticosteroids*. 2001, Erasmus Universiteit Rotterdam: Rotterdam.
163. Lapillonne, A., et al., *Lipid needs of preterm infants: updated recommendations*. J Pediatr, 2013. **162**(3 Suppl): p. S37-47.
164. Cristofalo, E.A., et al., *Randomized Trial of Exclusive Human Milk versus Preterm Formula Diets in Extremely Premature Infants*. J Pediatr, 2013.
165. Tantibhedhyangkul, P. and S.A. Hashim, *Clinical and physiologic aspects of medium-chain triglycerides: alleviation of steatorrhea in premature infants*. Bull N Y Acad Med, 1971. **47**(1): p. 17-33.
166. de Boo, H.A. and J.E. Harding, *Protein metabolism in preterm infants with particular reference to intrauterine growth restriction*. Arch Dis Child Fetal Neonatal Ed, 2007. **92**(4): p. F315-9.
167. Morgan, C., *Early amino acid administration in very preterm infants: Too little, too late or too much, too soon?* Semin Fetal Neonatal Med, 2013.
168. Stoll, B., et al., *Chronic parenteral nutrition induces hepatic inflammation, steatosis, and insulin resistance in neonatal pigs*. J Nutr, 2010. **140**(12): p. 2193-200.
169. Miloudi, K., et al., *The mode of administration of total parenteral nutrition and nature of lipid content influence the generation of peroxides and aldehydes*. Clin Nutr, 2012. **31**(4): p. 526-34.

170. Moyses, H.E., et al., *Early parenteral nutrition and growth outcomes in preterm infants: a systematic review and meta-analysis*. Am J Clin Nutr, 2013. **97**(4): p. 816-26.
171. Sala-Vila, A., V.M. Barbosa, and P.C. Calder, *Olive oil in parenteral nutrition*. Curr Opin Clin Nutr Metab Care, 2007. **10**(2): p. 165-74.
172. Ulrich, H., et al., *Parenteral use of medium-chain triglycerides: a reappraisal*. Nutrition, 1996. **12**(4): p. 231-8.
173. Kalish, B.T., et al., *A metabolomic analysis of two intravenous lipid emulsions in a murine model*. PLoS One, 2013. **8**(4): p. e59653.
174. Javid, P.J., et al., *The route of lipid administration affects parenteral nutrition-induced hepatic steatosis in a mouse model*. J Pediatr Surg, 2005. **40**(9): p. 1446-53.
175. Cowan, E., P. Nandivada, and M. Puder, *Fish oil-based lipid emulsion in the treatment of parenteral nutrition-associated liver disease*. Curr Opin Pediatr, 2013. **25**(2): p. 193-200.
176. Colomb, V., et al., *Role of lipid emulsions in cholestasis associated with long-term parenteral nutrition in children*. JPEN J Parenter Enteral Nutr, 2000. **24**(6): p. 345-50.
177. Xu, Z.W. and Y.S. Li, *Pathogenesis and treatment of parenteral nutrition-associated liver disease*. Hepatobiliary Pancreat Dis Int, 2012. **11**(6): p. 586-93.
178. Crook, M.A. and K. Sriram, *Successful treatment of parenteral nutrition-associated liver disease in an adult by use of a fish oil-based lipid source*. Nutrition, 2013. **29**(4): p. 700-1.
179. Granato, D., et al., *Effects of parenteral lipid emulsions with different fatty acid composition on immune cell functions in vitro*. JPEN J Parenter Enteral Nutr, 2000. **24**(2): p. 113-8.
180. van Kempen, A.A., et al., *Stimulation of gluconeogenesis by intravenous lipids in preterm infants: response depends on fatty acid profile*. Am J Physiol Endocrinol Metab, 2006. **290**(4): p. E723-30.
181. Dennison, A.R., et al., *The metabolic consequences of infusing emulsions containing medium chain triglycerides for parenteral nutrition: a comparative study with conventional lipid*. Ann R Coll Surg Engl, 1986. **68**(3): p. 119-21.
182. Stein, T.P., et al., *Comparison of glucose, LCT, and LCT plus MCT as calorie sources for parenterally nourished rats*. Am J Physiol, 1984. **246**(3 Pt 1): p. E277-87.
183. Jiang, Z.M., et al., *A comparison of medium-chain and long-chain triglycerides in surgical patients*. Ann Surg, 1993. **217**(2): p. 175-84.
184. Rayyan, M., et al., *Short-term use of parenteral nutrition with a lipid emulsion containing a mixture of soybean oil, olive oil, medium-chain triglycerides, and fish oil: a randomized double-blind study in preterm infants*. JPEN J Parenter Enteral Nutr, 2012. **36**(1 Suppl): p. 81S-94S.
185. Meyers, J.M., et al., *Potential Influence of Total Parenteral Nutrition on Body Composition at Discharge in Preterm Infants*. Journal of Maternal-Fetal and Neonatal Medicine, 2013: p. 1-23.
186. Ito, K., et al., *Lipid emulsion administered intravenously or orally attenuates triglyceride accumulation and expression of inflammatory markers in the liver of nonobese mice fed parenteral nutrition formula*. J Nutr, 2013. **143**(3): p. 253-9.
187. Poindexter, B.B., C.A. Karn, and S.C. Denne, *Exogenous insulin reduces proteolysis and protein synthesis in extremely low birth weight infants*. J Pediatr, 1998. **132**(6): p. 948-53.
188. Popp, M.B., S.D. Morrison, and M.F. Brennan, *Growth and body composition during long-term total parenteral nutrition in the rat*. Am J Clin Nutr, 1982. **36**(6): p. 1119-28.
189. Meehan, J.J. and K.E. Georgeson, *Prevention of liver failure in parenteral nutrition-dependent children with short bowel syndrome*. J Pediatr Surg, 1997. **32**(3): p. 473-5.
190. Matarese, L.E., et al., *Body composition changes in cachectic patients receiving home parenteral nutrition*. JPEN J Parenter Enteral Nutr, 2002. **26**(6): p. 366-71.
191. Dieterle, F., et al., *NMR and MS methods for metabolomics*. Methods Mol Biol, 2011. **691**: p. 385-415.
192. Seen, H., Schlotterbeck, G., *NMR Spectroscopy Techniques for Application to Metabolomics*, in *The Handbook of Metabolomics and Metabolomics*, J.C. Lindon, Nicholson, J.K., Holmes, E., Editor. 2007, Elsevier Amsterdam.
193. Waugh, J.S., *Principles of nuclear magnetic resonance in one and two dimensions*, R. R. Ernst, G. Bodenhausen, and A. Wokaun. Oxford Univ. Press, London/New York. 1987. 610 pp. \$98. Magnetic Resonance in Medicine, 1988. **7**(2): p. 253-253.
194. Friebolin, H., *Basic One- and Two-Dimensional NMR Spectroscopy*. 2010.
195. Jeener, J., et al., *Investigation of exchange processes by two-dimensional NMR spectroscopy*. The Journal of Chemical Physics, 1979. **71**(11): p. 4546-4553.
196. Bax, A. and R. Freeman, *Investigation of complex networks of spin-spin coupling by two-dimensional NMR*. Journal of Magnetic Resonance (1969), 1981. **44**(3): p. 542-561.
197. Bax, A., R. Freeman, and G. Morris, *Correlation of proton chemical shifts by two-dimensional Fourier transform NMR*. Journal of Magnetic Resonance (1969), 1981. **42**(1): p. 164-168.
198. Maudsley, A.A. and R.R. Ernst, *Indirect detection of magnetic resonance by heteronuclear two-dimensional spectroscopy*. Chemical Physics Letters, 1977. **50**(3): p. 368-372.
199. Nagayama, K., et al., *Experimental techniques of two-dimensional correlated spectroscopy*. Journal of Magnetic Resonance (1969), 1980. **40**(2): p. 321-334.
200. Trygg, J., Lundstedt, T., *Chemometrics Techniques for Metabolomics*, in *The Handbook of Metabolomics and Metabolomics*, J.C. Lindon, Nicholson, J.K., Holmes, E., Editor. 2007, Elsevier Amsterdam.
201. Ebbels, T.M.D., *Non-linear Methods for the Analysis of Metabolic Profiles*, in *The Handbook of Metabolomics and Metabolomics*, J.C. Lindon, Nicholson, J.K., Holmes, E., Editor. 2007, Elsevier Amsterdam.
202. Triba, M.N., et al., *PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross-validation quality parameters*. Mol Biosyst, 2015. **11**(1): p. 13-9.

203. Posma, J.M., et al., *Subset optimization by reference matching (STORM): an optimized statistical approach for recovery of metabolic biomarker structural information from 1H NMR spectra of biofluids*. *Anal Chem*, 2012. **84**(24): p. 10694-701.
204. MacGregor, B.J., *Molecular approaches to the study of aquatic microbial communities*. *Current Opinion in Biotechnology*, 1999. **10**(3): p. 220-224.
205. Muyzer, G., *DGGE/TGGE a method for identifying genes from natural ecosystems*. *Curr Opin Microbiol*, 1999. **2**(3): p. 317-22.
206. Muyzer, G., *DGGE/TGGE a method for identifying genes from natural ecosystems*. *Current Opinion in Microbiology*, 1999. **2**(3): p. 317-322.
207. Peterson, C., *Denaturing Gradient Gel Electrophoresis (DGGE)*. *Journal of Visualized Experiments : JoVE*, 2007(2): p. 164.
208. King, C. and T. Scott-Horton, *Pyrosequencing: A Simple Method for Accurate Genotyping*. *Journal of Visualized Experiments : JoVE*, 2008(11): p. 630.
209. Ansorge, W.J., *Next-generation DNA sequencing techniques*. *New Biotechnology*, 2009. **25**(4): p. 195-203.
210. Woo, P.C., et al., *Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories*. *Clin Microbiol Infect*, 2008. **14**(10): p. 908-34.
211. Fox, G.E., L.J. Magrum, and W.E. Balch, *Classification of methanogenic bacteria by 16S ribosomal RNA characterization*. *Proceedings of the National Academy of Sciences of the United States of America*, 1977. **74**(10): p. 4537-4541.
212. MoBio, *PowerSoil®DNA Isolation Kit Instruction Manual*.
213. Fouhy, F., et al., *The effects of freezing on faecal microbiota as determined using MiSeq sequencing and culture-based investigations*. *PLoS One*, 2015. **10**(3): p. e0119355.
214. Burns, J.L. and J.-M. Rolain, *Culture-based diagnostic microbiology in cystic fibrosis: Can we simplify the complexity?* *Journal of Cystic Fibrosis*, 2014. **13**(1): p. 1-9.
215. Elizaquível, P., et al., *Pyrosequencing vs. culture-dependent approaches to analyze lactic acid bacteria associated to chicha, a traditional maize-based fermented beverage from Northwestern Argentina*. *International Journal of Food Microbiology*, 2015. **198**(0): p. 9-18.
216. Gonzalez, R.A. and R.T. Noble, *Comparisons of statistical models to predict fecal indicator bacteria concentrations enumerated by qPCR- and culture-based methods*. *Water Research*, 2014. **48**(0): p. 296-305.
217. Ervin, J.S., et al., *Characterization of fecal concentrations in human and other animal sources by physical, culture-based, and quantitative real-time PCR methods*. *Water Research*, 2013. **47**(18): p. 6873-6882.
218. Melendez, J.H., et al., *Real-time PCR assays compared to culture-based approaches for identification of aerobic bacteria in chronic wounds*. *Clinical Microbiology and Infection*, 2010. **16**(12): p. 1762-1769.
219. Sjöberg, F., et al., *Comparison between terminal-restriction fragment length polymorphism (T-RFLP) and quantitative culture for analysis of infants' gut microbiota*. *J Microbiol Methods*, 2013. **94**(1): p. 37-46.
220. Rouge, C., et al., *Investigation of the intestinal microbiota in preterm infants using different methods*. *Anaerobe*, 2010. **16**(4): p. 362-70.
221. Osborne, C.A., et al., *PCR-generated artefact from 16S rRNA gene-specific primers*. *FEMS Microbiology Letters*, 2005. **248**(2): p. 183-187.
222. Chen, L., et al., *Development of a Lactobacillus Specific T-RFLP Method to Determine Lactobacilli Diversity in Complex Samples*. *Journal of microbiological methods*, 2012. **91**(2): p. 262-268.
223. Dickie, I.A. and R.G. FitzJohn, *Using terminal restriction fragment length polymorphism (T-RFLP) to identify mycorrhizal fungi: a methods review*. *Mycorrhiza*, 2007. **17**(4): p. 259-70.
224. Fredriksson, N.J., M. Hermansson, and B.-M. Wilén, *Impact of T-RFLP data analysis choices on assessments of microbial community structure and dynamics*. *BMC Bioinformatics*, 2014. **15**(1): p. 360.
225. Pires, A.C.C., et al., *Denaturing Gradient Gel Electrophoresis and Barcoded Pyrosequencing Reveal Unprecedented Archaeal Diversity in Mangrove Sediment and Rhizosphere Samples*. *Applied and Environmental Microbiology*, 2012. **78**(16): p. 5520-5528.
226. Dorigo, U., L. Volatier, and J.-F. Humbert, *Molecular approaches to the assessment of biodiversity in aquatic microbial communities*. *Water Research*, 2005. **39**(11): p. 2207-2218.
227. Ercolini, D., *PCR-DGGE fingerprinting: novel strategies for detection of microbes in food*. *Journal of Microbiological Methods*, 2004. **56**(3): p. 297-314.
228. Shimano, S., M. Sambe, and Y. Kasahara, *Application of Nested PCR-DGGE (Denaturing Gradient Gel Electrophoresis) for the Analysis of Ciliate Communities in Soils*. *Microbes and Environments*, 2012. **27**(2): p. 136-141.
229. Roudière, L., et al., *Optimized PCR-Temporal Temperature Gel Electrophoresis compared to cultivation to assess diversity of gut microbiota in neonates*. *Journal of Microbiological Methods*, 2009. **79**(2): p. 156-165.
230. Víglask, V., et al., *Early melting of supercoiled DNA topoisomers observed by TGGE*. *Nucleic Acids Research*, 2000. **28**(11): p. e51-e51.
231. Al-Sadi, A.M., S.S. Al-Mazroui, and A.J. Phillips, *Evaluation of culture-based techniques and 454 pyrosequencing for the analysis of fungal diversity in potting media and organic fertilizers*. *J Appl Microbiol*, 2015.
232. Ihle, M.A., et al., *Comparison of high resolution melting analysis, pyrosequencing, next generation sequencing and immunohistochemistry to conventional Sanger sequencing for the detection of p.V600E and non-p.V600E BRAF mutations*. *BMC Cancer*, 2014. **14**: p. 13-13.

