# Omics signatures variability in 156 children from five European countries

**by Marta Gallego Paüls**

**Directed by** Léa Maitre, ISGlobal

**Tutored by** Alberto Allepuz, Universitat Autònoma de Barcelona

**Official Master in Zoonosis and One Health**

Universitat Autònoma de Barcelona

Course 2018/2019

# OMICS SIGNATURES VARIABILITY IN 156 CHILDREN FROM FIVE EUROPEAN COUNTRIES

Official Master in Zoonosis and One Health

Universitat Autònoma de Barcelona

Course 2018/2019


Author:

Marta Gallego Paüls

Director:

Léa Maitre, ISGlobal

Tutor:

Alberto Allepuz, Universitat Autònoma de Barcelona

**Aknowledgements**

In the first place, I would like to say a big thank you to Léa Maitre from ISGlobal for her endless patience, help and guidance over these past months, and for trusting me to handle this research.

I would also like to thank all the families that took part in the HELIX project as well as ISGlobal for contributing to science by providing more knowledge about how environmental exposures affect population.

.

# Content

**ABSTRACT** ..... **1**

**INTRODUCTION** ..... **2**

**RESEARCH AIM AND SPECIFIC OBJECTIVES** ..... **5**

**MATERIALS AND METHODS** ..... **5**

Study design and population ..... 5
Sample Collection ..... 7
Measurement of omics signatures ..... 8
Explanatory variables ..... 9
Statistical analysis ..... 10

**RESULTS** ..... **12**

Population description ..... 12
Omics variability by intra-individual, inter-individual and cohort variability ..... 14
Can the intra-individual and inter-individual variability observed be explained by measured factors? ..... 15
Are we overestimating or underestimating the fixed effect proportion of variance explained by the explanatory variables? ..... 18
Are we overestimating the proportion of variance explained by the intra-individual variability due to technical effects? ..... 19

**DISCUSSION** ..... **20**

Limitations of our study ..... 22
Strengths of our study ..... 23

**CONCLUSION** ..... **23**

**BIBLIOGRAPHY** ..... **24**

**ANNEX 1** ..... **27**

A - Top 10 omics markers for intra-individual, inter-individual and cohort variability by omics ..... 27
B - Top 10 omics markers by variability of the explanatory variables by omics ..... 29

**ANNEX 2** ..... **30**

A - Mean over-/underestimation of the proportion of variance due to the explanatory variables ..... 30
B - Mean over-/underestimation of the proportion of variance due to the explanatory variables for the Top 10 intra- and inter-individual variability markers ..... 31

# Abstract

**Background:** Omics technologies are becoming increasingly useful at characterising health status. In terms of environmental epidemiology, omics profiles may confer a source of markers of environmental exposures, especially in the early-years of life, that can be of great use at understanding the molecular mechanisms that link the environmental exposures and the clinical outcomes of children.

**Objective:** To assess the intra- versus the inter-individual variability of omics signatures (proteome, transcriptome, serum and urine metabolome and miRNA) in the short term in 156 children from five European countries.

**Materials and methods:** The population under study are 156 children from the Child Panel Study that belongs to the HELIX project. Blood and urine samples of the children were collected at two different time-points (≈6 months), and urine samples were collected twice a day: first morning void and night-time. Proteins, transcripts, miRNA and serum metabolites were measured on blood samples (transcriptome: HTA V.2.0, Affymetrix; miRNA: SurePrint Human miRNA rel 21, Agilent; plasma proteins and metabolites: Luminex, cytokines 30-plex, apoliprotein 5-plex and adipokine 15-plex). Urine metabolome was measured using $^{1}$H NMR spectroscopy.

**Results:** Mean proportion of variance explained by intra-, inter-individual and variability differs among omics although all of them present the highest variability at the intra-individual level (59% for proteomics, 70% for urine metabolome, 50% for serum metabolome, 63% for miRNA and 84% for transcriptome). Also, within omics, there is diversity among different markers, ones being more explained by one level of variance or another. When biological traits and technical effects are included in the model as explanatory variables, little proportion of variance is attributed to each of them, except for seasonality, that generally accounts for a greater percentage of variation.

**Conclusion:** This study is the first to provide a multi-omics scope in omics variability in the short-term in children population. Intra- and inter-individual variability has been estimated for proteome, transcriptome, miRNA and serum and urine metabolome markers and differences on the behaviour between and within omics has been found. Further studies are needed in order to characterize the most meaningful markers in terms of environmental epidemiology and biological significance.

# Introduction

Omics technologies, such as transcriptomics, proteomics or metabolomics, are becoming increasingly useful at characterizing health status by providing highly detailed molecular information. As human diseases are complex at many biological levels, the information provided alone by one omics or another may not be enough to understand whole pattern of a disease. Therefore, the integration of multiple technologies may confer a more comprehensive view of human biology and disease. Additionally, omics could be measured at different stages of life to predict later health outcomes or to understand diseases occurring at different stages of life. This potential use of omics makes these technologies likely to become commonplace in the future clinical practice for diagnosis and prognosis (Karczewski & Snyder, 2018).

In environmental epidemiological studies, omics may provide biomarkers of key events, such as smoking, and indicate early molecular changes at low levels of environmental exposures. This statement comes from the idea that since the existence of low dose effects exists, a molecular change at those levels of exposure should also be able to be detected (Vineis, 2018).

The study of the temporal variability of omics can be implemented to characterise which omics markers behave in a more stable or dynamic way in order to compare them through time and determine possible associations with any clinical outcome. The determination of the levels and sources of variations among different omics could be essential in order to distinguish which alterations are really due to disease from those caused by physiological conditions (Shau et al., 2019). Therefore, systematic prospective and longitudinal biological sample collections in large-scale epidemiological studies open up new research opportunities. Especially during the early years of life, starting from (pre)conception, omics variability may be useful and play an essential role in understanding and discovering the impact of pre-natal and post-natal exposures on the onset of child and subsequent adult physiological conditions.

How omics vary over time and their variation patterns are thought to provide useful information for clinical purposes. The metabolome varies widely within individuals over time (Sampson et al., 2013). This is because metabolites are a result of all types of in vivo substances and are involved in a wide range of biological processes and network systems. Alterations in metabolites are linked to multiple factors that include genetic, environmental, drug, dietary factors and circadian rhythms (Zhan, Long & Lu, 2018); thus, we could expect metabolomics to be more dynamic in front of environmental events, which would make the metabolome worthy to reflect the status of physiological and pathological processes (Zhan, Long & Lu,

2018). The components of the proteome and their variations over time are also found to be influenced by many inherent and environmental factors that include age or medication, for example, and alterations in protein expression and/or content have been found to be useful at diagnosing many pathologies (Liu, Song, Guo, Sun & Liu, 2019). MicroRNAs (miRNAs), in turn, regulate gene expression by binding directly to specific mRNAs. Certain miRNAs appear to be highly variable because, depending on the physiological conditions, are selectively secreted to extracellular spaces (Yoon, Belmonte, Kasten, Bateman & Kim, 2017). Therefore, these regulatory elements have been found to be involved in a wide range of pathologies, including cancer or cardiovascular diseases, and may emerge as potential biomarkers (Keller et al. 2017). In the same way, transcriptomics (gene expression) variability is becoming increasingly acknowledged as a metric by which to evaluate transcriptome data, providing relevant information for human disease (Erickson, Otoupal & Chatterjee, 2017).

All omics, in terms of variation, have in common that their quantification is always influenced by technical effects such as experimental procedures that jeopardise the within-individual variability by reducing the power of epidemiological studies, especially in large-scale studies (Sampson et al., 2013; Liu, Song, Guo, Sun & Liu, 2019). Then, before any omics analysis that pursues a clinical application or biomarker research, it is necessary to understand the factors that introduce variability by themselves (Li-Gao et al., 2019).

Intra- and inter-individual variability regarding omics has been previously described in various studies, but these mainly focused in a particular omics. Kim et al. (2014), for example, assessed meal-time, temporal, and daily variability of the human urinary and plasma metabolomes in a tightly controlled environment by conducting Variance component analyses (VCA) to estimate the relative contributions of the temporal factors to the total variation in metabolite intensities. Shao et al. (2019) assessed individual variation in the urinary proteome and revealed gender-related and age-related differences in urinary proteins from healthy adult donors, and suggested that these factors should be considered as crucial factors in experimental design and data analyses of further studies. Hughes et al. (2015), in turn, evaluated intra- and inter-individual variation in the human placental transcriptome performing Analyses of variance (ANOVA), both incorporating environmental and biological factors such as the new-born weight and the maternal age. Keller et al. (2017) collected miRNA samples with a time interval of 5 years in order to determine sources of variability in circulating human miRNA, and found a group of 135 miRNAs that showed low variability between individuals and across the time-span, and were highly independent of the sampling process, which pointed them to be promising

biomarker candidates. Regarding multi-omic approaches, Piening et al. (2018) performed a controlled longitudinal weight perturbation study combining multiple omics strategies (genomics, transcriptomics, multiple proteomics assays, metabolomics, and microbiomics) during periods of weight gain and loss in humans and results demonstrated that omics signatures associated with insulin resistance that may serve as novel diagnostics. Chen et al. (2012), on the other hand, presented an integrative personal omics profile combining genomic, transcriptomic, proteomic, metabolomic and autoantibody profiles from a single individual over 14 months in order to reveal changes during healthy and disease states.

Nevertheless, none of the studies above-mentioned allowed for the integration of the whole omics profiles variability interpretation in the global population from an epidemiological scope, and only few of them addressed short or medium term variability. Maitre et al. (2017) applied Intra-class correlation coefficients (ICC) and mixed effect models to characterize short-term variability (6 days) in urinary metabolites measured from 20 children aged 8–9 years old. Using ICC as well, Peck et al. (2009) evaluated intra- and inter-individual variability of urinary phthalate metabolite concentrations in Hmong women of reproductive age during a month. Nagaraj & Mann (2011) characterised the urinary proteome of seven healthy human donors over three consecutive days and attributed 7.5%, 45.5% and 47.1% of the variability to technical, intrapersonal and interpersonal effects, respectively. These studies suggest that some specific markers might be useful at providing biological information over short periods of time due to their characteristic behaviour in terms of variability.

Hence, little is known about the changes related to short- and medium-term variability and seasonality in omics profiles, in particular in free-living human populations and child studies, and the concrete contribution (in terms of percentage) of different factors such as the BMI, the age or the ethnicity.

The HELIX subcohort study represents a multicentre European cohort of 1300 children and aims to assess the exposome. The exposome is defined as the totality of environmental exposures from conception to death, and it can be separated in the external exposome, the environmental exposures we are exposed to, and the internal exposome, the molecular responses of our organism to the external exposome to control homeostasis. It is precisely the internal exposome (or molecular/omics signatures) that might be used to: (1) identify novel biomarkers of exposures and (2) understand the molecular mechanism that link environmental exposures with health outcomes (Maitre et al., 2018). The children that took part in the HELIX

subcohort study present complete omics profiles for transcriptomics, epigenetics and metabolomics. Extensive data was collected about their environmental exposures and their lifestyle from pregnancy until childhood (7-11 years). Among these children, 157 did a second clinical visit, in average 6 months later, and omics analyses were repeated (Maitre et al., 2018).

The purpose of this study is to describe, in this repeated design, the inter-individual versus the intra-individual variability of the omics signatures in 156 children from the HELIX child panel study. In order to quantify this variability, a variance decomposition analysis has been performed using mixed effect models, taking into account intra- and inter-individual variability, cohort, and other technical effects and explanatory variables.

# Research aim and specific objectives

The present study aims to assess the omics signature variability in the short-term in 156 children from five European countries, by estimating the intra- versus the inter-individual variability in transcriptomics, metabolomics, miRNA and proteomics.

