THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Maternal and neonatal metabolomes and their associations to immune maturation and allergy in early life

OLLE HARTVIGSSON

Food and nutrition Science
Department of Biology and Biological Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2022

Maternal and neonatal metabolomes and their associations to immune maturation and allergy in early life

Olle Hartvigsson

ISBN: 978-91-7905-713-8

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 5179
ISSN 0346-718X

Chalmers University of Technology

Dept. of Biology and Biological Engineering

Food and Nutrition Science

SE-412 96 GOTHENBURG

Telephone: + 46 (0) 31 772 10 00

Cover: Illustrated by Olof Svanberg.

Printed by Chalmers Digitaltryck
Gothenburg, Sweden 2022

Maternal and neonatal metabolomes and their associations to immune maturation and allergy in early life

OLLE HARTVIGSSON

Department of Biology and Biological Engineering
Chalmers University of Technology, Gothenburg, Sweden

Abstract

Allergy, one of the most common chronic diseases worldwide, is caused by a dysregulated immune system reacting to normally harmless proteins. However, regulating mechanisms are not well understood. The aim of this thesis was to investigate if plasma and placenta metabolites associate prospectively to allergy development and immune maturation. The aim was further to explore differences between arterial and venous umbilical cord blood metabolomes, and if they associated with maternal or infant traits.

Placentas and plasma (maternal from pregnancy and delivery and from the umbilical cord) were obtained from the prospective NICE-cohort. Metabolites were measured by LC-MS and GC-MS.

None of the measured metabolomes associated with any of the investigated allergic outcomes (asthma, food allergy and eczema). Modest associations were observed between immune maturation (in particular memory B cells) and plasma and placenta metabolomes. Energy-related metabolites were higher in arterial cord blood, while amino acids were higher in venous cord blood. Amino acid and energy metabolites were higher in first born children compared to children with older siblings.

Overall, the results suggest that immunomodulatory metabolites might be transferred from mother to child during pregnancy, affecting the future production and maturation of immune cells. Studies involving umbilical cord should consider the differences in arterial and venous cord blood and the association of maternal parity in experimental design and data analysis.

Furthermore, algorithms for real-time quality monitoring in untargeted LC-MS metabolomics were developed to improve quality during data generation. Quality monitoring was based on general metrics (e.g. total intensity and number of peaks) and peak metrics from so-called landmark features (e.g. peak area and noise). Landmark features were discovered, validated and then extracted and used in procedures to automatically discover injections of poor data quality. The developed procedures show great promise for improved data generation in high-throughput metabolomics.

Keywords: Untargeted Metabolomics, LC-MS, plasma, placenta, allergy, immune system, umbilical cord, quality control, QC

List of publications

This thesis is based on work presented in 4 manuscripts, referred to by roman numerals

I. *Olle Hartvigsson, Malin Barman, Hardis Rabe, Anna Sandin, Agnes E. Wold, Carl Brunius & Ann-Sofie Sandberg.* Associations of maternal and infant metabolomes with immune maturation and allergy development at 12 months in the Swedish NICE-cohort. Scientific reports, 2021, 11.1: 1-12

II. *Olle Hartvigsson, Malin Barman, Hardis Rabe, Anna Sandin, Agnes E. Wold, Carl Brunius, Ann-Sofie Sandberg.* Associations of the placental metabolome with immune maturation up to one year of age in the Swedish NICE-cohort. *Manuscript in preparation*

III. *Olle Hartvigsson, Malin Barman, Otto Savolainen, Alastair B. Ross, Anna Sandin, Bo Jacobsson, Agnes E. Wold, Ann-Sofie Sandberg & Carl Brunius.* Differences between Arterial and Venous Umbilical Cord Plasma Metabolome and Association with Parity. Metabolites, 2022, 12.2: 175.

IV. *Olle Hartvigsson, Anton Ribbenstedt, Marina Armeni, Kristian Pirttilä, Rui Zheng, Otto Savolainen & Carl Brunius.* Continuous Quality Monitoring for Non-target MS-based Analysis. *Submitted*

Published papers not included in the thesis

- *Mia Stråvik, Karin Jonsson, Olle Hartvigsson, Anna Sandin, Agnes E. Wold, Ann-Sofie Sandberg & Malin Barman.* Food and nutrient intake during pregnancy in relation to maternal characteristics: Results from the NICE Birth Cohort in Northern Sweden. Nutrients, 2019, 11(7), 1680.

- *Alastair B Ross, Malin Barman, Olle Hartvigsson, Anna-Carin Lundell, Otto Savolainen, Bill Hesselmar, Agnes E. Wold & Ann-Sofie Sandberg.* Umbilical cord blood metabolome differs in relation to delivery mode, birth order and sex, maternal diet and possibly future allergy development in rural children. PloS one, 2021, 16(1), e0242978.

Contribution report

Paper I: Olle Hartvigsson (OH) conducted the laboratory work, pre-processed, analysed and interpreted the data and was responsible for writing the manuscript

Papers II and III: OH shared responsibility for pre-processing of data, was responsible for analysis and interpretations of the data and was responsible for writing the manuscript

Paper IV: OH participated in the conceptualization of the project, shared responsibility on writing algorithms and shared responsibility for writing manuscript and tutorial

Abbreviations

APC	Antigen presenting cell
CTL	Cytotoxic T lymphocyte
GC	Gas chromatography
HILIC	Hydrophilic interaction liquid chromatography
HMDB	Human metabolome database
HN	HILIC negative
HP	HILIC positive
Ig	Immunoglobulin
IPO	Isotopologue Parameter Optimization
KREC	κ-deleting recombination excision circle
LaMa	Landmark feature
LC	Liquid chromatography
MoNA	Massbank of North America
MS	Mass spectrometry
NMR	Nuclear magnetic resonance
PAG	Phenylacetylglutamine
PLS	Partial least squares
RF	Random Forest
PAMP	Pathogen associated molecular pattern
QA	Quality assurance
QC	Quality control
QTOF	Quadrupole/time-of-flight
RN	Reversed phase negative
RP	Reversed phase positive
SST	System suitability test
TIC	Total ion count
TOF	Time-of-flight
TREC	T cell receptor excision circle

Table of Contents

1 INTRODUCTION

Allergy is one of the most common chronic diseases in the western world, affecting between 10 - 40 % of the population in any country (1). Allergy is an umbrella term for several diseases including asthma, atopic eczema, rhinitis, conjunctivitis, and food allergy. These diseases are caused by a hypersensitive reaction by the immune system to normally harmless antigens, such as pollen, food or animal protein, commonly referred to as allergens (1). When the immune cells in tissues encounter these allergens, symptoms such as wheeze, dry and itchy skin, urticaria, swollen red eyes, anaphylaxis may occur (2). In addition to being a major issue for the individual, allergy is also costly from a societal perspective, since a large number of hospitalizations occur each year due to allergies. In 2014, avoidable costs related to allergies within the EU alone were estimated to range between a staggering €55 - €151 billion annually (3).

The leading hypothesis regarding allergy development is the so-called hygiene hypothesis (4). This states that the increased cleanliness of the Western world leads to a less stimulated and immature immune system prone to respond to harmless antigens that should normally be tolerated. Despite being well characterized disease conditions, with a well understood pathogenesis, much remains hidden as to the mechanisms by which allergies are developed.

Metabolomics is the study of small molecules in biological samples (e.g., blood, placenta or faeces), and is the 'omics'-discipline that is closest to the molecular phenotype. It has therefore been used for a wide range of applications, such as to analyse effects of exercise (5, 6), diet (7), pollution (8) and drug exposure (9). Metabolomics has further been extensively used to gain more in-depth knowledge of a variety of non-communicable diseases such as diabetes (10), cancer (11), cardiovascular diseases (12), and Alzheimer's disease (13).

Metabolomics is frequently divided into targeted and untargeted metabolomics. Targeted metabolomics is hypothesis driven, where the researcher has an idea of what metabolites to investigate and uses a method for the detection of tens up to a few hundred metabolites (14). Untargeted metabolomics on the other hand, rather focuses on being hypothesis generating, and aims to characterize all metabolites (biologically active molecules <1500 Da) in any given sample.

Metabolomics is heavily reliant on the use of analytical chemistry methods and the main methods are nuclear magnetic resonance (NMR) and mass spectrometry (MS). NMR-based metabolomics is highly reproducible and does not cause sample depletion. However, NMR has a large limitation in its metabolite coverage, can only pick up signals from molecules that are high in concentration and have issues relating to overlapping signals. MS-based metabolomics on the other hand is affected by reproducibility within and between labs and requires extensive sample preparation. On the other hand, it also provides a wider metabolite coverage than NMR. MS-based metabolomics is usually conducted using separation (e.g., chromatography or capillary electrophoresis) as a front end to the MS. The chromatography enables separation of molecules based on their physio-chemical properties and thus increases the resolution of

metabolic features significantly. The two most commonly used chromatographic methods in metabolomics are gas chromatography (GC) and liquid chromatography (LC). Advances over the past two decades in instrumentation have made LC-MS based metabolomics the dominating technique for metabolite profiling by enabling analysis of thousands of features simultaneously (15, 16). Although untargeted LC-MS based metabolomics is a well-established technique, it still has issues relating to inherent variation and instrument malfunction. Labs throughout the world are dealing with these issues in different ways (17) and these methodologies are often laborious for the operator and cost considerable resources.

Metabolomics has previously been used to study manifest allergy: e.g., in a study using exhaled breath condensate a classification rate of 81 % was achieved using NMR based metabolomics on healthy and asthmatic children (18). Another study found a significant decrease in sphingolipid levels in children with prevalent food allergy compared to healthy controls (19). While these studies have pointed towards aiding in assessing disease severity and diagnosis, they were not designed to provide insight on the causal effects leading to disease incidence. Instead, prospective studies where samples are acquired before disease onset hold higher potential for studying the origin of the disease. Recently, prospective studies have been published showing e.g., that decreased concentrations of unconjugated bilirubin in plasma at one year of age inversely associated with asthma at the age of six (20). Moreover, the coffee related metabolites: caffeine, theophylline, trigonelline, quinate, and 3-hydroxypyridine sulphate in maternal pregnancy metabolome or the umbilical cord metabolome were inversely associated with future asthma development (21). Several metabolites relating to steroid metabolism was further associated with an increased risk of asthma (21). However, these associations were moderate and require replication in independent cohorts.

2 AIMS & OBJECTIVES

The overall objective of this thesis work was to identify associations of metabolites before and during birth with future allergy development and immune maturation, identify key aspects affecting the umbilical cord blood metabolome and to develop procedures for automatic quality assessment of injections using untargeted LC-MS based metabolomics. The specific aims were to:

- Identify associations of future allergy development with metabolites in maternal blood during pregnancy and at delivery and infant blood at delivery. (Paper I)

- Identify associations between metabolites in umbilical cord blood, maternal blood and placenta with immune maturation in children during the first year of life (Papers I and II)

- Investigate how the sampling of umbilical cord blood (arterial vs venous) contributes to systematic differences in the measurable metabolome and their association to maternal and neonatal factors (Paper III)

- Improve quality control and assessment in untargeted LC-MS based metabolomics. (Paper IV)

3 BACKGROUND

The Immune System

The purpose of the immune system is to protect the organism from external and internal threats such as pathogens and cancer by distinguishing them from the organisms' own healthy cells. It consists of two parts, the innate, or non-specific immune system which responds quickly to intrusions in a repeatable fashion and the adaptive immune system which upon first exposure to the pathogen takes a longer time but can generate a more specific response to the allergen (22).

The innate immune system consists of physical barriers such as: the skin and various mucosal layers; the complement system which consists of several proteins that aid in the neutralization of pathogens by either marking them for destruction by other cells or causing the pathogen to undergo lysis; and certain leukocytes such as macrophages, dendritic cells, and mast cells. Macrophages, and to a lesser extent dendritic cells, are phagocytic cells which can engulf and lyse pathogens. Dendritic cells are also the main antigen presenting cells (APCs), which can react to pathogen associated molecular patterns (PAMPs) (23), e.g., glycoproteins from bacterial cell walls. When a PAMP is encountered, the APC engulfs the pathogen and can transport parts of the pathogens to the lymph nodes where it presents them to the cells involved in the adaptive immune system. Mast cells are covered in so-called IgE-antibodies and become activated when they encounter antigens matching the antibodies on its surface, whereupon histamine is released (24). The histamine, in turn, causes blood vessels to expand and blood to leak into the tissue, causing redness and swelling, as is commonly seen during an allergic reaction.

The adaptive immune system consists of lymphocytes, that include two main cell types: T and B cells. Both precursor T and B cells are produced in the bone marrow. Precursor T cells have to travel to the thymus where they mature into naïve T cells. These cell types have receptors on their surface that are each specific to one unique antigen. T cells can be further subdivided into categories; helper $CD4^+$ T cells (T_h cells), cytotoxic $CD8^+$ T lymphocytes (CTLs) and regulatory T cells (T_{regs}). As the name suggests, T_h cells help other cells in fighting pathogens, e.g., by activating B cells, thereby aiding in generating an inflammatory response. CTLs aids in the killing of the host's own cells that show foreign markers due to viruses or tumours. The main function of T_{regs} is to regulate and hinder an overactive immune response, by dampening the activity of T cells towards the end of a response, or by supressing autoreactive immune cells (i.e., cell that are reacting to the body's own cells).

When an APC from the innate immune system identifies and presents an antigen that corresponds to one of the specific T cell receptors present on the T cell surface, the T cell becomes activated and undergoes clonal expansion, i.e., making it rapidly proliferate and multiply itself (25). B cells in most cases require both its specific antigen and aid from T_h cells to become activated (26). They then undergo clonal expansion and differentiate into plasma cells. The plasma cells, in turn, start to produce immunoglobulins that enter the blood stream and aid in the removal of the pathogens (Fig. 1).

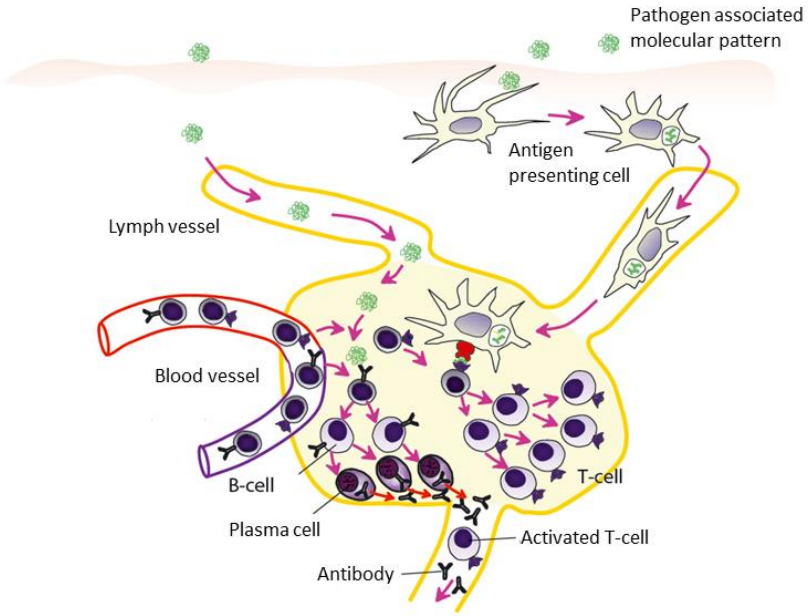

***Figure 1**. Activation of the adaptive immune system. A pathogen-associated molecular pattern is engulfed by an antigen presenting cell. The antigen presenting cell brings the antigen to the lymph node via lymph vessels where it presents the antigen to T cells. Antigens can further travel to the lymph node by themselves and there, bind to B cell receptors leading to B cell activation. Upon activation, both T and B cells proliferate and are transported to the rest of the body where they can aid in the immune response, by antibody production, cytotoxic activity, and cytokine release. Figure adapted from Inflammationsjukdommar by Agnes Wold and Johan Mölne (2).*

The process from infection to adaptive immune system response may take up to two weeks. Upon activation, some of the adaptive immune cells differentiate into memory cells that can stay in the body for months up to years after infection. This process ensures that the immune system can respond faster if the same antigen is encountered again, thus also ensuring that symptoms become significantly milder or even non-noticeable. Plasma cells are mainly responsible for producing antibodies, or immunoglobulins. Antibodies are proteins that, similar to the T cell receptors, have a unique specificity and affinity for a molecular pattern (antigen). There are five types of immunoglobulins, (IgD, IgM, IgG, IgA and IgE) and different types of B cells release different types of immunoglobulins. However, unlike T cells, a B cell can switch class, which enables it to change which immunoglobulin type it produces. IgE-antibodies is a particular class of antibody, exclusively found in mammals, and are related to allergy and the associated histamine release (27).

The types and number of T and B cells can be measured based on their cell surface markers using flow cytometry (Fig. 2). In flow cytometry, a cell suspension is mixed with an antibody cocktail, containing antibodies specific to various regions expressed on the cell surface. These antibodies are labelled with a fluorophore that can be detected using fluorescence spectrometry. The cell suspension is then passed through a narrow path where only one cell fits at a time and the excitation and emission wavelengths of each cell is measured. Using flow cytometry, several surface markers on immune cells can be used to identify what type each cell is, and their concentration in the original sample.

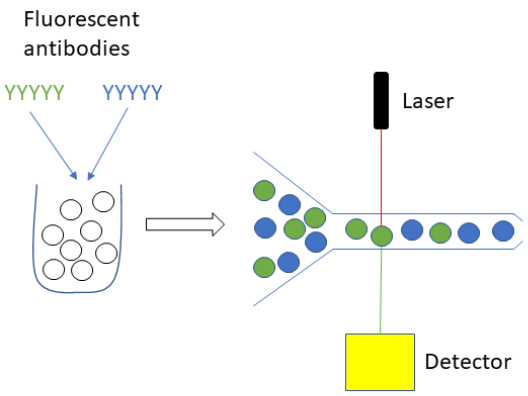

***Figure 2**. Schematic figure of flow cytometry. Cells are mixed with fluorescent antibodies, then forced through a narrow path where a laser excites fluorophores, and a detector measures the emitted and scattered light.*

Another measure for studying T and B cells is through T cell Receptor Excision Circles (TREC) and Kappa-deleting Recombination Excision Circles (KREC) (28). These are in essence a measure of newly formed T and B cells, respectively. A TREC is a small circular DNA fragment, that is created when genes in the T cell are rearranged in the thymus during its maturation in order to create its antigen specificity (29). When a T cell is cloned, however, the TREC is only kept in one of the T cells. Measuring the number of TRECs present in a blood sample therefore gives information about how many T cells are being created rather than cloned. Similar to TRECs, KRECs are created when newly created B cells undergo gene shuffling in the bone marrow in order to gain its antigen specificity and does not multiply when a cell clones itself (30) (Fig. 3). KRECs are therefore used to estimate B cell production. As the numbers of cells expressing either TREC or KREC are diluted during proliferation, low levels of TREC and KREC can also be a sign of high proliferation in T and B cells

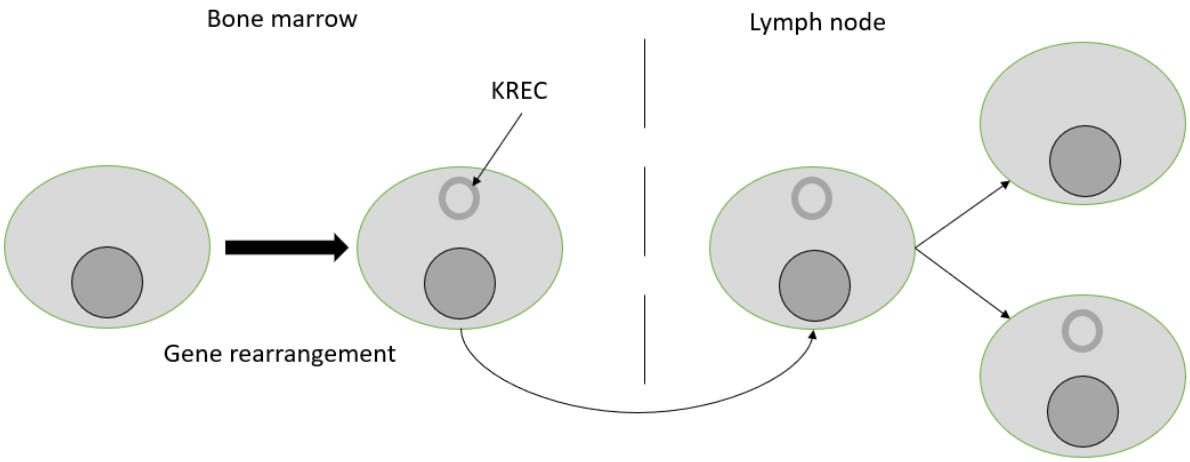

***Figure 3**. Schematic picture showing KREC formation during gene rearrangement in the bone marrow and how the KREC is only kept in one cell after cloning of itself.*

Allergy and allergy development

The immune system is exceptionally well constructed to manage potentially harmful insults. However, the immune system may become dysregulated and activated by otherwise harmless proteins (allergens), such as milk protein, pollen, or animal dander. The allergic disease progression starts with the cells in the adaptive immune system getting activated by an allergen leading to the creation of antibodies specific to the allergen. This process is known as sensitization (2). Sensitization is a prerequisite for subsequent IgE-mediated allergy, although not all people who are sensitized develop allergies.

IgE-mediated allergy is the most common form of allergy (31), where mast cells in local tissues (e.g., air ducts) are coated with IgE antibodies specific to such antigens, causing histamine release and associated inflammatory symptoms upon exposure (27). Allergy is classified into categories, depending on what symptoms are presented, or what causes the symptoms, with the most common allergic diseases in children being atopic dermatitis, food allergies, and asthma. Atopic dermatitis, or atopic eczema, is very prevalent in young children (up to 20 % in school children (32)) and affected children will often develop hay fever or asthma later in life, a phenomenon known as the atopic march (33). The cause of atopic dermatitis is unknown, but a combination of genetic factors and environmental exposures are likely (34). Food allergy is classified by the specific food compounds triggering the allergic reaction. Symptoms within this class can differ from stomach pains to severe anaphylaxis (35). Asthma is often concomitant with other allergic diseases, such as pollen allergy or eczema, and the definition of allergic asthma is in fact partially dependent on the presence of other allergic diseases (36-38).