233. Altimari, A., et al., *454 next generation-sequencing outperforms allele-specific PCR, Sanger sequencing, and pyrosequencing for routine KRAS mutation analysis of formalin-fixed, paraffin-embedded samples*. *OncoTargets and therapy*, 2013. **6**: p. 1057-1064.
234. Sanger, F. and A.R. Coulson, *A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase*. *Journal of Molecular Biology*, 1975. **94**(3): p. 441-448.
235. Pettersson, E., J. Lundeberg, and A. Ahmadian, *Generations of sequencing technologies*. *Genomics*, 2009. **93**(2): p. 105-111.
236. Mardis, E.R., *The impact of next-generation sequencing technology on genetics*. *Trends in Genetics*, 2008. **24**(3): p. 133-141.
237. Sanger, F., et al., *Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing*. *Journal of Molecular Biology*, 1980. **143**(2): p. 161-178.
238. Loman, N.J., et al., *Performance comparison of benchtop high-throughput sequencing platforms*. *Nat Biotechnol*, 2012. **30**(5): p. 434-9.
239. Diagnostics, R. *GS Junior System brochure*. 2011.
240. Zheng, Z., et al., *Titration-free massively parallel pyrosequencing using trace amounts of starting material*. *Nucleic Acids Research*, 2010. **38**(13): p. e137-e137.
241. Liu, L., et al., *Comparison of next-generation sequencing systems*. *J Biomed Biotechnol*, 2012. **2012**: p. 251364.
242. Shokralla, S., et al., *Next-generation DNA barcoding: using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens*. *Mol Ecol Resour*, 2014. **14**(5): p. 892-901.
243. Ram, D., et al., *Evaluation of GS Junior and MiSeq next-generation sequencing technologies as an alternative to Trugene population sequencing in the clinical HIV laboratory*. *J Virol Methods*, 2015. **212**: p. 12-6.
244. Morozova, O. and M.A. Marra, *Applications of next-generation sequencing technologies in functional genomics*. *Genomics*, 2008. **92**(5): p. 255-264.
245. Jervis-Bardy, J., et al., *Deriving accurate microbiota profiles from human samples with low bacterial content through post-sequencing processing of Illumina MiSeq data*. *Microbiome*, 2015. **3**: p. 19.
246. Zhang, X., et al., *Illumina MiSeq Sequencing Reveals Diverse Microbial Communities of Activated Sludge Systems Stimulated by Different Aromatics for Indigo Biosynthesis from Indole*. *PLoS One*, 2015. **10**(4): p. e0125732.
247. Ullmann, L.S., et al., *Comparative clinical sample preparation of DNA and RNA viral nucleic acids for a commercial deep sequencing system (Illumina MiSeq(R))*. *J Virol Methods*, 2015. **220**: p. 60-3.
248. Shokralla, S., et al., *Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform*. *Sci Rep*, 2015. **5**: p. 9687.
249. Unno, T., *Bioinformatic Suggestions on MiSeq-based Microbial Community Analysis*. *J Microbiol Biotechnol*, 2015.
250. Katsuoka, F., et al., *An efficient quantitation method of next-generation sequencing libraries by using MiSeq sequencer*. *Anal Biochem*, 2014. **466**: p. 27-9.
251. McElhoe, J.A., et al., *Development and assessment of an optimized next-generation DNA sequencing approach for the mtgenome using the Illumina MiSeq*. *Forensic Sci Int Genet*, 2014. **13**: p. 20-9.
252. Fadrosch, D.W., et al., *An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform*. *Microbiome*, 2014. **2**(1): p. 6.
253. Caporaso, J.G., et al., *Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms*. *ISME j*, 2012. **6**(8): p. 1621-4.
254. Schloss, P.D., et al., *Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities*. *Appl Environ Microbiol*, 2009. **75**(23): p. 7537-41.
255. Parks, D.H., Tyson, G.W., Hugenholtz, P., Beiko, R.G., *STAMP: Statistical analysis of taxonomic and functional profiles*. *Bioinformatics*, 2014. **30**: p. 3123-3124.
256. White, H., *A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity*. *Econometrica*, 1980. **48**(4): p. 817-838.
257. Welch, B.L., *The generalization of 'Student's' problem when several different population variances are involved*. *Biometrika*, 1947. **34**: p. 28-35.
258. J.H., M., *Handbook of Biological Statistics*. 2014: Sparky House Publishing.
259. Kruskal, W.H. and W.A. Wallis, *Use of Ranks in One-Criterion Variance Analysis*. *Journal of the American Statistical Association*, 1952. **47**(260): p. 583-621.
260. Chao, A., *Nonparametric Estimation of the Number of Classes in a Population*. *Scandinavian Journal of Statistics*, 1984. **11**(4): p. 265-270.
261. Jacobs, D.M., et al., *(1)H NMR metabolite profiling of feces as a tool to assess the impact of nutrition on the human microbiome*. *NMR Biomed*, 2008. **21**(6): p. 615-26.
262. Monleon, D., et al., *Metabolite profiling of fecal water extracts from human colorectal cancer*. *NMR Biomed*, 2009. **22**(3): p. 342-8.
263. Pettersson, J., et al., *NMR metabolomic analysis of fecal water from subjects on a vegetarian diet*. *Biol Pharm Bull*, 2008. **31**(6): p. 1192-8.
264. Romick-Rosendale, L.E., et al., *NMR-based metabonomics analysis of mouse urine and fecal extracts following oral treatment with the broad-spectrum antibiotic enrofloxacin (Baytril)*. *Magn Reson Chem*, 2009. **47** **Suppl 1**: p. S36-46.
265. Scientific, T., *260/280 and 260/230 Ratios T-009-Technical Bulletin*.
266. Saude, E., et al., *Variation of metabolites in normal human urine*. *Metabolomics*, 2007. **3**(4): p. 439-451.
267. Atzori, L., et al., *1H NMR-based metabolomic analysis of urine from preterm and term neonates*. *Front Biosci (Elite Ed)*, 2011. **3**: p. 1005-12.

268. Trump, S., et al., *1H-NMR metabolic profiling of human neonatal urine*. Magma, 2006. **19**(6): p. 305-12.
269. Maher, A.D., et al., *Experimental and analytical variation in human urine in 1H NMR spectroscopy-based metabolic phenotyping studies*. Anal Chem, 2007. **79**(14): p. 5204-11.
270. Walsh, M.C., et al., *Effect of acute dietary standardization on the urinary, plasma, and salivary metabolomic profiles of healthy humans*. The American Journal of Clinical Nutrition, 2006. **84**(3): p. 531-539.
271. Jeong, J.-Y., et al., *H NMR-Based Urinary Metabolic Profiling of Gender and Diurnal Variation in Healthy Korean Subjects*. J Environ Toxicol, 1986. **25**(4): p. 295-306.
272. Moltu, S.J., et al., *Urinary metabolite profiles in premature infants show early postnatal metabolic adaptation and maturation*. Nutrients, 2014. **6**(5): p. 1913-30.
273. Maeda, K., et al., *Perinatal bile acid metabolism: analysis of urinary unsaturated ketonic bile acids in preterm and full-term infants*. Acta Paediatr, 2003. **92**(2): p. 216-20.
274. Yamato, Y., et al., *Fetal bile acid metabolism: analysis of urinary 3beta-monohydroxy-delta(5) bile acid in preterm infants*. Biol Neonate, 2001. **80**(1): p. 19-25.
275. Hinchliffe, S.A., et al., *Human intrauterine renal growth expressed in absolute number of glomeruli assessed by the disector method and Cavalieri principle*. Lab Invest, 1991. **64**(6): p. 777-84.
276. Webb, M.E., A.G. Smith, and C. Abell, *Biosynthesis of pantothenate*. Nat Prod Rep, 2004. **21**(6): p. 695-721.
277. Genschel, U., *Coenzyme A biosynthesis: reconstruction of the pathway in archaea and an evolutionary scenario based on comparative genomics*. Mol Biol Evol, 2004. **21**(7): p. 1242-51.
278. Shenderov, B.A., *Gut indigenous microbiota and epigenetics*. Microbial Ecology in Health and Disease, 2012. **23**: p. 10.3402/mehd.v23i0.17195.
279. Kau, A.L., et al., *Human nutrition, the gut microbiome, and immune system: envisioning the future*. Nature, 2011. **474**(7351): p. 327-336.
280. Turrioni, F., et al., *Human gut microbiota and bifidobacteria: from composition to functionality*. Antonie Van Leeuwenhoek, 2008. **94**(1): p. 35-50.
281. Spry, C., K. Kirk, and K.J. Saliba, *Coenzyme A biosynthesis: an antimicrobial drug target*. FEMS Microbiol Rev, 2008. **32**(1): p. 56-106.
282. Ford, J.E., et al., *Comparison of the B vitamin composition of milk from mothers of preterm and term babies*. Archives of Disease in Childhood, 1983. **58**(5): p. 367-372.
283. Villasenor, A., et al., *Breast milk metabolome characterization in a single-phase extraction, multiplatform analytical approach*. Anal Chem, 2014. **86**(16): p. 8245-52.
284. *PART II: Investigations concerning pantothenic acid*. Acta Medica Scandinavica, 1950. **138**(S249): p. 51-65.
285. Mahan L.K., R.J.L., *Krause's Food & the Nutrition Care Process*. 2011: Saunders.
286. Tsuji, T., et al., *Urinary excretion of vitamin B1, B2, B6, niacin, pantothenic acid, folate, and vitamin C correlates with dietary intakes of free-living elderly, female Japanese*. Nutr Res, 2010. **30**(3): p. 171-8.
287. Bales, J.R., et al., *1H NMR studies of urine during fasting: excretion of ketone bodies and acetylcarnitine*. Magn Reson Med, 1986. **3**(6): p. 849-56.
288. Sanchez-Niño, M.D. and A. Ortiz, *Differential effects of oral and intravenous l-carnitine on serum lipids: is the microbiota the answer?* Clinical Kidney Journal, 2014. **7**(5): p. 437-441.
289. Bremer, J., *Carnitine--metabolism and functions*. Physiol Rev, 1983. **63**(4): p. 1420-80.
290. Novak, M., et al., *Acetylcarnitine and free carnitine in body fluids before and after birth*. Pediatr Res, 1979. **13**(1): p. 10-5.
291. Sandor, A., et al., *On carnitine content of the human breast milk*. Pediatr Res, 1982. **16**(2): p. 89-91.
292. Vaziri, N.D., Y.Y. Zhao, and M.V. Pahl, *Altered intestinal microbial flora and impaired epithelial barrier structure and function in CKD: the nature, mechanisms, consequences and potential treatment*. Nephrol Dial Transplant, 2015.
293. Ferreira, F., et al., *Trimethylaminuria (fish odor syndrome): genotype characterization among Portuguese patients*. Gene, 2013. **527**(1): p. 366-70.
294. Klepacki, J., et al., *A high-performance liquid chromatography - tandem mass spectrometry - based targeted metabolomics kidney dysfunction marker panel in human urine*. Clin Chim Acta, 2015. **446**: p. 43-53.
295. Tang, W.H., et al., *Intestinal microbiota-dependent phosphatidylcholine metabolites, diastolic dysfunction, and adverse clinical outcomes in chronic systolic heart failure*. J Card Fail, 2015. **21**(2): p. 91-6.
296. Zhang, A.Q., S.C. Mitchell, and R.L. Smith, *Dietary precursors of trimethylamine in man: a pilot study*. Food Chem Toxicol, 1999. **37**(5): p. 515-20.
297. Zeisel, S.H. and K.A. DaCosta, *Increase in human exposure to methylamine precursors of N-nitrosamines after eating fish*. Cancer Res, 1986. **46**(12 Pt 1): p. 6136-8.
298. Zeisel, S.H., S. Gettner, and M. Youssef, *Formation of aliphatic amine precursors of N-nitrosodimethylamine after oral administration of choline and choline analogues in the rat*. Food Chem Toxicol, 1989. **27**(1): p. 31-4.
299. Maitre, L., et al., *Urinary metabolic profiles in early pregnancy are associated with preterm birth and fetal growth restriction in the Rhea mother-child cohort study*. BMC Med, 2014. **12**: p. 110.
300. Ozand, P.T., et al., *Malonic aciduria*. Brain and Development, 1994. **16**, Supplement(0): p. 7-11.
301. Pollitt, R.J., et al., *Increased excretion of propan-1,3-diol and 3-hydroxypropionic acid apparently caused by abnormal bacterial metabolism in the gut*. Clin Chim Acta, 1987. **169**(2-3): p. 151-7.
302. Iacobelli, S., et al., *Factors affecting postnatal changes in serum creatinine in preterm infants with gestational age <32 weeks*. J Perinatol, 2009. **29**(3): p. 232-6.
303. Gubhaju, L., et al., *Assessment of renal functional maturation and injury in preterm neonates during the first month of life*. Am J Physiol Renal Physiol, 2014. **307**(2): p. F149-58.