The main objectives are:

1. To describe the child panel study population and repeated variables.
2. To estimate the overall variance explained by intra-, inter-individual and cohort effects in the five omics datasets.
3. To estimate the proportion of variance attributable to the measured factors (technical effects and explanatory variables).
4. To assess and determine whether the model under- or over-estimates the proportion of variance attributable to the measured factors.

# Materials and Methods

## Study design and population

The HELIX (Human Early Life Exposome) study is a collaborative project of six-population based birth cohorts in different European Countries: United Kingdom (BiB: Born in Bradford) (Wright et al., 2012), France (EDEN: Étude des Déterminants pré et postnatals du développement et de la santé de l'Énfant) (Heude et al., 2015) , Spain (INMA: Infancia y Medio Ambiente) (Guxens et al., 2011), Lithuania (KANC: Kaunus cohort) (Grazuleviciene et al.,

2009), Norway (MoBa: Norwegian Mother and Child Cohort Study) (Magnus et al., 2016) and Greece (RHEA: Mother Child Cohort study in Crete) (Chatzi et al., 2017) (Maitre et al., 2018).

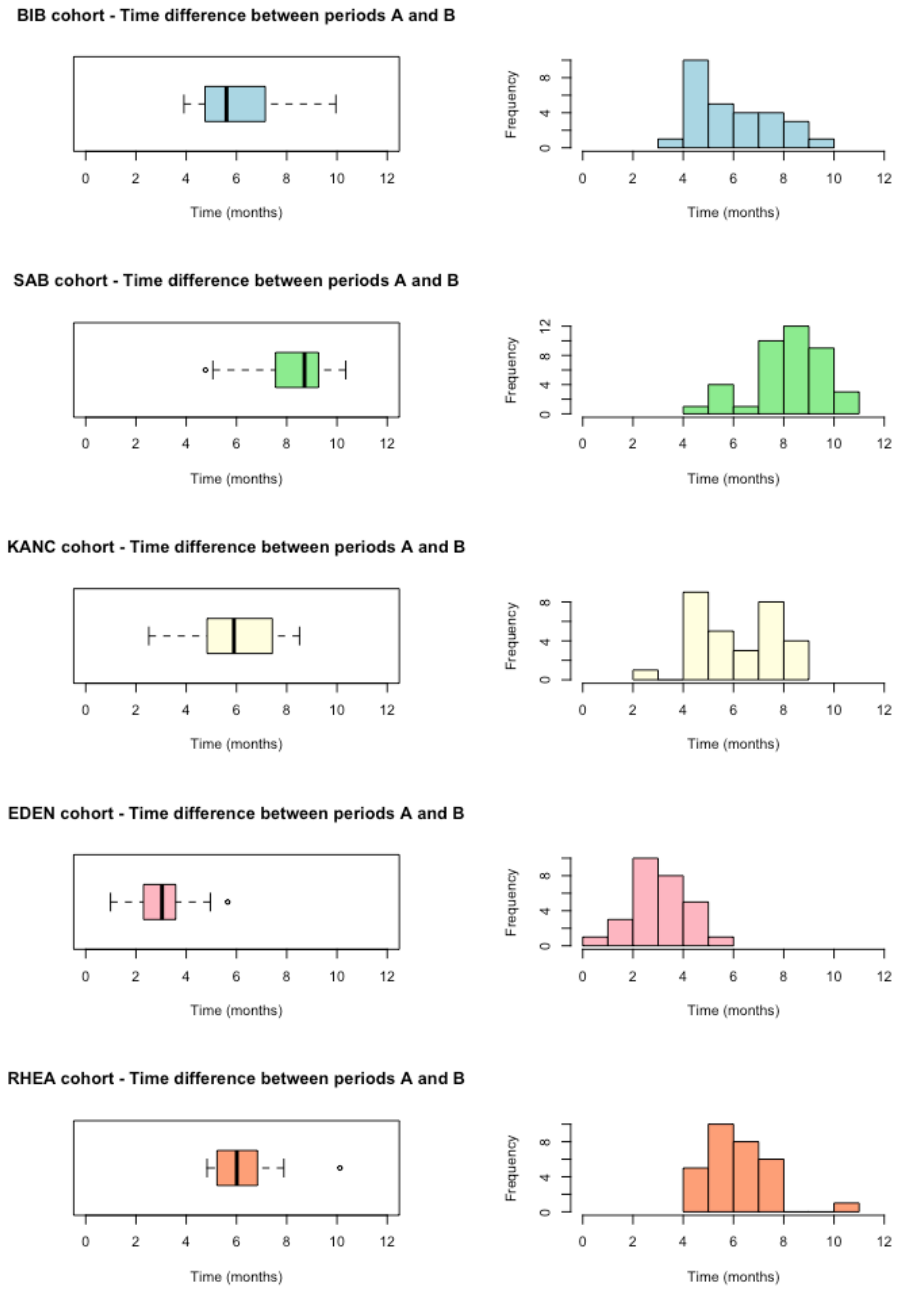
The aim of the Human Early Life Exposome (HELIX) study was to measure and describe multiple environmental exposures during early life (pregnancy and childhood) in a prospective cohort and associate these exposures with molecular omics signatures and child health outcomes. The project used a multilevel study design with the entire study population totalling 31 472 mother-child pairs, recruited during pregnancy, in the six existing cohorts (first level); a subcohort of 1301 mother-child pairs where biomarkers, omics signatures and child health outcomes were measured at age 6–11 years (second level) and repeat-sampling panel studies with around 150 children and 150 pregnant women aimed at collecting personal exposure data at two time points (third level) (Maitre et al., 2018).

Concretely, the panel-studies collected data on short-term (mean difference between the two visits: 6.11 months, SD: 2.18 months) (Figure 1) temporal variability in exposure biomarkers and omics biomarkers, individual behaviours (physical activity, mobility) and personal and indoor exposures (Donaire-Gonzalez et al., 2019). The child panel study included children from the HELIX sub-cohort (n=157, from all cohorts except MoBa) who lived in a first floor apartment or private house and were sampled following a maximum variation sampling strategy to high traffic-density exposure at home address. The present work focuses on the child panel study in order to characterise in depth the variability of the omics measurements (Maitre et al., 2018).

The study population has been restricted to children with complete information for both periods A and B for at least one of the omics (proteome, transcriptome, miRNA, serum metabolome or urine metabolome), which left 1 child out of the study (N=156). The final sample size, different for each omics, is 149, 105, 100, 154 and 154 children for proteome, transcriptome, miRNA, serum metabolome and urine metabolome, respectively.

Prior to the start of HELIX, national ethics committees had previously evaluated all participants as required and granted all the required permissions that allowed the participants cohort recruitment and follow-up visits. Additionally, all the participants were asked to sign a HELIX specific informed consent and new ethics approvals (Maitre et al., 2018).

*Figure 1.* ***Time difference between sample collection by cohort in the HELIX child panel study***




## Sample Collection

During the sub-cohort follow-up examination, new biological samples suitable for all planned omics analyses were collected using the same standardised protocols across all five cohorts. Urine samples were collected twice daily (first morning void and bedtime sample) in high-quality polypropylene tubes (Sarstedt: 75.9922.744). The two urine samples were brought by

the participants to the centre in cool packs and stored at −4°C until processing. After aliquoting, the urine samples were frozen at −80°C under optimized and standardised procedures. If the families did not bring urine samples with them, a new sample was collected on arrival at the centre (this occurred in 6.6% of the sub-cohort children). A pooled sample of both the morning and the night urine samples was used for analyses if available.

18mL of blood was collected at the end of the clinical examination of the child, ensuring an approximate 3hour fasting time since the last meal. Blood samples were collected using a 'butterfly' vacuum clip and local anaesthetic and processed into a variety of sample matrices for serum, plasma, whole blood for RNA extraction (Tempus tubes - Life Technologies Cat. No.: 4342792), red cells and a buffy coat for DNA extraction. After processing, these samples were frozen at −80°C under optimised and standardised procedures (Maitre et al., 2018).

## Measurement of omics signatures

In the sub-cohort, the following measurements of molecular omics signatures at the age of 6–11 years were obtained: blood leucocyte DNA methylation (450K, Illumina), whole blood transcription (HTA V.2.0, Affymetrix and SurePrint Human miRNA rel 21, Agilent), serum metabolites (AbsoluteIDQ p180 kit, Biocrates), urine metabolites (proton nuclear magnetic resonance ($^{1}$H NMR spectroscopy) and plasma proteins (Luminex, cytokines 30-plex, apoliprotein 5-plex and adipokine 15-plex). All the samples went through quality control procedures and only those with good quality markers were kept for study.

The number of omics markers varies greatly across the omics platforms: from 36 for proteomics to 58,254 for the transcriptome (Table 1). The platforms and data processing procedures selected for the proteins and serum metabolome were in fact targeted assays (<200 features) in order to obtain the best quality data for a large number of samples with fully annotated proteins and metabolites.

Further data filtering was applied to decrease the apparent complexity in the omics data. Proteins were filtered out if 30% of samples were outside of the linear range of quantification. Transcriptome and miRNA features with a call rate above 25% were kept for analysis. Serum metabolites were excluded following two criteria: i) %CV > 30% AND ii) % BLD+zeros >30%. In the urine metabolome, generated from untargeted NMR spectroscopic analysis from 128K spectral data points, 44 metabolite integrals were calculated only for resonances with high abundance and limited overlap with other metabolite signals. Urine metabolites were

normalized using the median fold change (Dieterle, Ross, Schlotterbeck & Senn, 2006) (Table 1).

*Table 1.* ***Omics analysis description in 156 children from the HELIX child panel study****.*

*LOQ1 and LOQ2: limits of quantification of the machine; CV: ratio of the standard deviation to the mean; BLD: "below limit of detection"; Call rate: proportion of reliable measures for a given sample or probe.*

| Omics | Matrix | Number of omics markers | Exclusion criteria | Laboratory processing |
|---|---|---|---|---|
| **Proteome** | Plasma | 36 | <30% LOQ1 or LOQ2 | Full random |
| **Urine metabolome** | Urine | 44 | Low abundance and limited overlap with other signals. | Full random |
| **Serum metabolome** | Serum | 177 | <30 %CV and %BLD+zeros | Full random |
| **miRNA** | Whole blood | 1117 | <25% call rate | Paired in batch (array) |
| **Transcriptome** | Whole blood | 58254 | <25% call rate | Paired in batch (array) |

## Explanatory variables

Explanatory variables considered for the statistical analyses were classified according to their different scopes: 1) omics platform-specific technical variables, 2) sample collection, 3) Intra-individual variability and 4) inter-individual variability. Table 2 shows the different explanatory variables that were included in the models.

*Table 2.* ***Explanatory variables considered for the variance decomposition analyses****.*

| | PROTEOME | TRANSCRIPTOME | MIRNA | SERUM METABOLOME | URINE METABOLOME |
|---|---|---|---|---|---|
| **Omics platform technical variables** | • Already corrected | • Already corrected | • Already corrected | • Plate batch | • Batch |
| **Sample collection variables** | • Time to last meal<br>• Hour of blood collection | • Time to last meal<br>• Hour of blood collection | • Time to last meal<br>• Hour of blood collection | • Time of collection (morning, night or pool) | • Time to last meal<br>• Hour of blood collection |
| **Inter-individual** | • Cohort<br>• Sex<br>• Age (months)<br>• BMI score<br>• Ethnicity | • Cohort<br>• Sex<br>• Age (months)<br>• BMI score<br>• Ethnicity | • Cohort<br>• Sex<br>• Age (months)<br>• BMI score<br>• Ethnicity | • Cohort<br>• Sex<br>• Age (months)<br>• BMI score<br>• Ethnicity | • Cohort<br>• Sex<br>• Age (months)<br>• BMI score<br>• Ethnicity |
| **Intra-individual** | • Season<br>• BMI change rate | • Season<br>• BMI change rate | • Season<br>• BMI change rate | • Season<br>• BMI change rate | • Season<br>• BMI change rate |

The explanatory variable "age", in months, was different at both visits, which allowed the model to take into account the time difference between the two visits, between subjects.