Although a dysregulated immune system is at the core of allergy, the aetiology is still unclear. It has been hypothesized that the increase of allergy in the western world is due to an increased cleanliness and therefore less exposure to microorganisms and other stimuli the immune system can react to. This is known as the ‘hygiene hypothesis’ and was first stipulated by Dr. David Strachan 1989 (4). The hygiene hypothesis is further supported by studies showing that children growing up on farms have less allergy than both their rural and city dwelling counterparts (39, 40). Pet ownership is another factor that seems to play a role in the future disease risk (41, 42). Studies have also shown that genetics plays a crucial role in the development of the diseases (43-47) and that the maternal heredity appears to have a stronger impact than the paternal on asthma development (48, 49). Several studies point towards food during early life having an influence on disease onset where consumption of butter, dairy products, and fish (50, 51) seem protective while consumption of margarine and oils is associated with increased risk of allergy development (39, 52). Additionally, there is evidence that food consumed by the mother during pregnancy might have an impact on the future well-being of the child (53, 54), although the evidence for this are not conclusive (55).

The Metabolome and Metabolomics

Metabolites are small molecules (<1500 Da) that are involved in the metabolism (56). Unlike genes and proteins, metabolites are subject to rapid changes throughout the day and thus provide the closest representation of the molecular phenotype (56). The metabolome is a term that encompasses all metabolites in a given biological sample (57). Along these lines, studying the metabolome offers the potential to elucidate mechanistic effects of e.g., a specific medication or diet, or to aid in disease diagnosis (58). Metabolomics aims to measure the entire or parts of the metabolome either quantitively or qualitatively. Metabolomics is consequently last in the line of the major 'omics' methodologies, preceded by genomics, transcriptomics, and proteomics (Fig. 4).

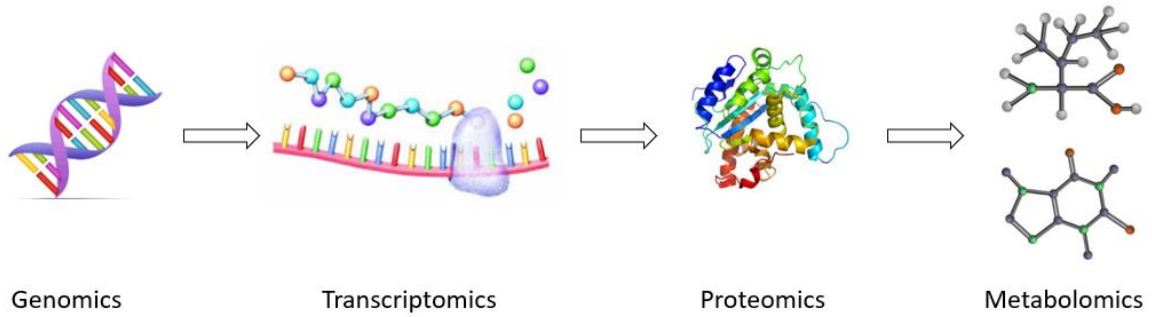

***Figure 4**. The four main 'omics': genomics, transcriptomics, proteomics, and metabolomics. Genomics describes the measurement of the genes present in DNA, transcriptomics the abundance of genes transcribed into RNA, proteomics the abundance of proteins that are created based on the RNA and metabolomics measures metabolites, stemming from processes catalysed by proteins (enzymes) and exposures.*

Advances in analytical chemistry techniques, such as NMR or MS, have facilitated the metabolomics research field to provide an overview of the measurable metabolome. The study of the metabolome has been widely used the past two decades and has been successfully applied to investigate many non-communicable diseases such as cancer (59-61), diabetes (10, 62, 63), and cardiovascular disease (64, 65). Metabolomics has further been used in the study of effect of exercise (5, 6), pollutant exposures (8, 66), and food on the molecular phenotype (67). It has also been applied to identify biomarkers of dietary exposures, such as rye (68, 69), citrus fruit (70-74), vegetables (7, 75, 76), and coffee (7).

Metabolomics can be divided into targeted and untargeted approaches (56): Targeted metabolomics is a hypothesis driven research field, where the aim is to quantitively determine the concentrations of a pre-determined set of metabolites, normally corresponding to a research question or hypothesis (77). The set of metabolites could for instance be based on previous studies which have shown associations to the outcome of interest, or based on available assays, such as the ones available via commercial actors such as Nightingale Health (https://nightingalehealth.com) and Biocrates (https://biocrates.com). The set of metabolites could also be determined from available in-house assays.

In contrast to targeted methods, in non-targeted methods, the peak height or area is calculated without comparing to any internal standards and therefore feature abundance is compared qualitatively, thus, untargeted metabolomics is predominantly hypothesis-generating (78).

Using this methodology, the aim is usually to identify new associations or mechanisms linking metabolites to exposure or health status. As these associations are often hypothesis-agnostic, this is performed by covering as large portion as possible of the measurable metabolome with little or no *a priori* assumptions on what to look for.

Non targeted profiling is data driven and is often reliant on machine learning techniques for data analysis. However, when attempting to maximize coverage, measured chemical entities (analytes) are frequently, or even predominantly, of unknown identity. The identification of molecular features of interest (i.e., associated with the research question) is thus required. However, this is also very resource-demanding and is widely considered one of the major bottlenecks within the metabolomics community to date (79).

Regardless of the choice between targeted or untargeted approaches, metabolomics can be performed using a variety of techniques, where the most common are based on NMR or MS. NMR is a spectroscopy-based method that measures the magnetic resonance at so-called chemical shifts (i.e. a quality of the atom nucleus, which changes based on proximity to other atoms) (80). The most common NMR methodology in metabolomics is ^{1}H-NMR, where the resonance of hydrogen atoms within molecules is measured quantitatively and shifts compared against reference values for identification. MS on the other hand measures the mass-to-charge ratio (m/z) of analytes (e.g., using quadrupoles or time-of-flight mass analysers). Metabolites are identified by comparing against reference databases and quantification is achievable by precise calibration for each metabolite. Alternatively, relative quantification is possible by comparing peak areas of the same analyte over different samples, without using a calibration curve. MS methods are most frequently coupled to a chromatographic separation step before mass analysis, most frequently GC or LC, both of which were used in this thesis. NMR has the advantages of an easier sample preparation, that it is non-destructive of the samples that absolute quantification of analytes is achievable. The MS based methods have a larger metabolome coverage, often need less sample volume, and can measure analytes present at lower concentrations than NMR-based methods. Below follows a brief introduction of LC-MS and GC-MS.

In both GC-MS and LC-MS, the chromatographic frontend serves to separate molecules based on their physicochemical properties. Consequently, using these hyphenated techniques, metabolites can be identified with an increased accuracy both using m/z and retention time. In chromatographic separation, a sample is transported through a stationary column by either a gas (for GC) or a liquid (for LC) mobile phase. In GC, the column the stationary phase normally consists of a liquid film layer and in LC, it normally consists of solid particles with a wide variety of surface chemistries for different chemical properties. The stationary phase thus retains different analytes for different amounts of time depending on e.g., boiling point (GC) and the size and polarity (LC) of the analyte.

In gas chromatography, analytes must be in gaseous phase. To facilitate this process, a derivatization step is usually required, e.g., using methoxymation or silylation, to make the molecules more volatile. Prior to injection, samples are incrementally heated to convert more analytes into gaseous phase. GCs have a stable retention time profile and due to the implementation of Kovats retention index (81), comparing identification and results between different labs is routinely performed. Furthermore, GC is easy to combine with an MS, as it,

unlike LC, does not require the use of high pressure and high mass flows in the column to obtain separation of analytes.

In LC, the mobile phase is in liquid phase. LC uses much shorter columns and applies a high pressure to force the liquid through the column. The most common LC method used in metabolomics is Ultra High-Performance Liquid Chromatography (UHPLC). UHPLC utilizes densely packed particles (~2µm) in the column which requires very high pressure, usually several hundred bars, to force liquid through the column. The small particle sizes used results in a high surface area of column particles, thus enhancing separation of analytes. The mobile phase is delivered in a gradient that changes the composition over time, in order to change its polarity and thereby cover a wider range of analytes. To further broaden the coverage, different combinations of column and mobile phase are used. The two most common methods are reversed phase and hydrophilic interaction liquid chromatography (HILIC). Reversed phase uses a hydrophobic column and polar solvents to elute analytes primarily of low polarity. HILIC uses a hydrophilic column and even stronger hydrophilic solvents to achieve separation primarily of high polarity.

The most common mass spectrometers used are orbitraps, quadrupoles and time-of-flight (TOF). Prior to entering the MS however, analytes must be ionized in the interface between the chromatographic separation and the MS. The ionization is important as MS measures the mass-to-charge ratio (m/z) of an analyte, and thus requires the analyte to be charged. Normally, soft ionization techniques such as electro spray ionization are employed to minimize fragmentation of the analytes. Quadrupoles consists of four parallel rods that by an alternating current can separate ionized analytes by their m/z ratio. Quadrupoles are highly selective but come with the downside of having a low mass accuracy (approximately 1 Da). Quadrupoles are frequently used either as single quadrupoles, or as triple quadrupoles (QqQ). For QqQ, the first is used to select the appropriate mass, the second is used as a collision cell to fragment the analyte and the third to select the mass fragments created in the collision cell in order to identify and quantitate the metabolite. TOFs work by measuring the time it takes for a molecule to be transported a certain length in an electric field, which correlates with the m/z. Orbitraps on the other hand measures oscillation frequency of analytes which in turn relate to their m/z (82). Quadrupoles are often coupled to TOFs (QTOF), resulting in an instrument that can achieve fragmentation spectra via the quadrupole and have the mass accuracy of a TOF. Both QTOFs and orbitraps are high-resolution mass spectrometers thus their mass accuracy is below 5ppm, and even below 1ppm for orbitraps (83). QTOFs have traditionally had higher acquisition speed than orbitraps making them more suitable when coupled to UHPLC. However, with ongoing instrument development, this distinction is subject to change.

The combination of a chromatograph and a MS gives separation between analytes both regarding their retention time in the column as well as through the specific m/z of the analyte (Fig. 5). Thus, analytes in LC-MS, frequently referred to as features, are attributed with m/z ratio and retention time. To elucidate what metabolite a feature corresponds to, fragmentation of the molecule is employed. This fragmentation is done by combining two MS together, with an intermediate collision chamber and is therefore called MS/MS or MS2. Once a feature is selected by the first MS, the molecule is struck by particles causing it to fragment. The formed

fragments are subsequently measured using the second MS, giving a molecular fingerprint for each molecule. This fingerprint can then be used to match against available libraries (either in-house or from online data bases) and a molecular entity can be assigned to the feature.

For targeted applications, triple quadrupoles are normally used, as they have the highest sensitivity and are, after development of dedicated methods, the most specific (77). For untargeted metabolomics, the most common choices are Orbitrap and QTOF, due to the high acquisition speed and high mass accuracy. In this work, QTOF has been used exclusively.

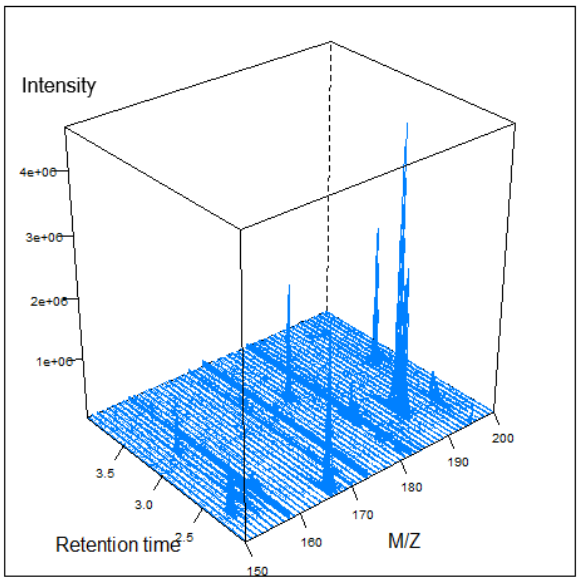

***Figure 5**. Three-dimensional representation of how analytes from one sample are separated both in retention time and m/z domain.*

As the aim of untargeted metabolomics is to cover as large part as possible of the metabolome, different combinations of LC and MS are frequently used. Reversed phase and HILIC are frequently used in combination to cover less and more polar analytes, respectively. Further, as some analytes are more prone to being negatively or positively charged, depending on their electron configuration, both positive and negative ionization are usually applied. This leads to a total of four different analytical modes for each sample: reversed phase positive (RP), reversed phase negative (RN), HILIC positive (HP), and HILIC negative (HN), although more options still are available (84).

Metabolomics in allergy

Metabolomics studies in both adults and children have shown molecules in different biospecimen that can distinguish allergic from non-allergic individuals and/or be used as a tool to assess disease severity (18, 85-90). These studies range from using blood or urine as a biospecimen for different allergic diseases, to studies made on tissues or sample matrices specific for the disease. Carraro et al. studied the metabolomes in exhaled breath condensate (EBC) in relation to asthma and found a higher presence of oxidized compounds in asthmatic children compared to healthy controls (18). The same group further observed that there was a clear association between the exhaled molecules and the severity of the disease, with

concentrations of vitamin D, adenosine and retinoic acid-relating compounds as the main drivers of the association (88). In another study, atopic dermatitis was studied by comparing metabolites measured in skin samples with or without eczema (90). Results indicated that several amino acids and glycerophospholipids were enriched in lesioned skin. A summary of metabolomics studies on allergic diseases is shown in Table 1.

***Table 1**. Studies performed in humans using metabolomics to investigate metabolic differences between healthy subjects and subjects with asthma, food allergy (FA) or atopic dermatitis (AD).*

First Author	Outcome	Subjects	Biospecimen	Method	Main findings
Carraro (2007) (18)	Asthma	17 persistent asthma 8 intermittent asthma 11 controls	Exhaled breath condensate	^{1}HNMR	Presence of acetylated and oxidized compounds in children with asthma
Saude (2011) (91)	Asthma	73 stable asthma 20 unstable asthma 42 controls	Urine	^{1}HNMR	Metabolites related to citric acid cycle, stress on energy metabolism, protein, and amino acids metabolism
Mattarucchi (2012) (92)	Asthma	41 asthma 12 controls	Urine	LC-MS	(↓) Urocanic acid in asthma
Assfalg (2012) (89)	AD	20 AD 12 controls	Urine	^{1}HNMR	(↑) Creatinine, creatine, citrate, formate, 2-hydroxybutyrate, dimethylglycine, and lactate in AD (↓) Betaine, glycine, and alanine in AD
Janssens (2012) (93)	AD	14 controls 14 atopic eczema + FLG mutation 14 atopic eczema - FLG mutation	Non-lesioned skin	LCMS	(↑) very short length ceramides in eczema
Caldeira (2012) (94)	Asthma	32 asthma 27 controls	Exhaled breath	GCxGC MS	(↑) 3,6-dimethyldecane, nonane, 2,2,4,6,6-pentamethylheptane, decane, dodecane, and tetradecane in asthma (↓) nonanal, decanal, and dodecanal in asthma
Carraro (2013) (88)	Asthma	31 non-severe asthma 11 severe asthma 15 controls	Exhaled breath condensate	LCMS	Retinoic acid (↑ in severe vs control) Deoxyadenosine (↑ in severe vs control) Vitamin D (↑ in non-severe asthma vs control)
Ibrahim (2013) (95)	Asthma	79 asthma 34 controls	Exhaled breath condensate	^{1}HNMR	No metabolites identified.
Jung (2013) (85)	Asthma	39 asthmatic 26 controls	Serum	^{1}HNMR	(↑) Methionine, glutamine, and histidine; (↓) Formate, methanol, acetate, choline, O-phosphocholine, arginine, and glucose (↑) VLDL/LDL products in severe asthma patients

Huang (2014) (87)	AD	45 controls 83 AD	Serum	LCMS	(↑) Several unsaturated fatty acids in AD. (↓) Taurine and glycine conjugated bile acids in AD
Checkley (2016) (96)	Asthma	25 asthma overweight 25 asthma normal weight 25 control overweight 25 control normal weight	Serum	LC-MS	(↓) ascorbic acid, 2-isopropylmalic acid, shikimate-3-phosphate, 6-phospho-d-gluconate, and reduced glutathione in asthma
Ottas (2017) (86)	AD	13 AD (untargeted) 15 controls (untargeted) 25 cases (targeted) 24 controls (targeted)	Serum	LCMS	(↑) PC (16:0–16:1) / (14:0–18:1) in AD (↓) PC (16:1/20:4), acetyl carnitine (C2), phosphatidylcholine diacyl C38:5, phosphatidylcholine diacyl C40:5 in AD
Chiu (2018) (97)	Asthma	30 asthma 30 paired controls	Urine	^{1}HNMR	(↓) 1-methylnicotinamide and allantoin in asthma
Carraro (2018) (98)	Asthma	16 transient wheezing 16 early-onset asthma 13 controls	Urine	LC-MS	Discrimination between transient wheezing and early-onset asthma prior to asthma diagnosis
Kelly (2018) (99)	Asthma	46 asthma 197 controls	Plasma	LC-MS	(↑) N1-Methyl-2-pyridone-5-carboxamide in asthma (↓) p-cresol sulphate in asthma
Lee-Sarwar (2018) (100)	FA	12 FA 32 food sensitization 37 controls	Faeces	LC-MS	Significant changes in sphingolipids, bile acids and diacylglycerols in FA
Tao (2019) (101)	Asthma	29 controls 37 uncontrolled asthma 43 controlled asthma	Urine	GC-MS	51 differentiating metabolites mainly belonging to amino acid, carbohydrate and purine metabolism
Chawes (2019) (102)	Asthma	20 asthma 151 controls	Urine	LC-MS	(↑) taurochenodeoxycholate-3-sulfate and 3-hydroxytetradecanedioic acid in subsequent asthma
Lee-Sarwar (2019) (103)	Asthma	85 asthma 276 controls	Faeces + Plasma	LC-MS	(↓) Faecal Docosapentaenoate, DG 16:0/18:1, and 17-alpha-hydroxypregnenolone 3-sulfate in asthma
Crestani (2019) (19)	FA + asthma	35 food allergy 35 Asthma 35 food allergy and asthma 20 controls	Serum	LCMS	Alterations in sphingolipid and ceramide metabolism in FA Less notable differences in asthma

Chiu (2019) (104)	Rhinitis and Asthma	27 rhinitis 34 asthma 24 controls	Faeces	^{1}HNMR	(↓) histidine and butyrate in rhinitis (↓) butyrate in asthma
Chiu (2020) (105)	Asthma	28 asthma 26 controls	Urine + Plasma	^{1}HNMR	(↑) plasma histidine in asthma (↓) urinary 1-methylnicotinamide and trimethylamine N-oxide (TMAO) in asthma
Ross (2021) (106)	Any allergy	14 allergic 30 controls	Umbilical cord plasma	GC-MS	(↑) ursodeoxycholic acid in subsequent allergy at any timepoint (↓) α-ketoglutaric acid in subsequent allergy at 18 and 36 months
Fujiogi (2021) (107)	Asthma	55 asthma 85 controls	Nasopharynx + Serum	LC-MS	No associations to asthma development
Turi (2021) (20)	Asthma	n=335 (INSPIRE) n=145 (WISC) n = 67 (COAST 1y) n = 126 (COAST 3y)	Plasma	LC-MS	(↓) Z,Z unconjugated bilirubin in asthma
Guo (2021) (108)	Asthma	51 asthma 9 controls	Serum	LC-MS	(↓) SM34:2, SM38:1, and SM40:1 in asthma
Sharma (2021) (109)	Asthma	30 asthma 8 atopic controls 35 healthy controls	Exhaled breath	Portable GC	Clear discrimination between groups
Huang (2021) (21)	Asthma	200 asthma 485 controls	Maternal pregnancy plasma	LC-MS	(↓) Caffeine, theophylline, trigonelline, quinate, and 3-hydroxypyridine in asthma
Chang-Chien (2021) (110)	Asthma	92 asthma 73 controls	Exhaled breath condensate	^{1}HNMR	(↓) lactate, formate, butyrate, and isobutyrate in asthma
Bao (2021) (111)	FA	23 allergic 12 controls	Faeces	LC-MS	(↑) secoisolariciresinol in FA (↓) linoleoyl-linolenoyl-glycerol (18:2/18:3) in FA
Qu (2022) (112)	Asthma	602 asthma 593 controls	Plasma	^{1}HNMR	(↓) citrate, ketone bodies, histidine, and glutamine in asthma
Li (2022) (113)	Asthma	19 asthma 21 controls	Urine	LC-MS	Alterations in phenylalanine, tyrosine, and beta-alanine metabolism
Zheng (2022) (114)	Asthma	27 allergic asthma 10 non-allergic asthma 20 controls	Faeces	LC-MS	42 metabolites associated with allergic asthma

Reinke (2022) (115)	Asthma	418 severe asthma 87 mild-moderate asthma 100 controls	Urine	LC-MS	(↓) Carnitine levels in severe asthma
Gürdeniz (2022) (116)	Asthma	n=677 (COPSAC2010) n=377 (COPSAC2000)	Dried blood spots from cord blood	LC-MS	(↓) tryptophan, bile acid and phenylalanine in caesarean section Caesarean section increased risk of asthma
Lee (2022) (117)	Asthma	514 asthma 2833 controls	Serum	LC-MS	(↑) 1-arachidonoyl-GPA (20:4), glutamate and tyrosine in asthma
Chang (2022) (118)	Asthma + Rhinitis	n = 111	Urine	^{1}HNMR	(↓) galactose glycine in rhinitis (↓) succinic acid in asthma
Jang (2021) (119)	FA	20 FA 20 controls	Serum	LC-MS	(↑) sphingolipid metabolites in FA (↓) acylcarnitine metabolites in FA

Pre-analytical and analytical quality in Metabolomics

During sampling, sample management, and sample preparation, there may be several sources of undesired variability, with the potential to obscure underlying associations in the data. The metabolome is affected by fasting state (120), as well as by the circadian rhythm (121). Furthermore, as samples used in metabolomics are biologically active, rapid changes can occur in blood and plasma due to presence of living cells and active enzymes. It is thus of high importance to have uniform and rapid sample management conditions to avoid unnecessary variation (122). MS-based metabolomics further suffers retention time drifts, stemming from column contamination from injections, as well as intensity drifts from decreased detector sensitivity over time (123). Both the retention time and intensity drifts can be reduced by regular instrument maintenance, but frequently also requires post analytical correction. There are thus several sources of variation that requires attention, and effects are not always obvious at an initial glance.