304. Sjolín, J., et al., *Evaluation of urinary 3-methylhistidine excretion in infection by measurements of 1-methylhistidine and the creatinine ratios*. Am J Clin Nutr, 1989. **49**(1): p. 62-70.
305. Mannan, M.A., et al., *Postnatal development of renal function in preterm and term neonates*. Mymensingh Med J, 2012. **21**(1): p. 103-8.
306. Shemesh, O., et al., *Limitations of creatinine as a filtration marker in glomerulopathic patients*. Kidney Int, 1985. **28**(5): p. 830-838.
307. Bruel, A., et al., *Critical Serum Creatinine Values in Very Preterm Newborns*. PLoS ONE, 2013. **8**(12): p. e84892.
308. Thayyil, S., et al., *A gestation- and postnatal age-based reference chart for assessing renal function in extremely premature infants*. J Perinatol, 2008. **28**(3): p. 226-229.
309. Miall, L.S., et al., *Plasma creatinine rises dramatically in the first 48 hours of life in preterm infants*. Pediatrics, 1999. **104**(6): p. e76.
310. Levine, J., *Controlled trials of inositol in psychiatry*. Eur Neuropsychopharmacol, 1997. **7**(2): p. 147-55.
311. Palatnik, A., et al., *Double-blind, controlled, crossover trial of inositol versus fluvoxamine for the treatment of panic disorder*. J Clin Psychopharmacol, 2001. **21**(3): p. 335-9.
312. Levine, J., et al., *Double-blind, controlled trial of inositol treatment of depression*. Am J Psychiatry, 1995. **152**(5): p. 792-4.
313. Barberini, L., et al., *Urinary metabolomics (GC-MS) reveals that low and high birth weight infants share elevated inositol concentrations at birth*. J Matern Fetal Neonatal Med, 2014. **27 Suppl 2**: p. 20-6.
314. Dessi, A., et al., *Investigation of the (1)H-NMR based urine metabolomic profiles of IUGR, LGA and AGA newborns on the first day of life*. J Matern Fetal Neonatal Med, 2014. **27 Suppl 2**: p. 13-9.
315. Dessi, A., et al., *Metabolomics in newborns with intrauterine growth retardation (IUGR): urine reveals markers of metabolic syndrome*. J Matern Fetal Neonatal Med, 2011. **24 Suppl 2**: p. 35-9.
316. Pereira, G.R., et al., *Serum myoinositol concentrations in premature infants fed human milk, formula for infants, and parenteral nutrition*. Am J Clin Nutr, 1990. **51**(4): p. 589-93.
317. Bromberger, P. and M. Hallman, *Myoinositol in Small Preterm Infants: Relationship Between Intake and Serum Concentration*. J Pediatr Gastroenterol Nutr, 1986. **5**(3): p. 455-458.
318. Clements, R.S., Jr. and R. Reynertson, *Myoinositol metabolism in diabetes mellitus. Effect of insulin treatment*. Diabetes, 1977. **26**(3): p. 215-21.
319. Baati, T., et al., *Quantification of fumaric acid in liver, spleen and urine by high-performance liquid chromatography coupled to photodiode-array detection*. J Pharm Biomed Anal, 2011. **56**(4): p. 758-62.
320. Walker, V. and G.A. Mills, *Urinary organic acid excretion by babies born before 33 weeks of gestation*. Clin Chem, 1989. **35**(7): p. 1460-6.
321. Wu, P.Y., et al., *Urinary excretion of citric acid cycle metabolites in premature newborn infants with and without a respiratory distress syndrome*. Pediatrics, 1965. **36**(6): p. 856-60.
322. Shimizu, S., et al., *Evaluation of two alternative metabolic pathways for creatinine degradation in microorganisms*. Archives of Microbiology, 1986. **145**(4): p. 322-328.
323. Szulmajster, J., *BACTERIAL FERMENTATION OF CREATININE I. : Isolation of N-Methyl-Hydantoin*. Journal of Bacteriology, 1958. **75**(6): p. 633-639.
324. Wyss, M. and R. Kaddurah-Daouk, *Creatine and creatinine metabolism*. Physiol Rev, 2000. **80**(3): p. 1107-213.
325. Kopper, P.H. and H.H. Beard, *Creatinase activity of a strain of Pseudomonas*. Arch Biochem, 1947. **15**(2): p. 195-9.
326. ten Krooden, E. and C.W. Owens, *Creatinine metabolism by Clostridium welchii isolated from human faeces*. Experientia, 1975. **31**(11): p. 1270.
327. Twort, F.W. and E. Mellanby, *On creatin-destroying bacilli in the intestine, and their isolation*. J Physiol, 1912. **44**(1-2): p. 43-9.
328. Dunn, S.R., et al., *Induction of creatininase activity in chronic renal failure: timing of creatinine degradation and effect of antibiotics*. Am J Kidney Dis, 1997. **29**(1): p. 72-7.
329. Jones, J.D. and P.C. Brunett, *Creatinine metabolism and toxicity*. Kidney Int Suppl, 1975(3): p. 294-8.
330. Deshpande, G., et al., *Fish Oil (SMOFlipid) and olive oil lipid (Clinoleic) in very preterm neonates*. J Pediatr Gastroenterol Nutr, 2014. **58**(2): p. 177-82.
331. Minh, D.P., et al., *Degradation of olive oil mill effluents by catalytic wet air oxidation: 2-Oxidation of p-hydroxyphenylacetic and p-hydroxybenzoic acids over Pt and Ru supported catalysts*. Applied Catalysis B: Environmental, 2007. **73**(3-4): p. 236-246.
332. Chalmers, R.A., H.B. Valman, and M.M. Liberman, *Measurement of 4-hydroxyphenylacetic aciduria as a screening test for small-bowel disease*. Clin Chem, 1979. **25**(10): p. 1791-4.
333. Ross, A.B., et al., *A whole-grain-rich diet reduces urinary excretion of markers of protein catabolism and gut microbiota metabolism in healthy men after one week*. J Nutr, 2013. **143**(6): p. 766-73.
334. Jacobs, D.M., et al., *Non-digestible food ingredients, colonic microbiota and the impact on gut health and immunity: a role for metabolomics*. Curr Drug Metab, 2009. **10**(1): p. 41-54.
335. Nicholson, J.K., et al., *Host-gut microbiota metabolic interactions*. Science, 2012. **336**(6086): p. 1262-7.
336. Li, M., et al., *Symbiotic gut microbes modulate human metabolic phenotypes*. Proc Natl Acad Sci U S A, 2008. **105**(6): p. 2117-22.
337. Kimura, A., et al., *Profile of Urinary Bile Acids in Infants and Children: Developmental Pattern of Excretion of Unsaturated Ketonic Bile Acids and 7 [beta]-Hydroxylated Bile Acids*. Pediatr Res, 1999. **45**(4, Part 1 of 2): p. 603-609.

338. Yousef, I.M., et al., *Urinary bile acid profile in children with inborn errors of bile acid metabolism and chronic cholestasis: screening technique using electrospray tandem mass-spectrometry (ES/MS/MS)*. Med Sci Monit, 2003. **9**(3): p. MT21-31.
339. Obatake, M., et al., *Urinary sulfated bile acids: A new simple urine test for cholestasis in infants and children*. Journal of Pediatric Surgery, 2002. **37**(12): p. 1707-1708.
340. Fulford, A., J.S. Scolapio, and J. Aranda-Michel, *Parenteral nutrition-associated hepatotoxicity*. Nutr Clin Pract, 2004. **19**(3): p. 274-83.
341. Calvani, R., et al., *Fecal and urinary NMR-based metabolomics unveil an aging signature in mice*. Exp Gerontol, 2014. **49**: p. 5-11.
342. Marchesi, J.R., et al., *Rapid and noninvasive metabolomic characterization of inflammatory bowel disease*. J Proteome Res, 2007. **6**(2): p. 546-51.
343. Weir, T.L., et al., *Stool Microbiome and Metabolome Differences between Colorectal Cancer Patients and Healthy Adults*. PLoS ONE, 2013. **8**(8): p. e70803.
344. Le Gall, G., et al., *Metabolomics of fecal extracts detects altered metabolic activity of gut microbiota in ulcerative colitis and irritable bowel syndrome*. J Proteome Res, 2011. **10**(9): p. 4208-18.
345. Wu, J., et al., *An optimised sample preparation method for NMR-based faecal metabolomic analysis*. Analyst, 2010. **135**(5): p. 1023-1030.
346. Claesson, M.J., et al., *Gut microbiota composition correlates with diet and health in the elderly*. Nature, 2012. **488**(7410): p. 178-184.
347. Saric, J., et al., *Species variation in the fecal metabolome gives insight into differential gastrointestinal function*. J Proteome Res, 2008. **7**(1): p. 352-60.
348. Hiele, M., et al., *Influence of nutritional substrates on the formation of volatiles by the fecal flora*. Gastroenterology, 1991. **100**(6): p. 1597-602.
349. Martin, F.P., et al., *Dietary modulation of gut functional ecology studied by fecal metabolomics*. J Proteome Res, 2010. **9**(10): p. 5284-95.
350. Ho, J.J., *Appropriate use of antibiotics in the NICU*. Eastern Journal of Medicine, 2010. **15**: p. 133-138.
351. Veerman-Wauters, G., *Neonatal gut development and postnatal adaptation*. European Journal of Pediatrics, 1996. **155**(8): p. 627-632.
352. Hopkins, M.J., R. Sharp, and G.T. Macfarlane, *Variation in human intestinal microbiota with age*. Dig Liver Dis, 2002. **34 Suppl 2**: p. S12-8.
353. Amer, M.D., et al., *Platelet-activating factor concentration in the stool of human newborns: effects of enteral feeding and neonatal necrotizing enterocolitis*. Biol Neonate, 2004. **85**(3): p. 159-66.
354. De Filippo, C., et al., *Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa*. Proc Natl Acad Sci U S A, 2010. **107**(33): p. 14691-6.
355. Ford, H.R., et al., *The microenvironment influences the pattern of bacterial translocation in formula-fed neonates*. J Pediatr Surg, 1996. **31**(4): p. 486-9.
356. Hascoet, J.M., et al., *Effect of formula composition on the development of infant gut microbiota*. J Pediatr Gastroenterol Nutr, 2011. **52**(6): p. 756-62.
357. Martin, F.-P.J., et al., *Dietary Modulation of Gut Functional Ecology Studied by Fecal Metabolomics*. Journal of Proteome Research, 2010. **9**(10): p. 5284-5295.
358. Gueimonde, M., et al., *Breast milk: a source of bifidobacteria for infant gut development and maturation?* Neonatology, 2007. **92**(1): p. 64-6.
359. Heinzmann, S.S., et al., *Stability and Robustness of Human Metabolic Phenotypes in Response to Sequential Food Challenges*. Journal of Proteome Research, 2011. **11**(2): p. 643-655.
360. Poroyko, V., et al., *Diet creates metabolic niches in the "immature gut" that shape microbial communities*. Nutr Hosp, 2011. **26**(6): p. 1283-95.
361. Chiang, S.-S. and T.-M. Pan, *Beneficial effects of <i>Lactobacillus paracasei</i> subsp. <i>paracasei</i> NTU 101 and its fermented products*. Applied Microbiology and Biotechnology, 2012. **93**(3): p. 903-916.
362. Fernandez, L., et al., *The human milk microbiota: Origin and potential roles in health and disease*. Pharmacol Res, 2012.
363. Hunt, K.M., et al., *Human milk oligosaccharides promote the growth of staphylococci*. Appl Environ Microbiol, 2012. **78**(14): p. 4763-70.
364. Martin, V., et al., *Sharing of bacterial strains between breast milk and infant feces*. J Hum Lact, 2012. **28**(1): p. 36-44.
365. Young, W., et al., *Changes in Bowel Microbiota Induced by Feeding Weanlings Resistant Starch Stimulate Transcriptomic and Physiological Responses*. Applied and Environmental Microbiology, 2012. **78**(18): p. 6656-6664.
366. Bezirtzoglou, E., M.B. Romond, and C. Romond, *Regulation of the bacterial intestinal implantation in infant born by caesarean section*. Comp Immunol Microbiol Infect Dis, 1992. **15**(1): p. 71-4.
367. Grönlund, M.-M., et al., *Fecal Microflora in Healthy Infants Born by Different Methods of Delivery: Permanent Changes in Intestinal Flora After Cesarean Delivery*. Journal of Pediatric Gastroenterology and Nutrition, 1999. **28**(1): p. 19-25.
368. Hanson, L.Å. and M. Korotkova, *The role of breastfeeding in prevention of neonatal infection*. Seminars in Neonatology, 2002. **7**(4): p. 275-281.
369. Huurre, A., et al., *Mode of delivery - effects on gut microbiota and humoral immunity*. Neonatology, 2008. **93**(4): p. 236-40.

370. Matsumiya, Y., et al., *Molecular epidemiological study of vertical transmission of vaginal Lactobacillus species from mothers to newborn infants in Japanese, by arbitrarily primed polymerase chain reaction*. J Infect Chemother, 2002. **8**(1): p. 43-9.
371. Morelli, L., *Postnatal development of intestinal microflora as influenced by infant nutrition*. J Nutr, 2008. **138**(9): p. 1791S-1795S.
372. Zeitlin, J., et al., *Variability in caesarean section rates for very preterm births at 28–31 weeks of gestation in 10 European regions: Results of the MOSAIC project*. European Journal of Obstetrics & Gynecology and Reproductive Biology, 2010. **149**(2): p. 147-152.
373. Ferraris, L., et al., *Clostridia in Premature Neonates' Gut: Incidence, Antibiotic Susceptibility, and Perinatal Determinants Influencing Colonization*. PLoS One, 2012. **7**(1): p. e30594.
374. Jauregui, F., et al., *Effects of intrapartum penicillin prophylaxis on intestinal bacterial colonization in infants*. J Clin Microbiol, 2004. **42**(11): p. 5184-8.
375. Swann, J.R., et al., *Variation in antibiotic-induced microbial recolonization impacts on the host metabolic phenotypes of rats*. J Proteome Res, 2011. **10**(8): p. 3590-603.
376. Kochhar, S., et al., *Probing gender-specific metabolism differences in humans by nuclear magnetic resonance-based metabolomics*. Anal Biochem, 2006. **352**(2): p. 274-81.
377. Modi, N., *Survival after extremely preterm birth*. BMJ, 2008. **336**(7655): p. 1199-200.
378. Seravalli, E. and R. Dubos, *Lasting biological effects of early environmental influences. II. Lasting depression of weight caused by neonatal contamination*. J Exp Med, 1968. **127**(4): p. 801-18.
379. Coates, M.E., J.E. Ford, and G.F. Harrison, *Intestinal synthesis of vitamins of the B complex in chicks*. Br J Nutr, 1968. **22**(3): p. 493-500.
380. Mayo B., v., *Bifidobacteria: Genomics and Molecular Aspects*. Caister Academic Press.
381. Gate, C. *Datacard: Nutriprem 1 low birthweight formula*. 2015.
382. Williams, R.R. and J.K. Cline, *SYNTHESIS OF VITAMIN B1*. Journal of the American Chemical Society, 1936. **58**(8): p. 1504-1505.
383. Kochlar S., M.F.-P., *Metabonomics and Gut Microbiota in Nutrition and Disease*. 2014: Springer.
384. Tang, W.H.W. and S.L. Hazen, *The contributory role of gut microbiota in cardiovascular disease*. The Journal of Clinical Investigation, 2014. **124**(10): p. 4204-4211.
385. Brown, C.T., et al., *Genome resolved analysis of a premature infant gut microbial community reveals a Varibaculum cambriense genome and a shift towards fermentation-based metabolism during the third week of life*. Microbiome, 2013. **1**: p. 30-30.
386. Sato, H., *Fecal trimethylamine, lactate and volatile fatty acid concentrations and their relationships with diarrhea in newborn dairy calves*. J Vet Med Sci, 2010. **72**(12): p. 1551-5.
387. He, Q., et al., *Intrauterine growth restriction alters the metabonome of the serum and jejunum in piglets*. Mol Biosyst, 2011. **7**(7): p. 2147-55.
388. Chatterjea, M.N., *Textbook of Medical Biochemistry*. 2011, Jaypee Brothers Medical Publishers.
389. Vogt, J.A. and T.M. Wolever, *Fecal acetate is inversely related to acetate absorption from the human rectum and distal colon*. J Nutr, 2003. **133**(10): p. 3145-8.
390. Miller, T.L. and M.J. Wolin, *Pathways of acetate, propionate, and butyrate formation by the human fecal microbial flora*. Appl Environ Microbiol, 1996. **62**(5): p. 1589-92.
391. Louis, P., G.L. Hold, and H.J. Flint, *The gut microbiota, bacterial metabolites and colorectal cancer*. Nat Rev Micro, 2014. **12**(10): p. 661-672.
392. Tremaroli, V. and F. Backhed, *Functional interactions between the gut microbiota and host metabolism*. Nature, 2012. **489**(7415): p. 242-249.
393. Fernandes, J., et al., *Adiposity, gut microbiota and faecal short chain fatty acids are linked in adult humans*. Nutr Diab, 2014. **4**: p. e121.
394. Rasmussen, H.S., et al., *Faecal concentrations and production rates of short chain fatty acids in normal neonates*. Acta Paediatr Scand, 1988. **77**(3): p. 365-8.
395. Marcobal, A., et al., *Tyramine and phenylethylamine biosynthesis by food bacteria*. Crit Rev Food Sci Nutr, 2012. **52**(5): p. 448-67.
396. Landete, J.M., I. Pardo, and S. Ferrer, *Tyramine and phenylethylamine production among lactic acid bacteria isolated from wine*. Int J Food Microbiol, 2007. **115**(3): p. 364-8.
397. Masson, F., R. Talon, and M.C. Montel, *Histamine and tyramine production by bacteria from meat products*. International Journal of Food Microbiology, 1996. **32**(1–2): p. 199-207.
398. Kim, M.-J. and K.-S. Kim, *Tyramine production among lactic acid bacteria and other species isolated from kimchi*. LWT - Food Science and Technology, 2014. **56**(2): p. 406-413.
399. Suzzi, G. and S. Torriani, *Editorial: Biogenic amines in foods*. Front Microbiol, 2015. **6**: p. 472.
400. Kennedy D.O., *Plants and the Human Brain*. 2014: Oxford University Press.
401. Ozogul, F., et al., *The Impact of Cell-Free Supernatant of Lactic Acid Bacteria on Putrescine and other Polyamine Formation by Food-Borne-Pathogens in Ornithine Decarboxylase Broth*. J Agric Food Chem, 2015.
402. Tomar, P.C., N. Lakra, and S.N. Mishra, *Cadaverine: A lysine catabolite involved in plant growth and development*. Plant Signaling & Behavior, 2013. **8**(10): p. e25850.
403. Hussain, A., et al., *High-affinity olfactory receptor for the death-associated odor cadaverine*. Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(48): p. 19579-19584.
404. Romano, A., et al., *Three-Component Lysine/Ornithine Decarboxylation System in Lactobacillus saerimneri 30a*. Journal of Bacteriology, 2013. **195**(6): p. 1249-1254.