## Statistical analysis

We count on a three-level hierarchical clustered data:

- Level 1: *i*= 1,… I cohorts (I=5),
- Level 2: *j*= 1,…J individuals within cohort i,
- Level 3: *k*=1,2 (unbalanced) measurements within each individual.

The outcome of the model ($Y_{ijk}$) is the value of a specific omics marker, Y, at the moment of measurement *k* for individual *j* within cohort *i*.

The covariates have been selected based on a review of the literature (Hughes et al., 2016; Maitre et al., 2017; Piening et al., 2018) and are the following:

- $Age_{ijk}$: age (continuous) of individual *j* in cohort *i* at the moment of measurement *k*.
- $Sex_{ij}$: sex (categorical) of individual *j* in cohort *i*.
- $Ethnicity_{ij}$: ethnic group (categorical) of individual *j* in cohort *i* (as Caucasian or non-Caucasian).
- Moment of sample collection: moment (continuous) of the day in which measurement *k* in individual *i* of cohort *i* was taken.
- zBMI: BMI score (continuous) of individual *j* in cohort *i* at the moment of measurement.
- BMI change rate: how fast or slow BMI from individual *j* of cohort *I* changed between the two measurements *k*. This value was calculated as the zBMI difference divided by the age difference between the two visits:

$$BMI\ change\ rate = \frac{zBMI_{ij2} - zBMI_{ij1}}{Age_{ij2} - Age_{ij1}}$$

- Time to last meal: hours (continuous) of fasting since the last meal to the moment of the day in which measurement *k* in individual *i* of cohort *j* was taken.
- Season: season (categorical) in which measurement *k* in individual *i* of cohort *j* was taken.

- Urine sample type: for urine metabolome, moment (categorial) in which measurement *k* in individual *i* of cohort *j* was taken.
- Run order: for urine metabolome, the order in which the sample was run.
- Plate batch: for serum metabolome, the concrete batch at which the sample was analysed (from 1 to 19).

Taking all this into account, the study aims to carry out a variance partition estimation: inter-cohort, Inter-individual and intra-individual variance, modelled as random effects and of known factors: age, sex, ethnicity, BMI score, BMI change rate, season, moment of sample collection and time to last meal, modelled as fixed effects.

We have considered two mixed effect linear models, the first being a model that does not adjust for the explanatory variables (except for age at both visits) which has been used to estimate the proportion of variance attributed to intra-individual, inter-individual and cohort variability,

$$\boldsymbol{Y_{ijk}} = \beta_0 + \beta_1 Age_{ijk} + \beta Cohort_i + \beta Individual_j + \epsilon_{ijk}$$

and the latter adjusting for the above-mentioned explanatory variables to determine the proportion of variance that can be attributed to these.

$$\boldsymbol{Y_{ijk}} = \beta_0 + \beta_1 Sex_{ij} + \beta_2 Age_{ijk} + \beta_3 Ethnicity_{ij} + \beta_4 zBMI_{ijk}$$

$$+ \beta_5 \left( \frac{zBMI_{ij2} - zBMI_{ij1}}{Age_{ij2} - Age_{ij1}} \right) + \beta_6 Season_{ijk} + \beta_7 Moment\ of\ sample\ collection_{ijk}$$

$$+ \beta_8 Time\ to\ last\ meal_{ijk} + \beta Cohort_i + \beta Individual_j + \epsilon_{ijk}$$

The models were considered by assuming:

1. There is a specific baseline level of the outcome in each cohort, but potential effects of covariates are the same in all cohorts. Therefore, a random effect of the cohort in the intercept is included in the model.
2. There is a specific baseline level of the outcome in each individual but potential effects of explanatory variables are the same in all individuals. Therefore, a random effect of the individual in the intercept is included as well.
3. There is a possible intra-individual effect of the rate of change of BMI in the outcome. Therefore, a fixed effect of the intra-individual rate of change of zBMI is added.

4. Adjustment for: Age, Sex, Ethnicity, Hour of blood sampling, Hours to last meal, zBMI, BMI change rate, seasonality.

5. There is a possible effect of the varying time difference between visit A and B across individual. However, since time is actually measured by age, if we adjust for age, we are indeed adjusting for time.

These regression models were calculated for each omics marker using the R package "lme4" and the function "lmer" (Bates et al., 2015). The proportion of variance explained, or multiple $R^2$, was calculated as ratios of the total variance and represented as the per cent of variability because of differences between children and between cohorts. In addition, the residual variance of the model contains the intra-individual, which includes the variability attributed to the technical effects and the explanatory variables, this latter being added in the second model.

Firstly, the proportion of variance explained by each of the explanatory variables independently was calculated as ratios of the total variance. Parallel to this, the total variance attributable to these was extracted using the function "r.squaredGLMM" of the package "MuMIn" (Bartón, 2013), which gives the proportion of variance explained by the totality of the fixed effects, including their respective positive or negative covariances). For the categorical explanatory variables, the ratio was calculated separately for each comparison between categories and summed in order to obtain the whole proportion of variance attributed to such variable.

# Results

## Population description

From visit A to visit B sample collection and across the five cohorts, some of the characteristics of the children under study varied. Among the changes between the cohorts, we found children from EDEN cohort were considerably older (10.83 years old at visit A and 11.08 years old at visit B) compared to the mean age of the other cohorts (7.84 years old at visit A and 8.35 years old at visit B), being almost three years older than the total population average. Regarding zBMI, this was in average 0.27 points higher in period B, although most of the population was in the healthy range based on the zBMI categories stablished by the

World Health Organization (growth reference 5-19 years) ("BMI-for-age (5-19 years)", 2019). Also, time of fasting was shorter in an average of 0.99 hours for period B.

*Table 3.* ***Population description***

| | | Period A | Period B |
|---|---|---|---|
| **Number of children** | | 156 | |
| | Male | 89 | |
| | Female | 67 | |
| **Ethnicity** | | | |
| | Caucasian | 145 | |
| | Other | 11 | |
| **Cohort** | | | |
| | BIB | 28 | |
| | EDEN | 28 | |
| | KANC | 30 | |
| | RHEA | 30 | |
| | SAB | 40 | |
| **Age** (years) | | | |
| Total | mean (SD) | 7.84 (1.70) | 8.35 (1.65) |
| BIB | mean (SD) | 6.69 (0.23) | 7.20 (0.31) |
| EDEN | mean (SD) | 10.83 (0.46) | 11.08 (0.48) |
| KANC | mean (SD) | 6.65 (0.45) | 7.16 (0.46) |
| RHEA | mean (SD) | 6.31 (0.15) | 6.83 (0.16) |
| SAB | mean (SD) | 8.6 (0.52) | 9.28 (0.51) |
| **zBMI** | | | |
| | mean (SD) | 0.392 (1.21) | 0.419 (1.32) |
| **zBMI categories** | | | |
| | Thinness (<2) | 1 | 2 |
| | Healthy | 111 | 108 |
| | Overweight (>1) | 27 | 27 |
| | Obese (>2) | 17 | 19 |
| **Time to last meal** (hours) | | | |
| | mean (SD) | 3.57 (1.19) | 2.58 (1.51) |
| **Time of sample collection** (hours) | | | |
| | mean (SD) | 16.9 (2.85) | 16.3 (3.05) |
| **zBMI change between time points** | | | |
| | mean (SD) | 0.027 (0.525) | |
| **Age difference between time points (months)** | | | |
| | mean (SD) | 6.11 (2.18) | |
| **zBMI change rate** | | | |
| | mean (SD) | 0.00114 (0.104) | |

## Omics variability by intra-individual, inter-individual and cohort variability

The proportion of mean variance explained by the intra-individual, inter-individual and cohort variability varied across the different omics (Table 4), based on the results obtained by the mixed effect linear model without taking into account the explanatory variables. Intra-individual variability appeared to explain higher proportion of variation in the transcriptome, whereas serum metabolome markers were the most influenced by the inter-individual variability and proteome features accounted for the highest inter-cohort variability.

*Table 4.* ***Summary of the mean explained variance by omics (model without fixed effects)***

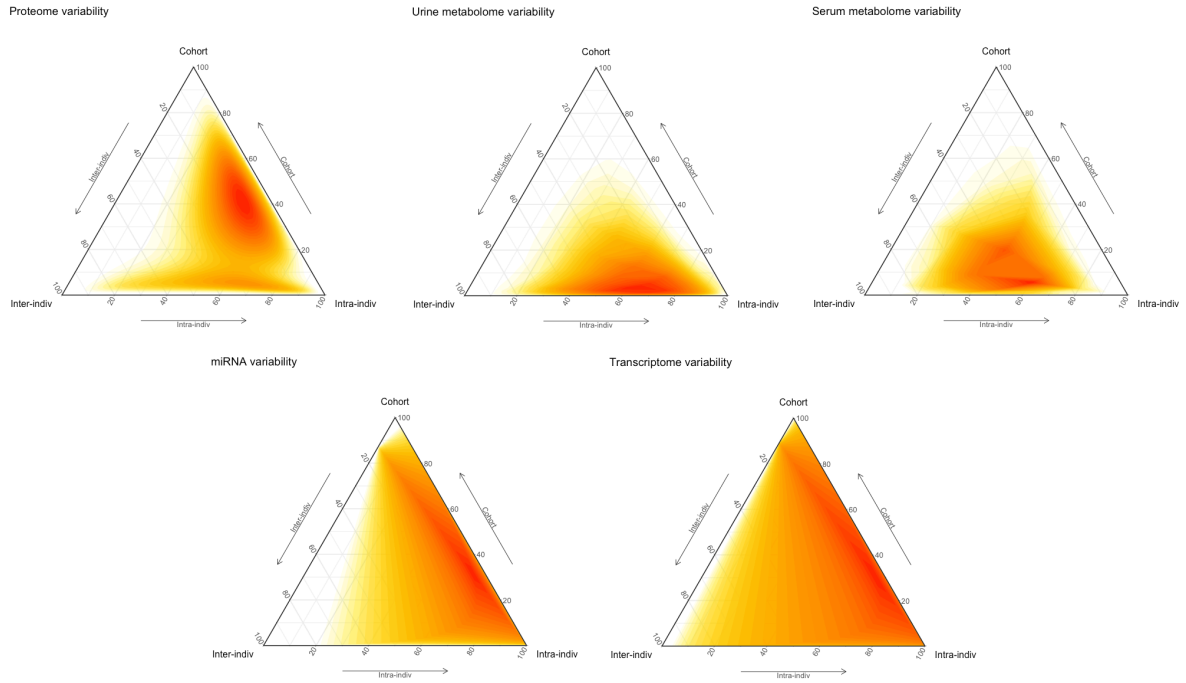
| | Proteome | Urine metabolome | Serum metabolome | miRNA | Transcriptome |
|---|---|---|---|---|---|
| **Cohort** | 23.78 | 3.43 | 8.72 | 11.32 | 1.128 |
| **Inter-individual** | 18.00 | 26.88 | 41.67 | 25.33 | 14.04 |
| **Intra-individual** | 59.98 | 69.69 | 49.60 | 63.35 | 84.83 |

Also, within each particular omics, biomarkers behaved in different ways (Figure 3). For instance, proteome markers mostly contained highest proportions of intra-individual and cohort variability, but there were also another group of proteins that were situated towards the inter-individual variability corner, which indicated that their variability was mostly inter-individual. Urine and serum metabolites variability was mostly due to either intra- or inter-individual variability since no points appeared to tend to the cohort variability corner. miRNA markers mostly contained higher proportions of cohort and intra-individual variability. In contrast, the transcriptome was the omics that showed the most variety among its omics markers as it is shown in the density plot, where there were dots spread all along the area going from one very corner to the other, although there was more density in-between the cohort and the intra-individual variability corners.