Quality assessment (QA) and quality control (QC) are vital parts of any untargeted metabolomics protocol (17): The concept of quality assessment includes the development of standard operating procedures and training of researchers in using these. Quality control on the other hand refers to procedures used on a day-to-day basis to ensure that data generation is of adequate quality. Untargeted LC-MS metabolomics is especially prone to instrument-related issues relating to e.g., pressure and mass flux. These issues include for example leakages, needle clogging, column- and detector deterioration, in addition to general metabolomics-issues with e.g., sample preparation. In addition, variation in sampling and sample management might have detrimental effects on the measurable metabolome as biological samples contain degradable metabolites, living cells and active enzymes that may alter the composition of metabolites in the sample even after it is removed from the host (124). Sampling and sample handling pose a significant difficulty in clinical settings, where the primary aim of the clinician is to look after the patient rather than to ensure perfect sample treatment. To combat these issues, rigid QA and QC procedures should be in place to minimize and detect problematic injections so that they do not end up in the final data analysis (125). Further, QA and QC practices differ widely between different labs, and in an effort to harmonize and improve general procedures, the Metabolomics Quality Assurance & Quality Control Consortium (mQACC) was founded in 2018 (126). They have since conducted surveys and published several manuscripts (126-129) in their ongoing effort to identify, harmonize and disseminate best practices for QA and QC.

For QA practices, maintenance, calibration and tuning of instruments are integral to keep an instrument up and running with accurate results. For QC practices, standard procedures today include the use of system suitability tests (SSTs), solvent blanks and QC samples (128). SSTs are injected prior to the analysis of study samples and contain a set of well-defined analytes that give an indication of instrument performance at time of testing. Solvent blanks are injections where only the solvent (e.g., acetonitrile or methanol) are injected without any sample, and are used to check if there is any contamination from previous injections on the column. QC samples are samples that are identical to each other, often created by combining small aliquots of several samples from the study or a reference population. These are injected throughout an experiment to for several purposes: to monitor injection quality, to assess feature quality, to adjust for retention time drift and to adjust for intensity drift. In addition, injections are frequently inspected manually based on a few well-known metabolites to diagnose injection

quality. This is however very time consuming for the operator. Moreover, although this procedure might indicate sufficient injection quality, this procedure is not always adequate.

Particular emphasis is placed on QC-samples, both for monitoring, correction, and exclusion of poor-quality features (123, 130). In brief, within an experiment, identical QC-samples are injected between every few samples (e.g., every 10-12th injection) so that these can be compared to each other. Furthermore, as QC-samples should have the same analyte intensities, they can be used both to correct for instrument related drift issues in both intensity and retention time as well as for exclusion of features that do not give reproducible intensities in the QC-samples. Several QC-procedures have been developed during the past years (123, 130-136). One major drawback of these methodologies is that they are applied after an entire batch has been run or even after an entire experiment, causing a delay in potential discovery of instrument or quality issues. If injections, or even batches, are found to be of poor quality, the corresponding data must be discarded and samples re-injected. However, this causes complications since instrument time is a contested resource. Moreover, it introduces additional complexity in adjusting data for drift in intensity and retention time. In addition, samples are frequently a scarce resource and further aliquots may no longer be available which could result in valuable data being lost. Consequently, it would prove beneficial to have a monitoring system that could immediately identify and notify users on potential instrumental issues. If combined with a high degree of automation, such a system could provide a more resource efficient workflow for operators.

4 METHODS AND METHODOLOGICAL CONSIDERATIONS

Study participants and sampling in the NICE-cohort (papers I, II and III)

All samples and data for papers I-III were from the Nutritional impact on Immunological maturation during Childhood in relation to the Environment (NICE) cohort, that has been thoroughly described elsewhere (137). In brief, the NICE-cohort is a prospective parent/child population-based cohort, where 655 pregnant women and their partners were recruited between the years 2015 and 2018 (Table 2).

***Table 2**. Basic characteristics of new-borns and mothers in the NICE-study presented as median (IQR).*

Infant characteristics	
Birth weight (grams)	3565 (3234 – 3955)
Gestational length (days)	281 (274 - 287)
Length (cm)	50 (49 - 52)
Caesarean Section (%, yes)	13.4
Sex (%, female)	54
Maternal characteristics	
Parity (%, nulliparous)	47
Age (years)	30 (27 - 34)
BMI (kg/m^2)	24.35 (22.13 – 27.81)

All prospective parents within the catchment area of Sunderby Hospital in northern Sweden were given brief information about the cohort during their routine visit in their local maternity clinic at gestational weeks 10-12. Recruitment took place during routine ultrasound visits at gestational week 18-19, at which time more detailed information was provided to the prospective study participants. Parents who wished to participate in the study were asked to send in their informed consent after taking the time to properly consider participation. Inclusion criteria were: ability to understand oral and written Swedish, intent to give birth at Sunderby Hospital as well as residency in the county of Norrbotten. Initially 655 pregnant women and their partners were recruited. Due to drop-out, stillbirth and withdrawal from the study, 601 families remained when the children were one year of age (Fig. 6).

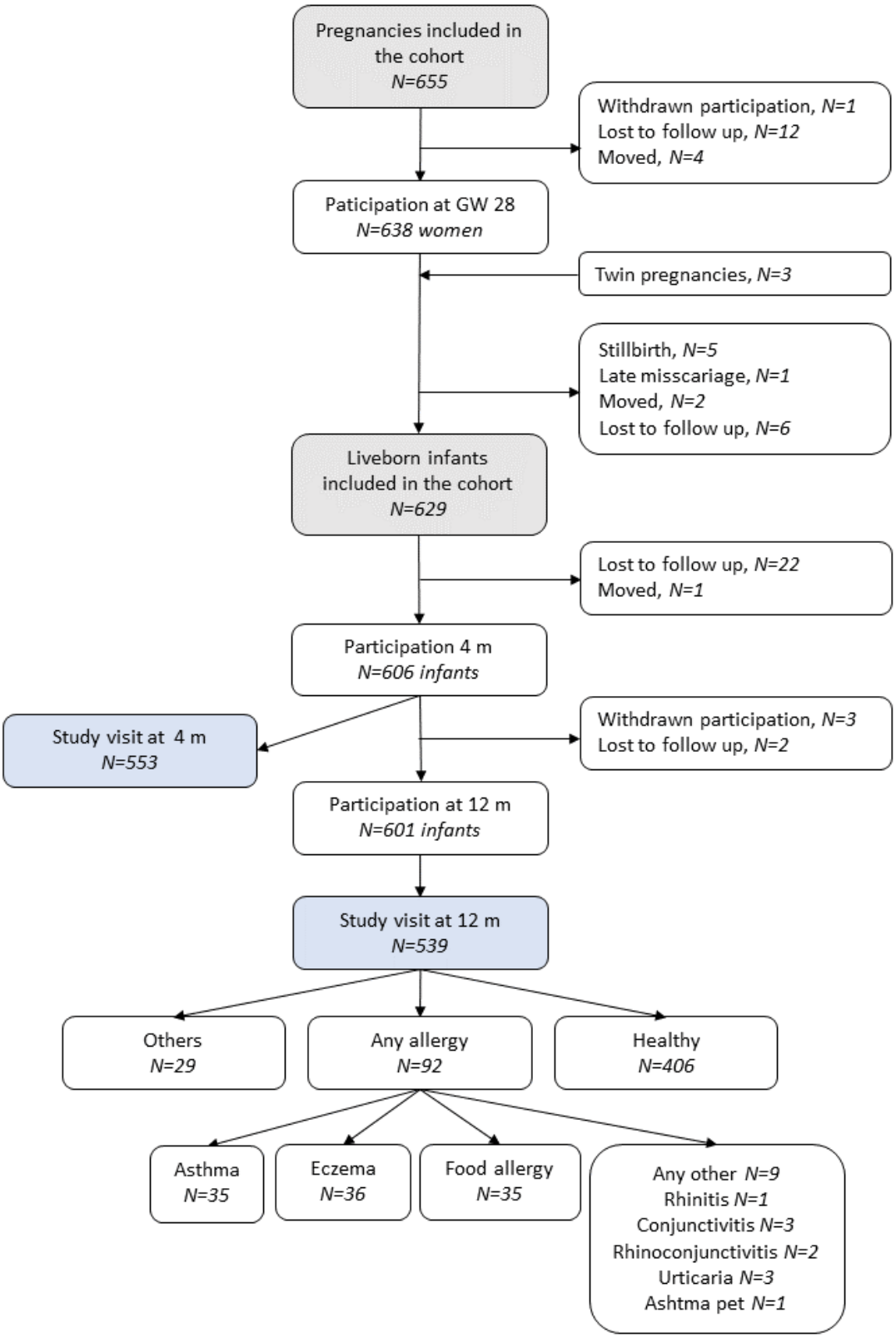

***Figure 6**. Flow chart of families participating in the NICE study at various time points up to 12 months of age including the number of families present at the 4- and 12-month study visits and allergy diagnoses at 12 months of age.*

The NICE-study collects several sample types (Fig. 7), and only a subset of these are used within this thesis, namely: peripheral blood plasma of the mother at gestational week 28 and in connection to delivery, umbilical cord blood plasma (venous, arterial and a mix between the two) and placenta from the delivery as well as peripheral whole blood from the child at 48 h, 1 month, 4 month and 1 year of age.

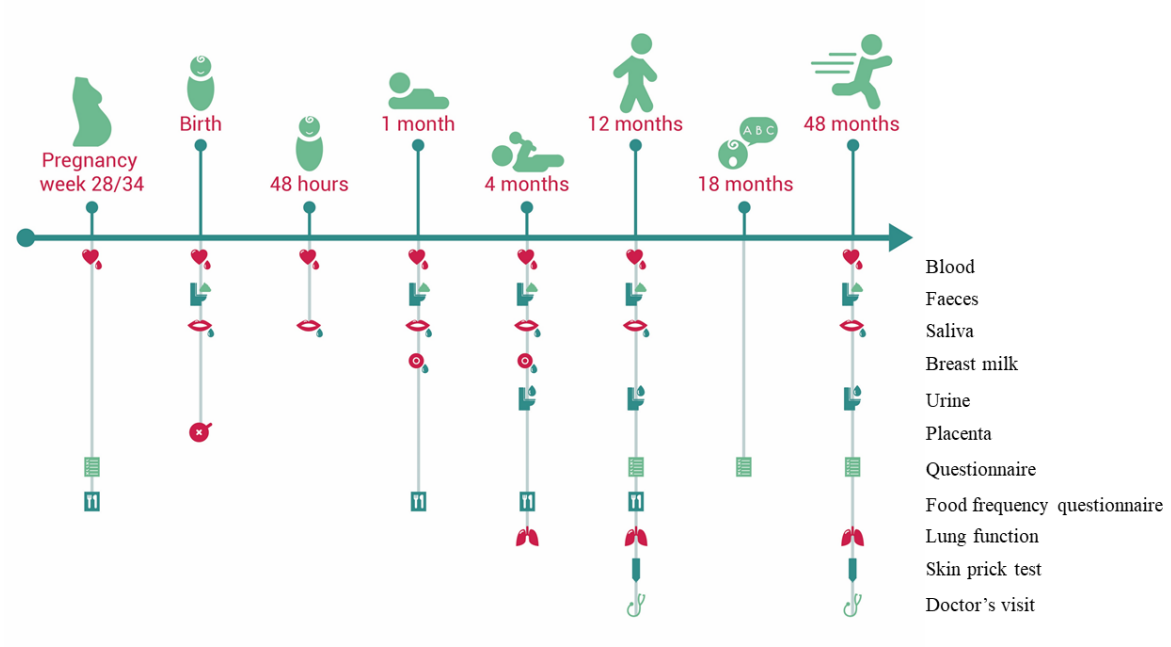

***Figure 7**. Timeline of samples and questionnaires obtained in the NICE-cohort.*

Maternal blood was sampled via venepuncture in EDTA-tubes for samples taken both at gestation week 28 and in connection to delivery. A majority of gestation week 28 samples were taken by midwifes at the local maternity clinics throughout the country of Norrbotten and some were taken at Sunderby hospital. Samples were left for 30 minutes prior to centrifugation and subsequent storage at 4 °C before transportation to Sunderby Hospital where samples were aliquoted and frozen at -80 °C. Due to ethical considerations, mothers were not asked to fast prior to sampling in connection to delivery but were encouraged to fast prior to sampling at gestation week 28.

Venous and arterial umbilical cord blood samples (Paper III) were collected in connection with sampling for blood gas analysis in the new-born. This was performed using heparinized syringes in immediate connection to delivery. Excess blood that was not needed for blood gas analysis were put into EDTA-tubes for subsequent, centrifugation and freezing. Mixed umbilical cord blood was sampled in EDTA-tubes after the cord was cut, by squeezing blood out of the umbilical cord. Due to research personnel not being available every hour of the week, some samples were left to stand in the delivery ward for an extended time until samples could be centrifuged, aliquoted and frozen (Full information available in Paper I).

Chorionic villous placental tissue was collected from placentas within 4 hours after delivery. 1-2 cm^3 pieces were obtained from four different locations on the placenta with the goal of the umbilical cord being equidistant from the sampling locations (Fig. 8). The four pieces were washed in ice cold PBS to remove potentially contaminating blood. The four pieces were further subdivided into four even smaller pieces (3-4 mm^3), placed into microtubes and frozen at -80 °C.

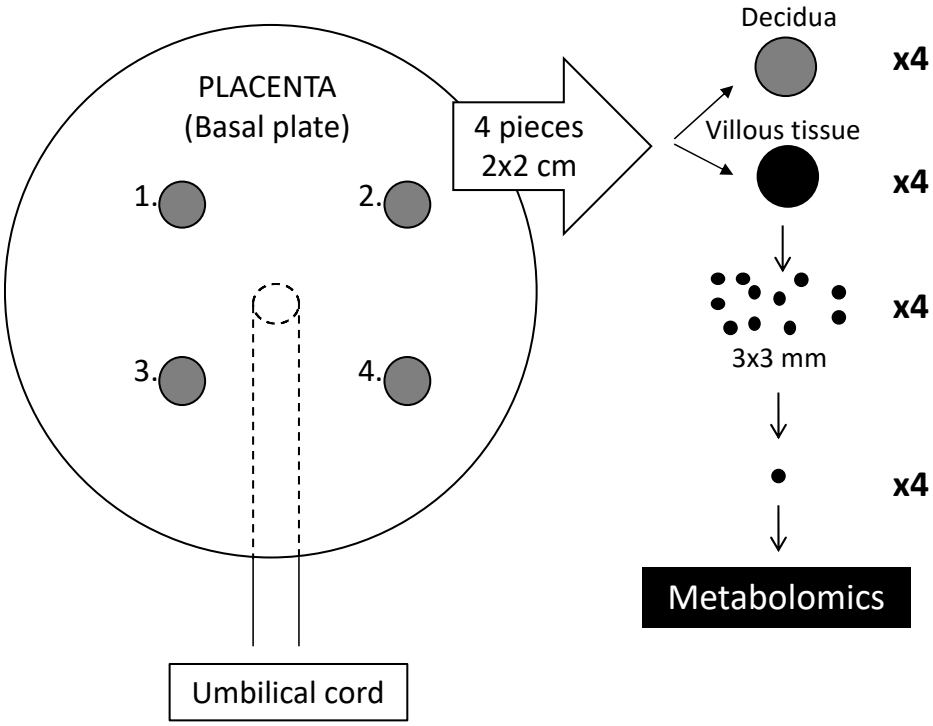

***Figure 8**. Sampling of the placenta. Four pieces of chorionic villous tissue were cut out and each piece was further separated into four smaller pieces and put into microtubes for subsequent freezing.*

Whole blood samples for flow cytometry, TREC and KREC analyses were taken at birth, 48 h, one month, four months and one year of age. Samples were drawn into TruCount™ tubes and immediately stored in a dark room at room temperature prior to analysis that was conducted within 48 hours of sampling.

Allergy diagnoses were performed according to predefined protocols (137) by a single allergologist (Anna Sandin, M.D.). In brief, food allergy was diagnosed if an immediate reaction after ingestion had been observed, or if there was a delayed reaction with improved conditions once the food item was removed from the diet. Unless first symptoms were acute and severe, diagnosis was confirmed by provocation of the food item that was suspected to cause symptoms. Sensitization and specific IgE antibodies were used as support for food allergy diagnosis but were not considered mandatory. Asthma was diagnosed if the child met one or more of the following criteria: wheezing between infections, persistent wheeze for four weeks or more, wheezing during infection combined with concomitant allergic disease, or three episodes of wheezing during an infection, without concomitant allergic disease. Atopic dermatitis (eczema) was diagnosed according to the Williams criteria (36-38), in which the

subject needs to have an itchy skin condition as well as three or more of the following criteria: 1) involvement of the skin creases, 2) a history of hay fever or asthma, 3) problems with dry skin during the past year, 4) visible flexural eczema (e.g., eczema in knees or elbows), and 5) onset before 2 years of age (the last one not used during work in this thesis as all participants were under the age of 2 at time of diagnosis).

Data for development of quality monitoring procedures (Paper IV)

For development and testing of procedures for continuous quality monitoring in untargeted LCMS metabolomics data generation, 3 data sets were used: the *MedGICarb*, *SMC-C* and *QC* data sets. In this thesis, these data sets are consistently referred to in italics for clarity.

The *MedGICarb* data was generated on samples from the MedGICarb study, which is described in detail elsewhere (138). In short, the MedGICarb study is a multicentre dietary intervention study consisting of adults from the US, Italy, and Sweden at risk for type 2 diabetes. The study aims to investigate the effect of low vs high Glycaemic Index foods within a Mediterranean-style healthy eating pattern. Blood was drawn at 8 time points (4 at baseline and 4 post intervention) from 145 individuals with at least two traits of the metabolic syndrome, resulting in 1160 samples that were analysed using untargeted LC-MS based metabolomics for the analytical modes RP, RN, HP, and HN.

The Swedish Mammography Cohort – Clinical (SMC-C) is a sub-cohort of the larger Swedish Mammography Cohort (SMC) (139) taking place in Uppsala and Västmanland counties and recruiting all women in the counties born in 1917 through 1948. SMC-C is an ongoing population-based longitudinal cohorts designed to study dietary exposures in relation to chronic disease outcomes. SMC-C entails around 8000 women, out of which 5022 samples from baseline were available for untargeted LC-MS based metabolomics for the analytical modes RP and RN.

The *QC-data* set consists of repeated injections of three different quality control samples available at the Chalmers Mass Spectrometry Infrastructure (CMSI) during the time of analysis in 2017. In total, 527 injections over 6 batches were analysed in the analytical modes RP and RN.

Measurements of immune maturation

T and B cell populations were measured by flow cytometry. To ensure that the proportion of different cell populations are representative for the circulation, phenotypical analyses requires fresh blood and viable cells, achieved either by analysis within 48 h or by specific cryopreservation protocols (140). It is thus imperative to maintain proper sampling protocols for analytical quality. However, this can be difficult to coordinate in a clinical setting and is a major contributing factor to why the number of flow cytometry analyses is often lower than the total amount of available samples, as is the case in the present work. Furthermore, depending on the flow cytometer used, there may be limitations in the number of fluorophores and consequently, the number of possible combination of fluorophores used is not sufficient to

cover all subpopulations of the T and B cells. Thus, the selection of surface markers and fluorophores might not be large enough to quantify every cell type, and potential information on immune maturation might have been missed. In the NICE-cohort, we could identify all steps of the circulating T and B cell populations, i.e. transitional B cells, naïve and memory B cells, recent thymic emigrants, as well as naïve, central- and effector memory T cells. TREC and KREC, measures of newly formed T and B cells respectively, were measured using real-time PCR on whole blood samples.

The interpretation of immune data is difficult. For instance, naïve cells could be present at higher numbers for several reasons. These could be caused by high production of immune cells, which would indicate a more mature immune system. However, these could also relate to fewer immune cells being activated, thus staying in their naïve form. This, in turn, could indicate that the immune system has encountered fewer antigens, and thereby not required activation. Whilst perhaps memory cells are easier to interpret, with higher proportion or total number indicating a more mature immune system (141).

Metabolomics method selection

Untargeted LC-MS based metabolomics has the widest coverage, is the method with the overall lowest detection limit and has a comparatively easy sample preparation. Drawbacks of this method includes complicated pre-processing of instrument data, in order to retain as much of the relevant metabolite signals as possible, while keeping noise features to a minimum. Further, LC-MS systems comes with an inherent variability both with regards to retention time and intensity drifts and struggles with reproducibility between labs. Identification of measured features is the major bottle neck for the LC-MS community (79).

In comparison to the MS based methods, NMR has high reproducibility, and the results are easy to compare between different labs. Also, it only requires simple sample preparation, the sample is not spent analysis and it offers the possibility of absolute quantification. However, it also has limitations with regards to sensitivity and deconvolution, reflecting in a lower metabolome coverage and higher detection limit. GC-MS based metabolomics can analyse more analytes and detect compounds of lower concentration than NMR and is also robust in terms of retention stability (142). Furthermore, extensive libraries exist for GC-MS based metabolomics making metabolite identification comparatively easy. A drawback is that analytes need to be volatile to be analysed, either inherently or through derivatization, which normally involves more complex sample preparation. Thus, while the metabolome coverage is normally larger than for NMR, it is still smaller than for LCMS. In addition, different analytes have different response factors, which incurs a higher demand on calibration for absolute quantification. Further, the detector suffers from sensitivity deterioration over time making it necessary to monitor and adjust for intensity drift.