405. Wolrath, H., et al., *Analysis of Bacterial Vaginosis-Related Amines in Vaginal Fluid by Gas Chromatography and Mass Spectrometry*. Journal of Clinical Microbiology, 2001. **39**(11): p. 4026-4031.
406. Benner, R.A., Jr., W.F. Staruszkiewicz, and W.S. Otwell, *Putrescine, cadaverine, and indole production by bacteria isolated from wild and aquacultured penaeid shrimp stored at 0, 12, 24, and 36 degrees C*. J Food Prot, 2004. **67**(1): p. 124-33.
407. Ben-Gigirey, B., et al., *Histamine and cadaverine production by bacteria isolated from fresh and frozen albacore (Thunnus alalunga)*. J Food Prot, 1999. **62**(8): p. 933-9.
408. Lohinai, Z., et al., *Bacterial Lysine Decarboxylase Influences Human Dental Biofilm Lysine Content, Biofilm Accumulation and Sub-Clinical Gingival Inflammation*. Journal of periodontology, 2012. **83**(8): p. 1048-1056.
409. Kamio, Y. and K. Nakamura, *Putrescine and cadaverine are constituents of peptidoglycan in Veillonella alcalescens and Veillonella parvula*. Journal of Bacteriology, 1987. **169**(6): p. 2881-2884.
410. Watanabe, M., et al., *GABA and GABA receptors in the central nervous system and other organs*. Int Rev Cytol, 2002. **213**: p. 1-47.
411. Goudet, C., et al., *Metabotropic receptors for glutamate and GABA in pain*. Brain Res Rev, 2009. **60**(1): p. 43-56.
412. Wisnowski, J.L., et al., *Altered Glutamatergic Metabolism Associated with Punctate White Matter Lesions in Preterm Infants*. PLoS ONE, 2013. **8**(2): p. e56880.
413. Jasmin, L., M.V. Wu, and P.T. Ohara, *GABA puts a stop to pain*. Curr Drug Targets CNS Neurol Disord, 2004. **3**(6): p. 487-505.
414. Hemarajata, P. and J. Versalovic, *Effects of probiotics on gut microbiota: mechanisms of intestinal immunomodulation and neuromodulation*. Therapeutic Advances in Gastroenterology, 2013. **6**(1): p. 39-51.
415. Su, M.S., S. Schlicht, and M.G. Ganzle, *Contribution of glutamate decarboxylase in Lactobacillus reuteri to acid resistance and persistence in sourdough fermentation*. Microb Cell Fact, 2011. **10 Suppl 1**: p. S8.
416. Li, H. and Y. Cao, *Lactic acid bacterial cell factories for gamma-aminobutyric acid*. Amino Acids, 2010. **39**(5): p. 1107-16.
417. Higuchi, T., H. Hayashi, and K. Abe, *Exchange of glutamate and gamma-aminobutyrate in a Lactobacillus strain*. J Bacteriol, 1997. **179**(10): p. 3362-4.
418. Owens, D.F. and A.R. Kriegstein, *Is there more to GABA than synaptic inhibition?* Nat Rev Neurosci, 2002. **3**(9): p. 715-27.
419. Li, K. and E. Xu, *The role and the mechanism of gamma-aminobutyric acid during central nervous system development*. Neurosci Bull, 2008. **24**(3): p. 195-200.
420. Fiumelli, H. and M.A. Woodin, *Role of activity-dependent regulation of neuronal chloride homeostasis in development*. Curr Opin Neurobiol, 2007. **17**(1): p. 81-6.
421. Ben-Ari, Y., et al., *GABA: a pioneer transmitter that excites immature neurons and generates primitive oscillations*. Physiol Rev, 2007. **87**(4): p. 1215-84.
422. Ben-Ari, Y., et al., *gamma-Aminobutyric acid (GABA): a fast excitatory transmitter which may regulate the development of hippocampal neurones in early postnatal life*. Prog Brain Res, 1994. **102**: p. 261-73.
423. Leinekugel, X., et al., *GABA is the principal fast-acting excitatory transmitter in the neonatal brain*. Adv Neurol, 1999. **79**: p. 189-201.
424. McLean, H.A., et al., *Spontaneous release of GABA activates GABAB receptors and controls network activity in the neonatal rat hippocampus*. J Neurophysiol, 1996. **76**(2): p. 1036-46.
425. Back, S.A., et al., *Late oligodendrocyte progenitors coincide with the developmental window of vulnerability for human perinatal white matter injury*. J Neurosci, 2001. **21**(4): p. 1302-12.
426. Braat, S. and R.F. Kooy, *The GABA Receptor as a Therapeutic Target for Neurodevelopmental Disorders*. Neuron, 2015. **86**(5): p. 1119-1130.
427. Kamyar, M., et al., *Magnesium sulfate, chorioamnionitis, and neurodevelopment after preterm birth*. Bjog, 2015.
428. Rodrigues, A.N., et al., *Neurodevelopmental outcome of extremely preterm infants born to rural and urban residents' mothers in Australia*. Early Hum Dev, 2015. **91**(8): p. 437-443.
429. Koc, O., et al., *School Performance and Neurodevelopment of Very Low Birth Weight Preterm Infants: First Report From Turkey*. J Child Neurol, 2015.
430. Shaw, J.C., et al., *Preterm birth affects GABAA receptor subunit mRNA levels during the foetal-to-neonatal transition in guinea pigs*. J Dev Orig Health Dis, 2015. **6**(3): p. 250-60.
431. Corpeleijn, W.E., et al., *Almost all enteral aspartate is taken up in first-pass metabolism in enterally fed preterm infants*. Clin Nutr, 2010. **29**(3): p. 341-6.
432. Chen, P.E., et al., *Structural Features of the Glutamate Binding Site in Recombinant NR1/NR2A N-Methyl-D-aspartate Receptors Determined by Site-Directed Mutagenesis and Molecular Modeling*. Molecular Pharmacology, 2005. **67**(5): p. 1470-1484.
433. Ueland, P.M., P.I. Holm, and S. Hustad, *Betaine: a key modulator of one-carbon metabolism and homocysteine status*. Clin Chem Lab Med, 2005. **43**(10): p. 1069-75.
434. Lever, M. and S. Slow, *The clinical significance of betaine, an osmolyte with a key role in methyl group metabolism*. Clin Biochem, 2010. **43**(9): p. 732-44.
435. Kikugawa, Y., et al., *N-methoxydiacetamide: A new selective acetylating agent*. Tetrahedron Letters, 1990. **31**(2): p. 243-246.
436. Kobata, A., *Structures and application of oligosaccharides in human milk*. Proceedings of the Japan Academy. Series B, Physical and Biological Sciences, 2010. **86**(7): p. 731-747.
437. Martin, R., et al., *Human milk is a source of lactic acid bacteria for the infant gut*. J Pediatr, 2003. **143**(6): p. 754-8.

438. Ogawa, K., et al., *Volatile fatty acids, lactic acid, and pH in the stools of breast-fed and bottle-fed infants*. J Pediatr Gastroenterol Nutr, 1992. **15**(3): p. 248-52.
439. Tomaszewski, L., et al., *[Lactic acid excretion with stools by premature babies, newborns, infants and healthy children]*. Pediatr Pol, 1971. **46**(1): p. 13-9.
440. Jenness, R., *The composition of human milk*. Semin Perinatol, 1979. **3**(3): p. 225-39.
441. Lonnerdal, B., *Effects of maternal dietary intake on human milk composition*. J Nutr, 1986. **116**(4): p. 499-513.
442. Wagner, C.L., et al., *High-dose vitamin D3 supplementation in a cohort of breastfeeding mothers and their infants: a 6-month follow-up pilot study*. Breastfeed Med, 2006. **1**(2): p. 59-70.
443. Cederlund, A., et al., *Lactose in human breast milk an inducer of innate immunity with implications for a role in intestinal homeostasis*. PLoS One, 2013. **8**(1): p. e53876.
444. Zivkovic, A., et al., *Human milk glycomiome and its impact on the infant gastrointestinal microbiota*. Proc Natl Acad Sci U S A, 2011. **108**(Suppl 1): p. 4653 - 4658.
445. Urbaniak, C., J. Burton, and G. Reid, *Breast, milk and microbes: a complex relationship that does not end with lactation*. Womens Health (Lond Engl), 2012. **8**: p. 385 - 398.
446. Lepage, P. and P. Van de Perre, *The immune system of breast milk: antimicrobial and anti-inflammatory properties*. Adv Exp Med Biol, 2012. **743**: p. 121 - 137.
447. Jeurink, P., et al., *Human milk: a source of more life than we imagine*. Benef Microbes, 2013. **4**: p. 17 - 30.
448. Hunt, K., et al., *Human milk oligosaccharides promote the growth of Staphylococci*. Appl Environ Microbiol, 2012. **78**: p. 4763 - 4770.
449. Goldman, A., *The immune system of human milk: antimicrobial, antiinflammatory and immunomodulating properties*. Pediatr Infect Dis J, 1993. **12**: p. 664 - 671.
450. Engfer, M., et al., *Human milk oligosaccharides are resistant to enzymatic hydrolysis in the upper gastrointestinal tract*. Am J Clin Nutr, 2000. **71**: p. 1589 - 1596.
451. Ward, T., et al., *Human milk metagenome: a functional capacity analysis*. BMC Microbiology, 2013. **13**(1): p. 116.
452. Meinen-Derr, J., et al., *Role of human milk in extremely low birth weight infants' risk of necrotizing enterocolitis or death*. J Perinatol, 2009. **29**: p. 57 - 62.
453. Martin, V., et al., *Streptococcus lactarius sp. nov., isolated from breast milk of healthy women*. Int J Syst Evol Microbiol, 2011. **61**: p. 1048 - 1052.
454. Heikkila, M. and P. Saris, *Inhibition of Staphylococcus aureus by the commensal bacteria of human milk*. J Appl Microbiol, 2003. **95**: p. 471 - 478.
455. Parm, U., et al., *Risk factors associated with gut and nasopharyngeal colonization by common Gram-negative species and yeasts in neonatal intensive care units patients*. Early Hum Dev, 2011. **87**(6): p. 391-9.
456. Dogra, S., et al., *Dynamics of infant gut microbiota are influenced by delivery mode and gestational duration and are associated with subsequent adiposity*. MBio, 2015. **6**(1).
457. Kinross, J.M., et al., *The human gut microbiome: implications for future health care*. Curr Gastroenterol Rep, 2008. **10**(4): p. 396-403.
458. Lindberg, E., et al., *High rate of transfer of Staphylococcus aureus from parental skin to infant gut flora*. J Clin Microbiol, 2004. **42**(2): p. 530-4.
459. Henderson, J.J., et al., *Effect of preterm birth and antenatal corticosteroid treatment on lactogenesis II in women*. Pediatrics, 2008. **121**(1): p. e92-100.
460. Boehm, G., et al., *Supplementation of a bovine milk formula with an oligosaccharide mixture increases counts of faecal bifidobacteria in preterm infants*. Archives of Disease in Childhood Fetal and Neonatal Edition, 2002. **86**(3): p. F178-F181.
461. Khodayar-Pardo, P., et al., *Impact of lactation stage, gestational age and mode of delivery on breast milk microbiota*. J Perinatol, 2014. **34**(8): p. 599-605.
462. Underwood, M.A., et al., *A comparison of two probiotic strains of bifidobacteria in premature infants*. The Journal of pediatrics, 2013. **163**(6): p. 10.1016/j.jpeds.2013.07.017.
463. Hopkins, M.J., et al., *Characterisation of intestinal bacteria in infant stools using real-time PCR and northern hybridisation analyses*. FEMS Microbiol Ecol, 2005. **54**(1): p. 77-85.
464. Cabral, M.E., et al., *Citrate metabolism by Enterococcus faecium and Enterococcus durans isolated from goat's and ewe's milk: influence of glucose and lactose*. Can J Microbiol, 2007. **53**(5): p. 607-15.
465. Vaningelgem, F., et al., *Cometabolism of Citrate and Glucose by Enterococcus faecium FAIR-E 198 in the Absence of Cellular Growth*. Applied and Environmental Microbiology, 2006. **72**(1): p. 319-326.
466. Barrett, E., et al., *gamma-Aminobutyric acid production by culturable bacteria from the human intestine*. J Appl Microbiol, 2012. **113**(2): p. 411-7.
467. Lee, J.-H. and D.J. O'Sullivan, *Genomic Insights into Bifidobacteria*. Microbiology and Molecular Biology Reviews, 2010. **74**(3): p. 378-416.
468. Guaraldi, F. and G. Salvatori, *Effect of Breast and Formula Feeding on Gut Microbiota Shaping in Newborns*. Frontiers in Cellular and Infection Microbiology, 2012. **2**: p. 94.
469. Matamoros, S., et al., *Development of intestinal microbiota in infants and its impact on health*. Trends in Microbiology. **21**(4): p. 167-173.
470. Putignani, L., *Human gut microbiota: onset and shaping through life stages and perturbations, 2nd Edition*. 2014: Frontiers E-books.
471. Collado, M.C., et al., *Microbial ecology and host-microbiota interactions during early life stages*. Gut Microbes, 2012. **3**(4): p. 352-65.