Boxplots in Figure 4A reflect how distant from the mean were some groups of markers. The results determined, in average, that more than a half of the proteome variability was explained by the intra-individual effect (55.30%; SD: 17.41). Urine metabolites presented higher proportion of variability explained by intra-individual effect (69.69%; SD: 17.16) and a small percentage of variance explained by cohort (3.34%; SD: 5.65). Serum metabolites were the markers that obtained the highest mean of variability explained by inter-individual effect (41.67%; SD: 15.02). miRNAs and transcripts variabilities, in turn, were the most explained by

intra-individual effects (78.90%; SD: 20.66 for miRNAs, 87.13%; SD: 17.33 for transcripts) and transcripts were almost not influenced by cohort effect (1.02%; SD: 2.52).

*Figure 3.* ***Density plot of the omics markers in terms of intra-, inter-individual and cohort variability percentage by omic***

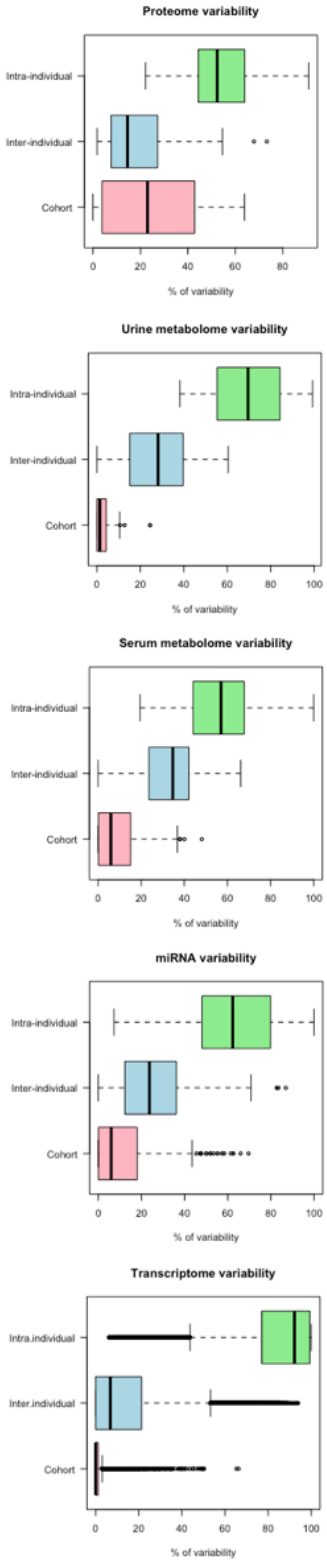



## Can the intra-individual and inter-individual variability observed be explained by measured factors?

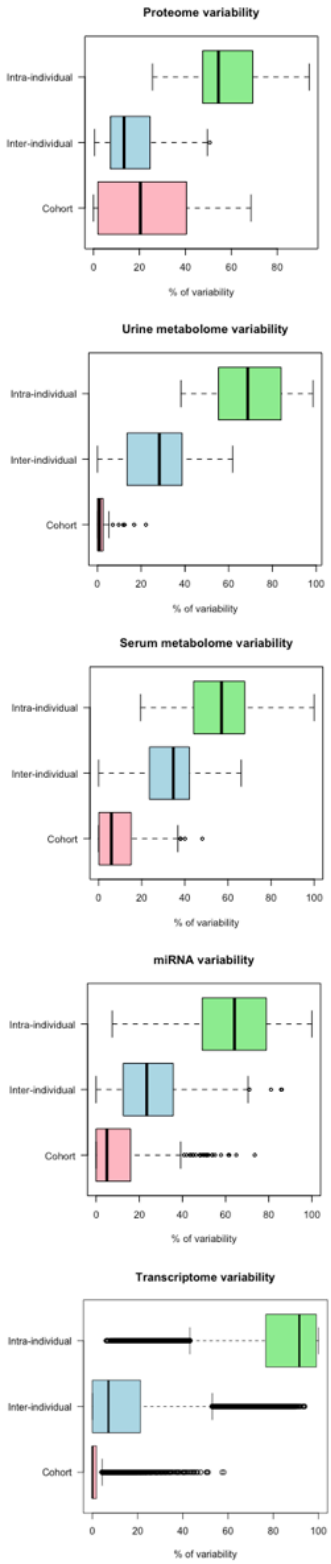
In order to correct for the explanatory variables, these were added as fixed effects to the model (Figure 4B). The distribution of variabilities differed slightly from the model without fixed effects (Figure 4A). The intra-individual variability in the proteome increased in a 4.68% (59.98%; SD: 17.05). Urine metabolites were ones that varied less from one model to the other by only increasing a 0.67% their intra-individual variability and diminishing a 0.62% their cohort variability. Serum metabolites, in contrast, presented a decrease of 8.4% in their inter-individual variability (33.27%; SD: 14.42) and an increase of 7.55% in their intra-individual variability (from 49.60%; SD: 15.20, to 57.15%; SD: 17.44). miRNAs and transcripts continued having the highest intra-individual variabilities (79.43%; SD: 20.12 for miRNAs, 84.47%; SD: 18.59 for transcripts), although transcripts increased in a 2.17% their inter-individual variability (from 11.85%; SD: 16.78, to 14.02%; SD: 17.94) by decreasing a 2.66% their intra-individual variability.

*Figure 4.* ***Omics variability without fixed effects in the model (A) and with fixed effects in the model (B)***

*A) Without fixed effects*




*B) With fixed effects*



Overall, the proportion of variance attributed to the explanatory variables (assuming they were independent) was low for all omics (Figure 5), although some differences were found among them. The variation attributed to both biological and technical effects accounted for 17.2%, 18.3%, 13.7%, 24.2% and 20.9% for proteome, urine metabolome, serum metabolome, miRNA and transcriptome, respectively, of the total variation. Most of this variance was due to seasonality in all omics.

The zBMI (4.62%; SD: 10.36) and seasonality (6.94%; SD: 7.93) were the explanatory variables that explained most of the variability in the proteome, whereas the urine sample type (morning, night or pool) (3.59%; SD: 4.67) and season (10.07%; SD: 9.89) were the more relevant ones in the urine metabolome. Apart from seasonality (5.37%, SD: 5.46), serum metabolome variability appeared to be mostly explained by the age of the individuals (2.57%; SD: 3.17). Except for seasonality (16.64%; SD: 17.19 for miRNA, 16.50%; SD: 17.01 for transcriptome), the explanatory variables explained little miRNA and transcriptome variability (Figure 5).

In order to know which omics markers presented the highest intra-, inter-individual and cohort variabilities, we sorted them and represented a "top 10" for each of the three effects by omics (see Annex 1). Annex 1A shows each top 10 by proportion of variability explained by intra-, inter-individual and cohort effect and Annex 1B represents, for the top 10 markers, the amount of variability attributed to the different explanatory variables.

Among the urine metabolites with the highest intra-individual variability, N-methylpicolinic acid, trimethylamine oxide and dimethylamine were found (see Annex 1A).

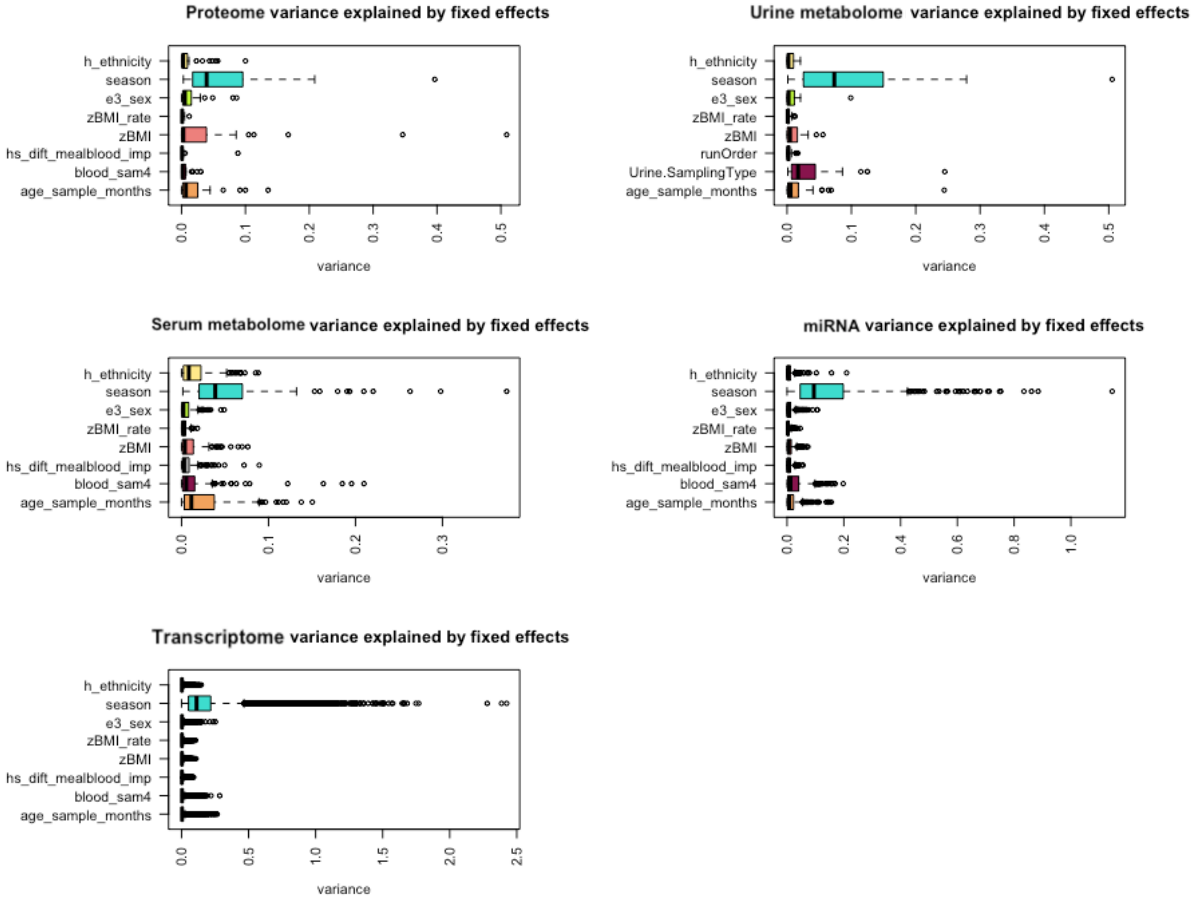
C14:1 (Tetradecenoylcarnitine) and C16:1 (Hexadecenoylcarnitine) were among the top 10 intra-individual variability serum metabolites that presented the most variability due to sex differences. Regarding the top 10 inter-individual variability serum metabolites, PC aa C40:4 (a glycerophospholipid) was the metabolite most influenced by child zBMI (see Annex 1B).

Among the proteins with the highest inter-individual variation, we found that variation in il6 (Isoleucine 6) was largely explained by zBMI, as well as the time of fasting for insulin, which was one of the proteins that presented more intra-individual variation (see Annex 1B).

Hsa-miR-100-5p was found to be one of the miRNAs presenting the most inter-individual variability (see Annex 1A).

*Figure 5.* ***Variance explained by fixed effects.***

*h_ethnicity: ethnicity; e3_sex: sex; zBMI ra te: zBMI change rate; hs_dift_mealblood_imp: time to last meal; blood_sam4: time of sample collection; age_sample_months: age in months at the time of sampling.*




## Are we overestimating or underestimating the fixed effect proportion of variance explained by the explanatory variables?

The total proportion of variance explained by the fixed effects of the model was calculated, taking into account both the negative and positive covariances between them. These results were compared with the previous models were we independently calculated the proportion of variance explained by each explanatory variable and summed all the values. In the previous models we were assuming that the fixed effects were independent and that they did not covary. The total variance explained by fixed effects differed according to the model assumptions based on "r.squaredGLMM" package.