During the work of this thesis, untargeted LC-MS based metabolomics was chosen as the main analytical strategy (Paper I, II and IV). This was primarily due to the exploratory nature of the research and the corresponding need to maximize the metabolome coverage since little is known about prospective markers of allergy and immune maturation. Further, the procedures

for continuous quality monitoring software were developed mainly with LC-MS in mind, since this is the most common analytical technique utilized in our lab. In the study on arterial and venous umbilical cord plasma (Paper III), GC-MS based metabolomics was chosen, primarily since the same methodology had been used in a previous study (106), and the possibility of replication was desired.

All plasma samples (Papers I and IV) were prepared according to the same protocol (SMCC-metabolomic SOP v010, available upon request). In short, frozen plasma samples were thawed for 15 h at 4 °C. Samples were then vortexed and pipetted into 96 deep well plates, after which acetonitrile was added to remove protein in the samples. Samples were shaken, centrifuged and the supernatant was filtered under water vacuum using a filter plate and collected into 96 well plates followed by a final shaking and centrifugation and covered prior to analysis in the LC-MS.

The placenta metabolomics (Paper II) was performed in a similar way to the above-described procedure, with some exceptions: Placenta samples were ground using stainless steel beads and methanol was used to precipitate proteins instead of acetonitrile. In addition, since the placenta is not a homogenous tissue, and some features might not be reproducible throughout the same placenta, 4 pieces per placenta were analysed as biological replicates and downstream data analysis limited to stable features, i.e. where the coefficient of variation was < 0.3 for at least 50 % of the placentas.

Study-specific quality control (QC) samples were used in all analyses. These were obtained by combining aliquots from all samples within the first analytical batch and were prepared as described above for each sample matrix. QC samples were injected 5 times at the beginning of each batch for conditioning, and then consecutively every 12th injection for downstream retention time and intensity drift correction and feature assessment.

All samples from both plasma and placenta were analysed using UHPLC-Electrospray Ionization-Quadrupole-Time of Flight systems. All analyses employed identical LC instrumentation and settings (Agilent 1290). In brief, samples were injected onto a UPLC HSS T3 column kept at 45 °C. Methanol was used as a polar solvent for placenta samples while acetonitrile was used for all plasma samples. QTOF settings varied between instrumentation, using an Agilent 6520 QTOF for maternal and neonatal plasma samples (Paper I) and a 6550 QTOF for placenta, *MedGICarb*, *SMC-C* and *QC*-data (Papers III and IV).

MS2 data were obtained by identifying samples with the highest concentrations of the analytes of interest. These samples were reinjected and subjected to MS/MS fragmentation at 10, 20 and 40 eV. MS/MS fragmentation patterns were matched against an inhouse library of authentic standards for level 1 ID, and to the Massbank of North America (MoNA) using an in-house developed script (SLIM, to be published) and the human metabolome database (HMDB) (143) for level 2 ID. Further, the Sirius GUI (144) including use of CSI:fingerID (145) and CANOPUS (146, 147) were used to further aid identification at lower levels of confidence. Initiatives have been carried out in an attempt to harmonize methodologies and reporting standards between labs (148, 149). All identification performed in this work are reported as suggested by the Metabolomics Standard Initiative (MSI).

In the work presented in this thesis using GC-MS (Paper III), derivatization was performed by silylation and methoxymation. This is in essence performed by attaching either tri-methyl-silyl or methoxyamine to the analyte to make it more stable and to facilitate the transition into gaseous phase (150). The method used in this thesis combines untargeted (n = 268), targeted (n = 118), and multiple reaction monitoring (n = 30), meaning that the method is targeting a set of known metabolites, whilst simultaneously trying to find molecular entities not in the list of targeted metabolites.

In short, plasma samples were mixed with methanol and internal standards, followed by shaking, incubation and centrifugation. The supernatant was evaporated followed by methoxymation and subsequent silylation before analysed in the GC-MS/MS system. Spectral information from MRM was acquired using data from pre-determined transitions (150), targeted spectral information was acquired using a MATLAB script provided by the Swedish Metabolomics Centre that automatically scans spectral data based on MS1 and MS2 matches and integrates peak areas to annotated compound identifiers, and untargeted spectra were obtained using MS-Dial (described below).

Pre-processing of metabolomics data

Spectral data from the instrument need pre-processing into tabularized format that can be used in downstream statistical data analysis. Especially for LC-MS data that normally contains thousands of peaks per injection, this is not trivial. Pre-processing can be subdivided into the following discrete tasks (151) (Fig. 9): 1) peak picking, where spectral peaks are determined and integrated, 2) retention time alignment, where drift in the chromatographic retention is adjusted for, 3) feature correspondence, where features in different samples are matched to their corresponding features in other samples, 4) consensus integration, where peaks not found by the peak picking algorithm are integrated, 5) imputation, where data that are not present are filled in with plausible values, 6) intensity drift correction, where drift in the mass spectrometer response is adjusted for, and 7) feature clustering, where features likely belonging to the same metabolite are clustered together to one feature.

Two different approaches for pre-processing were used in the work reflected in this thesis: One centred around the XCMS package (152) in R (153) (Papers I and IV) and the other using the MS-Dial (154) software (Papers II and III).

Parameters for XCMS-based peak picking were optimized by a combination of the Isotopologue Parameter Optimization (IPO)-algorithm (155) and manual assessment of peak quality. Due to retention time shift in the LC-MS system, retention time adjustment is necessary ensure that peaks corresponding to the same analyte elute at the same time in all the different injections. Post retention time alignment, peak correspondence is performed, where the peaks from each injection are grouped into features. After correspondence, the ensuing consensus peak filling is performed. Both XCMS and MS-dial do this by checking every injection and every feature for missing values. If a value is missing, it performs integration in the injections where the features are missing and assigns the integrated area for the corresponding feature (152, 154).

For various reasons, some injections will not have any signal at the given location (156) (157). These reasons include that there are no measurable metabolites in the sample, that the peak is misaligned in RT or m/z space or that the instrument or peak picking software failed to recognise a peak. However, missing values are a nuisance that contribute to loss of power in statistical analysis. Moreover, several machine learning methods will not accept missingness (156). To compensate for this, imputation is usually performed, where so-called dynamic imputation is usually considered to be the best choice (156, 157). Imputation was performed through the use of a partial least squares (PLS)-based two-step imputation, where the first iteration uses slack tolerance settings, in order to provide a good starting guess for the second, stricter round of imputation using an in-house algorithm (function mvImpWrap, available at https://gitlab.com/CarlBrunius/StatTools).

After imputation, intensity drift correction was carried out. This was done using the batchCorr (123) algorithm, which has been shown to outperform several of the more common batch correction methods (158). Following the intensity drift correction, the RamClust (159) algorithm was used to cluster features likely corresponding to the same molecule. RamClust parameters were optimized based on visual inspection of a random subset of clusters, to check whether features not belonging together were present in the same cluster or not according to in-house procedures.

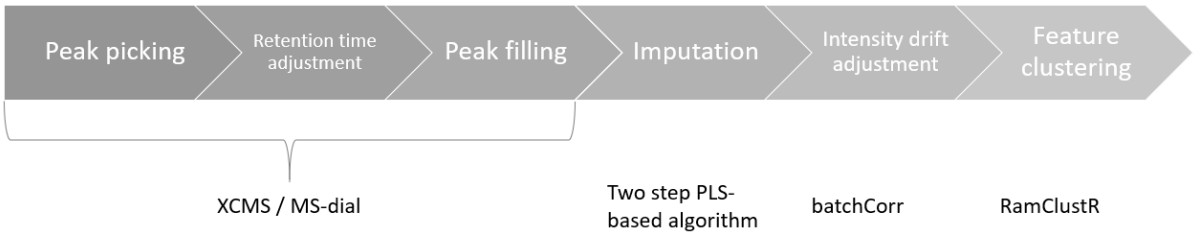

***Figure 9**. Workflow of pre-processing for all untargeted data sets used in this thesis.*

Data Analysis

All analyses containing sensitive personal data were carried out using resources provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX, partially funded by the Swedish Research Council through grant agreement no. 2018-05973. For all analyses on the NICE cohort data (Papers I, II and III), initial analyses were carried out using Random Forest (RF) modelling (160) within a repeated double Cross Validation scheme (161, 162) with integrated variable selection implemented via the MUVR R package (163). Random forest is one of many possible machine learning options when dealing with metabolomics data, where other possible options include e.g. PLS (including PLS-based methods such as OPLS, PLS-DA etc.) and support vector machines. Random forest was chosen as this is a non-linear method, that does not rely on specific distributions of the input variables, and it requires relatively little parameter optimization to work well for metabolomics data (164). As with every method, random forest also comes with drawbacks. One of the major drawbacks is that it is difficult to produce intuitive and accurate visualisations of the decision process, thereby impeding interpretation and making it somewhat of a black box method, although all the information is

theoretically possible to obtain. Furthermore, RF is unable to extrapolate values outside of the available range in the training data (165).

Since multiple RF models were made for each research question in this thesis, *a priori* limits for fitness metrics were decided at Q2 > 0.2 and classification rate > 66 %. These limits were chosen based on experience from working with predictive modelling, with the intent to only evaluate models containing potentially relevant information with regards to the outcomes studied.

When attempting to discern differences between arterial and venous umbilical cord plasma (Paper III), arterial and venous cord blood samples were drawn from the sample individuals, producing non-independent samples. To take this into account, a so-called multilevel approach (163, 166) was utilized, where an effect matrix was used as the predictor data in the RF analysis against a dummy response variable. The effect matrix was calculated as the $\log_2$ fold change of the measured arterial and venous cord blood metabolomes (equation 1).

$$Effect\ Matrix = \ log_2\left(\frac{X_A}{X_V}\right) \tag{1}$$

Features selected from the MUVR models were then further tested using more traditional feature-wise methods, such as Spearman correlation (for T and B cell markers and continuous demographic variables (e.g., maternal BMI); Papers I, II and III), and the Mann-Whitney U-test (for dichotomous demographic variables (e.g., parity); Papers II and III). Adjustment for multiple testing was generally not carried out as the number of variables to adjust for was difficult to estimate. The number of features in an untargeted metabolomics experiment can reach tens of thousands, and some of these features will not correspond to real molecules originally present in the sample (e.g., instrument noise, artefacts from sample management and fragments of metabolites) and thus including these in an adjustment scheme may obfuscate true underlying associations.

In addition to the general strategy described above, analyses were also performed to investigate whether metabolites previously associated to manifest allergic disease were also prospectively associated to incidence of allergy (Paper I). This analysis was performed using a combination of criteria: i) a potential match of exact mass (MS1) of metabolite candidates from literature (with adducts $[M + H]^+$, $[M + Na]^+$, $[M + K]^+$, $[M + NH4]^+$, $[M + CH3OH + H]^+$, $[M + ACN + H]^+$ and $[M + 2H]^{2+}$ in positive mode and $[M–H]^–$, $[M–H2O–H]^–$, $[M + Na–2H]^–$, $[M + K–2H]^–$, $[M + Cl]^–$, $[M + FA–H]^–$, $[M + HAc–H]^–$ and $[M–2H]^{2–}$ in negative mode) with the generated metabolomics data (within 20 ppm); ii) Paired t-tests of matched case-control pairs and generalized linear models (unmatched, but adjusting for gender, gestational length, caesarean section, age of mother, parity and BMI of the mother), where $p < 0.05$; and association to iii) the same allergic disease and; iv) in the same effect direction as the original study.

Tests and metrics used in *QualiMon*

Quality Monitoring using landmark features *(QualiMon)* uses several metrics to assess the data quality of an injection. These include metrics that are judged based on the whole injection, hence referred to as injection-based metrics, but also metrics that are specific to individual landmark (*LaMa*) features, referred to as *LaMa*-based metrics.

The injection-based metrics include Total Ion Count (TIC), number of peaks (using predefined peak picking parameters), IPO-score and number of landmarks found. TIC was chosen as this gives a good overall measure of detector sensitivity. Number of peaks was chosen as this could indicate whether something went wrong during injection of sample, or during sample preparation as a significant drop in number of peaks indicates that no or little sample was injected. The IPO-score measures the number of peaks with corresponding isotopic peaks (i.e., peaks containing one or more ^{2}H or ^{13}C) and correlates with the number of peaks but provides further information about general peak quality for the injection. Number of *LaMas* was chosen as these are less sensitive to intensity drift and can provide a measure of instrument stability even though other injection-based metrics fluctuate between injections. Once a value has been assigned to an injection for each of these metrics, this value will remain constant, even as more injections are added. For each metric, a students' t-test is thus performed to compare the value of the injection with the current distribution of all other injections performed within the experiment. If a metric deviates enough from the total population, according to a pre-determined p-value threshold, the injection is then flagged for that particular metric. In addition to testing deviations from the sample population, *QualiMon* also assesses whether injections deviate according to pre-determined (absolute) limits. The injection-based metric values are compared to soft and hard limits, reflecting user defined limits determined based on usual instrument performance. Injections are then further flagged if passing these thresholds.

The *LaMa*-based metrics are as follows: Full width at half maximum, tailing factor, intensity, retention time deviation, scans per peak, peak height and number of *LaMas* found. Most of these metrics are generally used within traditional analytical chemistry when determining peak quality (167). For each *LaMa*-based metric, the number of outliers is calculated. The number of outlying *LaMas* is further compared to the same value for the other injections in the database.

When doing this many tests to a sample (one test for each *LaMa* and metric), it is very likely that several of these will end up as significant outliers by chance alone. We therefore further test if the number of outliers per sample and metric differ from the total distribution of outliers in the previous samples. Furthermore, as distributions change over time as more injections are added, an iterative testing of previous injections is also performed. This iterative testing is often useful in the beginning of a new batch, where previous injections might not be representative for the new batch. A new injection might thus be initially classified as deviating, but after a few further injections be perceived as 'normal' by the algorithm, through the addition of samples with similar characteristics of the new batch.

At the end of the *QualiMon* testing, an accumulative quality score (the so-called *status* score) is calculated, where every flagged metric is tallied, effectively summing outliers from the sample population as well as broken limits (equation 2). Similar to the tests for the *LaMa*-based

metrics, the status score is highly dependent on the distributions of previous injections and are similarly recalculated for each sample after each new injection is added to the distribution.

$$\frac{Score = \sum flags\ from\ distributions + \sum broken\ soft\ limits + 2 * \sum broken\ hard\ limits}{Maximum\ possible\ score} \quad (2)$$

Associations of maternal and neonatal metabolomes with allergy development (Paper I)

Random forest analysis did not show any associations of either the maternal metabolome during pregnancy, the maternal metabolome at delivery or the umbilical cord metabolome with any of the four allergy outcomes (asthma, eczema, food allergy, and any allergy) at one year of age (Table 3). Classification rates ranged between 32 % and 61 %, indicating that the models were on par with random (permuted) models at predicting allergies. Further inference from these models regarding mechanisms or metabolites-of-interest were consequently not drawn. Our null results regarding future allergy development could be due to several reasons. For one, the allergy diagnoses were performed at one year of age, meaning that there is ample time for children currently diagnosed as healthy to develop allergies in the future, which would obfuscate potential findings relating this. Another possible explanation could be that there might be several endotypes relating to these diseases depending on origin, severity, and trigger. Several studies in manifest disease have shown that asthma is a heterogenous disease (168, 169) and that the metabolome changes depending on severity of asthma (88, 170, 171), and it is not impossible that these metabolic differences could be observed already prior to disease onset, and thus contribute to additional variation in the allergic groups. In addition, the criteria used for asthma diagnosis could reflect both atopic and non-atopic asthma, increasing the heterogeneity in this group. Furthermore, as the metabolome represents a snapshot of the phenotype and the samples were taken from the umbilical cord and the mother in immediate connection to delivery, the stressful situation for both the child and the mother might obscure potential underlying associations.

***Table 3**. Classification rates from random forest analysis of all plasma metabolomes and all allergy outcome*

	CR[a] (%) Maternal pregnancy metabolome	**CR[a] (%) Maternal metabolome at delivery**	**CR[a] (%) Umbilical cord blood metabolome**
Asthma	32	33	42
Food allergy	58	56	40
Eczema	36	59	36
Any allergy (Food allergy or eczema)	39	61	32

[a] CR = Classification Rate

In a targeted search for metabolites previously reported in the literature of manifest allergy, only hypoxanthine showed an association consistent with the disease in question and direction of the association (i.e., higher in asthma (96)). It should be noted that all candidate markers that

we identified from literature were from studies investigating markers of manifest disease, potentially reflecting metabolite associations at later stages of disease progression or symptoms rather than mechanistic metabolites causal for disease development. Consequently, it is not surprising that most of the metabolite-allergy associations could not be replicated in a prospective setting. No adjustment for multiple testing was performed, and false positive replications cannot be ruled out. However, filtering metabolite candidates by disease and direction of association in practice limited the number of actual tests performed.

Associations of maternal and neonatal metabolomes with immune maturation (Paper I & II)

Modest associations were found for memory and naïve lymphocytes, measured by flow cytometry, with the investigated metabolomes, i.e., the plasma metabolome of the mother during pregnancy and delivery, the umbilical cord metabolome, and the placenta metabolome (Table 4). Unfortunately, most of the features driving these associations could not be identified. Identified associations further suffers from low sample sizes, as missing values were present both in relation to the metabolomics measurements as well as the flow cytometry measurements.

Table 4. *All discovered associations (Q2 ≥ 0.2) using random forest analysis on both plasma and placenta metabolomes with regards to immune maturation parameters at 48 hours, 1 month, 4 months and 12 months of age.*

Outcome	**n**	**Q2**	**p[a]**
Maternal metabolome during pregnancy			
CD27+ B cell count at 4 months	24	0.20	0.023
Maternal metabolome at delivery			
CD24+CD38low B cell count at 4 months	26	0.21	0.033
Umbilical cord metabolome			
CD24+CD38low B cell count at 4 months	22	0.24	0.009
CD5+ B cell count at 4 months	22	0.23	0.037
Placenta metabolome			
KREC (at birth)	36	0.33	0.003
CCR7+CD45RA- T cell count at 4 months	22	0.23	0.028
Proportion of CCR7+ CD45RA- T cells of CD4+ T cells at 4 months	25	0.26	0.002
Proportion of CD24low CD38low B cells of total B cells at 12 months	23	0.26	0.006

Plasma metabolomes and immune maturation

Associations were observed between the maternal metabolome at gestational week 28 and the number of $CD27^+$ B cells at 4 months of age (Table 4), suggesting a link specifically to memory B cells (172). This association indicates the possibility of immunomodulatory compounds being transferred from the mother to the child already at mid-pregnancy. The association was mainly driven by 4 features, a diglyceride (32:1), an unknown carnitine, and an unknown phosphatidylserine. Further associations were found between another marker of memory B cells, $CD24^+CD38^{low}$ B cells and both the blood plasma from mothers during delivery (Q2 = 0.21, p = 0.33) and the umbilical cord plasma (Q2 = 0.24, p = 0.009), further strengthening the link to memory B cells. These associations were primarily driven by 4 features from the maternal plasma and 3 features from the umbilical cord plasma. Most notable among these features, phenylacetylglutamine (PAG) was inversely associated with $CD24^+CD38^{low}$ memory B cells in both the maternal and umbilical cord plasma. PAG has previously been reported to have anti-inflammatory properties in mice, where it reduced the endogenous production of tumour necrosis factor-α and interleukin-6 (IL-6) (173). IL-6 in turn is a cytokine that stimulates differentiation in B cells (174). Our findings thus suggest that PAG, transferred from the maternal circulation to the foetus, could inhibit B cell differentiation. Among the other features, only one putative annotation could be made of a triglyceride (45:7) in the maternal plasma at delivery, which associated negatively with the formation of memory B cells. This finding strengthens the findings of previous research, where polyunsaturated fatty acids have been shows to have immune supressing effects (175, 176). However, as the triglyceride could not be identified, further interpretation of the results is speculative. The umbilical cord metabolome further associated with the number of $CD5^+$ immature B cells at 4 months of age (Q2 = 0.26, p = 0.006), suggesting a link to naïve B cells that have not undergone differentiation into a more mature form.

The placenta metabolome and immune maturation

The association of the placenta metabolome with KRECs (Q2 = 0.33, p = 0.003) indicate an association to newly formed B cells, and thereby to B cell production in the bone marrow. Further, the proportion of $CD24^{low}$ $CD38^{low}$ (i.e. naïve) B cells at one year of age associated with the placenta metabolome (Q2 = 0.26, p = 0.006), thereby potentially pointing towards a less mature immune system. Both the number of $CCR7^+CD45RA^-$ T cells within the $CD4^+$ T cells (Q2 = 0.23, p = 0.028) and the proportion (Q2 = 0.26, p =0.002) at one year of age was found to associate with the placenta metabolome. $CCR7^+CD45RA^-$ is a marker of central memory T_h cells (177), meaning that they are memory cells that reside in the lymph nodes. Once its specific antigen is presented, central memory T cells can rapidly undergo clonal expansion, and differentiate into effector T cells that aid in B cell activation as well as stimulate the innate immune cells.

The features driving the associations found in the placenta metabolome were unfortunately of low intensity, and MS2 fragmentation spectra were not obtainable. An attempt to circumvent this issue was performed in which features were filtered based on their signal intensity, resulting in exclusion of approximately half of the features, prior to reperforming RF analysis. Most of the previously observed associations were no longer present, highlighting that

associations were primarily driven by low intensity features that are susceptible to background instrument noise. In fact, all the features of interest selected from the original analyses were of low signal intensity. Thus, we speculate that any potentially immunomodulatory metabolites likely do not originate from the placenta but rather are transferred from the maternal blood stream.

We further examined whether the associations observed in the maternal and umbilical cord metabolomes could be found in the placenta metabolome, and if the associations observed in the placenta metabolome could be found in the plasma metabolomes. However, none of these associations could be observed. This lack of replication between the two sample types might indicate that the findings in one or both sample matrices could be spurious. Another possible explanation for the lack of overlapping features could be that the features found in maternal and umbilical cord blood might easily pass across the placenta, and thus be below the limit of detection in the placenta. As the placenta is a metabolically active organ (178) the features found in the placenta could be metabolized further prior to entering the maternal or neonatal blood streams, and thereby not be present in either blood stream.