472. Marincola, F.C., et al., *A metabolomic study of preterm human and formula milk by high resolution NMR and GC/MS analysis: preliminary results*. J Matern Fetal Neonatal Med, 2012. **25**(Suppl 5): p. 62-7.
473. Turroni, F., et al., *Diversity of Bifidobacteria within the Infant Gut Microbiota*. PLoS ONE, 2012. **7**(5): p. e36957.
474. Palmer, C., et al., *Development of the human infant intestinal microbiota*. PLoS Biol, 2007. **5**(7): p. e177.
475. Saavedra, J.M. and A.M. Dattilo, *Early development of intestinal microbiota: implications for future health*. Gastroenterol Clin North Am, 2012. **41**(4): p. 717-31.
476. Berrington, J.E., et al., *Gut microbiota in preterm infants: assessment and relevance to health and disease*. Arch Dis Child Fetal Neonatal Ed, 2013. **98**(4): p. F286-90.
477. Hopkins, M.J., R. Sharp, and G.T. Macfarlane, *Age and disease related changes in intestinal bacterial populations assessed by cell culture, 16S rRNA abundance, and community cellular fatty acid profiles*. Gut, 2001. **48**(2): p. 198-205.
478. Eggesbo, M., et al., *Development of gut microbiota in infants not exposed to medical interventions*. APMIS, 2011. **119**(1): p. 17-35.
479. Schwartz, A., et al., *Development of the Intestinal Bacterial Composition in Hospitalized Preterm Infants in Comparison with Breast-Fed, Full-Term Infants*. Pediatr Res, 2003. **54**(3): p. 393-399.
480. France de La Cochetiere, M., et al., *Intestinal Microbiota in Neonates and Preterm Infants: A Review*. Current Pediatric Reviews, 2007. **3**(1): p. 21-34.
481. Wang, Y., et al., *16S rRNA gene-based analysis of fecal microbiota from preterm infants with and without necrotizing enterocolitis*. ISME J, 2009. **3**(8): p. 944-954.
482. Chang, J.Y., et al., *Pyrosequencing-based molecular monitoring of the intestinal bacterial colonization in preterm infants*. J Pediatr Gastroenterol Nutr, 2011. **53**(5): p. 512-9.
483. De Palma, G., et al., *Intestinal dysbiosis and reduced immunoglobulin-coated bacteria associated with coeliac disease in children*. BMC Microbiol, 2010. **10**: p. 63.
484. Dong, P., Y. Yang, and W.P. Wang, *The role of intestinal bifidobacteria on immune system development in young rats*. Early Hum Dev, 2010. **86**(1): p. 51-8.
485. Donnet-Hughes, A., et al., *Potential role of the intestinal microbiota of the mother in neonatal immune education*. Proc Nutr Soc, 2010. **69**: p. 407 - 415.
486. Forsberg, G., et al., *Presence of bacteria and innate immunity of intestinal epithelium in childhood celiac disease*. Am J Gastroenterol, 2004. **99**(5): p. 894-904.
487. Kelly, D., T. King, and R. Aminov, *Importance of microbial colonization of the gut in early life to the development of immunity*. Mutat Res, 2007. **622**(1-2): p. 58-69.
488. Manco, M., L. Putignani, and G.F. Bottazzo, *Gut Microbiota, Lipopolysaccharides, and Innate Immunity in the Pathogenesis of Obesity and Cardiovascular Risk*. Endocrine Reviews, 2010. **31**(6): p. 817-844.
489. McElroy, S.J. and J.H. Weitkamp, *Innate Immunity in the Small Intestine of the Preterm Infant*. Neoreviews, 2011. **12**(9): p. e517-e526.
490. Morin, S., et al., *Delayed bacterial colonization of the gut alters the host immune response to oral sensitization against cow's milk proteins*. Mol Nutr Food Res, 2012. **56**(12): p. 1838-47.
491. El Aidy, S., et al., *Gut bacteria-host metabolic interplay during conventionalisation of the mouse germfree colon*. ISME J, 2013. **7**(4): p. 743-55.
492. Rezzonico, E., et al., *Bacterial adaptation to the gut environment favors successful colonization: microbial and metatranscriptomic characterization of a simplified microbiota mouse model*. Gut Microbes, 2011. **2**(6): p. 307-18.
493. Arbolea, S., et al., *Facultative to strict anaerobes ratio in the preterm infant microbiota: a target for intervention?* Gut Microbes, 2012. **3**(6): p. 583-8.
494. Benach, J.L., E. Li, and M.M. McGovern, *A microbial association with autism*. MBio, 2012. **3**(1).
495. Dumas, M.E., et al., *Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice*. Proc Natl Acad Sci U S A, 2006. **103**(33): p. 12511-6.
496. Gonzalez, A., et al., *The mind-body-microbial continuum*. Dialogues Clin Neurosci, 2011. **13**(1): p. 55-62.
497. Gracey, M., *Intestinal microflora and bacterial overgrowth in early life*. J Pediatr Gastroenterol Nutr, 1982. **1**(1): p. 13-22.
498. Burcelin, R., et al., *Gut microbiota and diabetes: from pathogenesis to therapeutic perspective*. Acta Diabetologica, 2011. **48**(4): p. 257-273.
499. Cani, P.D., et al., *Changes in Gut Microbiota Control Metabolic Endotoxemia-Induced Inflammation in High-Fat Diet-Induced Obesity and Diabetes in Mice*. Diabetes, 2008. **57**(6): p. 1470-1481.
500. Devaraj, S., P. Hemarajata, and J. Versalovic, *The human gut microbiome and body metabolism: implications for obesity and diabetes*. Clin Chem, 2013. **59**(4): p. 617-28.
501. Bäckhed, F., et al., *Mechanisms underlying the resistance to diet-induced obesity in germ-free mice*. Proceedings of the National Academy of Sciences, 2007. **104**(3): p. 979-984.
502. de La Serre, C.B., et al., *Propensity to high-fat diet-induced obesity in rats is associated with changes in the gut microbiota and gut inflammation*. American Journal of Physiology - Gastrointestinal and Liver Physiology, 2010. **299**(2): p. G440-G448.
503. Ding, S., et al., *High-Fat Diet: Bacteria Interactions Promote Intestinal Inflammation Which Precedes and Correlates with Obesity and Insulin Resistance in Mouse*. PLoS One, 2010. **5**(8): p. e12191.
504. Duncan, S.H., et al., *Human colonic microbiota associated with diet, obesity and weight loss*. Int J Obes (Lond), 2008. **32**(11): p. 1720-4.
505. Flier, J.S. and J.J. Mekalanos, *Gut Check: Testing a Role for the Intestinal Microbiome in Human Obesity*. Science Translational Medicine, 2009. **1**(6): p. 6ps7.

506. Haemer, M.A., T.T. Huang, and S.R. Daniels, *The effect of neurohormonal factors, epigenetic factors, and gut microbiota on risk of obesity*. Prev Chronic Dis, 2009. **6**(3): p. A96.
507. Harris, K., et al., *Is the Gut Microbiota a New Factor Contributing to Obesity and Its Metabolic Disorders?* Journal of Obesity, 2012. **2012**: p. 14.
508. Ji, Y., et al., *Modulation of the murine microbiome with a concomitant anti-obesity effect by <i>Lactobacillus rhamnosus</i> GG and <i>Lactobacillus sakei</i> NR28*. Beneficial Microbes, 2012. **3**(1): p. 13-22.
509. Ley, R.E., et al., *Microbial ecology: human gut microbes associated with obesity*. Nature, 2006. **444**(7122): p. 1022-3.
510. Murphy, E.F., et al., *Composition and energy harvesting capacity of the gut microbiota: relationship to diet, obesity and time in mouse models*. Gut, 2010. **59**(12): p. 1635-1642.
511. Patil, D.P., et al., *Molecular analysis of gut microbiota in obesity among Indian individuals*. J Biosci, 2012. **37**(4): p. 647-57.
512. Tehrani, A.B., et al., *Obesity and its associated disease: a role for microbiota?* Neurogastroenterology & Motility, 2012. **24**(4): p. 305-311.
513. Tilg, H., A.R. Moschen, and A. Kaser, *Obesity and the microbiota*. Gastroenterology, 2009. **136**(5): p. 1476-83.
514. Turnbaugh, P.J., et al., *Diet-Induced Obesity Is Linked to Marked but Reversible Alterations in the Mouse Distal Gut Microbiome*. Cell Host & Microbe, 2008. **3**(4): p. 213-223.
515. Zhang, H., et al., *Human gut microbiota in obesity and after gastric bypass*. Proc Natl Acad Sci U S A, 2009. **106**(7): p. 2365-70.

Appendix

The permission for the reprint of the figures number 2 and 3 was obtained in accordance with the requirements of the New England Journal of Medicine: “Content (full-text or portions thereof) may be used in print and electronic versions of your dissertation or thesis without formal permission from the Massachusetts Medical Society, Publisher of the New England Journal of Medicine”.

Urinary Metabonome

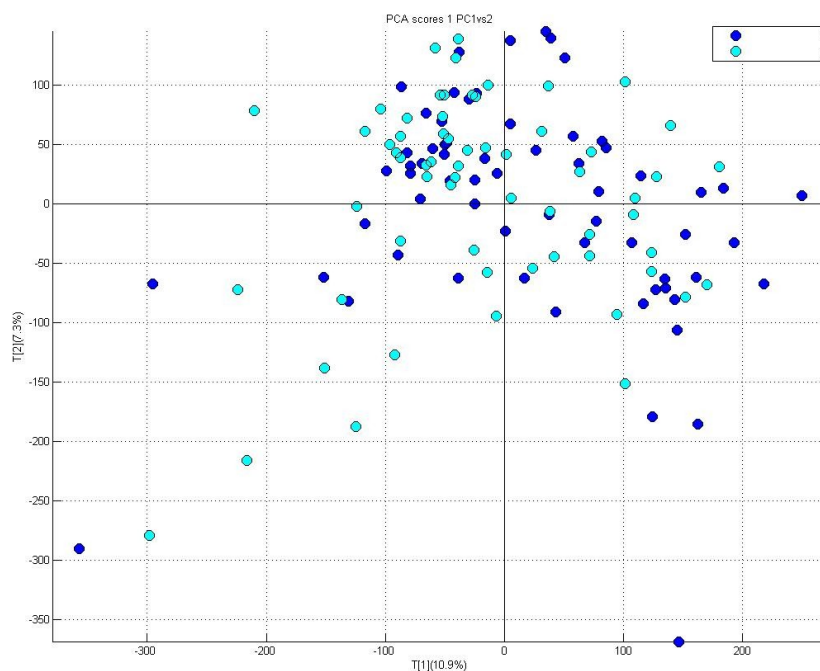


Figure 78: PCA Scores Scatter Plot of NMR spectra of urine samples collected in the 1st week of life. Samples are coloured according to type of intravenous lipid solution. Dark blue (1) represents Intralipid while light blue (2) corresponds to SMOF lipid.

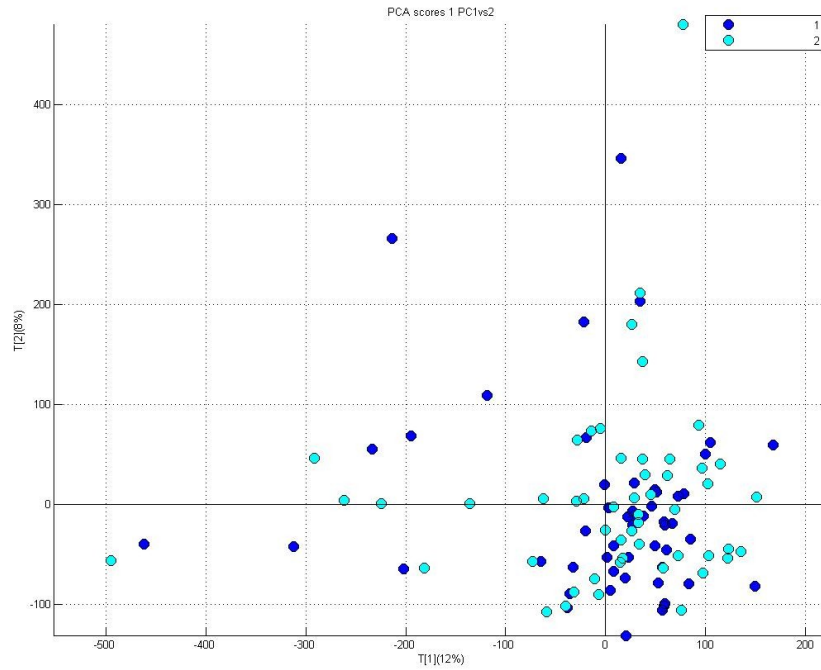


Figure 79: PCA Scores Scatter Plot of NMR spectra of urine samples collected in the 3rd week of life. Samples are coloured according to type of intravenous lipid solution. Dark blue (1) represents Intralipid while light blue (2) corresponds to SMOF lipid.

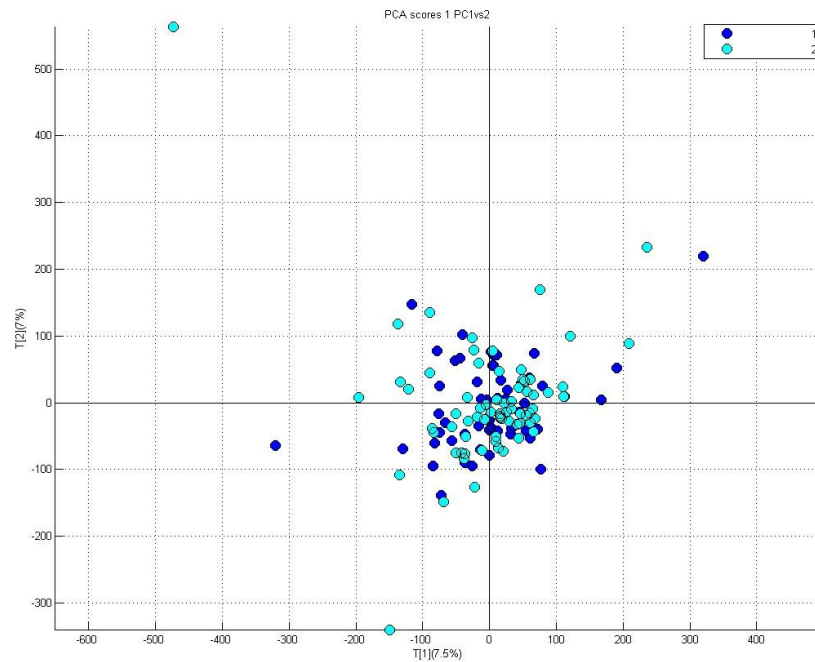


Figure 80: PCA Scores Scatter Plot of NMR spectra of the last urine samples collected. Samples are coloured according to type of intravenous lipid solution. Dark blue (1) represents Intralipid while light blue (2) corresponds to SMOF lipid.

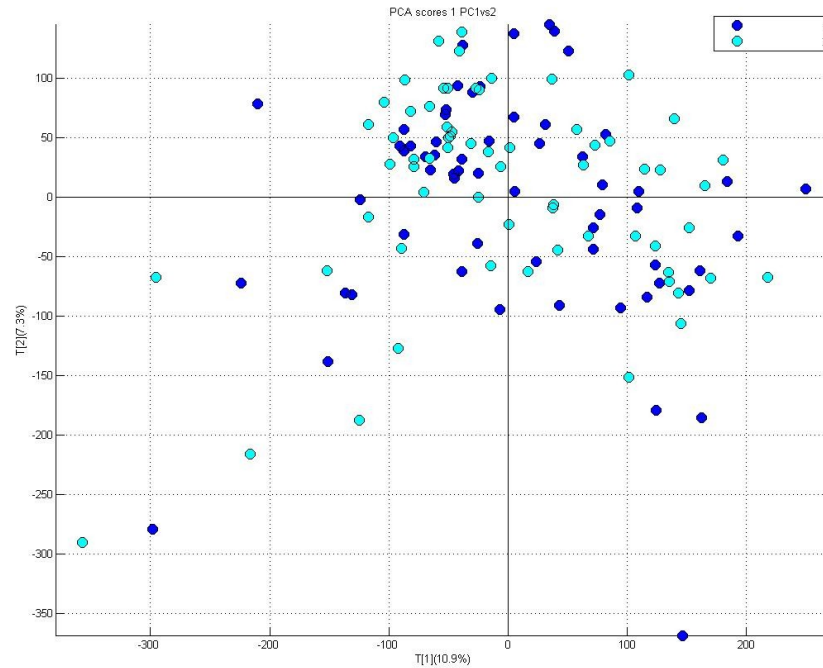


Figure 81: PCA Scores Scatter Plot of NMR spectra of urine samples collected in the 1st week of life. Samples are coloured according to concentration of amino acids in parenteral nutrition. Dark blue (1) represents increasing concentration while light blue (2) corresponds to recommended daily intake.