For proteome, urine metabolome, miRNA and transcriptome, the proportion of variability explained by the fixed effects was overestimated in a mean of 7.4, 12.5, 14.7 and 15.7%, respectively, since the values of our sum were substantially higher than the total value,

whereas for serum metabolome, the variability attributed to the explanatory variables was underestimated in a 8.4% (see Annex 2A).

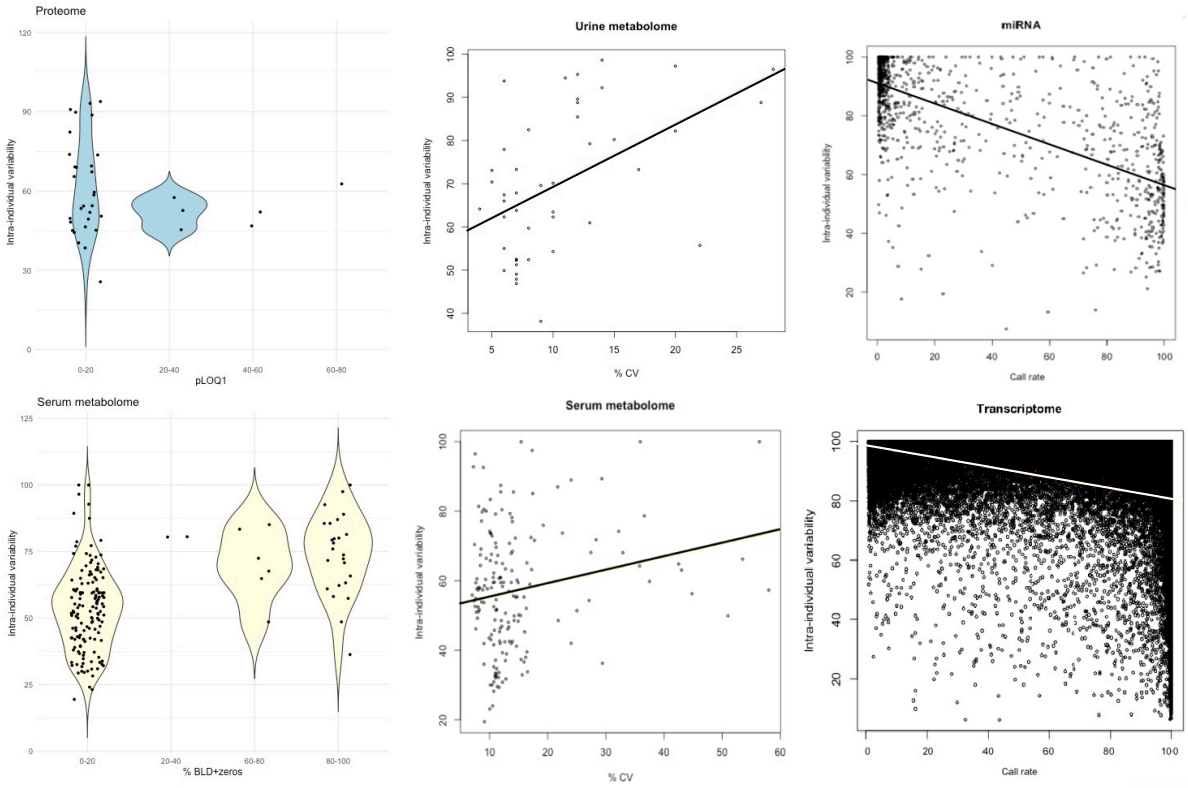
When comparing the over- and underestimation for the markers with highest intra- and inter-variability (see Annex 2B), in the case of the omics that were overestimated (proteome, urine metabolome, miRNA and transcriptome), a pattern could be observed: those markers that were highly explained by season (e.g. adiponectin protein and succinate urine metabolite) had, in general, an overestimation of the variability explained by the fixed effects.

## Are we overestimating the proportion of variance explained by the intra-individual variability due to technical effects?

For each particular omics, figure 6 shows the association between the intra-individual variability for each markers and a parameter representative of the quality of measurement of the marker or its level of detection. We used the percentage of limit of quantification 1 (pLOQ1) for the proteome, the percentage of coefficient of variation (% CV) for urine and serum metabolome, and the percentage of call rate (% call rate) for miRNA and transcriptome.

*Figure 6.* ***Association between intra-individual variability and Quality Control parameters***

*pLOQ1: percentage under limit of quantification; %CV: percentage of coefficient of variation; %BLD+zeros: percentage below limit of detection and zeros.*



In terms of pLOQ1, most of the proteome markers were around the first quartile and its intra-individual variability varied along the intra-individual variability axis.

For both urine and serum metabolome, there was a tendency of the markers with the higher coefficient of variation presenting a higher intra-individual variability.

Transcriptome and miRNA markers with the lower call rate, in turn, appeared to have higher intra-individual variability, while those with the highest quality (highest call rate) tended to be more stable.

# Discussion

The current study modelled for the first time the variability of five different types of omics data for 156 children from five European countries at two time points of 6-month interval, providing a multi-omics perspective of short-term variability in the early years of life. We found a disparity within each omics in terms of their stability over time and their specificity among children.

We estimated the proportion of variance attributed to intra-individual, inter-individual and cohort variability correcting for those explanatory variables we had complete information about, and we thought to be of greater importance. Although technical effects had been previously minimized through the quality control procedures, some of the intra-individual variability might still be due to technical difficulty to measure those omics markers with low concentrations at the time of analysis.

In general, proportions of variability attributed to cohort effects were low, in average, but some particular markers stood out to higher levels of cohort variability. It would be interesting to analyse more in depth which markers are these as well as their function since they could provide valuable epidemiological information at a large geographical scale level.

To our knowledge, only two previous studies have assessed omic markers variability in a way that can be commensurable to ours, regarding short-term and early years of life, in spite of the fact that they focus in a particular omics. Maitre et al. (2017) performed the characterization of short-term variability in urinary metabolites. The study focused on a subset of 20 children from the INMA cohort (Guxens et al., 2011) in Spain, aged 8–9 years old. They collected two daily urine samples during six consecutive days, first morning void and night-time, and they performed the analyses with the pooled samples, just the way our

urine metabolome analysis was carried out. The effects of inter-/intra-individual differences on the child urinary metabolome were estimated through mixed effect models. Among the metabolites that were found to have the highest intra-individual variability, N-methylpicolinic acid, trimethylamine oxide and dimethylamine were found. Interestingly, these also appear among the top ten metabolites with highest intra-individual variability in our study. Interestingly these metabolites were identified as representative of seafood consumption and gut microbial metabolism.

On the other hand, Hughes et al. (2015) estimated the gene expression variation of 66 human placentas (transcriptomics) as well as determined the contribution of environmental, technical and biological factors to the total variation. In their study, technical and biological traits effects accounted for roughly 2% of total gene expression variation (Hughes et al., 2015). This value agrees with the proportion of variation found to be attributable to most of the explanatory variables in our study (a clear exception would be seasonality). Remarkably, the transcriptome variation in our study is mostly due to intra-individual variation (84.8%) whereas Hughes et al. (2016) attributed 58.9% of the variability to inter-individual variation. We could expect these differences to be due to differences in tissue and differences in life-stage of the study, being our study conducted in children aged 8 years, approximately, and theirs in placental tissue.

Our results were consistent with previous studies. For example, among the proteins with the highest intra- and inter-individual variability we found Insulin and IL6, respectively. When we assessed the proportion of variance attributed to the explanatory variables in the proteins that presented highest intra- and inter-individual variability, we found these protein variabilities highly depended on “Time to last meal” and “zBMI”, respectively. This fact strengthens our results since (1) Insulin is a hormone related to the intake of food and (2) IL6 is a glycoprotein expressed by immune cells, and the body mass index and the immune response have been found to be strongly associated (Ilavská et al., 2012). For serum metabolites, as well, C14:1 (T etradecenoylcarnitine) and C16:1 (Hexadecenoylcarnitine) had been previously positively correlated with the female sex, and PC aa C40:4 (a glycerophospholipid) was reported to be negatively associated to zBMI in the HELIX subcohort (Lau et al., 2018). Hsa-miR-100-5p miRNA, for instance, is involved in the negative regulation of vascular smooth muscle cell differentiation, so further study on its potential role as a biomarker could be interesting.

De Candia, Torri, Pagani & Abrignani, 2014, suggested serum miRNAs as biomarkers of lymphocyte activation since lymphocytes release a great number of nano-sized vesicles

containing T-cell characteristic miRNAs. Interestingly, transcripts in our study were assessed mostly from immune cells and miRNA from circulating blood. Therefore, it would be useful for further analyses in the HELIX child panel study to see if the children had had or were having a cold prior to the sample collections, in order to assess if some of the variability seen could be attributable to a lymphocyte activation.

## Limitations of our study

In the first place, the proteome and the serum analysis were in fact targeted assays, which therefore did not cover a wide range of markers and so there might be potential important information that we missed by only analyzing a narrow spectrum of proteins and metabolites.

Secondly, despite we added many explanatory variables to correct for in the model, there are still more interesting variables that could be taken into account such as diet or physical activity, which are expected to influence the variability of omics such as the metabolome. In general, it would be interesting to study if and how other factors account for variability within omics.

Thirdly, our study is based on data collected at two time points. A more exhaustive collection of repeated samples across short periods of time would provide more robust and meaningful information of omics variability over time.

Of course, a higher population size would provide higher statistical power to our model given the fact that our study was conducted with 156 children and not all of them had complete data for all omics.

Finally, regarding the variation due to the explanatory variables, a great limitation of our study is the fact that the values obtained for each variable were calculated by assuming independence and no covariation among them. However, by running "r.squaredGLMM" function of the package "MuMIn" (Bartón, 2013), we could measure how over- or underestimated was the proportion of variance attributable to the fixed effects of the model in each omics. A reason of the overestimation in proteome, urine metabolome and miRNA might be due to an overestimation of the season effect, since, interestingly, those markers that are highly explained by seasonality are also really overestimated. Deeper analysis on the seasonality would be needed in order to understand this phenomenon and determine how seasonality really affects variability. This factor is often overlooked in clinical or epidemiological settings measuring omics and could confer large uncertainty in the measurements.

### Strengths of our study

This is the first study that assessed omics variability in the short-term in general populations from a multi-omics approach, concretely in child cohorts. Especially for miRNA, there is scarce previous data on its variability over time. In any case, the analysis of the different omics altogether is already useful to have global vision, all at once, of how different omics vary (1) over time and among them and (2) within, between individuals and between cohorts from different countries.

The fact that children from five European countries took part in the study allow the results to be more generalized and meaningful in terms of environmental epidemiology, as well as more generalizable.

By having these data on variability in different omics, this study could be the base of a deeper and exhaustive research in order to analyze which are the markers that present more inter-individual and cohort variability and how these can provide useful information for epidemiological purposes.

## Conclusion

Omics technologies are of valuable use at characterizing health status and omics markers may serve as indicators of environmental exposures through time, especially during the early-life years of life. This study characterizes intra-, inter-individual and cohort variability in children from five European child cohorts and finds that most of the variability within omics is due to intra-individual variability, despite this proportion varies among omics and within omics. On the other hand, the explanatory variables that have been included in the model explain low proportions of variability separately. As omics markers behave in different ways within omics, and omics themselves show different patterns of variability, future studies will look at the most outlying markers and will deepen into their biological function in order to give biological meaning to the research and find potential biomarkers for early-life environmental exposures.