Associations of arterial and venous umbilical cord metabolomes with maternal and neonatal traits (Paper III)

Systematic differences were found between arterial and venous umbilical cord blood as well as for the parity of the mother with the venous umbilical cord plasma metabolome. No associations were found relating either the venous or the arterial cord blood with gestational length, birth weight, sex, maternal age, or maternal BMI.

Metabolites differing between venous and arterial umbilical cord blood

RF modelling showed systematic differences between the arterial and the venous umbilical cord metabolomes measured using GC-MS (Classification Rate = 79%**,** $p_{permutation}$ = 0.004). The differences between arterial and venous umbilical cord plasma metabolomes were primarily driven by 11 metabolites. Amongst these metabolites, a hexose, two deoxy-hexoses, glutamic acid and two organic acids pertaining to energy metabolism were the most prominent (Fig. 10).

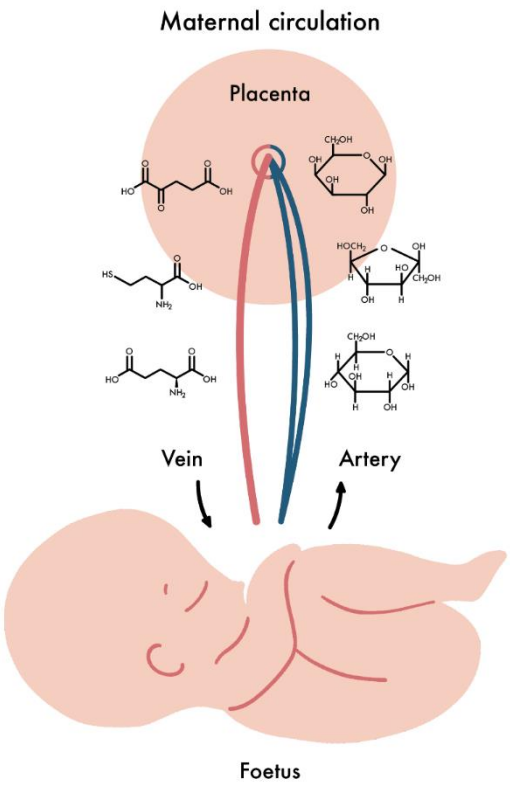

***Figure 10**. Graphical representation of organic acids travelling from the placenta to the child via the umbilical artery and de-oxy hexoses and a hexose travelling to the placenta from the child.*

The hexose and the deoxy-hexoses were higher in the arterial cord plasma, indicating a negative energy balance in the child during delivery. This condition likely reflects the high energy expenditure in the mother, shown by the higher levels of the TCA-cycle intermediates α-keto glutaric acid and succinic acid in the venous cord blood. A possible explanation for the child transferring energy metabolites to the mother could be that the mothers are in a stressful state at time of delivery, thus not having any surplus for transferring these to the child. This is further supported by the metabolites being transferred from the mother indicating an active energy metabolism. This is however, in contrast to a previous study, that found higher concentrations of glucose, lactate and catecholamines in the venous cord blood in children born via elective caesarean section (179). As most of the children in the NICE-cohort were born via vaginal delivery, this difference could be explained by the different energy expenditures between an elective caesarean section and a vaginal delivery. The amino acid glutamic acid was further found to be higher concentration in the venous cord blood, supporting previous findings by Holm et al., that found several amino acids to be higher concentration in the venous cord blood (180).

Taken together, these results suggests that the sampling of umbilical cord blood (i.e., arterial, venous, or mixed) has implications on the measured metabolome, which could contribute to obfuscating underlying associations of interest. Thus, it is important to select a sampling protocol in accordance with the research question at hand. If an association is expected to be

related to the child's metabolism, arterial cord blood likely represents the better choice while venous umbilical cord blood might be more suited if investigating effects mediated by the mothers' exposures and/or metabolism. It should however be noted that as the umbilical arteries are much smaller than the vein, it might be difficult to obtain large sample volumes of arterial cord blood, thus making this biospecimen less attractive for certain analyses.

Associations of umbilical cord plasma metabolomes and neonatal and maternal factors (Paper III)

Among the neonatal and maternal traits (gestational length, birth weight, sex, parity, maternal age, and BMI), only parity associated with the venous umbilical cord metabolome (Classification rate 77 %, $p_{permutation} = 0.004$). 14 features were selected from the RF modelling, 13 of which were in higher concentration in children with nulliparous mothers. Most of the identified metabolites related to energy and amino acid metabolism (Fig 11). The higher concentration of energy metabolites in the children with nulliparous mothers could relate to longer delivery times for the first delivery compared to subsequent ones (181). A longer delivery time would in turn mean that more energy has been expended during the labour process and thus being in a more exhausted metabolic state.

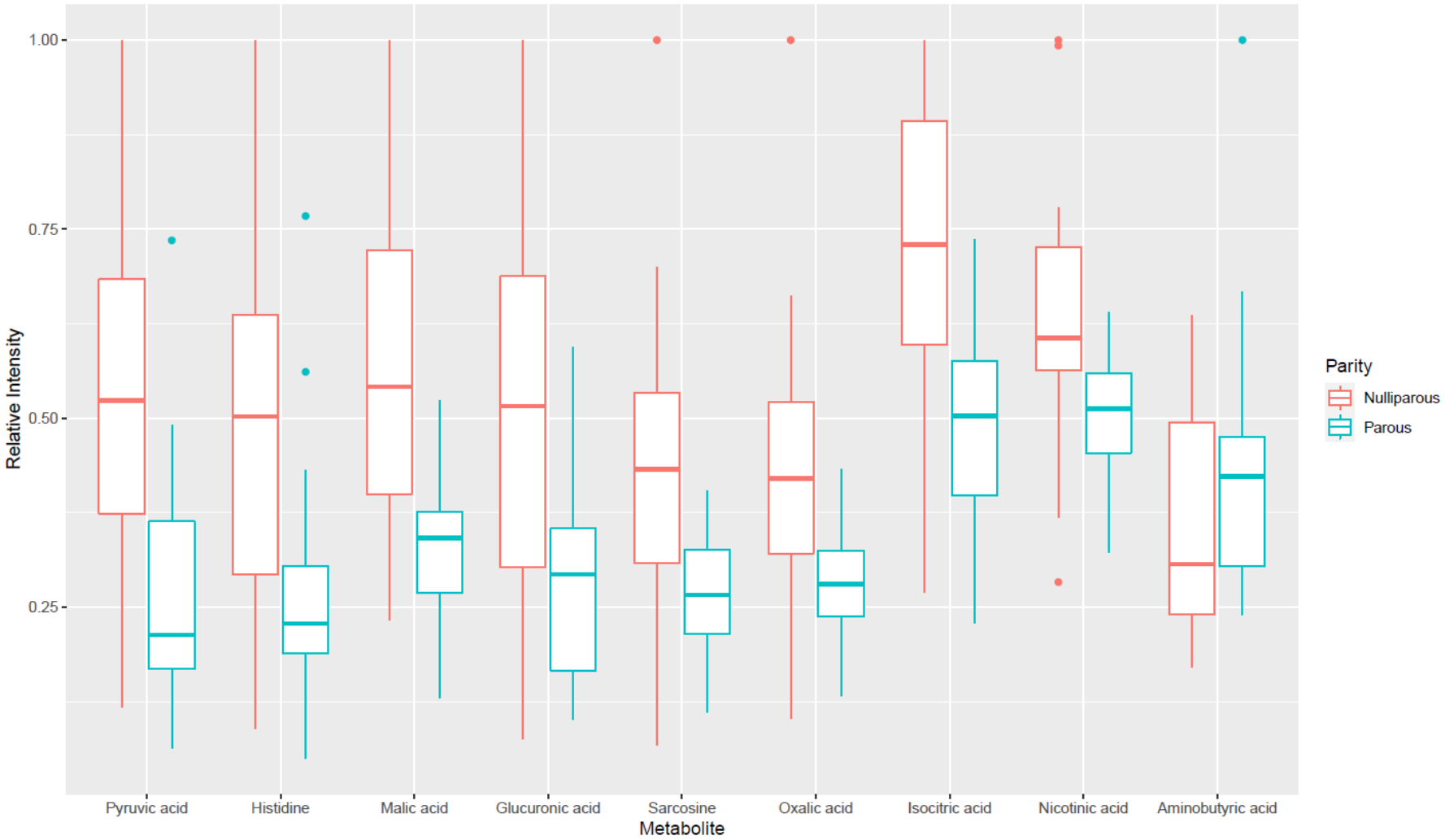

Figure 11. *Boxplots of all annotated metabolites differing between children with nulliparous and parous mothers in venous umbilical cord blood*

The association to parity was more pronounced in the venous blood compared to the arterial, which indicates that it is determined more by the mother's metabolism rather than that of the child's. The results further show that it is important to consider parity already in the study design, either through matching or through statistical adjustment, in addition to choosing the proper sample matrix.

Associations of the placenta metabolome with neonatal and maternal traits

Weak associations were found in the placenta metabolome with parity and sex (CR = 65 %, p = 0.04, CR = 65 %, p = 0.03 respectively). As neither of these associations were above the *a priori* threshold of 66 %, no further investigations into the molecular entities of the features driving these associations were performed. Although associations were weak, there appears to be some discriminating features between these factors and the placenta metabolome. However, the placenta appears to have a weaker association to parity than venous umbilical cord blood does.

Continuous monitoring of injection quality in non-target analysis (Paper IV)

To address the quality of data at the point of instrumental analysis in untargeted metabolomics, procedures for continuous quality monitoring were developed. These procedures were made freely available as a fully automated software solution under MIT licence in an R/Shiny implementation '*QualiMon*' (http://github.com/MetaboComp/QualiMon). In addition to monitoring newly generated instrument data, the *QualiMon* software also allows review of quality of previously generated data.

LandMark features

LaMas are features that are specific for the sample matrix, instrument and to each analytical mode and are present in nearly every injection, at high intensity and distant from any other features. We identified 128 and 160 landmarks for RP and RN respectively using the *MedGICarb* data. These landmarks were further validated in RP mode using two separate data sets that were analysed on the same instrument, showing that 83.2 (sd = 8.42), 81.8 (sd = 16.5) and 95.7 (sd = 31.2) *LaMas* were found in the *QC*-data, *SMCC* and *MedGICarb* data sets respectively (Fig. 12).

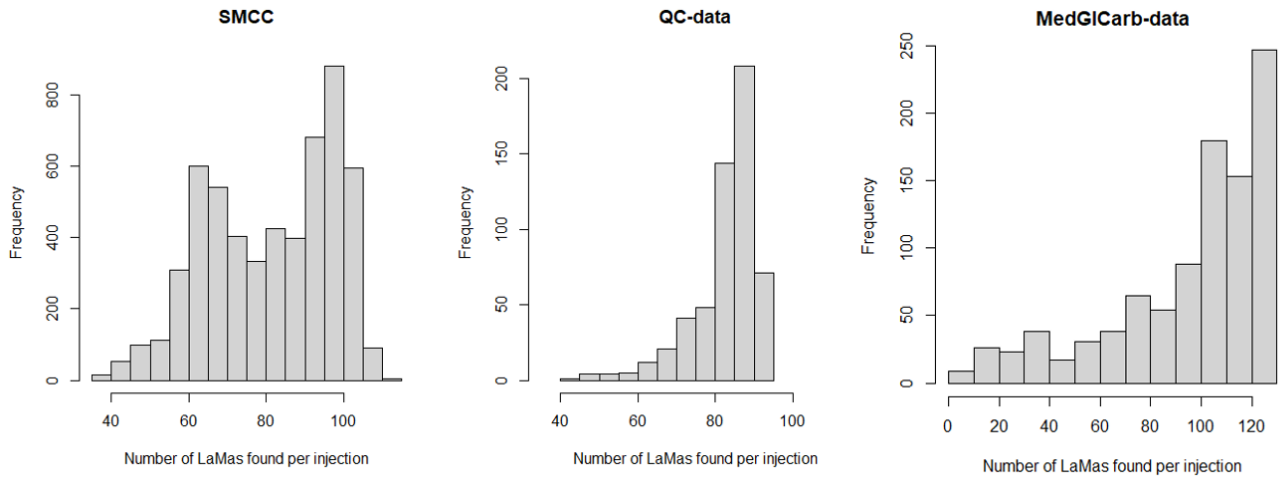

***Figure 12**. Histograms of LaMas found per sample for the SMC-C (left) and the QC (middle) and MedGICarb (right) data sets.*

We developed automated methods to aid users in acquiring *LaMas*. However, as experience working with the system grows, users should refine the *LaMa* selection to ensure that they are stable over longer time periods and properly reflect a general sample of the given sample matrix. We speculate that the *LaMas* can be categorized into three different categories: 1) sample specific metabolites, 2) artefacts from the sample preparation and 3) instrument artefacts. Although it may be tempting to focus specifically on sample specific metabolites, keeping *LaMas* originating from sample preparation and instrumental artefacts should have high value, since these might contain potential information on why injections are of subpar quality.

Assessment of injection quality

Injection quality is assessed based both on metrics that are pertaining to the injection as a whole, such as number of peaks and TIC and on the *LaMa*-based metrics. The injection-based metrics are assessed using a t-test comparing the value to the distribution of other injections, while the *LaMa*-based metrics are assessed based on the if proportion of the observed *LaMas* are significantly different from the proportion of previous injections. These two different ways of assessing quality provide a broad coverage of potential errors.

With only a few exceptions, almost all injections in the *QC*-data performed well regarding all quality metrics (Fig. 13), indicating that the instrument was in a good condition and little variation was found between the injections. This is to be expected as the *QC*-data represents optimal injection conditions: For one, as the QC-samples are injected from the same vials over and over, this reduces all variation in sample handling. It further does not represent any biological variation, as all the injections should be identical.

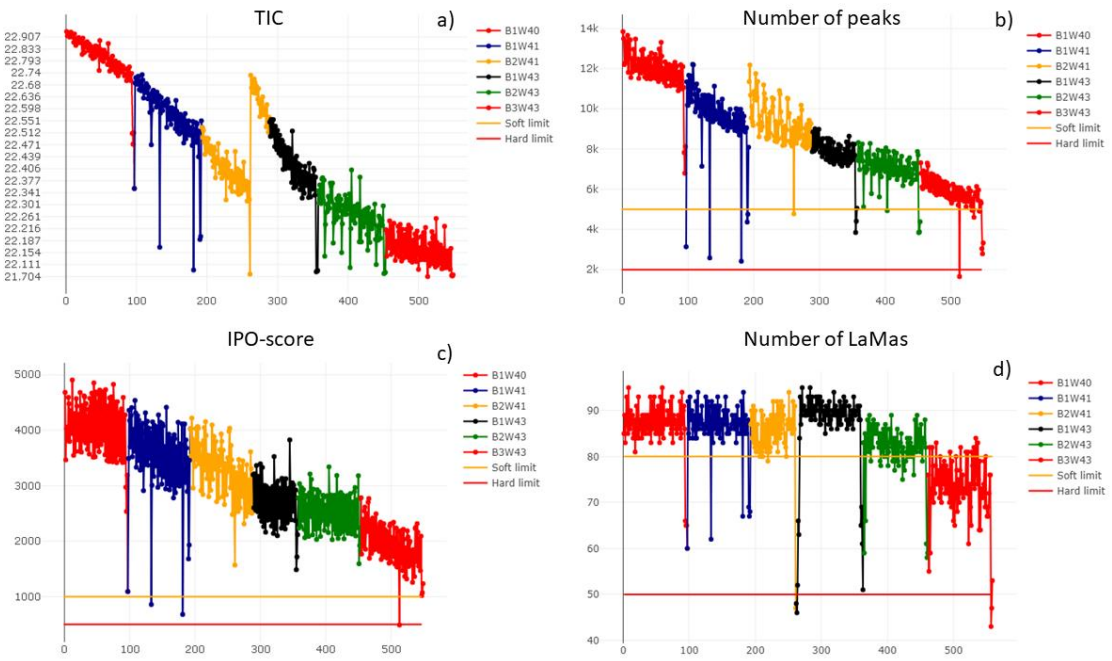

***Figure 13**. Injection dependent metrics generated by QualiMon using the QC-data in RP mode. Nearly all injections exhibit good quality properties in all metrics, with a few outlying drops in quality randomly distributed throughout the run. Injection quality seems to decrease with the number of injections for all metrics except for number of LaMas, which seems more stable over time, with sudden drops in the last two batches.*

Some injections did show significant differences in all the injection-based metrics and a clear deviation can be seen in the plots (Fig. 13). Furthermore, a clear decline in TIC, number of peaks and IPO-score can be seen as the injection progresses, showing the need of instrument maintenance after these six batches were completed. A sudden increase in TIC can be observed in Fig. 13a, corresponding to some unknown event, likely related to detector sensitivity or interface efficiency. In Fig. 13d, the number of *LaMas* is shown to decrease below the soft limits towards the end of the experiment. This behaviour highlights the importance of the soft and hard limits that can be used to pick up on instrument decline, as a steadily decreasing instrumental output will change the distribution to encompass injection of subpar quality. Furthermore, the strict limits are needed if the distribution of previous injections to a large extent consists of faulty injections, thus making a faulty injection seem as good

Due to the risk of potential variation in instrumental conditions (e.g., due to instrument maintenance or prolonged time between two batches), the *LaMa*-based metrics may entail artefactual incidence of low-quality injections. A sample in the beginning of a new batch with differing instrumental conditions should therefore be able to be re-evaluated as new injections from the current batch is expanding the reference population. Owing to this, a functionality to continuously re-evaluate injection quality with every addition of a new injection within an experiment. To properly visualize this re-evaluation, what is referred to as triangle plots were introduced (Fig. 14). These triangle plots contain one injection for each row, and the updated evaluation of that injection along the horizontal axis. Thus, the diagonal represents the initial evaluation of each injection at the time of injection and every subsequent point to the right represents how that injection behaves as more injection are added to the distribution. This

operation and plot are used to interpret and visualize all the *LaMa*-based metrics and for the final scoring summary.

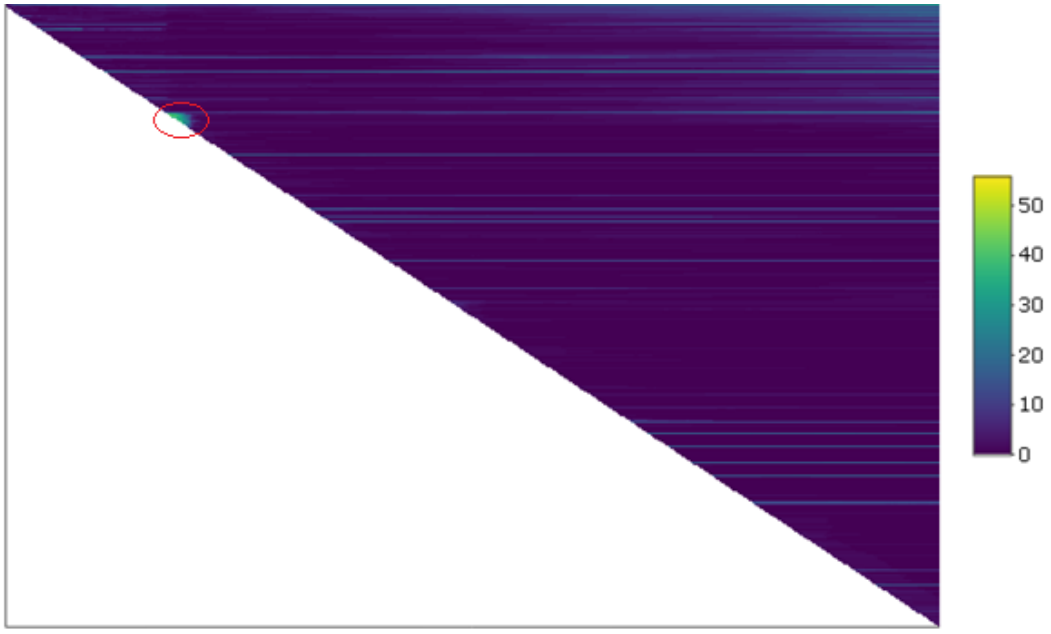

Figure 14. *Triangle plot showing % of LaMas that are outliers in intensity compared to the rest of the population using the 'optimal case' QC-data in the RP analytical mode. Each row corresponds to the performance of a single injection, while every column represents the time point of every injection. Almost every injection was classified as being of good quality (indicated by blue colour). Samples at the beginning of batch 2 (indicated by a red circle) were initially evaluated as of potentially poor quality (lighter colour). However, as the batch progressed, more samples from the same batch were added to the distribution and the evaluation updated towards well-behaved.*

Finally, the scoring function is visualized in a similar manner to that of the injection-based metrics. Figure 15 shows final scoring plots for a) the 'optimal case' data represented by the *QC*-data set.