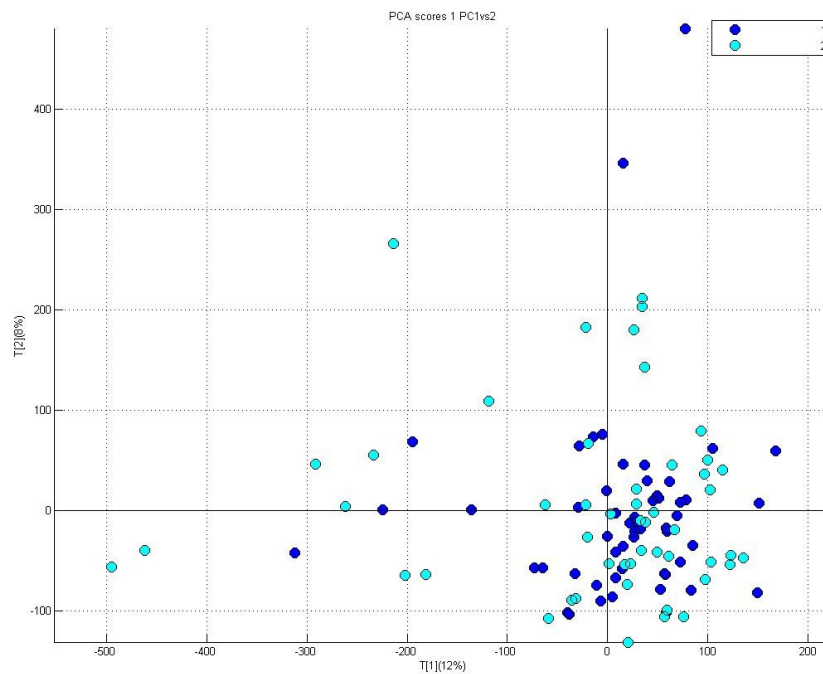


Figure 82: PCA Scores Scatter Plot of NMR spectra of urine samples collected in the 3rd week of life. Samples are coloured according to concentration of amino acids in parenteral nutrition. Dark blue (1) represents increasing concentration while light blue (2) corresponds to recommended daily intake.

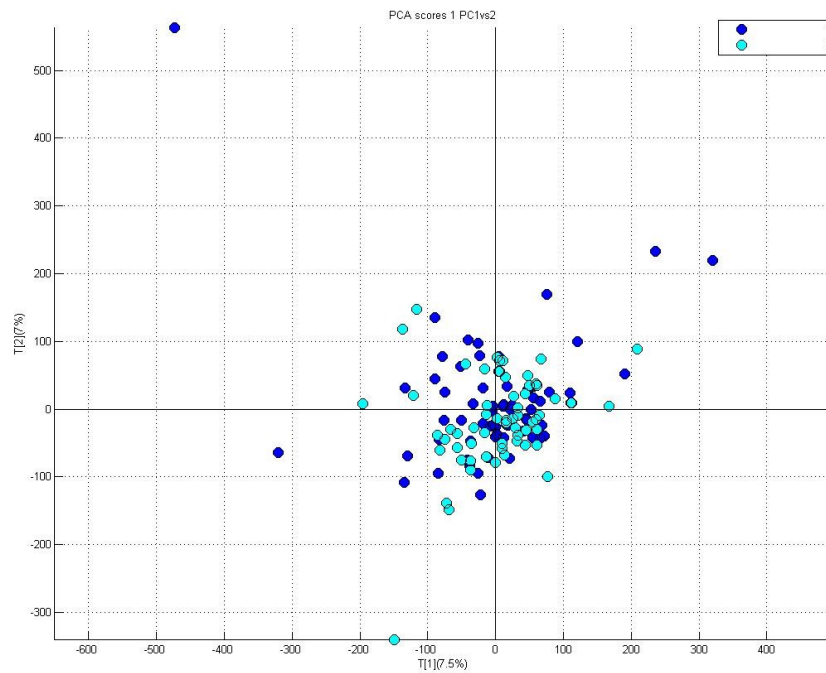


Figure 83: PCA Scores Scatter Plot of NMR spectra of the last urine samples collected. Samples are coloured according to concentration of amino acids in parenteral nutrition. Dark blue (1) represents increasing concentration while light blue (2) corresponds to recommended daily intake.

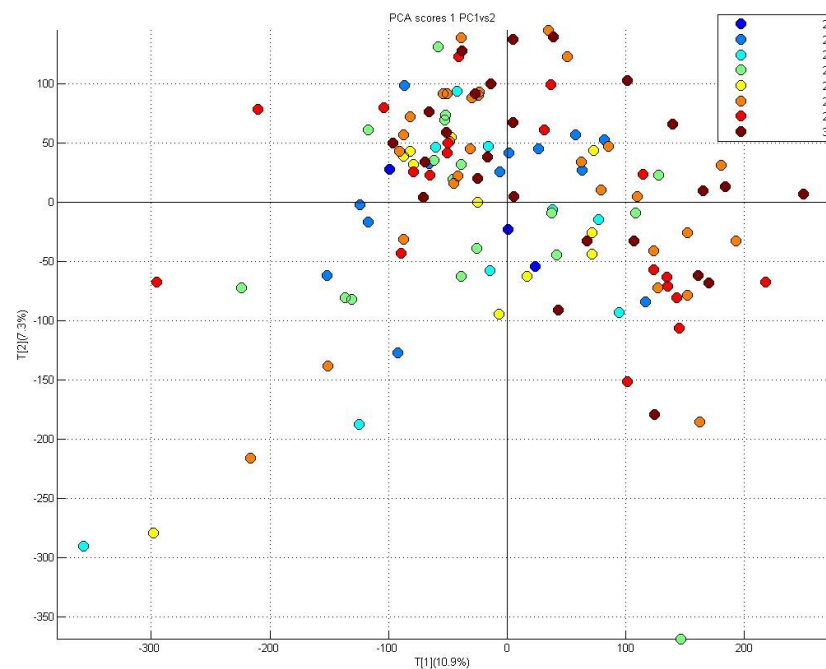


Figure 84: PCA Scores Scatter Plot of NMR spectra of urine samples collected in the 1st week of life. Samples are coloured according to the gestational age at birth (in weeks).

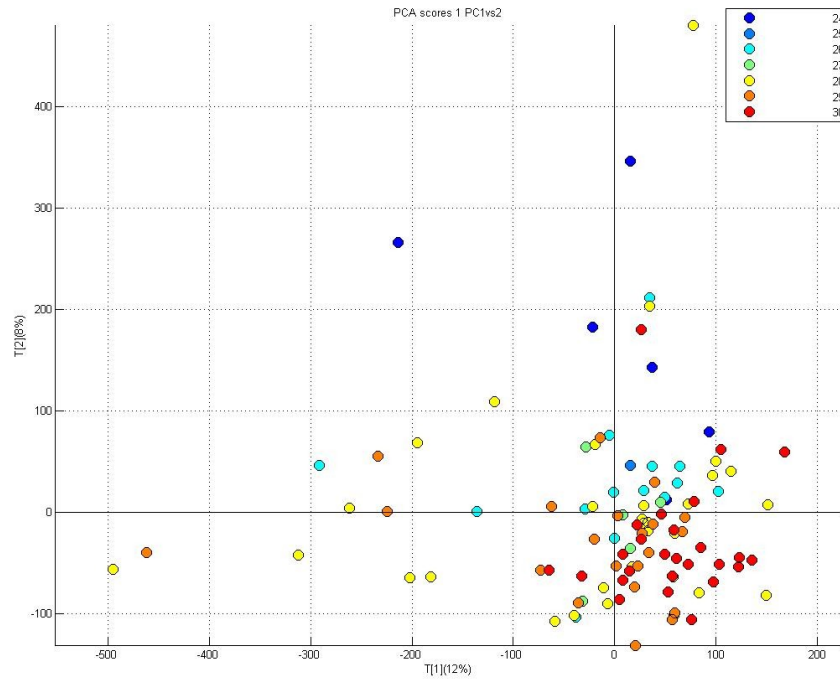


Figure 85: PCA Scores Scatter Plot of NMR spectra of urine samples collected in the 3rd week of life. Samples are coloured according to the gestational age at birth (in weeks).

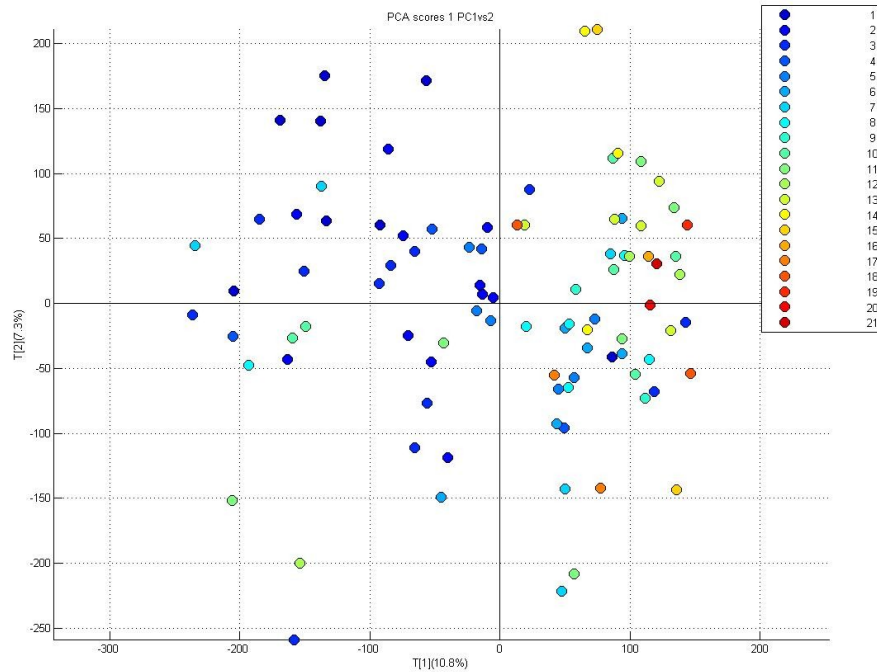


Figure 86: PCA Scores Scatter Plot of NMR spectra of urine samples collected from babies born in 24th week of gestation. Samples are coloured according to the postnatal age given in weeks.

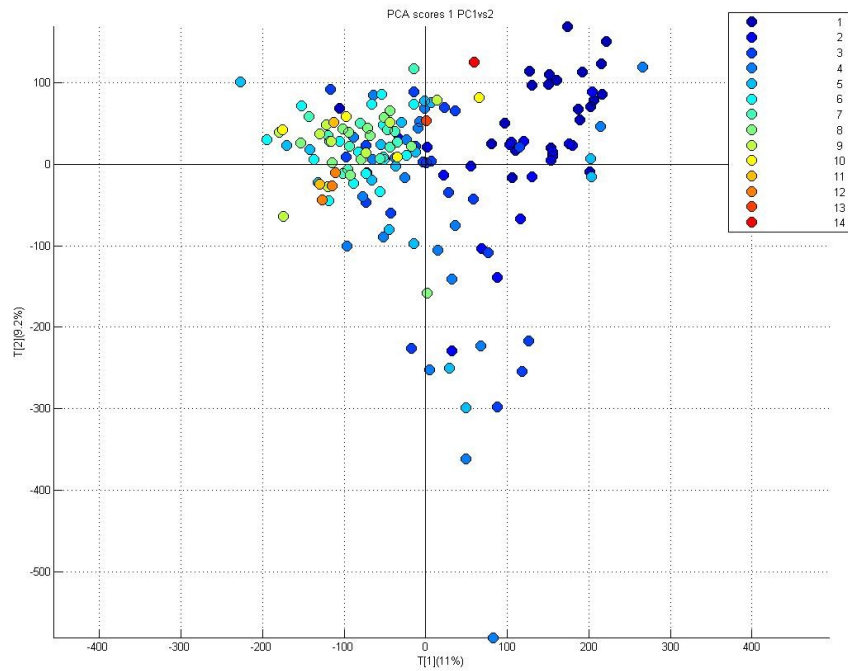


Figure 87: PCA Scores Scatter Plot of NMR spectra of urine samples collected from babies born in 28th week of gestation. Samples are coloured according to the postnatal age given in weeks.

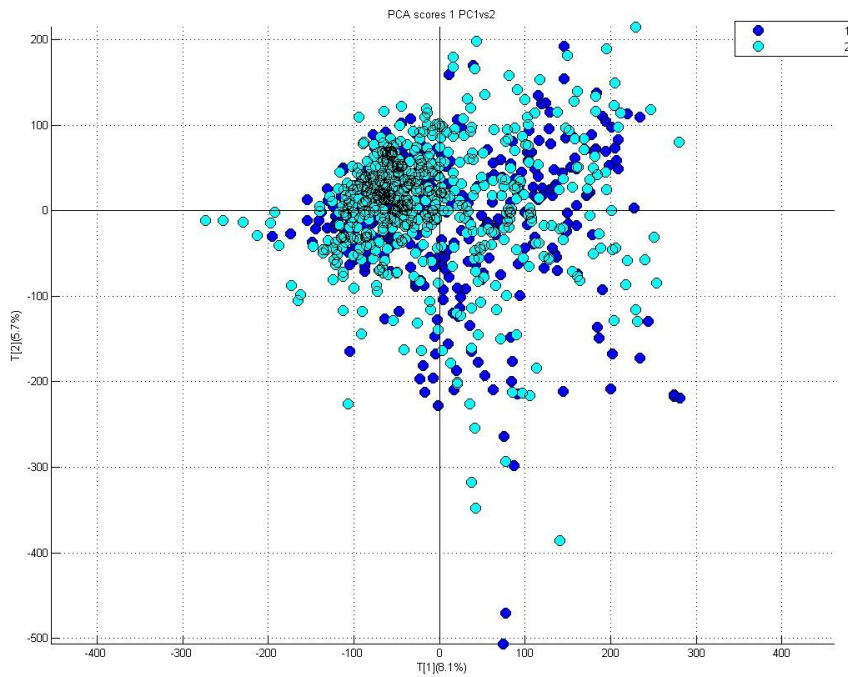


Figure 88: PCA Scores Scatter Plot of whole data set of NMR urine spectra. Samples are coloured according to the sex of the baby with dark blue (1) representing females while light blue (2) males.

