Omics Modules for Exposome-Health Associations (OMEXA)

MSc Tessa Schillemans (KI)
MSc Elise Nordin (Chalmers)
PhD Anton Ribbenstedt (Chalmers)

Carl Brunius, Associate Professor | Computational Metabolomics | carl.brunius@chalmers.se
Dept of Biology and Biological Engineering
Food Science & Nutrition
Chalmers University of Technology

CHALMERS
UNIVERSITY OF TECHNOLOGY

Introducing the Data Challenge Team

MSc Elise Nordin

Irritable Bowel Syndrome, FODMAPs & Gluten

Metabolomics

Microbiota

PhD Anton Ribbenstedt

Exposome – Health

LongITools

Metabolomics

Carl Brunius

Comput. Metabolomics

Exposures & Health

Molecular epidemiology

MultiOmics

MultiMethods

MSc Tessa Schillemans

Pollutants & health

Metabolomics

Epidemiology

Computational Metabolomics Group

Metabolomics Data Generation

- Improved pre-processing
- Improved identification
- Increased thruput

Associations Exposure-Omics-Health

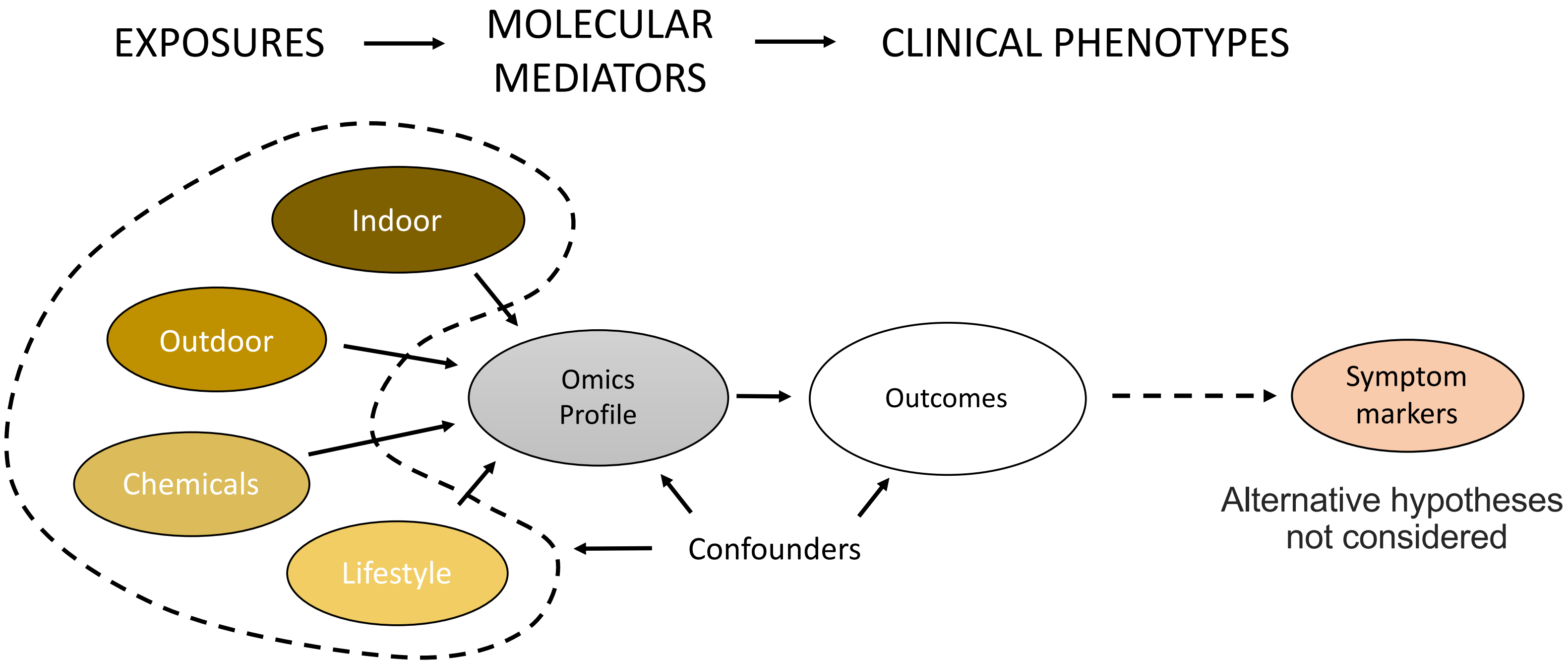
- Machine learning-based
  - molecular epidemiology
- Explorative analysis for
  - hypothesis generation
- Multi-exposure / multi-outcome modelling
  - using ”meet-in-the-middle” approach