# Bibliography

Bartón, K. (2013). MuMIn: Multi-modal inference. Model selection and model averaging based on information criteria (AICc and alike). http://cran.r-project.org/web/packages/MuMIn/index.html

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal Of Statistical Software, 67(1). doi: 10.18637/jss.v067.i01

BMI-for-age (5-19 years). (2019). Retrieved from https://www.who.int/growthref/who2007_bmi_for_age/en/

Chatzi, L., Leventakou, V., Vafeiadi, M., Koutra, K., Roumeliotaki, T., & Chalkiadaki, G. et al. (2017). Cohort Profile: The Mother-Child Cohort in Crete, Greece (Rhea Study). International Journal Of Epidemiology, 46(5), 1392-1393k. doi: 10.1093/ije/dyx084

Chen, R., Mias, G., Li-Pook-Than, J., Jiang, L., Lam, H., & Chen, R. et al. (2012). Personal Omics Profiling Reveals Dynamic Molecular and Medical Phenotypes. Cell, 148(6), 1293-1307. doi: 10.1016/j.cell.2012.02.009

de Candia, P., Torri, A., Pagani, M., & Abrignani, S. (2014). Serum microRNAs as Biomarkers of Human Lymphocyte Activation in Health and Disease. Frontiers In Immunology, 5. doi: 10.3389/fimmu.2014.00043

Dieterle, F., Ross, A., Schlotterbeck, G., & Senn, H. (2006). Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application in1H NMR Metabonomics. Analytical Chemistry, 78(13), 4281-4290. doi: 10.1021/ac051632c

Donaire-Gonzalez, D., Valentín, A., van Nunen, E., Curto, A., Rodriguez, A., & Fernandez-Nieto, M. et al. (2019). ExpoApp: An integrated system to assess multiple personal environmental exposures. Environment International, 126, 494-503. doi: 10.1016/j.envint.2019.02.054

Erickson, K., Otoupal, P., & Chatterjee, A. (2017). Transcriptome-Level Signatures in Gene Expression and Gene Expression Variability during Bacterial Adaptive Evolution. Msphere, 2(1). doi: 10.1128/msphere.00009-17

Grazuleviciene, R., Danileviciute, A., Nadisauskiene, R., & Vencloviene, J. (2009). Maternal Smoking,GSTM1 and GSTT1 Polymorphism and Susceptibility to Adverse Pregnancy Outcomes. International Journal Of Environmental Research And Public Health, 6(3), 1282-1297. doi: 10.3390/ijerph6031282

Guxens, M., Ballester, F., Espada, M., Fernández, M., Grimalt, J., & Ibarluzea, J. et al. (2011). Cohort Profile: The INMA—INfancia y Medio Ambiente—(Environment and Childhood) Project. International Journal Of Epidemiology, 41(4), 930-940. doi: 10.1093/ije/dyr054

Heude, B., Forhan, A., Slama, R., Douhaud, L., Bedel, S., & Saurel-Cubizolles, M. et al. (2015). Cohort Profile: The EDEN mother-child cohort on the prenatal and early postnatal determinants

of child health and development. International Journal Of Epidemiology, 45(2), 353-363. doi: 10.1093/ije/dyv151

Hughes, D., Kircher, M., He, Z., Guo, S., Fairbrother, G., & Moreno, C. et al. (2015). Evaluating intra- and inter-individual variation in the human placental transcriptome. Genome Biology, 16(1). doi: 10.1186/s13059-015-0627-z

Ilavská, S., Horváthová, M., Szabová, M., Nemessányi, T., Jahnová, E., & Tulinská, J. et al. (2012). Association between the human immune response and body mass index. Human Immunology, 73(5), 480-485. doi: 10.1016/j.humimm.2012.02.023

Karczewski, K., & Snyder, M. (2018). Integrative omics for health and disease. Nature Reviews Genetics, 19(5), 299-310. doi: 10.1038/nrg.2018.4

Keller, A., Rounge, T., Backes, C., Ludwig, N., Gislefoss, R., & Leidinger, P. et al. (2017). Sources to variability in circulating human miRNA signatures. RNA Biology, 14(12), 1791-1798. doi: 10.1080/15476286.2017.1367888

Kim, K., Mall, C., Taylor, S., Hitchcock, S., Zhang, C., & Wettersten, H. et al. (2014). Mealtime, Temporal, and Daily Variability of the Human Urinary and Plasma Metabolomes in a Tightly Controlled Environment. Plos ONE, 9(1), e86223. doi: 10.1371/journal.pone.0086223

Lau, C., Siskos, A., Maitre, L., Robinson, O., Athersuch, T., & Want, E. et al. (2018). Determinants of the urinary and serum metabolome in children from six European populations. BMC Medicine, 16(1). doi: 10.1186/s12916-018-1190-8

Li-Gao, R., Hughes, D., le Cessie, S., de Mutsert, R., den Heijer, M., & Rosendaal, F. et al. (2019). Assessment of reproducibility and biological variability of fasting and postprandial plasma metabolite concentrations using 1H NMR spectroscopy. PLOS ONE, 14(6), e0218549. doi: 10.1371/journal.pone.0218549

Liu, X., Song, Y., Guo, Z., Sun, W., & Liu, J. (2019). A comprehensive profile and inter-individual variations analysis of the human normal amniotic fluid proteome. Journal Of Proteomics, 192, 1-9. doi: 10.1016/j.jprot.2018.04.023

Magnus, P., Birke, C., Vejrup, K., Haugan, A., Alsaker, E., & Daltveit, A. et al. (2016). Cohort Profile Update: The Norwegian Mother and Child Cohort Study (MoBa). International Journal Of Epidemiology, 45(2), 382-388. doi: 10.1093/ije/dyw029

Maitre, L., Lau, C., Vizcaino, E., Robinson, O., Casas, M., & Siskos, A. et al. (2017). Assessment of metabolic phenotypic variability in children's urine using 1H NMR spectroscopy. Scientific Reports, 7(1). doi: 10.1038/srep46082

Maitre, L., de Bont, J., Casas, M., Robinson, O., Aasvang, G., & Agier, L. et al. (2018). Human Early Life Exposome (HELIX) study: a European population-based exposome cohort. BMJ Open, 8(9), e021311. doi: 10.1136/bmjopen-2017-021311

Nagaraj, N., & Mann, M. (2011). Quantitative Analysis of the Intra- and Inter-Individual Variability of the Normal Urinary Proteome. Journal Of Proteome Research, 10(2), 637-645. doi: 10.1021/pr100835s

Peck, J., Sweeney, A., Symanski, E., Gardiner, J., Silva, M., Calafat, A., & Schantz, S. (2009). Intra- and inter-individual variability of urinary phthalate metabolite concentrations in Hmong women of reproductive age. Journal Of Exposure Science & Environmental Epidemiology, 20(1), 90-100. doi: 10.1038/jes.2009.4

Piening, B., Zhou, W., Contrepois, K., Röst, H., Gu Urban, G., & Mishra, T. et al. (2018). Integrative Personal Omics Profiles during Periods of Weight Gain and Loss. Cell Systems, 6(2), 157-170.e8. doi: 10.1016/j.cels.2017.12.013

Sampson, J., Boca, S., Shu, X., Stolzenberg-Solomon, R., Matthews, C., & Hsing, A. et al. (2013). Metabolomics in Epidemiology: Sources of Variability in Metabolite Measurements and Implications. Cancer Epidemiology Biomarkers & Prevention, 22(4), 631-640. doi: 10.1158/1055-9965.epi-12-1109

Shao, C., Zhao, M., Chen, X., Sun, H., Yang, Y., & Xiao, X. et al. (2019). Comprehensive Analysis of Individual Variation in the Urinary Proteome Revealed Significant Gender Differences. Molecular & Cellular Proteomics, 18(6), 1110-1122. doi: 10.1074/mcp.ra119.001343

Vineis, P. (2018). From John Snow to omics: the long journey of environmental epidemiology. European Journal Of Epidemiology, 33(4), 355-363. doi: 10.1007/s10654-018-0398-4

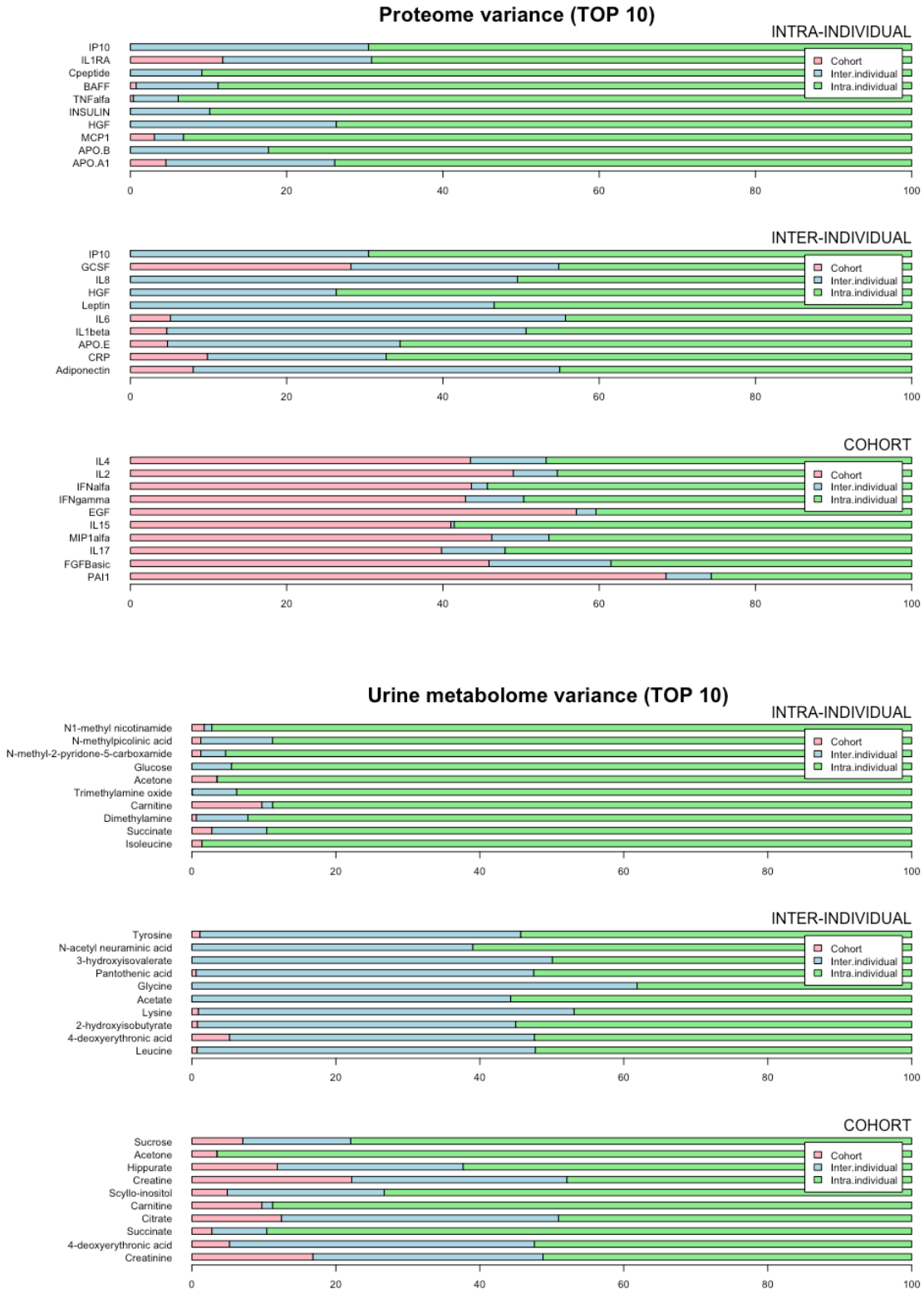
Wright, J., Small, N., Raynor, P., Tuffnell, D., Bhopal, R., & Cameron, N. et al. (2012). Cohort Profile: The Born in Bradford multi-ethnic family cohort study. International Journal Of Epidemiology, 42(4), 978-991. doi: 10.1093/ije/dys112