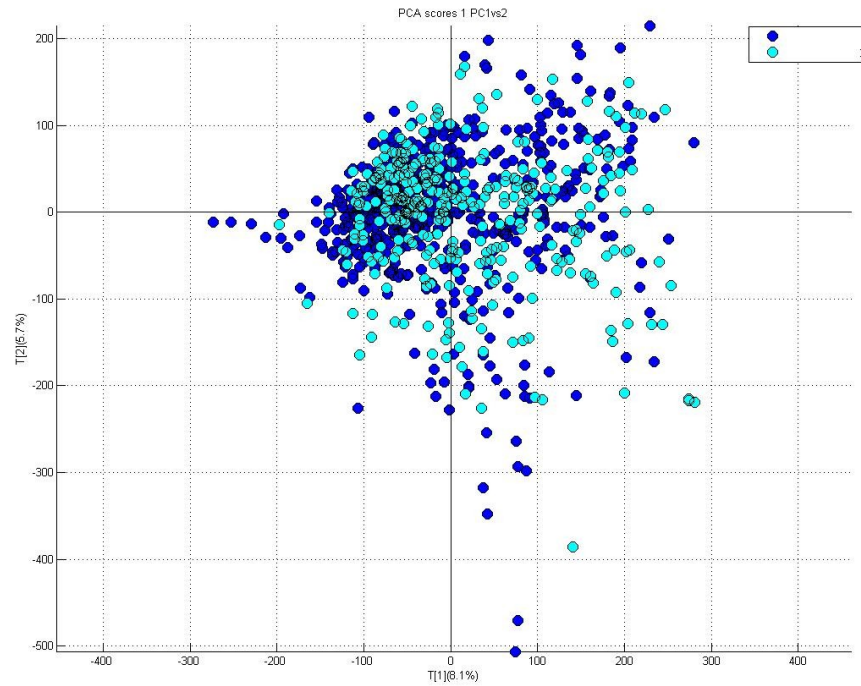


Figure 89: PCA Scores Scatter Plot of whole data set of NMR urine spectra. Samples are coloured according to the mode of delivery with dark blue (1) vaginal delivery while light blue (2) caesarean section.

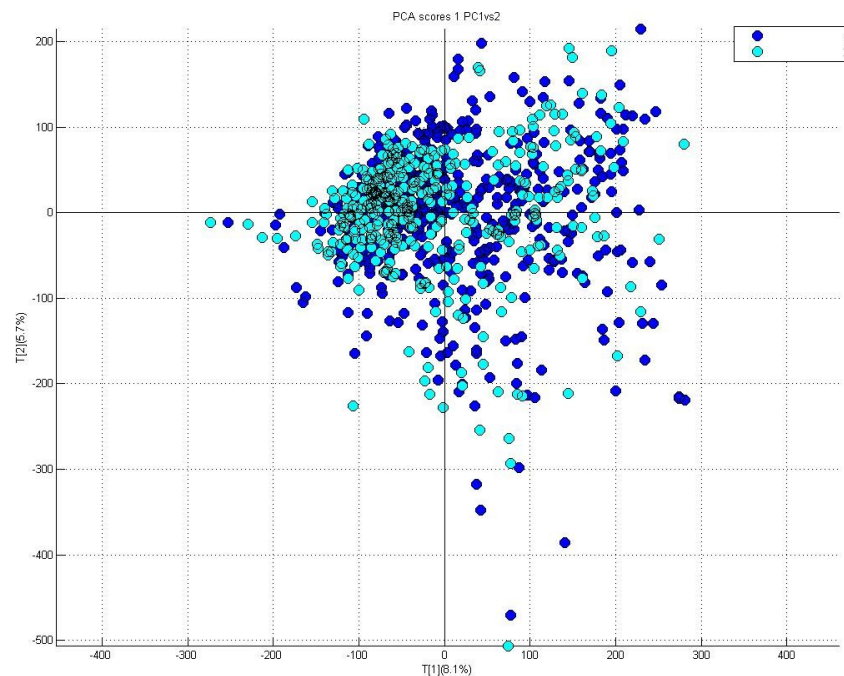


Figure 90: PCA Scores Scatter Plot of whole data set of NMR urine spectra. Samples are coloured according to the presence of labour during delivery with dark blue (1) representing labour while light blue (2) lack of thereof.

Stool Metabonome

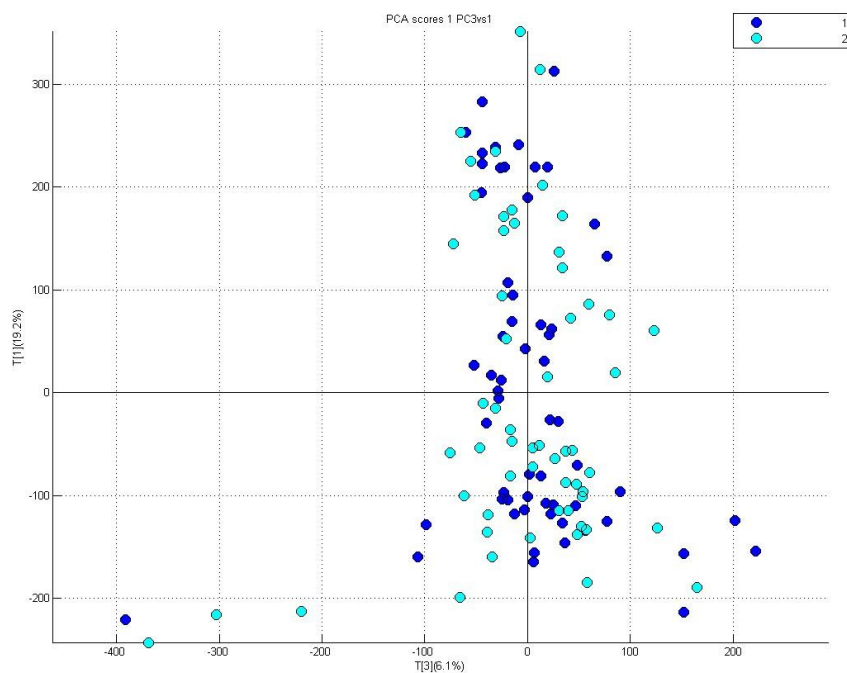


Figure 91: PCA Scores Scatter Plot of NMR spectra of stool samples collected in the 1st week of life. Samples are coloured according to type of intravenous lipid solution. Dark blue (1) represents Intralipid while light blue (2) corresponds to SMOF lipid.

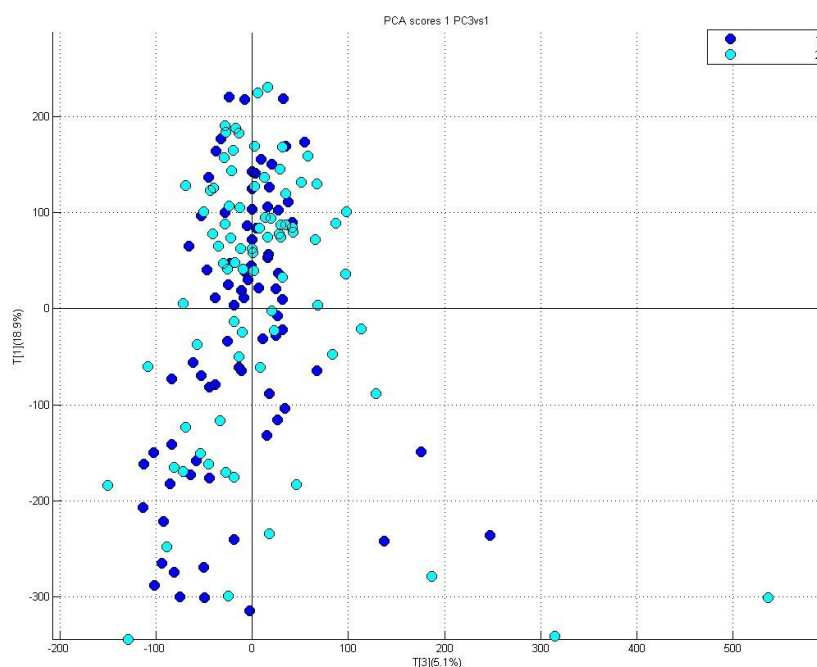


Figure 92: PCA Scores Scatter Plot of NMR spectra of stool samples collected in the 3rd week of life. Samples are coloured according to type of intravenous lipid solution. Dark blue (1) represents Intralipid while light blue (2) corresponds to SMOF lipid.

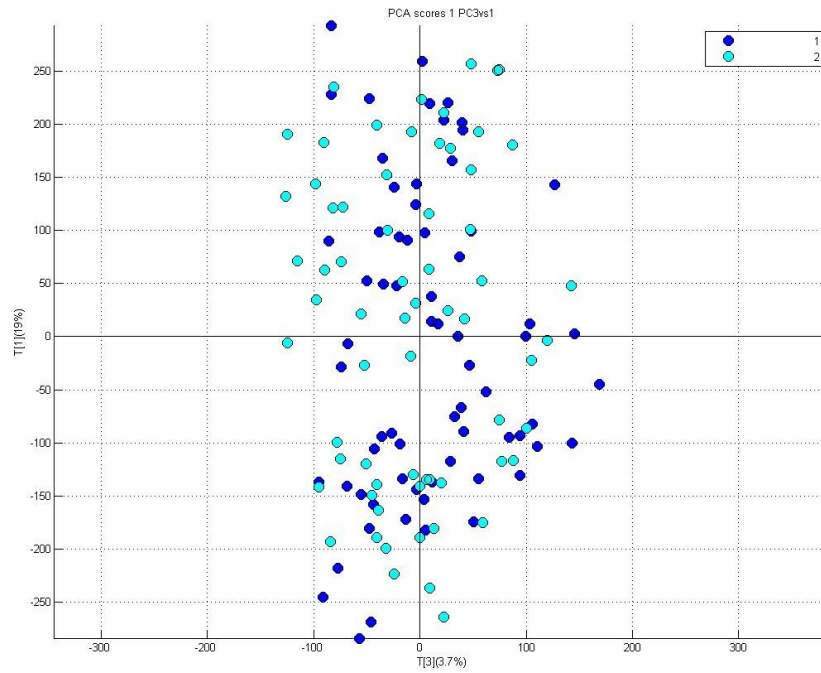


Figure 93: PCA Scores Scatter Plot of NMR spectra of the last stool samples collected. Samples are coloured according to type of intravenous lipid solution. Dark blue (1) represents Intralipid while light blue (2) corresponds to SMOF lipid.

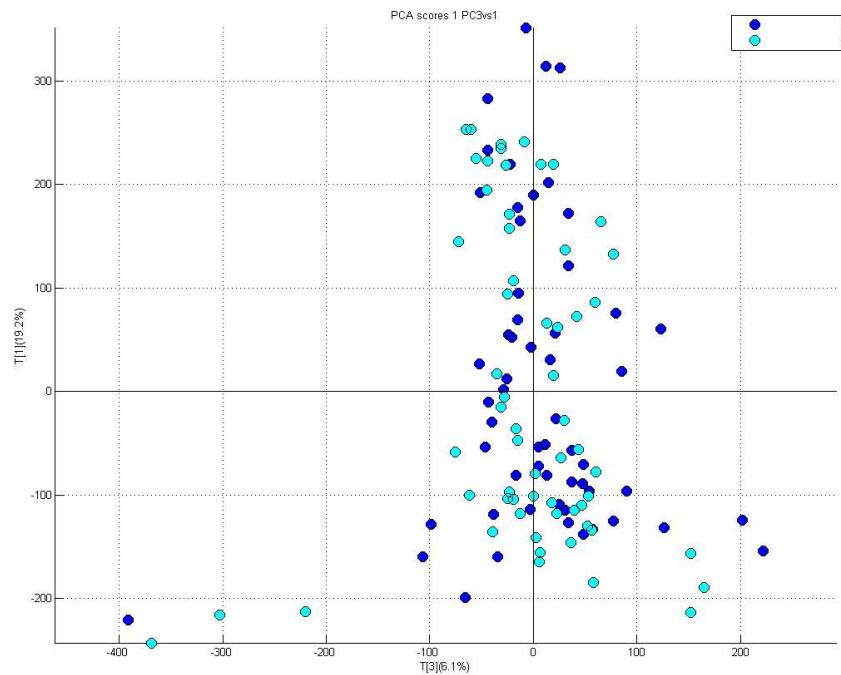


Figure 94: PCA Scores Scatter Plot of NMR spectra of stool samples collected in the 1st week of life. Samples are coloured according to concentration of amino acids in parenteral nutrition. Dark blue (1) represents increasing concentration while light blue (2) corresponds to recommended daily intake.

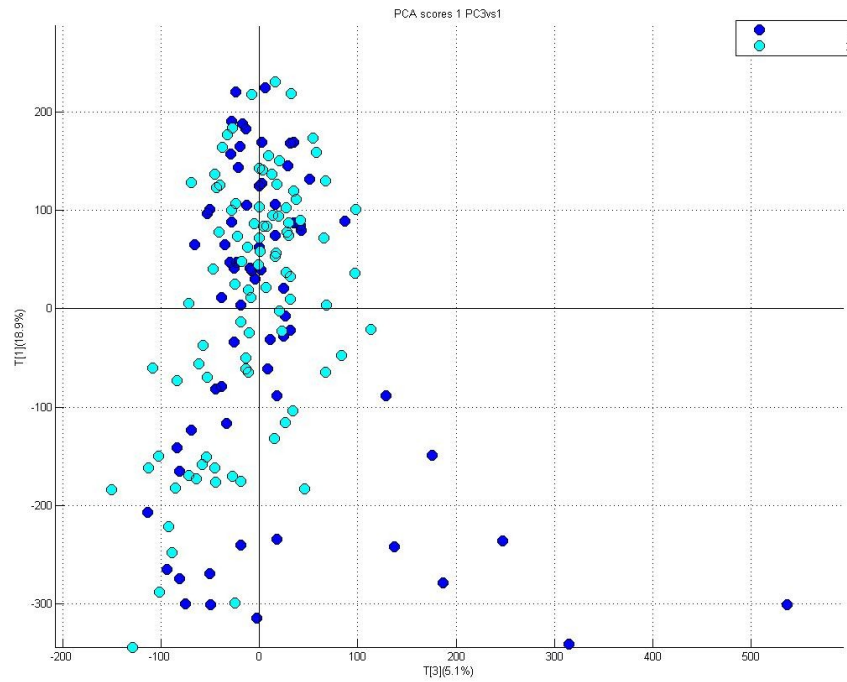


Figure 95: PCA Scores Scatter Plot of NMR spectra of stool samples collected in the 3rd week of life. Samples are coloured according to concentration of amino acids in parenteral nutrition. Dark blue (1) represents increasing concentration while light blue (2) corresponds to recommended daily intake.

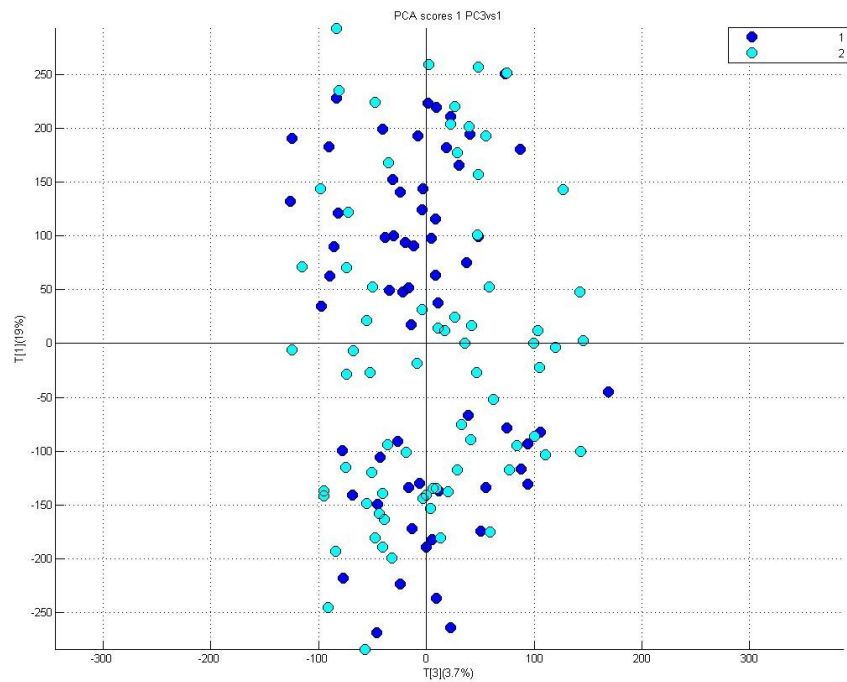


Figure 96: PCA Scores Scatter Plot of NMR spectra of the last stool samples collected. Samples are coloured according to concentration of amino acids in parenteral nutrition. Dark blue (1) represents increasing concentration while light blue (2) corresponds to recommended daily intake.