Omics Data Analysis

- Machine learning methods
  - Reduce bias
  - Variable selection
  - Adjustment for covariates
- Data analysis architecture
  - Meeting multiple demands
  - by combining analytical approaches

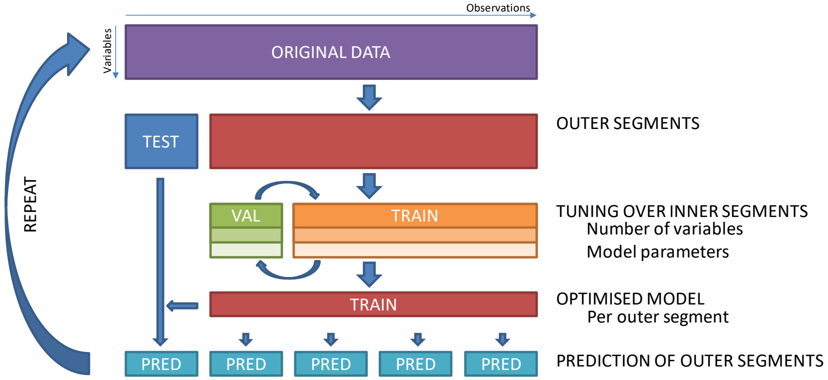
Impact of Combined Exposures on Metabolic Health

A meet-in-the-middle approach



Core methods

- MUVR – Machine learning to manage multi-collinear variabels and implicit interactions
  - Random Forest
  - Repeated double cross-validation
  - Minimally biased variable selection
  - Shi et al 2019. Bioinformatics
- Partial correlation
  - Further examine variables selected from MUVR
  - Adjusted for covariates (cohort, sex and age)
- GLM
  - Associate Omics variables with outcomes
  - Adjusted for covariates (cohort, sex and age)
- Latent variable analysis
  - Identification of ”Exposure modules” (Factor Analysis)
  - Identification of ”Omics modules” (PCA)
- TriPlot
  - Visualize ”everything” ☺
  - Schillemans et al 2019. Metabolites