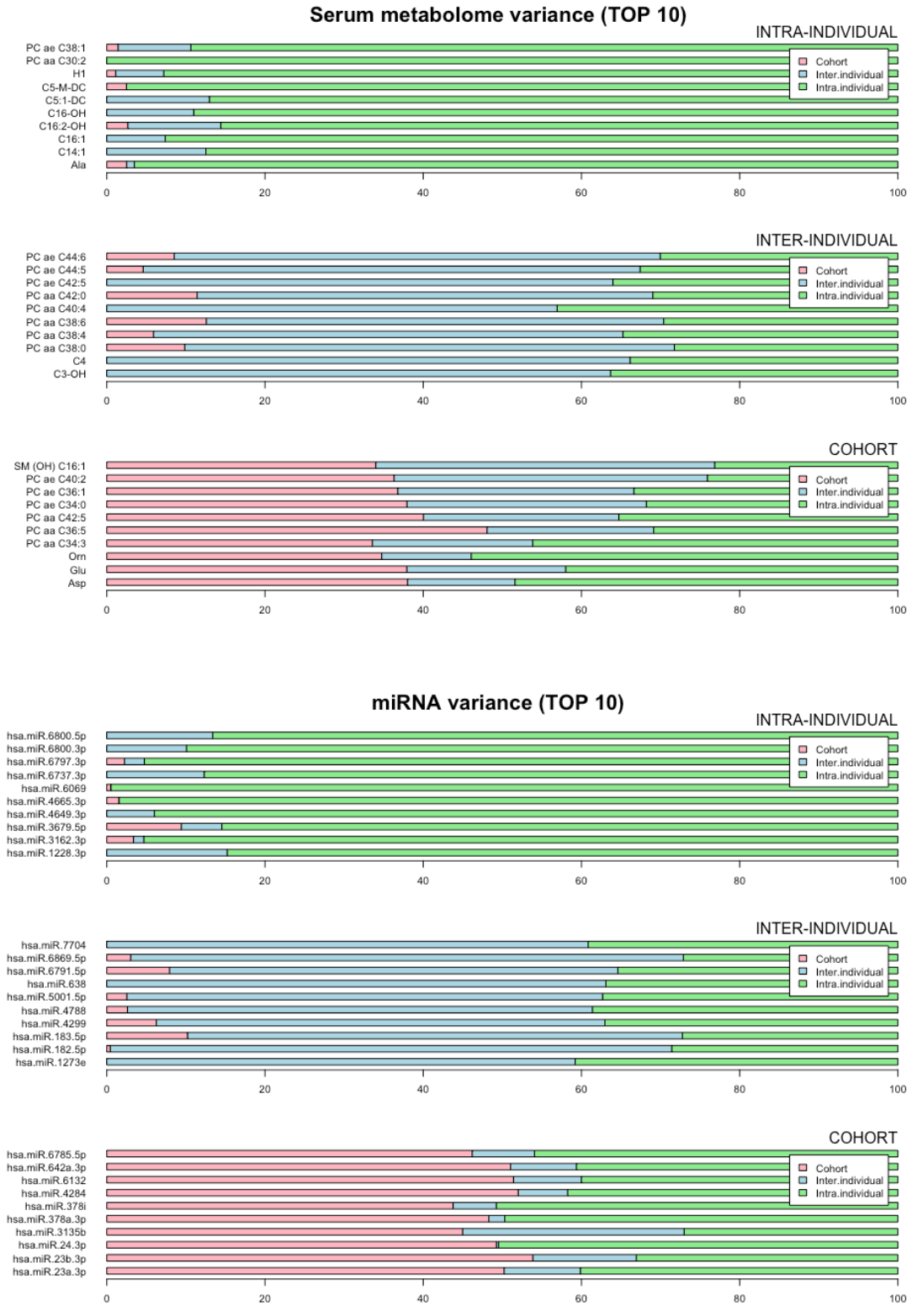
Yoon, H., Belmonte, K., Kasten, T., Bateman, R., & Kim, J. (2017). Intra- and Inter-individual Variability of microRNA Levels in Human Cerebrospinal Fluid: Critical Implications for Biomarker Discovery. Scientific Reports, 7(1). doi: 10.1038/s41598-017-13031-w

Zhan, X., Long, Y., & Lu, M. (2018). Exploration of variations in proteome and metabolome for predictive diagnostics and personalized treatment algorithms: Innovative approach and examples for potential clinical application. Journal Of Proteomics, 188, 30-40. doi: 10.1016/j.jprot.2017.08.020

# Annex 1

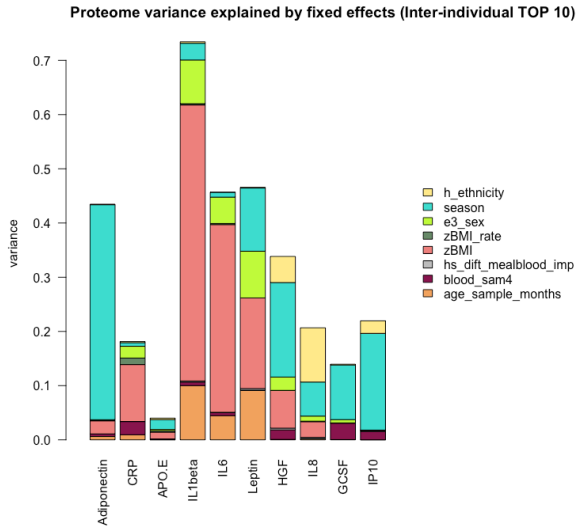
## A - Top 10 omics markers for intra-individual, inter-individual and cohort variability by omics

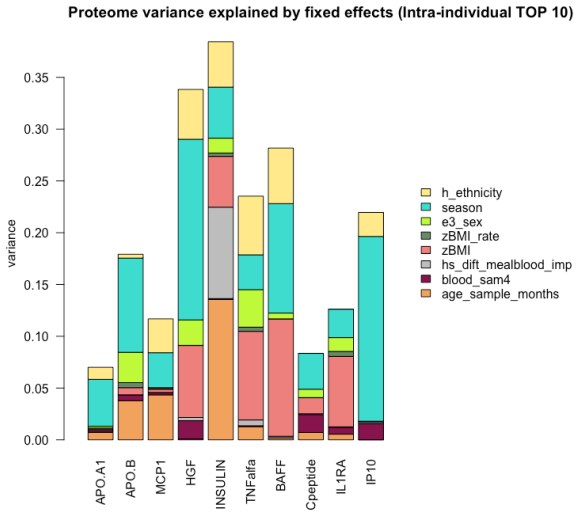


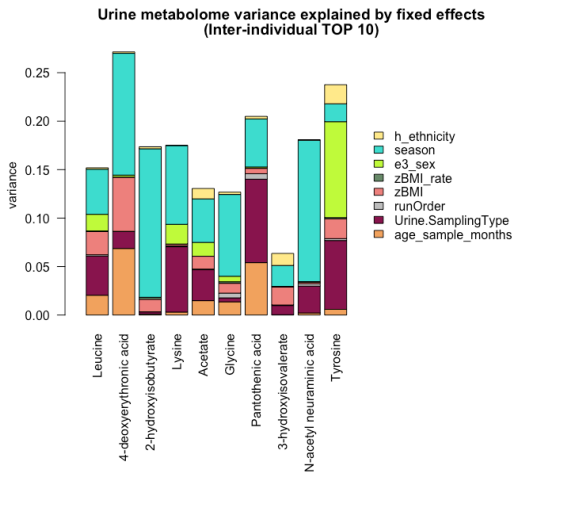


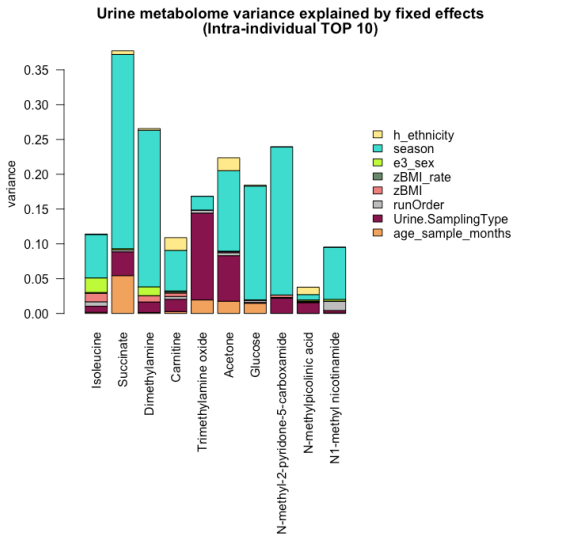

*Top 10 not available for transcriptome since thousands of markers have the highest values.

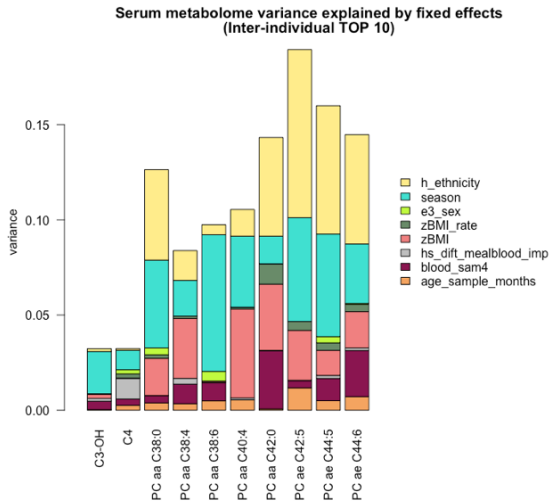
# B - Top 10 omics markers by variability of the explanatory variables by omics