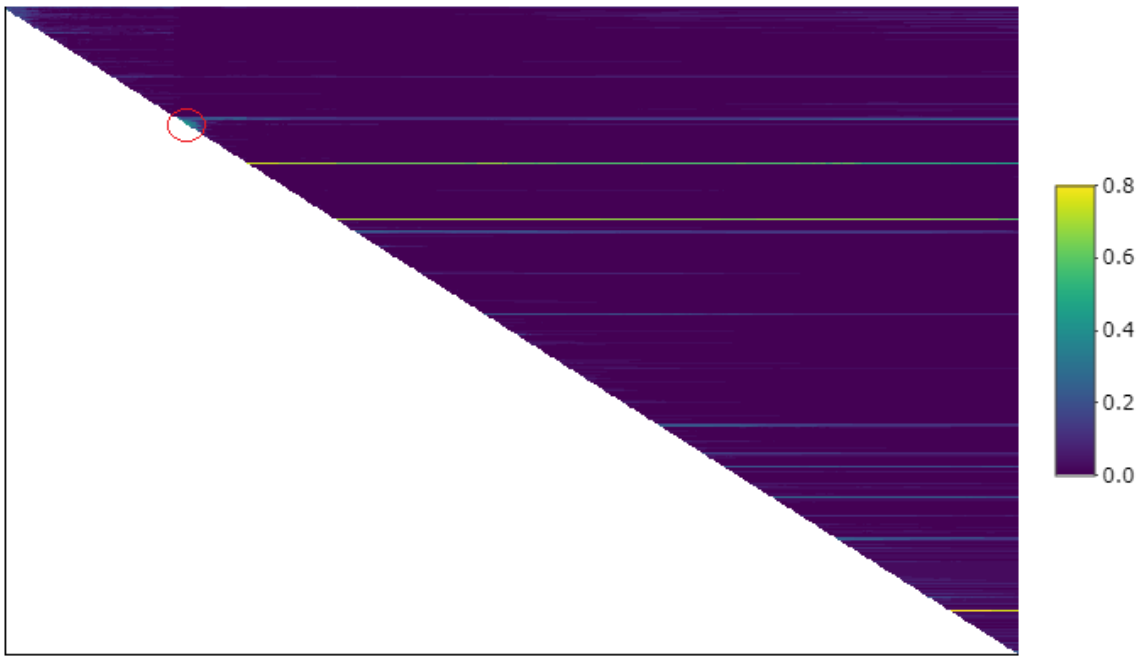

Figure 15*. Final scoring plot for 'optimal case' data represented by the QC data set in the RP analytical mode. Nearly all injections were rated as high quality, with a few erroneous injections, likely caused by instrument malfunction. Similar to Fig. 14, samples at the beginning of batch 2 (indicated by a red circle) were initially evaluated as of potentially poor quality (lighter colour). However, as the batch progressed, more samples from the same batch were added to the distribution and the evaluation updated towards well-behaved.*

When instead applying *QualiMon* to the *MedGICarb* data, i.e., an authentic dataset with all aspects of undesired variability from sampling, pre-analytical sample management as well as sample preparation, the variation in injection quality greatly increases (Fig. 16, 17, and 18).

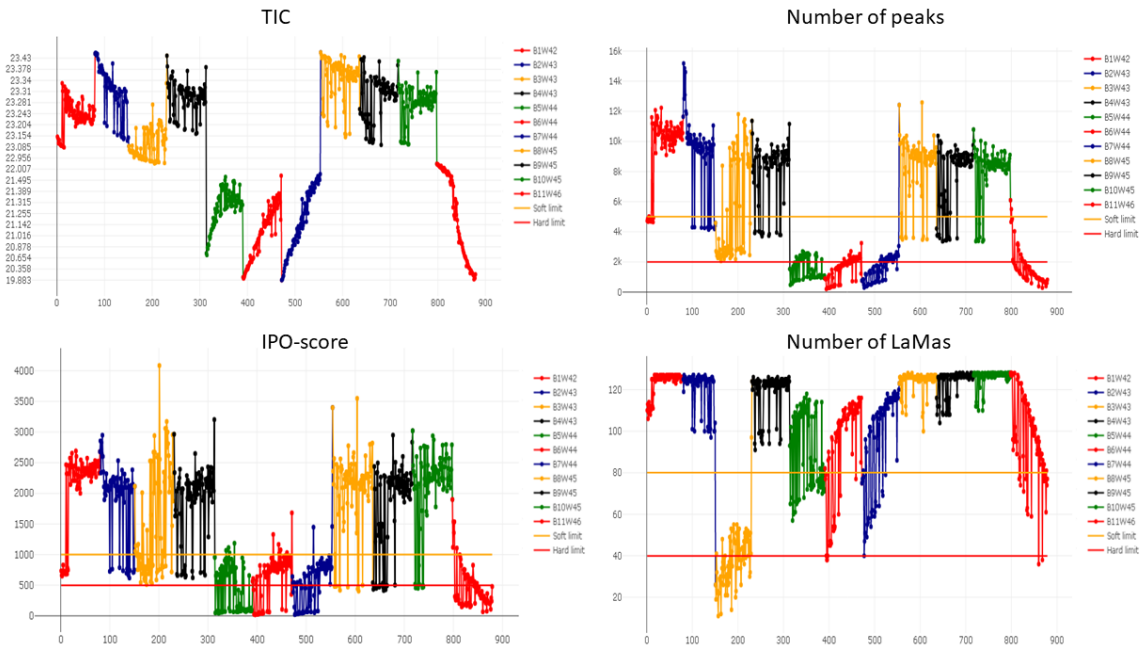

***Figure 16**. Injection-based metrics for real-case data represented by the MedGICarb data set. Note that batches 3, 5, 6, 7 and 11 shown in these plots were rerun as they were found to be of poor quality during manual checking.*

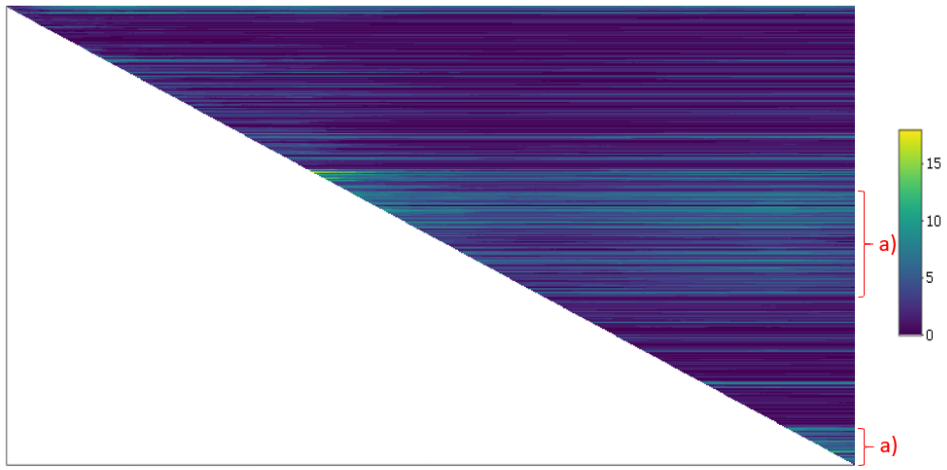

***Figure 17.** Triangle plot showing % of LaMas that are outliers for the tailing factor metric compared to the rest of the population using the 'real case' MedGICarb data in the RP analytical mode. Several low scoring injections can be found in the middle and end of the experiment (brackets marked with a, i.e., batches 5, 6, 7 and 11)*

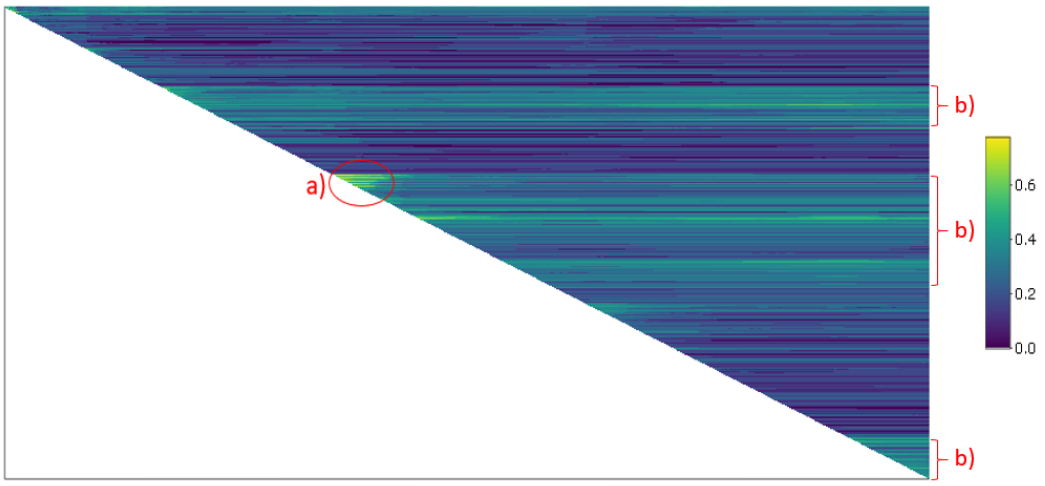

***Figure 18**. Status plot for MedGICarb-data in RP analytical mode. The red circle indicated by a) shows a typical behaviour in the beginning of a new batch where samples are first classified as very poor, to increase in quality score as more injections from the same batch are added to the distribution. Batches 3, 5, 6, 7 and 11 marked by b) were deemed to be of poor quality by manual checking at time of analysis and were rerun. While from the plots presented in this thesis, it is difficult to see which injections are of poor quality, the zooming, and mouse-over features in QualiMon makes it easy to quickly identify which injections to give further attention to.*

Several reasons exist to why the quality differs greatly between the 'optimal case' data (*QC-data set*) and authentic data, represented by the *MedGICarb-data set*: The *MedGICarb*-data stems from a real population and the biological variation is greater compared to the *QC*-data where only 3 samples were injected over and over. Even if standard operating procedures are in place and adhered to, an increase in variation will also come from the sampling, pre-analytical sample management and the sample preparation. Figures 16, 17 and 18 clearly show that 5 batches had severe problem related to data quality. These batches were consequently rerun.

The large number of *LaMa* peaks examined by *QualiMon* gives a much wider basis for assessing peak and data quality control compared to manual inspection where in reality only a few (normally 5-10) peaks are checked. This extensive testing is made possible by the automated metric extraction and assessment for a large number of *LaMas*. *QualiMon* further excels in injection quality assessment compared to post-batch QC as the quality is assessed immediately post-injection and not post-batch or even post-experiment. Moreover, it does not rely on QC-samples as a proxy for quality, but instead checks the quality of every single injection in and of itself. It should however be stressed that, *QualiMon* does not work as a replacement for QC-samples in an experiment, but rather as a complement: QC-samples are still required for adjustment of instrument drift and to exclude features that cannot be measured in a stable way.

QC-samples are further encouraged to combine with *QualiMon* usage, as the algorithm handles QC-injections separately from the normal injections. As the QC-samples within an experiment are *de facto* replicates, the reference populations will have less variability than the study

samples (i.e., they should be similar to the output of the *QC*-data set) and any abnormalities relating to instrumentation should be easier to identify in this sample type.

Automated monitoring

Following the set-up of the *QualiMon* software, the entire procedure is fully automated, from reading generated instrument files, optional data backup and quality assessment to sending real-time notifications to instrument operators via slack. *QualiMon* uses the open file format mzML to read instrument files which in turn are generated using Proteowizard (182). This makes the process vendor-agnostic which has a large positive impact on the availability and contributes further to open science. Open science is further promoted through making all code publicly available and modifications or adaptions into other usages from the community is greatly encouraged.

The automated file backup system addresses a major security issue that enables to not be reliant on manual batch-wise backups and the automated notification system leads to operators getting real-time notifications about injections potentially of poor quality resulting in an immediate possibility to investigate the issue hands on. This immediate response can reduce sample waste by not having to inject samples whilst the instrumental conditions are poor and can reduce the amount of valuable instrumental time generating data of poor quality. As both samples and instrument are scarce resources, *QualiMon* can reduce man hours required by an operator as well as streamline the instrumental usage.

QualiMon is and ongoing project and several updates regarding both functionality, efficiency and user friendliness are planned over the course of the coming year.

6 GENERAL DISCUSSION

Prospective cohort studies that collect information prior to disease onset have unique potential to study mechanisms in disease pathogenesis (183). However, an issue with prospective sampling is that it is difficult to estimate how many people will develop the disease in question in the sample population. In the available data, no associations could be identified between metabolome and allergy. This was true for all measured metabolomes (maternal metabolome during gestational week 28, maternal metabolome during delivery and umbilical cord metabolome) with all allergy outcomes (atopic dermatitis, food allergy and asthma). This lack of associations could possibly be due to the early timepoint of diagnosis, i.e., 12 months of age. Prolonging the follow-up time could give a different distribution of allergy cases in the sample population, which could give a two-fold advantage from a data analytical perspective: first, it would give more power to the analyses due to a larger case population and second, it would remove prospective cases from the control population. With this change of basis for analysis, other results could be obtained and potential associations with allergy development in children could be identified. As the metabolome is vast, there could also be metabolic features not covered by the methods used in this study that associates to the future allergy development. Another explanation to the null associations could relate to allergy being a multifaceted disease: There are several determining factors as to why allergy emerges, and there may not be sufficient homogeneity in potentially affected mechanisms or pathways to obtain meaningful predictions (184). The major strength of the present studies and the NICE-cohort as a whole is that allergy (atopic eczema, food allergy and asthma) was physician-diagnosed based on strict criteria and that all the children met with the same paediatrician specialized in allergy. Furthermore, the collection of biologic specimens was extensive, including blood samples from mothers during pregnancy and at delivery, umbilical cord blood (separated in arterial and venous in a subgroup of children), placenta, breast milk, urine, faeces, and saliva, enabling studies combining metabolomics from several time points and sample matrices, but also offers the potential to combine several different ‘omics’ methodologies.

Some associations were observed for both the plasma and placenta metabolomes with immune maturation markers. However, these associations were not very strong and could not easily be translated from the maternal and umbilical cord plasma metabolomes to the placenta metabolome or vice versa. It should also be noted that several hundred models were created to test for associations between the metabolomes and the different stages of immune maturation. However, adjustment for multiple testing in this scenario is difficult since models were tested for several outcomes. However, not all of these were independent (e.g., the total number of B cells are dependent on the number of memory B cells and the proportion of memory B cells of total B cells are dependent on both). Hence, we chose to report nominal p-values from permutation tests for all multivariate models. Among the observed associations, only memory B cells could be linked to more than one of the metabolomes (maternal plasma in connection to delivery and umbilical cord plasma), indicating that these associations represent actual

relationships. However, neither of the identified associations in the plasma and the placenta metabolome could be replicated in the other sample matrix, suggesting a substantial risk of the other observed associations being spurious.

Placenta was chosen as a biospecimen from the assumption that localized tissue could be advantageous as the biochemical processes in situ likely have a stronger finger print than metabolites in circulation (185). Moreover, the placenta is the organ that constitutes the border between the maternal and neonatal circulatory systems, which made it a relevant candidate to investigate potential associations with immune maturation. However, this hypothesis did not appear to hold as we observed that the intensities of potential features of interest were low. Moreover, none of the features of interest found in plasma could be found within the placenta. Together, these findings suggest that potential metabolites of interest do not stem from nor are enriched in the placenta. Furthermore, BMI and age are commonly used in our lab as sanity checks for blood plasma, but no such associations were found in the placenta metabolome. This lack of associations could point either towards that the placenta metabolome contains little valuable information about the measured outcomes or towards a lack of established best practices for working with placenta metabolomics and might stem from problems relating to not knowing how to manage samples, sample matrix heterogeneity or how to normalize data.

As touched upon previously, one of the major limitations in this work relates to small sample size. The delivery of a child is something that can occur at any hour. However, research staff is usually only available for sample management during normal office hours. In an effort to reduce the contribution from undesired variability from varying sample management conditions (predominantly related to prolonged pre-centrifugation delay times), only placentas that were obtained within 4 hours of delivery were used for metabolomics analyses, causing only 96 out of originally 414 placentas to be suitable for analysis. Although plasma samples were obtained from almost all mothers during pregnancy, fewer were available from time of delivery and fewer still from the umbilical cord blood. Since delivery is a stressful situation for both the mother and the staff, sampling for research purpose becomes a secondary issue if complications arise. In addition, for the immune maturation analyses, several of the measured cell types were too few in many of the samples to enable counting, leading to a substantial loss of statistical power for several of these measures. Further, taking blood from infants can be a daunting task, and for several children at varying time points it was not possible to take blood samples. For the allergy diagnoses, there were few children with allergy diagnoses at one year of age. When combining loss of obtainable information from both the outcomes and the samples, the end result is a lower than desired statistical power in the data analysis.

Major differences were observed between the arterial and the venous umbilical cord blood metabolomes. Moreover, the observed association between the venous umbilical cord blood plasma and parity and corresponding null association for arterial cord blood further supports the notion of the two different umbilical cord sample types being systematically different, showing that the choice of sample matrix is of high importance when studying the umbilical cord metabolome. Furthermore, the association with parity was weak for the placenta metabolome. This suggests that parity is mainly determined by the maternal metabolism and that parity as a covariate may be more important to adjust for when working with venous

umbilical cord plasma than when using placenta or arterial umbilical cord plasma. The weak association between the placenta metabolome and parity further supports that the lack of experience working with placenta metabolomics in our lab might have affected the results negatively. The weak association could also be explained by the fact that the metabolites driving the association in the venous blood were mainly related to energy metabolism. These energy metabolites are essential for the child and might thus not be retained by the placenta but rather can pass through quickly and thus not be identified.

Despite the weak results in relation to allergy and immune maturation, untargeted metabolomics has a considerable potential when it comes to evaluating health outcomes and other biological phenomena. Furthermore, recent advances in both the analytical and the computational fields are continuing to push the limits of metabolomics (186). Along these lines, new and vastly improved versions of open-source software have been released (e.g., MZmine 3.0 and MS-Dial 5.0 (manuscripts expected to be released shortly)). Although there is much progress in the field, major issues remain concerning e.g., reproducibility, comparability and identification (187). The studies conducted in this thesis points towards several of these difficulties: results from previous studies can frequently not be replicated [Papers I and III], partially due to the lack of overlapping metabolite coverage. Metabolites are difficult to identify [Papers I and II]. The data quality of injections in untargeted LC-MS based metabolomics is frequently uneven or even poor [paper IV].

To address this last issue, we developed procedures and an accompanying R-based software *QualiMon* that analyses several hundred traits on both the level of injection as a whole as well as specific peak characteristics from so-called landmark features. *QualiMon* serves as a complement to the current best QC/QA practices, and it should be stressed that using *QualiMon* alone does not ensure perfect quality: Rigid QA practices are still needed, and QC should be performed at the end of a batch or experiment to ensure the validity of the data. The developed software is capable of continuous monitoring to enable near real-time detection of injections of low quality and thereby alert operators of potential instrument-related deviations. In line with the strong movement towards open science, including sharing of protocols, methods, data, and algorithms, *QualiMon* is delivered through an open licence (MIT), and may therefore contribute to an increased awareness of quality issues and hopefully to improved data quality in untargeted metabolomics. Implementation of *QualiMon* can free up operator time from manually assessing peak and injection quality, improve sample and instrument usage by not having to re-perform low quality injections at a later stage and, importantly, improve data quality for downstream data analysis. Although functional, several improvements to *QualiMon* are planned for an enhanced user experience, to broaden its usability (including e.g., support for different types of reference materials (128) as well as MS2-level data) and to improve upon existing metrics.

7 MAIN FINDINGS AND CONCLUSIONS

The work in this thesis aimed to identify associations of maternal and neonatal metabolomes before and during birth with allergy development at one year of age and to associate these metabolomes with immune maturation. It further aimed to explore differences in arterial and venous umbilical cord blood as well as to see how these relate to maternal and neonatal demographics. Finally, the work aimed to develop procedures for automated quality monitoring in untargeted LC-MS metabolomics.

Even though the study had a large metabolite coverage, in a relatively large cohort and used robust statistical methodology, no associations were found for allergy development with either the maternal or neonatal metabolomes. The null findings can likely be attributed to low power due to few cases, diluted associations due to potential future allergy development or that no associations can be found this early in life. Although no associations were found, this is one of the largest prospective cohorts performed on prospective allergy development, and future analyses of the different collected biospecimen not yet analysed might still reveal associations to future allergy development.

Modest associations were found when relating various subpopulations of T and B cells to both the plasma and placenta metabolomes. Most notably, phenylacetylglutamine in both the maternal plasma at delivery and the umbilical cord metabolome associated with the formation of memory B cells at 4 months of age, suggesting that potentially immunomodulatory metabolites are transferred from the mother to the child.

Major differences between arterial and venous umbilical cord blood were found, mainly pertaining to energy and amino acid metabolism. This finding highlights the need of a proper study design and sample strategy prior to the start of a project. Furthermore, the venous cord blood associated with the maternal parity status, indicating that this factor is important to take into consideration when adjusting statistical models or when matching controls to cases.

Due to issues raised with injection quality during this work, this thesis further aimed to develop a quality monitoring aimed to combat the inherent variation in untargeted LC-MS based metabolomics, by developing procedures based on landmark features for automated real-time quality monitoring of injections. When used on data previously run on our instruments, it could distinguish batches and single injections of poor and good quality from each other, and the open source and open licence offer the possibility of enhancing quality monitoring procedures for the metabolomics community.

8 FUTURE RESEARCH

Given that allergy diagnoses were only available at one year of age, and that the diagnoses will likely change over the following years, it would be of great interest to reperform analyses regarding allergy development when diagnoses from 4 years are available or at even later time points.

As the data on immune maturation suffered from low sample sizes for several of the outcomes, it would be of great interest to investigate these markers in a larger study population and with more overlapping immune measurements. Furthermore, attempted replication of observed associations would be needed to confirm whether associations are real or if they are stemming from spurious findings. As potential immunomodulatory metabolites were only present at low concentrations, other omics methodologies could offer better insights into mechanisms at the maternal-foetal interface regulating immune maturation.

QualiMon, the software developed for automated quality monitoring developed in this thesis, is currently at a relatively early stage in its development, and several improvements and additional features are planned. For one, the current processing speed of the procedures is too slow to handle large scale metabolomics experiments with a large *LaMa* coverage (number of *LaMas* > 100), thus more efficient computing is needed to facilitate this. Currently, if several different sample matrices are being analysed on the same instrument, manual intervention is needed to switch between configuration files. For future versions, an option to identify what sample matrix is being analysed based on the project name will be added.

At present, the quality monitoring procedure is entirely data driven. However, options to include reference materials in quality monitoring would further increase the performance of the quality assessment as well as reduce the need for *a priori* existing data to set up the procedures. *QualiMon* was designed with untargeted LC-MS metabolomics in mind, but the concept can be applied to targeted methods as well as to GC-MS based methods as well and development for a broadened use is planned for the future.

9 ACKNOWLEDGEMENTS

First of all I would like to thank my supervisors. Calle, your knowledge, enthusiasm and passion has been a constant aid and motivation during these past four years. The devise “if you’re not having fun, you’re probably doing something wrong” is something that I will always carry with me. Malin, for contributing with a lot of accurate insights, even though the topic has not always been within your general competence area and for plenty of fun moments throughout these years. And Assi, for contributing with your experience and insights into the academic world, and some unforgettable moments including breaking taxi doors.