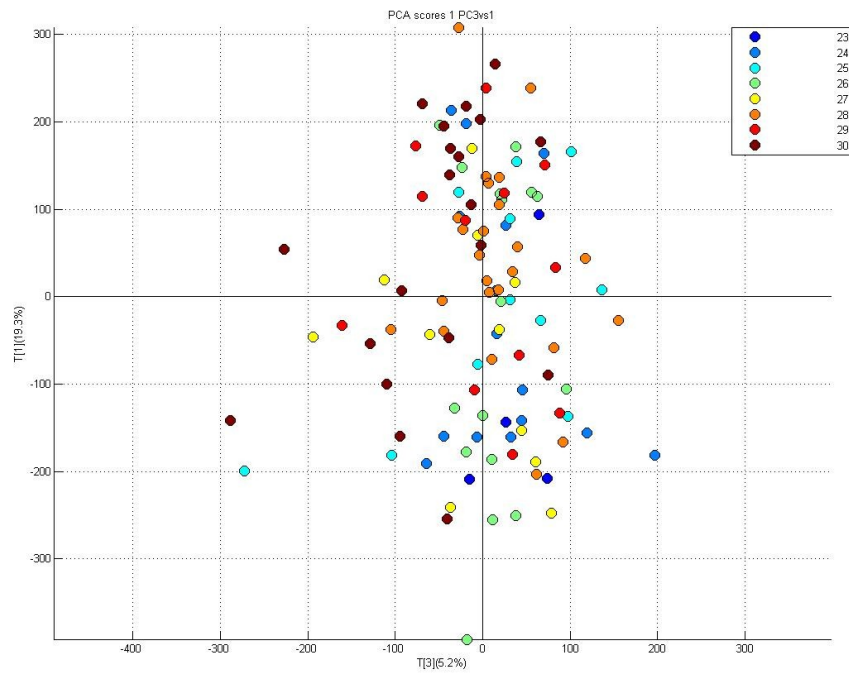


Figure 97: PCA Scores Scatter Plot of NMR spectra of stool samples collected in the 1st week of life. Samples are coloured according to the gestational age at birth (in weeks).

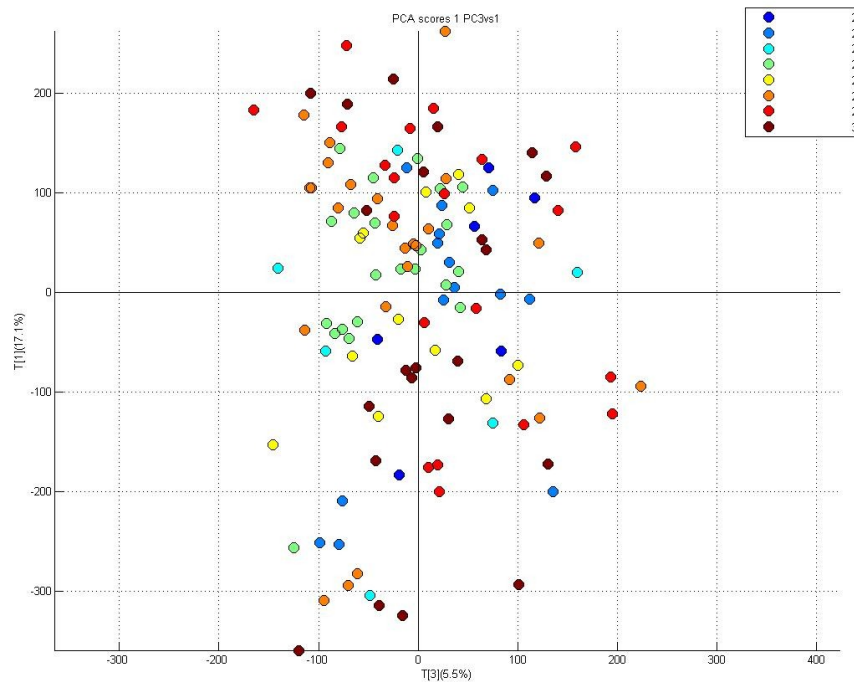


Figure 98: PCA Scores Scatter Plot of NMR spectra of stool samples collected in the 3rd week of life. Samples are coloured according to the gestational age at birth (in weeks).

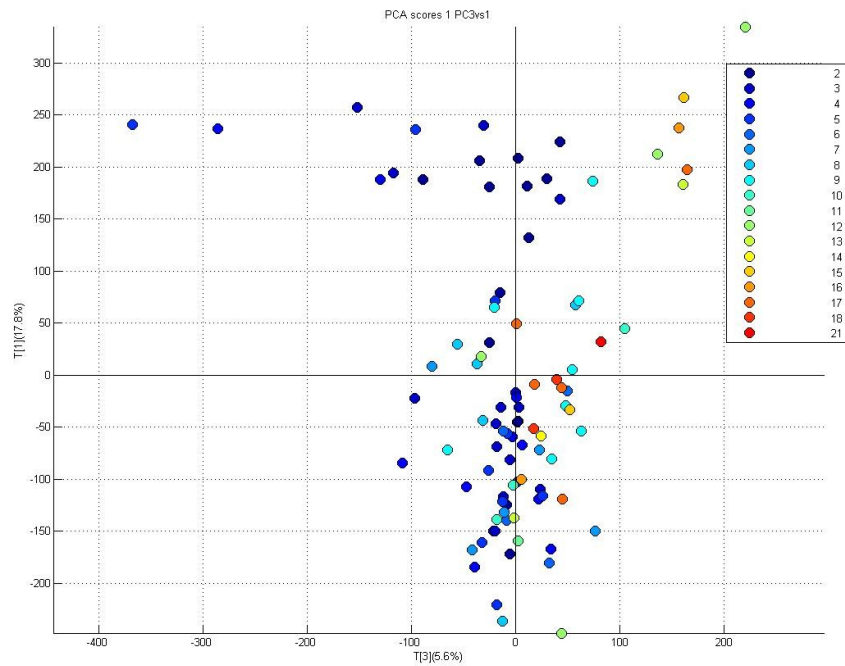


Figure 99: PCA Scores Scatter Plot of NMR spectra of stool samples collected from babies born in 24th week of gestation. Samples are coloured according to the postnatal age given in weeks.

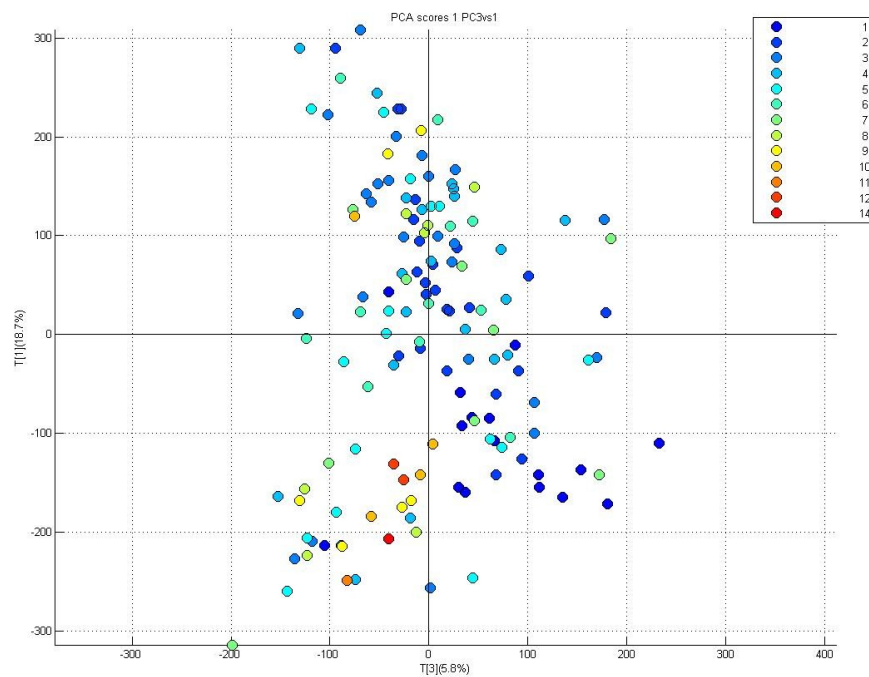


Figure 100: PCA Scores Scatter Plot of NMR spectra of stool samples collected from babies born in 28th week of gestation. Samples are coloured according to the postnatal age given in weeks.

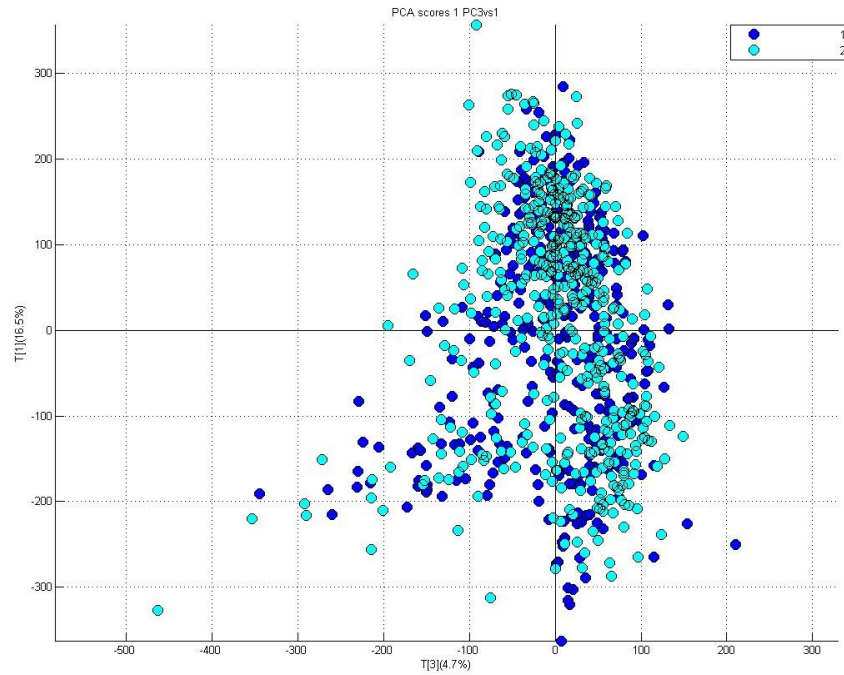


Figure 101: PCA Scores Scatter Plot of whole data set of NMR stool spectra. Samples are coloured according to the sex of the baby with dark blue (1) representing females while light blue (2) males.

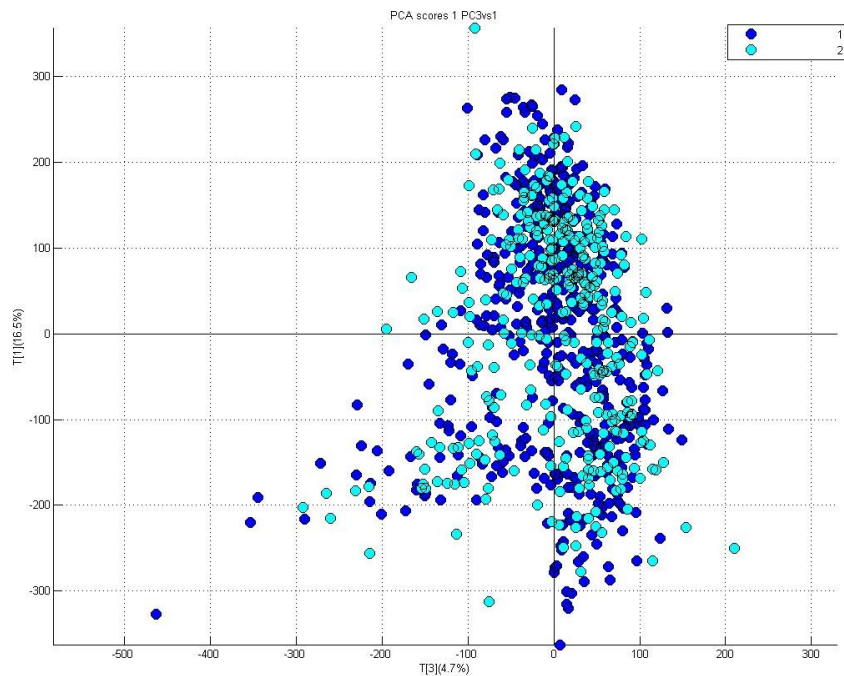


Figure 102: PCA Scores Scatter Plot of whole data set of NMR stool spectra. Samples are coloured according to the mode of delivery with dark blue (1) vaginal delivery while light blue (2) caesarean section.

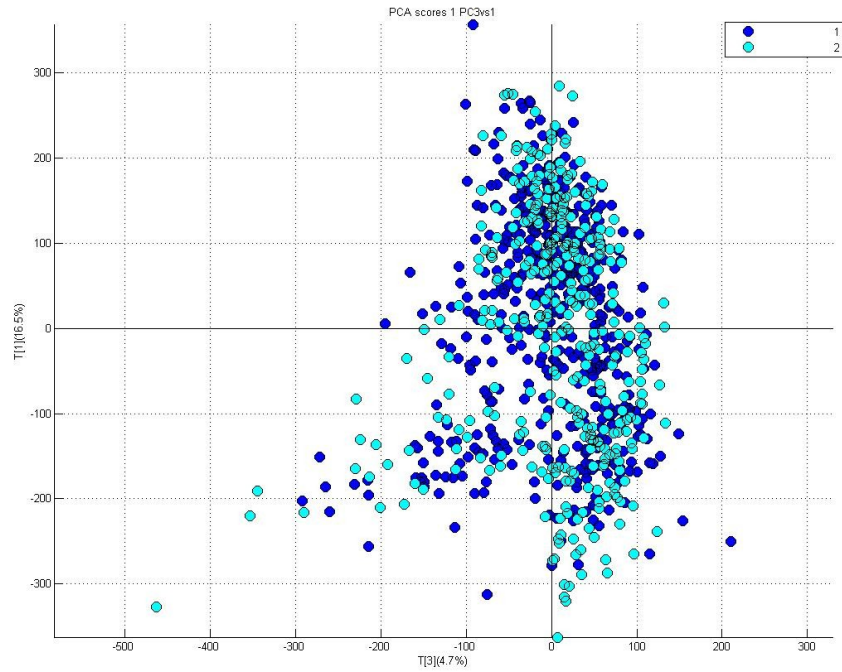


Figure 103: PCA Scores Scatter Plot of whole data set of NMR stool spectra. Samples are coloured according to the presence of labour during delivery with dark blue (1) representing labour while light blue (2) lack thereof.