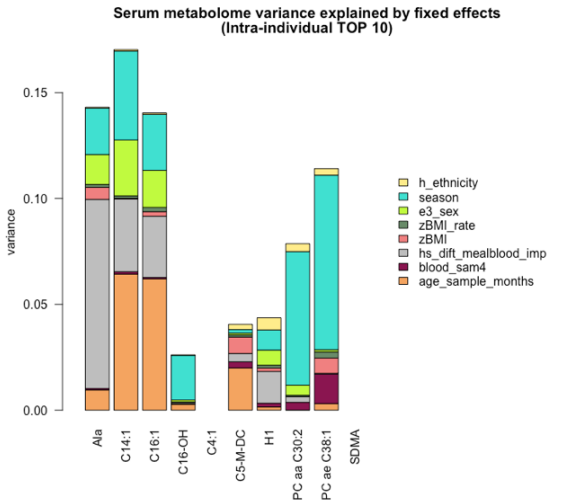


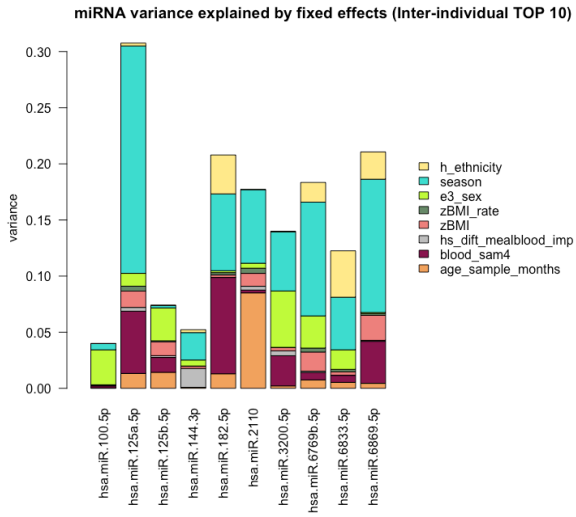



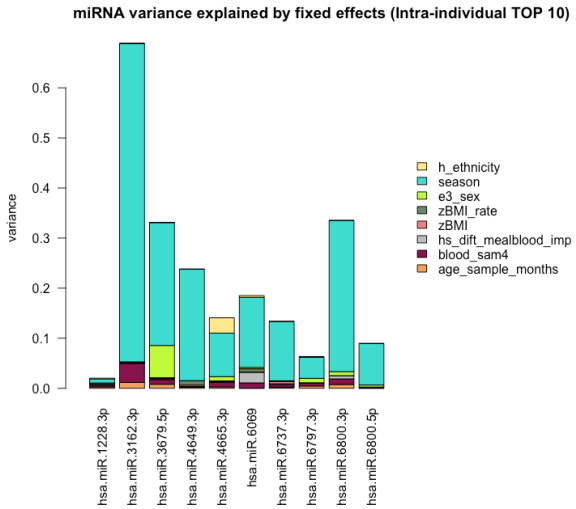



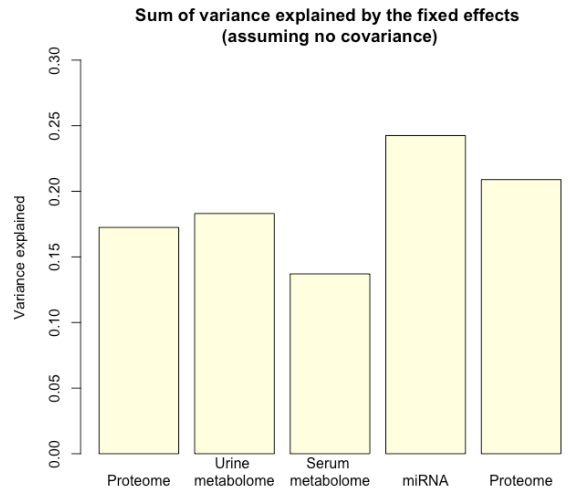



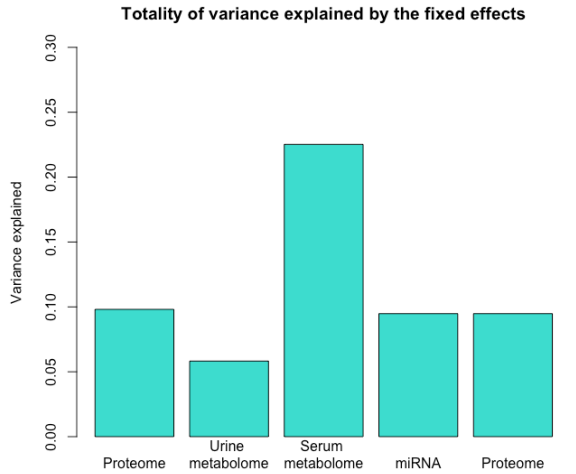


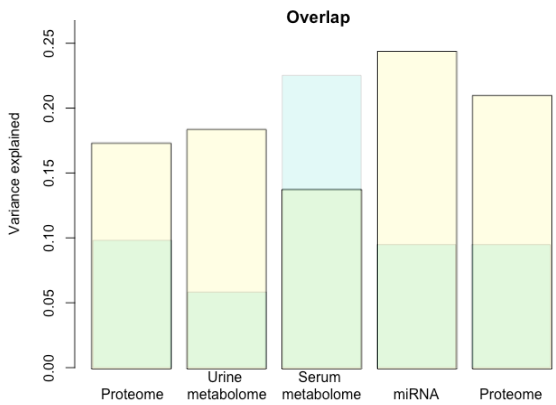




# Annex 2

## A - Mean over-/underestimation of the proportion of variance due to the explanatory variables

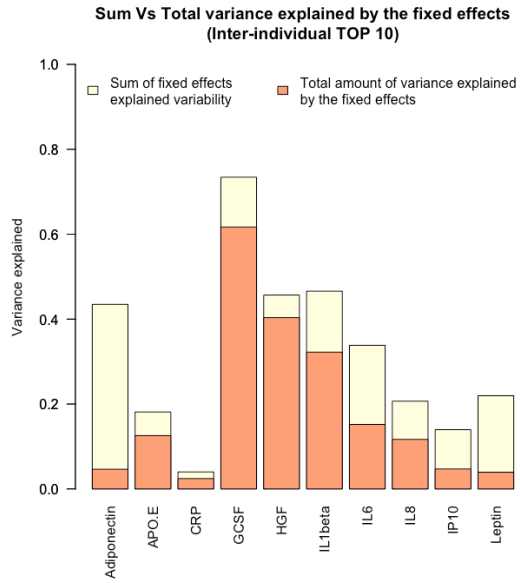


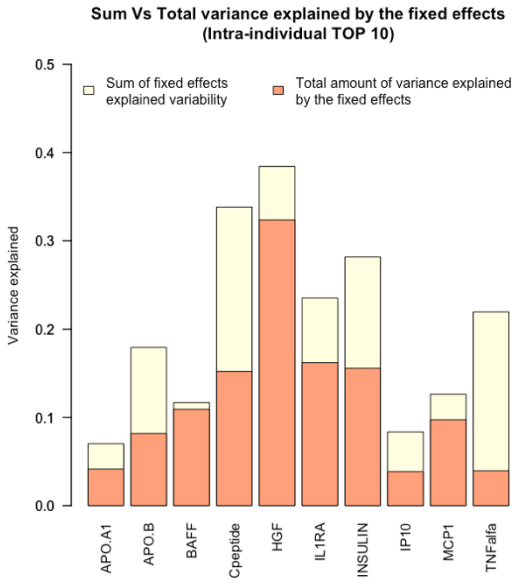



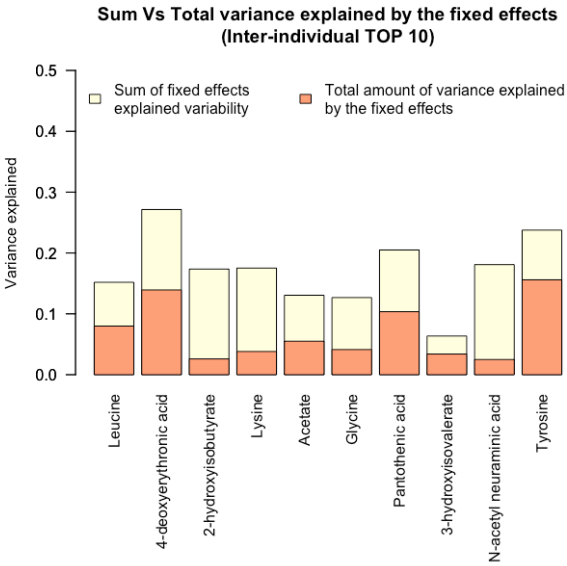
## B - Mean over-/underestimation of the proportion of variance due to the explanatory variables for the Top 10 intra- and inter-individual variability markers

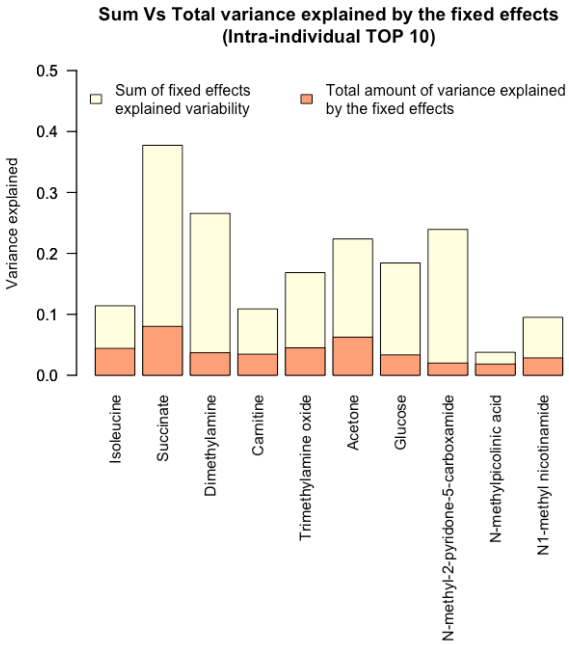
Proteome



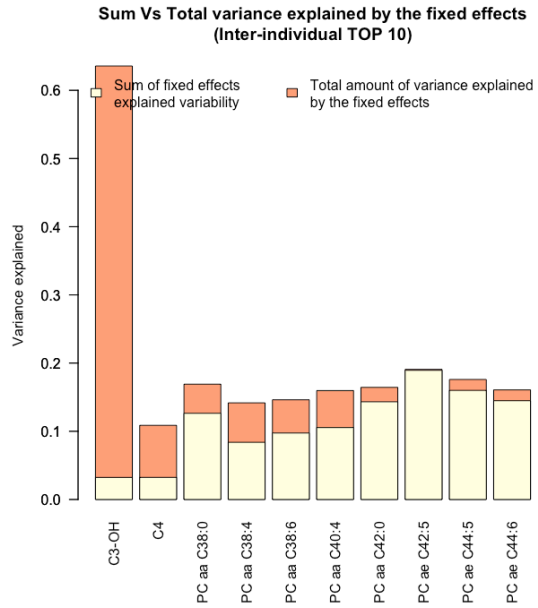



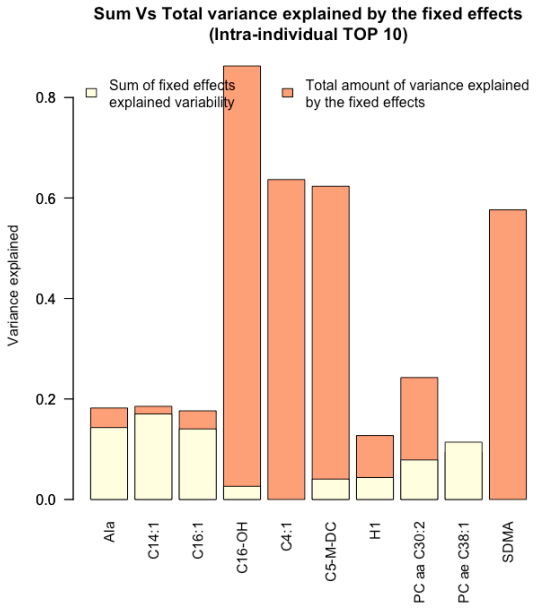

Urine metabolome



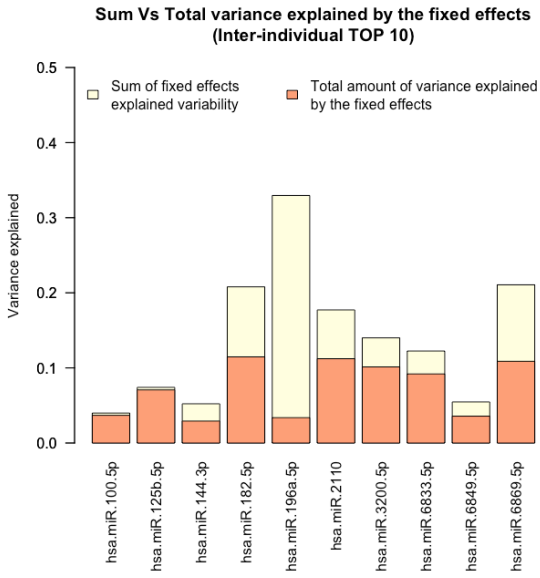



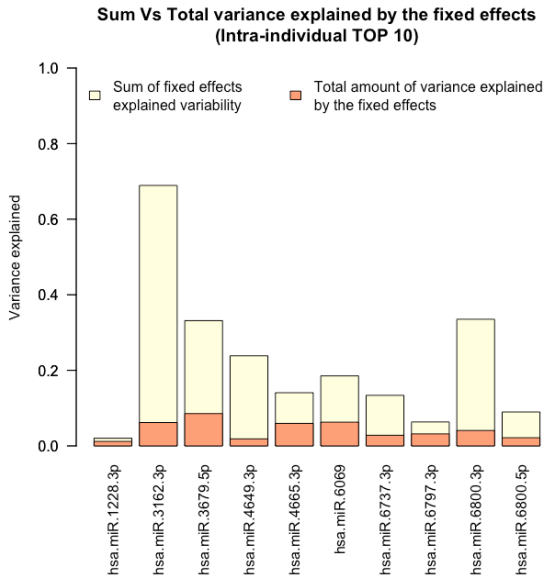
Serum metabolome




miRNA




*Top 10 not available for transcriptome since thousands of markers have the highest values.