I would like to thank all the funders who have been a part of funding the various projects I have had the privilege to be involved in over these years: Research and Innovation Unit at Region Norrbotten and Swedish Research Council for Health, Working Life and Welfare (FORTE) (Grant No. 2018-00485), Swedish Research Council (Grant No. 521-2013-3154 and Grant No 2019-0137, 2019-01264), Swedish Research Council Formas (Grant No. 2018-02275), Västra Götaland Region (RUN) (Grant No. 612-0618-15), Jane och Dan Olssons stiftelse (Grant No. 2020-23).

I further extend my gratitude to all my office mates throughout the years. Most notably Mia S who has been with me both as a part of the NICE-group and as an office mate since we did our master’s theses, without you, this journey would not have been half as fun as it has been. A further thanks to all the people in the computational metabolomics group at food and nutrition science. The group mentality and sharing of knowledge, ideas and services represents exactly what research is all about! Here I would like to give a special thanks to Anton R, without whom, this last year would not have been possible. Working together to develop *QualiMon* has been the most fun I’ve ever had working, so I guess we must have done something right.

Thank you to all the people working in the NICE-group, for being welcoming and for putting the research into a broader perspective with a lot of good times intermingled with the science. A special thanks to Agnes Wold and Hardis Rabe, for all the help regarding the immune system.

Thank you, Alastair Ross, for introducing me to the wonderous world that is Metabolomics. Thank you to Professor Rikard Landberg for very insightful and challenging questions over the course of the years. And a very special thank you to all the colleagues at the food and nutrition division, you are all amazing people, and it has been a pleasure to work amongst such competent and genuinely nice people.

To all the people at CMSI, for all the help with the chemical analyses, thank you so much, I would not have managed without you guys!

To all my friends, I cannot express how much all of you have meant to me, you have all been constant reminders that work is merely work, and that there are things more important than working throughout the evenings. I will not name any names, as the list would grow too long, but you are all amazing people, and I am incredibly lucky to be surrounded by so many fantastic individuals.

My family. My mother and father, thank you for the continuous support and encouragement to do whatever I feel like doing. To my brother, for the general science discussions, moral support, and the terrible puns that you have provided, thank you. To Vanessa, my dear partner, who would also classify under the friend, former office mate and CMSI categories, I think this is where you belong. For being my closest family during these last years, for the support, the love and the joy that you have brought, you have my deepest gratitude. And to all of my family, the knowledge of you always being there and will continue to be so in the future, has made everything so much easier. Thank you all so much.

10 REFERENCES

1. Pawankar R, Canonica G, Holgate S, Lockey R, Blaiss M. WAO white book on allergy: update 2013. World Allergy Organization. 2013;248.
2. Wold A, Mölne J. Inflammationssjukdomar: Liber AB; 2011.
3. Zuberbier T, Lötvall J, Simoens S, Subramanian S, Church MK. Economic burden of inadequate management of allergic diseases in the European Union: a GA2LEN review. Allergy. 2014;69(10):1275-9.
4. Strachan DP. Hay fever, hygiene, and household size. BMJ: British Medical Journal. 1989;299(6710):1259.
5. Berton R, Conceição MS, Libardi CA, Canevarolo RR, Gáspari AF, Chacon-Mikahil MPT, et al. Metabolic time-course response after resistance exercise: A metabolomics approach. Journal of sports sciences. 2017;35(12):1211-8.
6. Andersson Hall U, Edin F, Pedersen A, Madsen K. Whole-body fat oxidation increases more by prior exercise than overnight fasting in elite endurance athletes. Applied Physiology, Nutrition, and Metabolism. 2016;41(4):430-7.
7. Guertin KA, Moore SC, Sampson JN, Huang W-Y, Xiao Q, Stolzenberg-Solomon RZ, et al. Metabolomics in nutritional epidemiology: identifying metabolites associated with diet and quantifying their potential to uncover diet-disease relations in populations. The American journal of clinical nutrition. 2014;100(1):208-17.
8. Jin L, Godri Pollitt KJ, Liew Z, Rosen Vollmar AK, Vasiliou V, Johnson CH, et al. Use of untargeted metabolomics to explore the air pollution-related disease continuum. Current Environmental Health Reports. 2021;8(1):7-22.
9. Markin PA, Brito A, Moskaleva NE, Tagliaro F, Tarasov VV, La Frano MR, et al. Short- and medium-term exposures of diazepam induce metabolomic alterations associated with the serotonergic, dopaminergic, adrenergic and aspartic acid neurotransmitter systems in zebrafish (Danio rerio) embryos/larvae. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics. 2021;38:100816.
10. Tan YM, Gao Y, Teo G, Koh HW, Tai ES, Khoo CM, et al. Plasma metabolome and lipidome associations with type 2 diabetes and diabetic nephropathy. Metabolites. 2021;11(4):228.
11. Schmidt DR, Patel R, Kirsch DG, Lewis CA, Vander Heiden MG, Locasale JW. Metabolomics in cancer research and emerging applications in clinical oncology. CA: A Cancer Journal for Clinicians. 2021;71(4):333-58.
12. McGarrah RW, Crown SB, Zhang G-F, Shah SH, Newgard CB. Cardiovascular metabolomics. Circulation research. 2018;122(9):1238-58.
13. Wilkins JM, Trushina E. Application of metabolomics in Alzheimer’s disease. Frontiers in neurology. 2018;8:719.
14. Dudley E, Yousef M, Wang Y, Griffiths W. Targeted metabolomics and mass spectrometry. Advances in protein chemistry and structural biology. 2010;80:45-83.
15. Want EJ, O'Maille G, Smith CA, Brandon TR, Uritboonthai W, Qin C, et al. Solvent-dependent metabolite distribution, clustering, and protein extraction for serum profiling with mass spectrometry. Analytical chemistry. 2006;78(3):743-52.
16. Ren J-L, Zhang A-H, Kong L, Wang X-J. Advances in mass spectrometry-based metabolomics for investigation of metabolites. RSC advances. 2018;8(40):22335-50.

17. Dunn WB, Broadhurst DI, Edison A, Guillou C, Viant MR, Bearden DW, et al. Quality assurance and quality control processes: summary of a metabolomics community questionnaire. Metabolomics. 2017;13(5):1-6.
18. Carraro S, Rezzi S, Reniero F, Héberger K, Giordano G, Zanconato S, et al. Metabolomics applied to exhaled breath condensate in childhood asthma. American Journal of Respiratory and Critical Care Medicine. 2007;175(10):986-90.
19. Crestani E, Harb H, Charbonnier L-M, Leirer J, Motsinger-Reif A, Rachid R, et al. Untargeted metabolomic profiling identifies disease-specific signatures in food allergy and asthma. Journal of Allergy and Clinical Immunology. 2019.
20. Turi KN, McKennan C, Gebretsadik T, Snyder B, Seroogy CM, Lemanske Jr RF, et al. Unconjugated bilirubin is associated with protection from early-life wheeze and childhood asthma. Journal of Allergy and Clinical Immunology. 2021;148(1):128-38.
21. Huang M, Kelly RS, Chu SH, Kachroo P, Gürdeniz G, Chawes BL, et al. Maternal metabolome in pregnancy and childhood asthma or recurrent wheeze in the Vitamin D Antenatal Asthma Reduction Trial. Metabolites. 2021;11(2):65.
22. Delves PJ, Roitt IM. The immune system. New England journal of medicine. 2000;343(1):37-49.
23. Kumar H, Kawai T, Akira S. Pathogen recognition by the innate immune system. International reviews of immunology. 2011;30(1):16-34.
24. Rao KN, Brown MA. Mast cells: multifaceted immune cells with diverse roles in health and disease. Annals of the New York Academy of Sciences. 2008;1143(1):83-104.
25. Adams NM, Grassmann S, Sun JC. Clonal expansion of innate and adaptive lymphocytes. Nature Reviews Immunology. 2020;20(11):694-707.
26. Crotty S. A brief history of T cell help to B cells. Nature Reviews Immunology. 2015;15(3):185-9.
27. Brown J, Wilson T, Metcalfe D. The mast cell and allergic diseases: role in pathogenesis and implications for therapy. Clinical & Experimental Allergy. 2008;38(1):4-18.
28. Serana F, Chiarini M, Zanotti C, Sottini A, Bertoli D, Bosio A, et al. Use of V (D) J recombination excision circles to identify T-and B-cell defects and to monitor the treatment in primary and acquired immunodeficiencies. Journal of translational medicine. 2013;11(1):119.
29. Douek DC, McFarland RD, Keiser PH, Gage EA, Massey JM, Haynes BF, et al. Changes in thymic function with age and during the treatment of HIV infection. Nature. 1998;396(6712):690-5.
30. Van Zelm MC, Szczepański T, Van Der Burg M, Van Dongen JJ. Replication history of B lymphocytes reveals homeostatic proliferation and extensive antigen-induced B cell expansion. The Journal of experimental medicine. 2007;204(3):645-55.
31. Fiocchi A, Brozek J, Schünemann H, Bahna SL, Von Berg A, Beyer K, et al. World Allergy Organization (WAO) diagnosis and rationale for action against cow's milk allergy (DRACMA) guidelines. World Allergy Organization Journal. 2010;3(4):57-161.
32. Bylund S, von Kobyletzki LB, Svalstedt M, Svensson Å. Prevalence and incidence of atopic dermatitis: a systematic review. Acta dermato-venereologica. 2020;100(12):320-9.
33. Spergel JM, Paller AS. Atopic dermatitis and the atopic march. Journal of Allergy and Clinical Immunology. 2003;112(6):S118-S27.
34. Otsuka A, Nomura T, Rerknimitr P, Seidel JA, Honda T, Kabashima K. The interplay between genetic and environmental factors in the pathogenesis of atopic dermatitis. Immunological reviews. 2017;278(1):246-62.
35. Anvari S, Miller J, Yeh C-Y, Davis CM. IgE-mediated food allergy. Clinical reviews in allergy & immunology. 2019;57(2):244-60.

36. Williams H, Jburney P, Hay R, Archer C, Shipley M, Ahunter J, et al. The UK Working Party's Diagnostic Criteria for Atopic Dermatitis. I. Derivation of a minimum set of discriminators for atopic dermatitis. British journal of dermatology. 1994;131(3):383-96.
37. Williams H, Jburney P, Pembroke A, Hay R, Party ADDCW. The UK Working Party's diagnostic criteria for atopic dermatitis. III. Independent hospital validation. British journal of dermatology. 1994;131(3):406-16.
38. Williams H, JBURNEY P, Strachan D, Hay R, Party ADDCW. The UK Working Party's Diagnostic Criteria for Atopic Dermatitis II. Observer variation of clinical diagnosis and signs of atopic dermatitis. British journal of dermatology. 1994;131(3):397-405.
39. Jonsson K, Barman M, Moberg S, Sjöberg A, Brekke HK, Hesselmar B, et al. Fat intake and breast milk fatty acid composition in farming and nonfarming women and allergy development in the offspring. Pediatric research. 2016;79(1):114-23.
40. Loss G, Apprich S, Waser M, Kneifel W, Genuneit J, Büchele G, et al. The protective effect of farm milk consumption on childhood asthma and atopy: the GABRIELA study. Journal of Allergy and Clinical Immunology. 2011;128(4):766-73. e4.
41. Hesselmar B, Hicke-Roberts A, Lundell A-C, Adlerberth I, Rudin A, Saalman R, et al. Pet-keeping in early life reduces the risk of allergy in a dose-dependent fashion. PloS one. 2018;13(12).
42. Al-Tamprouri C, Malin B, Bill H, Lennart B, Anna S. Cat and dog ownership during/after the first year of life and risk for sensitization and reported allergy symptoms at age 13. Immunity, inflammation and disease. 2019;7(4):250-7.
43. Allen M, Heinzmann A, Noguchi E, Abecasis G, Broxholme J, Ponting CP, et al. Positional cloning of a novel gene influencing asthma from chromosome 2q14. Nature genetics. 2003;35(3):258-63.
44. Palmer CN, Irvine AD, Terron-Kwiatkowski A, Zhao Y, Liao H, Lee SP, et al. Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. Nature genetics. 2006;38(4):441-6.
45. Baurecht H, Irvine AD, Novak N, Illig T, Bühler B, Ring J, et al. Toward a major risk factor for atopic eczema: meta-analysis of filaggrin polymorphism data. Journal of Allergy and Clinical Immunology. 2007;120(6):1406-12.
46. Amoli M, Hand S, Hajeer A, Jones K, Rolf S, Sting C, et al. Polymorphism in the STAT6 gene encodes risk for nut allergy. Genes & Immunity. 2002;3(4):220-4.
47. Sénéchal H, Geny S, Desvaux F-X, Busson M, Mayera C, Arond Y, et al. Genetics and specific immune response in allergy to birch pollen and food: evidence of a strong, positive association between atopy and the HLA class II allele HLA-DR7. Journal of allergy and clinical immunology. 1999;104(2):395-401.
48. Litonjua AA, Carey VJ, Burge HA, Weiss ST, Gold DR. Parental history and the risk for childhood asthma: does mother confer more risk than father? American journal of respiratory and critical care medicine. 1998;158(1):176-81.
49. Wright AL, Holberg C, Taussig L, Martinez F. Factors influencing the relation of infant feeding to asthma and recurrent wheeze in childhood. Thorax. 2001;56(3):192-7.
50. Stråvik M, Barman M, Hesselmar B, Sandin A, Wold AE, Sandberg A-S. Maternal Intake of Cow’s Milk during Lactation Is Associated with Lower Prevalence of Food Allergy in Offspring. Nutrients. 2020;12(12):3680.
51. Calvani M, Alessandri C, Sopo SM, Panetta V, Pingitore G, Tripodi S, et al. Consumption of fish, butter and margarine during pregnancy and development of allergic sensitizations in the offspring: role of maternal atopy. Pediatric Allergy and Immunology. 2006;17(2):94-102.

52. Sausenthaler S, Kompauer I, Borte M, Herbarth O, Schaaf B, von Berg A, et al. Margarine and butter consumption, eczema and allergic sensitization in children. The LISA birth cohort study. Pediatric allergy and immunology. 2006;17(2):85-93.
53. de LSGSSKSSBLIBMHOvBAWH-EHJjhg. Maternal diet during pregnancy in relation to eczema and allergic sensitization in the offspring at 2 y of age. The American journal of clinical nutrition. 2007;85(2):530-7.
54. Nwaru BI, Ahonen S, Kaila M, Erkkola M, Haapala AM, Kronberg-Kippilä C, et al. Maternal diet during pregnancy and allergic sensitization in the offspring by 5 yrs of age: a prospective cohort study. Pediatric Allergy and Immunology. 2010;21(1-Part-I):29-37.
55. Netting MJ, Middleton PF, Makrides M. Does maternal diet during pregnancy and lactation affect outcomes in offspring? A systematic review of food-based approaches. Nutrition. 2014;30(11-12):1225-41.
56. Patti GJ, Yanes O, Siuzdak G. Metabolomics: the apogee of the omics trilogy. Nature reviews Molecular cell biology. 2012;13(4):263-9.
57. Oliver SG, Winson MK, Kell DB, Baganz F. Systematic functional analysis of the yeast genome. Trends in biotechnology. 1998;16(9):373-8.
58. Ramautar R, Berger R, van der Greef J, Hankemeier T. Human metabolomics: strategies to understand biology. Current opinion in chemical biology. 2013;17(5):841-6.
59. Deja S, Porebska I, Kowal A, Zabek A, Barg W, Pawelczyk K, et al. Metabolomics provide new insights on lung cancer staging and discrimination from chronic obstructive pulmonary disease. Journal of pharmaceutical and biomedical analysis. 2014;100:369-80.
60. Peng J, Chen Y-T, Chen C-L, Li L. Development of a universal metabolome-standard method for long-term LC–MS metabolome profiling and its application for bladder cancer urine-metabolite-biomarker discovery. Analytical chemistry. 2014;86(13):6540-7.
61. Qiu Y, Cai G, Zhou B, Li D, Zhao A, Xie G, et al. A Distinct Metabolic Signature of Human Colorectal Cancer with Prognostic PotentialMetabolic Signature of Human Colorectal Cancer. Clinical cancer research. 2014;20(8):2136-46.
62. La Torre D, Seppänen-Laakso T, Larsson HE, Hyötyläinen T, Ivarsson SA, Lernmark Å, et al. Decreased cord-blood phospholipids in young age–at–onset type 1 diabetes. Diabetes. 2013;62(11):3951-6.
63. Wang-Sattler R, Yu Z, Herder C, Messias AC, Floegel A, He Y, et al. Novel biomarkers for pre-diabetes identified by metabolomics. Molecular systems biology. 2012;8(1):615.
64. Zordoky BN, Sung MM, Ezekowitz J, Mandal R, Han B, Bjorndahl TC, et al. Metabolomic fingerprint of heart failure with preserved ejection fraction. PloS one. 2015;10(5):e0124844.
65. Wang J, Li Z, Chen J, Zhao H, Luo L, Chen C, et al. Metabolomic identification of diagnostic plasma biomarkers in humans with chronic heart failure. Molecular BioSystems. 2013;9(11):2618-26.
66. Surowiec I, Karimpour M, Gouveia-Figueira S, Wu J, Unosson J, Bosson JA, et al. Multi-platform metabolomics assays for human lung lavage fluids in an air pollution exposure study. Analytical and bioanalytical chemistry. 2016;408(17):4751-64.
67. O'Sullivan A, Gibney MJ, Brennan L. Dietary intake patterns are reflected in metabolomic profiles: potential role in dietary assessment studies. The American journal of clinical nutrition. 2011;93(2):314-21.
68. Aubertin-Leheudre M, Koskela A, Samaletdin A, Adlercreutz H. Plasma alkylresorcinol metabolites as potential biomarkers of whole-grain wheat and rye cereal fibre intakes in women. British journal of nutrition. 2010;103(3):339-43.
69. Landberg R, Townsend MK, Neelakantan N, Sun Q, Sampson L, Spiegelman D, et al. Alkylresorcinol metabolite concentrations in spot urine samples correlated with whole grain

and cereal fiber intake but showed low to modest reproducibility over one to three years in US women. The Journal of nutrition. 2012;142(5):872-7.
70. Atkinson W, Downer P, Lever M, Chambers ST, George PM. Effects of orange juice and proline betaine on glycine betaine and homocysteine in healthy male subjects. European Journal of Nutrition. 2007;46(8):446-52.
71. Pujos-Guillot E, Hubert J, Martin J-F, Lyan B, Quintana M, Claude S, et al. Mass spectrometry-based metabolomics for the discovery of biomarkers of fruit and vegetable intake: citrus fruit as a case study. Journal of proteome research. 2013;12(4):1645-59.
72. Heinzmann SS, Brown IJ, Chan Q, Bictash M, Dumas M-E, Kochhar S, et al. Metabolic profiling strategy for discovery of nutritional biomarkers: proline betaine as a marker of citrus consumption. The American journal of clinical nutrition. 2010;92(2):436-43.
73. May DH, Navarro SL, Ruczinski I, Hogan J, Ogata Y, Schwarz Y, et al. Metabolomic profiling of urine: response to a randomised, controlled feeding study of select fruits and vegetables, and application to an observational study. British journal of nutrition. 2013;110(10):1760-70.
74. Lloyd AJ, Beckmann M, Favé G, Mathers JC, Draper J. Proline betaine and its biotransformation products in fasting urine samples are potential biomarkers of habitual citrus fruit consumption. British Journal of Nutrition. 2011;106(6):812-24.
75. Edmands WM, Beckonert OP, Stella C, Campbell A, Lake BG, Lindon JC, et al. Identification of human urinary biomarkers of cruciferous vegetable consumption by metabonomic profiling. Journal of proteome research. 2011;10(10):4513-21.
76. Andersen M-BS, Reinbach HC, Rinnan Å, Barri T, Mithril C, Dragsted LO. Discovery of exposure markers in urine for Brassica-containing meals served with different protein sources by UPLC-qTOF-MS untargeted metabolomics. Metabolomics. 2013;9(5):984-97.
77. Roberts LD, Souza AL, Gerszten RE, Clish CB. Targeted metabolomics. Current protocols in molecular biology. 2012;98(1):30.2. 1-.2. 24.
78. Schrimpe-Rutledge AC, Codreanu SG, Sherrod SD, McLean JA. Untargeted metabolomics strategies—challenges and emerging directions. Journal of the American Society for Mass Spectrometry. 2016;27(12):1897-905.
79. Kodra D, Pousinis P, Vorkas PA, Kademoglou K, Liapikos T, Pechlivanis A, et al. Is current practice adhering to guidelines proposed for metabolite identification in lc-ms untargeted metabolomics? a meta-analysis of the literature. Journal of Proteome Research. 2021;21(3):590-8.
80. Mlynárik V. Introduction to nuclear magnetic resonance. Analytical Biochemistry. 2017;529:4-9.
81. Kovats vE. Gas-chromatographische charakterisierung organischer verbindungen. Teil 1: retentionsindices aliphatischer halogenide, alkohole, aldehyde und ketone. Helvetica Chimica Acta. 1958;41(7):1915-32.
82. Ghaste M, Mistrik R, Shulaev V. Applications of fourier transform ion cyclotron resonance (FT-ICR) and orbitrap based high resolution mass spectrometry in metabolomics and lipidomics. International journal of molecular sciences. 2016;17(6):816.
83. Knolhoff AM, Callahan JH, Croley TR. Mass accuracy and isotopic abundance measurements for HR-MS instrumentation: capabilities for non-targeted analyses. Journal of The American Society for Mass Spectrometry. 2014;25(7):1285-94.
84. Haggarty J, Burgess KE. Recent advances in liquid and gas chromatography methodology for extending coverage of the metabolome. Current Opinion in Biotechnology. 2017;43:77-85.
85. Jung J, Kim SH, Lee HS, Choi G, Jung YS, Ryu D, et al. Serum metabolomics reveals pathways and biomarkers associated with asthma pathogenesis. Clinical & Experimental Allergy. 2013;43(4):425-33.

86. Ottas A, Fishman D, Okas T-L, Püssa T, Toomik P, Märtson A, et al. Blood serum metabolome of atopic dermatitis: Altered energy cycle and the markers of systemic inflammation. PloS one. 2017;12(11).
87. Huang Y, Chen G, Liu X, Shao Y, Gao P, Xin C, et al. Serum metabolomics study and eicosanoid analysis of childhood atopic dermatitis based on liquid chromatography–mass spectrometry. Journal of proteome research. 2014;13(12):5715-23.
88. Carraro S, Giordano G, Reniero F, Carpi D, Stocchero M, Sterk P, et al. Asthma severity in childhood and metabolomic profiling of breath condensate. Allergy. 2013;68(1):110-7.
89. Assfalg M, Bortoletti E, D'Onofrio M, Pigozzi R, Molinari H, Boner A, et al. An exploratory 1H-nuclear magnetic resonance metabolomics study reveals altered urine spectral profiles in infants with atopic dermatitis. British Journal of Dermatology. 2012;166(5):1123-5.
90. Ilves L, Ottas A, Kaldvee B, Abram K, Soomets U, Zilmer M, et al. Metabolomic analysis of skin biopsies from patients with atopic dermatitis reveals hallmarks of inflammation, disrupted barrier function and oxidative stress. Acta Dermato-Venereologica. 2021;101(2):adv00407-adv.
91. Saude EJ, Skappak CD, Regush S, Cook K, Ben-Zvi A, Becker A, et al. Metabolomic profiling of asthma: diagnostic utility of urine nuclear magnetic resonance spectroscopy. Journal of allergy and clinical immunology. 2011;127(3):757-64. e6.
92. Mattarucchi E, Baraldi E, Guillou C. Metabolomics applied to urine samples in childhood asthma; differentiation between asthma phenotypes and identification of relevant metabolites. Biomedical Chromatography. 2012;26(1):89-94.
93. Janssens M, van Smeden J, Gooris GS, Bras W, Portale G, Caspers PJ, et al. Increase in short-chain ceramides correlates with an altered lipid organization and decreased barrier function in atopic eczema patients. Journal of lipid research. 2012;53(12):2755-66.
94. Caldeira M, Perestrelo R, Barros AS, Bilelo M, Morete A, Câmara J, et al. Allergic asthma exhaled breath metabolome: a challenge for comprehensive two-dimensional gas chromatography. Journal of Chromatography A. 2012;1254:87-97.
95. Ibrahim B, Marsden P, Smith J, Custovic A, Nilsson M, Fowler S. Breath metabolomic profiling by nuclear magnetic resonance spectroscopy in asthma. Allergy. 2013;68(8):1050-6.
96. Checkley W, Deza MP, Klawitter J, Romero KM, Klawitter J, Pollard SL, et al. Identifying biomarkers for asthma diagnosis using targeted metabolomics approaches. Respiratory medicine. 2016;121:59-66.
97. Chiu CY, Lin G, Cheng ML, Chiang MH, Tsai MH, Su KW, et al. Longitudinal urinary metabolomic profiling reveals metabolites for asthma development in early childhood. Pediatric Allergy and Immunology. 2018;29(5):496-503.
98. Carraro S, Bozzetto S, Giordano G, El Mazloum D, Stocchero M, Pirillo P, et al. Wheezing preschool children with early-onset asthma reveal a specific metabolomic profile. Pediatric Allergy and Immunology. 2018;29(4):375-82.
99. Kelly RS, Sordillo JE, Lasky-Su J, Dahlin A, Perng W, Rifas-Shiman SL, et al. Plasma metabolite profiles in children with current asthma. Clinical & Experimental Allergy. 2018;48(10):1297-304.
100. Lee-Sarwar K, Kelly RS, Lasky-Su J, Moody DB, Mola AR, Cheng T-Y, et al. Intestinal microbial-derived sphingolipids are inversely associated with childhood food allergy. Journal of Allergy and Clinical Immunology. 2018;142(1):335-8. e9.
101. Tao JL, Chen YZ, Dai QG, Tian M, Wang SC, Shan JJ, et al. Urine metabolic profiles in paediatric asthma. Respirology. 2019;24(6):572-81.
102. Chawes BL, Giordano G, Pirillo P, Rago D, Rasmussen MA, Stokholm J, et al. Neonatal urine metabolic profiling and development of childhood asthma. Metabolites. 2019;9(9):185.

103. Lee-Sarwar KA, Kelly RS, Lasky-Su J, Zeiger RS, O'Connor GT, Sandel MT, et al. Integrative analysis of the intestinal metabolome of childhood asthma. Journal of Allergy and Clinical Immunology. 2019;144(2):442-54.
104. Chiu CY, Cheng ML, Chiang MH, Kuo YL, Tsai MH, Chiu CC, et al. Gut microbial-derived butyrate is inversely associated with IgE responses to allergens in childhood asthma. Pediatric Allergy and Immunology. 2019;30(7):689-97.
105. Chiu C-Y, Cheng M-L, Chiang M-H, Wang C-J, Tsai M-H, Lin G. Metabolomic Analysis Reveals Distinct Profiles in the Plasma and Urine Associated with IgE Reactions in Childhood Asthma. Journal of clinical medicine. 2020;9(3):887.
106. Ross AB, Barman M, Hartvigsson O, Lundell A-C, Savolainen O, Hesselmar B, et al. Umbilical cord blood metabolome differs in relation to delivery mode, birth order and sex, maternal diet and possibly future allergy development in rural children. Plos one. 2021;16(1):e0242978.
107. Fujiogi M, Camargo Jr CA, Raita Y, Zhu Z, Celedón JC, Mansbach JM, et al. Integrated associations of nasopharyngeal and serum metabolome with bronchiolitis severity and asthma: A multicenter prospective cohort study. Pediatric Allergy and Immunology. 2021;32(5):905-16.
108. Guo C, Sun L, Zhang L, Dong F, Zhang X, Yao L, et al. Serum sphingolipid profile in asthma. Journal of Leukocyte Biology. 2021;110(1):53-9.
109. Sharma R, Zang W, Zhou M, Schafer N, Begley LA, Huang YJ, et al. Real time breath analysis using portable gas chromatography for adult asthma phenotypes. Metabolites. 2021;11(5):265.
110. Chang-Chien J, Huang HY, Tsai HJ, Lo CJ, Lin WC, Tseng YL, et al. Metabolomic differences of exhaled breath condensate among children with and without asthma. Pediatric Allergy and Immunology. 2021;32(2):264-72.
111. Bao R, Hesser LA, He Z, Zhou X, Nadeau KC, Nagler CR. Fecal microbiome and metabolome differ in healthy and food-allergic twins. The Journal of clinical investigation. 2021;131(2).
112. Qu H-Q, Glessner J, Qu J, Gilhool S, Mentch F, Campbell I, et al. Metabolomic profiling of samples from pediatric patients with asthma unveils deficient nutrients in African Americans. iScience. 2022;25(7):104650.
113. Li J, Li X, Liu X, Wang X, Li J, Lin K, et al. Untargeted metabolomic study of acute exacerbation of pediatric asthma via HPLC-Q-Orbitrap-MS. Journal of Pharmaceutical and Biomedical Analysis. 2022;215:114737.
114. Zheng P, Zhang K, Lv X, Liu C, Wang Q, Bai X. Gut Microbiome and Metabolomics Profiles of Allergic and Non-Allergic Childhood Asthma. Journal of Asthma and Allergy. 2022;15:419.
115. Reinke SN, Naz S, Chaleckis R, Gallart-Ayala H, Kolmert J, Kermani NZ, et al. Urinary metabotype of severe asthma evidences decreased carnitine metabolism independent of oral corticosteroid treatment in the U-BIOPRED study. European respiratory journal. 2022;59(6).
116. Gürdeniz G, Ernst M, Rago D, Kim M, Courraud J, Stokholm J, et al. Neonatal metabolome of caesarean section and risk of childhood asthma. European Respiratory Journal. 2022;59(6).
117. Lee Y, Chen H, Chen W, Qi Q, Afshar M, Cai J, et al. Metabolomic Associations of Asthma in the Hispanic Community Health Study/Study of Latinos. Metabolites. 2022;12(4):359.
118. Chang YH, Yeh KW, Huang JL, Su KW, Tsai MH, Hua MC, et al. Metabolomics analysis reveals molecular linkages for the impact of vitamin D on childhood allergic airway diseases. Pediatric Allergy and Immunology. 2022;33(5):e13785.

119. Jang H, Kim EG, Kim M, Kim SY, Kim YH, Sohn MH, et al. Metabolomic profiling revealed altered lipid metabolite levels in childhood food allergy. Journal of Allergy and Clinical Immunology. 2022;149(5):1722-31. e9.
120. Townsend MK, Bao Y, Poole EM, Bertrand KA, Kraft P, Wolpin BM, et al. Impact of pre-analytic blood sample collection factors on metabolomics. Cancer Epidemiology, Biomarkers & Prevention. 2016;25(5):823-9.
121. Dyar KA, Eckel-Mahan KL. Circadian metabolomics in time and space. Frontiers in neuroscience. 2017;11:369.
122. Zheng R, Brunius C, Shi L, Zafar H, Paulson L, Landberg R, et al. Prediction and evaluation of the effect of pre-centrifugation sample management on the measurable untargeted LC-MS plasma metabolome. Analytica Chimica Acta. 2021;1182:338968.
123. Brunius C, Shi L, Landberg R. Large-scale untargeted LC-MS metabolomics data correction using between-batch feature alignment and cluster-based within-batch signal intensity drift correction. Metabolomics. 2016;12(11):173.
124. Yin P, Lehmann R, Xu G. Effects of pre-analytical processes on blood samples used in metabolomics studies. Analytical and bioanalytical chemistry. 2015;407(17):4879-92.
125. Dunn WB, Wilson ID, Nicholls AW, Broadhurst D. The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans. Bioanalysis. 2012;4(18):2249-64.
126. Beger RD, Dunn WB, Bandukwala A, Bethan B, Broadhurst D, Clish CB, et al. Towards quality assurance and quality control in untargeted metabolomics studies. Metabolomics. 2019;15(1):4.
127. Evans AM, O'Donovan C, Playdon M, Beecher C, Beger RD, Bowden JA, et al. Dissemination and analysis of the quality assurance (QA) and quality control (QC) practices of LC–MS based untargeted metabolomics practitioners. Metabolomics. 2020;16(10):1-16.
128. Lippa KA, Aristizabal-Henao JJ, Beger RD, Bowden JA, Broeckling C, Beecher C, et al. Reference materials for MS-based untargeted metabolomics and lipidomics: a review by the metabolomics quality assurance and quality control consortium (mQACC). Metabolomics. 2022;18(4):1-29.
129. Kirwan JA, Gika H, Beger RD, Bearden D, Dunn WB, Goodacre R, et al. Quality assurance and quality control reporting in untargeted metabolic phenotyping: mQACC recommendations for analytical quality management. Metabolomics. 2022;18(9):1-16.
130. Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N, et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. Nature protocols. 2011;6(7):1060-83.
131. Van Der Kloet FM, Bobeldijk I, Verheij ER, Jellema RH. Analytical error reduction using single point calibration for accurate and precise metabolomic phenotyping. Journal of proteome research. 2009;8(11):5132-41.
132. Sangster T, Major H, Plumb R, Wilson AJ, Wilson ID. A pragmatic and readily implemented quality control strategy for HPLC-MS and GC-MS-based metabonomic analysis. Analyst. 2006;131(10):1075-8.
133. Fernández-Albert F, Llorach R, Garcia-Aloy M, Ziyatdinov A, Andres-Lacueva C, Perera A. Intensity drift removal in LC/MS metabolomics by common variance compensation. Bioinformatics. 2014;30(20):2899-905.
134. Kamleh MA, Ebbels TM, Spagou K, Masson P, Want EJ. Optimizing the use of quality control samples for signal drift correction in large-scale urine metabolic profiling studies. Analytical chemistry. 2012;84(6):2670-7.

135. Ranjbar MRN, Zhao Y, Tadesse MG, Wang Y, Ressom HW. Gaussian process regression model for normalization of LC-MS data using scan-level information. Proteome Science. 2013;11(1):1-12.
136. Wehrens R, Hageman JA, van Eeuwijk F, Kooke R, Flood PJ, Wijnker E, et al. Improved batch correction in untargeted MS-based metabolomics. Metabolomics. 2016;12(5):88.
137. Barman M, Murray F, Bernardi AI, Broberg K, Bölte S, Hesselmar B, et al. Nutritional impact on Immunological maturation during Childhood in relation to the Environment (NICE): a prospective birth cohort in northern Sweden. BMJ open. 2018;8(10):e022013.
138. Bergia III RE, Biskup I, Giacco R, Costabile G, Gray S, Wright A, et al. The MEDGICarb-Study: Design of a multi-center randomized controlled trial to determine the differential health-promoting effects of low-and high-glycemic index Mediterranean-style eating patterns. Contemporary clinical trials communications. 2020;19:100640.
139. Harris H, Håkansson N, Olofsson C, Stackelberg O, Julin B, Åkesson A, et al. The Swedish mammography cohort and the cohort of Swedish men: study design and characteristics of two population-based longitudinal cohorts. Age (years). 1997(61,433):38,984.
140. Kilani RT, Delehanty M, Shankowsky HA, Ghahary A, Scott P, Tredget EE. Fluorescent-activated cell-sorting analysis of intracellular interferon-γ and interleukin-4 in fresh and frozen human peripheral blood T-helper cells. Wound repair and regeneration. 2005;13(4):441-9.
141. Basha S, Surendran N, Pichichero M. Immune responses in neonates. Expert review of clinical immunology. 2014;10(9):1171-84.
142. Markley JL, Brüschweiler R, Edison AS, Eghbalnia HR, Powers R, Raftery D, et al. The future of NMR-based metabolomics. Current opinion in biotechnology. 2017;43:34-40.
143. Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, Vázquez-Fresno R, et al. HMDB 4.0: the human metabolome database for 2018. Nucleic acids research. 2018;46(D1):D608-D17.
144. Dührkop K, Fleischauer M, Ludwig M, Aksenov AA, Melnik AV, Meusel M, et al. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. Nature methods. 2019;16(4):299-302.
145. Dührkop K, Shen H, Meusel M, Rousu J, Böcker S. Searching molecular structure databases with tandem mass spectra using CSI: FingerID. Proceedings of the National Academy of Sciences. 2015;112(41):12580-5.
146. Dührkop K, Nothias L-F, Fleischauer M, Reher R, Ludwig M, Hoffmann MA, et al. Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra. Nature Biotechnology. 2020:1-10.
147. Feunang YD, Eisner R, Knox C, Chepelev L, Hastings J, Owen G, et al. ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. Journal of Cheminformatics. 2016;8(1):61.
148. Schymanski EL, Jeon J, Gulde R, Fenner K, Ruff M, Singer HP, et al. Identifying small molecules via high resolution mass spectrometry: communicating confidence. ACS Publications; 2014.
149. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al. Proposed minimum reporting standards for chemical analysis. Metabolomics. 2007;3(3):211-21.
150. Savolainen OI, Sandberg A-S, Ross AB. A simultaneous metabolic profiling and quantitative multimetabolite metabolomic method for human plasma using gas-chromatography tandem mass spectrometry. Journal of proteome research. 2016;15(1):259-65.

151. Tian H, Li B, Shui G. Untargeted LC–MS data preprocessing in metabolomics. Journal of Analysis and Testing. 2017;1(3):187-92.
152. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Analytical chemistry. 2006;78(3):779-87.
153. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2019.
154. Tsugawa H, Cajka T, Kind T, Ma Y, Higgins B, Ikeda K, et al. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. Nature methods. 2015;12(6):523-6.
155. Libiseller G, Dvorzak M, Kleb U, Gander E, Eisenberg T, Madeo F, et al. IPO: a tool for automated optimization of XCMS parameters. BMC bioinformatics. 2015;16(1):118.
156. Di Guida R, Engel J, Allwood JW, Weber RJ, Jones MR, Sommer U, et al. Non-targeted UHPLC-MS metabolomic data processing methods: a comparative investigation of normalisation, missing value imputation, transformation and scaling. Metabolomics. 2016;12(5):1-14.
157. Do KT, Wahl S, Raffler J, Molnos S, Laimighofer M, Adamski J, et al. Characterization of missing values in untargeted MS-based metabolomics data and evaluation of missing data handling strategies. Metabolomics. 2018;14(10):1-18.
158. Ribbenstedt A, Ziarrusta H, Benskin JP. Development, characterization and comparisons of targeted and non-targeted metabolomics methods. PLoS One. 2018;13(11):e0207082.
159. Broeckling CD, Afsar F, Neumann S, Ben-Hur A, Prenni J. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. Analytical chemistry. 2014;86(14):6812-7.
160. Breiman L. Random forests. Machine learning. 2001;45(1):5-32.
161. Westerhuis JA, Hoefsloot HC, Smit S, Vis DJ, Smilde AK, van Velzen EJ, et al. Assessment of PLSDA cross validation. Metabolomics. 2008;4(1):81-9.
162. Filzmoser P, Liebmann B, Varmuza K. Repeated double cross validation. Journal of Chemometrics: A Journal of the Chemometrics Society. 2009;23(4):160-71.
163. Shi L, Westerhuis JA, Rosén J, Landberg R, Brunius C. Variable selection and validation in multivariate modelling. Bioinformatics. 2018;35(6):972-80.
164. Chen T, Cao Y, Zhang Y, Liu J, Bao Y, Wang C, et al. Random forest in clinical metabolomics for phenotypic discrimination and biomarker selection. Evidence-Based Complementary and Alternative Medicine. 2013;2013.
165. Hengl T, Nussbaum M, Wright MN, Heuvelink GB, Gräler B. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ. 2018;6:e5518.
166. van Velzen EJ, Westerhuis JA, van Duynhoven JP, van Dorsten FA, Hoefsloot HC, Jacobs DM, et al. Multilevel data analysis of a crossover designed human nutritional intervention study. Journal of proteome research. 2008;7(10):4483-91.
167. Pirttilä K, Balgoma D, Rainer J, Pettersson C, Hedeland M, Brunius C. Comprehensive Peak Characterization (CPC) in Untargeted LC–MS Analysis. Metabolites. 2022;12(2):137.
168. Kelly RS, Mendez KM, Huang M, Hobbs BD, Clish CB, Gerszten R, et al. Metabo-Endotypes of Asthma Reveal Differences in Lung Function: Discovery and Validation in Two TOPMed Cohorts. American Journal of Respiratory and Critical Care Medicine. 2022;205(3):288-99.
169. Lötvall J, Akdis CA, Bacharier LB, Bjermer L, Casale TB, Custovic A, et al. Asthma endotypes: a new approach to classification of disease entities within the asthma syndrome. Journal of Allergy and Clinical Immunology. 2011;127(2):355-60.

170. Reinke SN, Gallart-Ayala H, Gómez C, Checa A, Fauland A, Naz S, et al. Metabolomics analysis identifies different metabotypes of asthma severity. European Respiratory Journal. 2017;49(3).
171. Ntontsi P, Ntzoumanika V, Loukides S, Benaki D, Gkikas E, Mikros E, et al. EBC metabolomics for asthma severity. Journal of Breath Research. 2020;14(3):036007.
172. Lundell A-C, Björnsson V, Ljung A, Ceder M, Johansen S, Lindhagen G, et al. Infant B cell memory differentiation and early gut bacterial colonization. The Journal of Immunology. 2012;188(9):4315-22.
173. Hazekawa M, Ono K, Nishinakagawa T, Kawakubo-Yasukochi T, Nakashima M. In vitro anti-inflammatory effects of the phenylbutyric acid metabolite phenylacetyl glutamine. Biological and Pharmaceutical Bulletin. 2018:b17-00799.
174. Chen-Kiang S. Regulation of terminal differentiation of human B-cells by IL-6. Mechanisms in B-Cell Neoplasia 1994. 1995:189-98.
175. Svahn SL, Väremo L, Gabrielsson BG, Peris E, Nookaew I, Grahnemo L, et al. Six tissue transcriptomics reveals specific immune suppression in spleen by dietary polyunsaturated fatty acids. PLoS One. 2016;11(5):e0155099.
176. Soni NK, Ross AB, Scheers N, Savolainen OI, Nookaew I, Gabrielsson BG, et al. Splenic immune response is down-regulated in C57BL/6J mice fed eicosapentaenoic acid and docosahexaenoic acid enriched high fat diet. Nutrients. 2017;9(1):50.
177. Sallusto F, Lenig D, Förster R, Lipp M, Lanzavecchia A. Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. Nature. 1999;401(6754):708-12.
178. Burton GJ, Fowden AL. The placenta: a multifaceted, transient organ. Philosophical Transactions of the Royal Society B: Biological Sciences. 2015;370(1663):20140066.
179. Koh DK, Hume R, Eisenhofer G, Watson J, Williams FL. Arterio-venous differences in cord levels of catecholamines, glucose, lactate and blood gases. Journal of perinatal medicine. 2016;44(6):695-704.
180. Holm MB, Bastani NE, Holme AM, Zucknick M, Jansson T, Refsum H, et al. Uptake and release of amino acids in the fetal-placental unit in human pregnancies. PloS one. 2017;12(10):e0185760.
181. Vahratian A, Hoffman MK, Troendle JF, Zhang J. The impact of parity on course of labor in a contemporary population. Birth. 2006;33(1):12-7.
182. Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, et al. A cross-platform toolkit for mass spectrometry and proteomics. Nature biotechnology. 2012;30(10):918.
183. Euser AM, Zoccali C, Jager KJ, Dekker FW. Cohort studies: prospective versus retrospective. Nephron Clinical Practice. 2009;113(3):c214-c7.
184. von Mutius E. Influences in allergy: epidemiology and the environment. Journal of allergy and clinical immunology. 2004;113(3):373-9.
185. Saigusa D, Matsukawa N, Hishinuma E, Koshiba S. Identification of biomarkers to diagnose diseases and find adverse drug reactions by metabolomics. Drug metabolism and pharmacokinetics. 2021;37:100373.
186. Li D, Gaquerel E. Next-generation mass spectrometry metabolomics revives the functional analysis of plant metabolic diversity. Annual Review of Plant Biology. 2021;72:867-91.
187. Spicer RA, Salek R, Steinbeck C. A decade after the metabolomics standards initiative it's time for a revision. Scientific Data. 2017;4(1):1-3.