Observations

Variables

ORIGINAL DATA

TEST

OUTER SEGMENTS

VAL

TRAIN

TUNING OVER INNER SEGMENTS
Number of variables
Model parameters

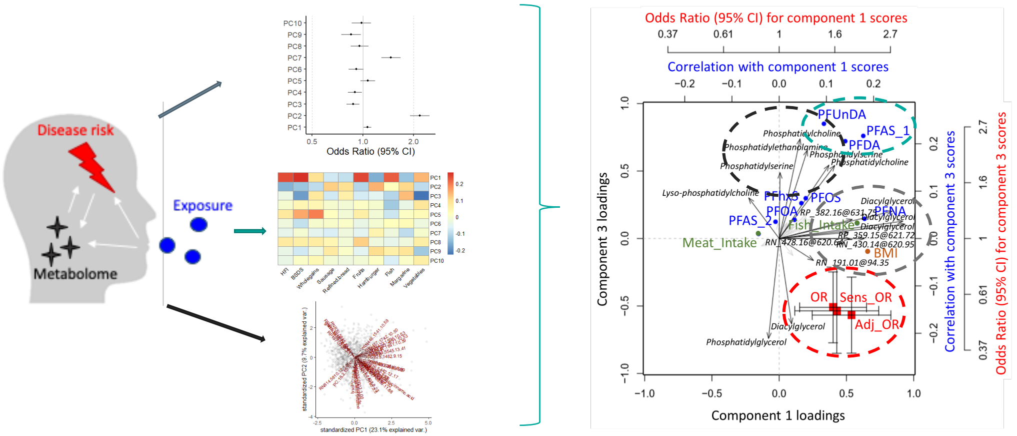
TRAIN

OPTIMISED MODEL
Per outer segment

REPEAT

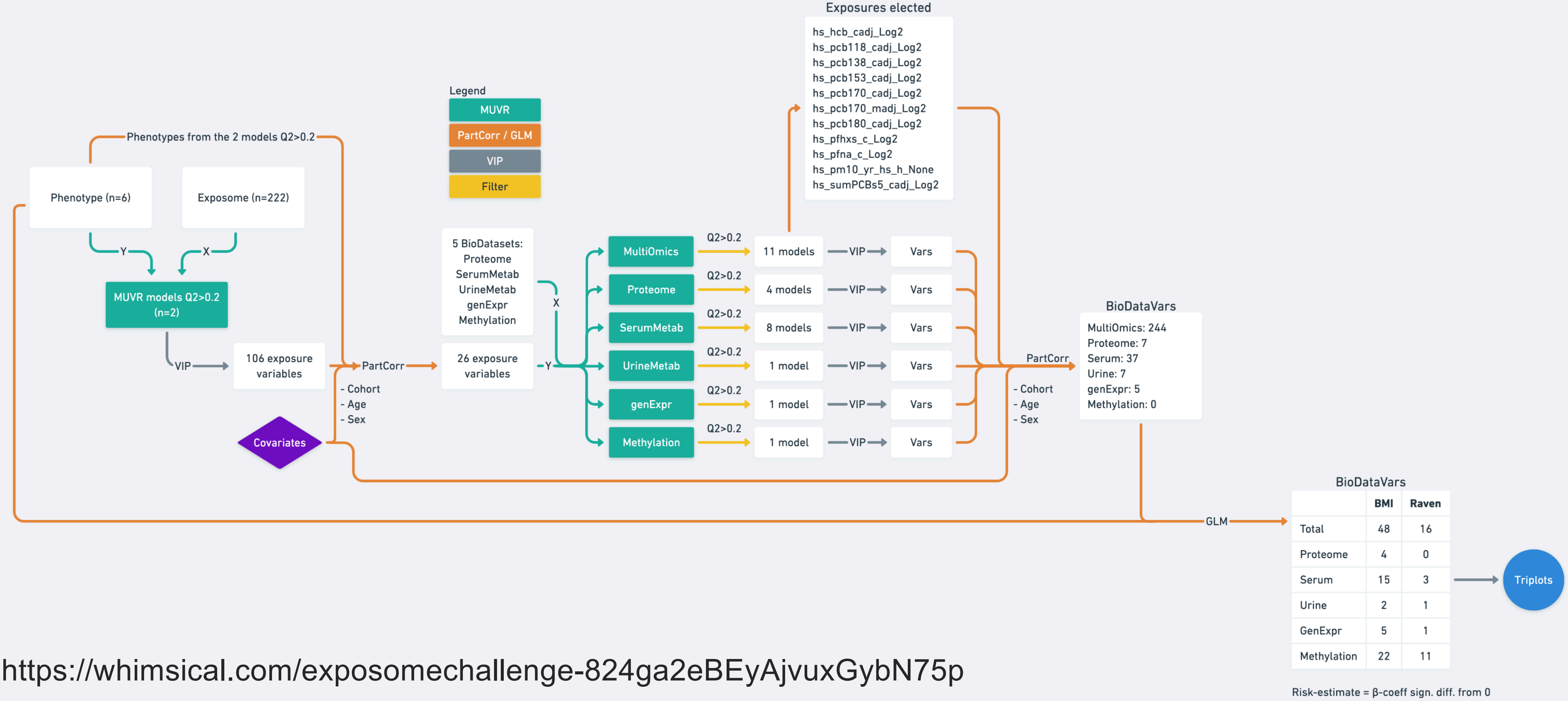
PRED PRED PRED PRED PRED

PREDICTION OF OUTER SEGMENTS



Analytical pipeline

CHALMERS
UNIVERSITY OF TECHNOLOGY



Legend
- MUVR
- PartCorr / GLM
- VIP
- Filter

Phenotypes from the 2 models Q2>0.2

Phenotype (n=6)

Exposome (n=222)

Y

X

MUVR models Q2>0.2
(n=2)

VIP

106 exposure variables

Covariates

PartCorr
- Cohort
- Age
- Sex

26 exposure variables

5 BioDatasets:
Proteome
SerumMetab
UrineMetab
genExpr
Methylation

X

Y

MultiOmics — Q2>0.2 → 11 models — VIP → Vars

Proteome — Q2>0.2 → 4 models — VIP → Vars

SerumMetab — Q2>0.2 → 8 models — VIP → Vars

UrineMetab — Q2>0.2 → 1 model — VIP → Vars

genExpr — Q2>0.2 → 1 model — VIP → Vars

Methylation — Q2>0.2 → 1 model — VIP → Vars

Exposures elected
hs_hcb_cadj_Log2
hs_pcb118_cadj_Log2
hs_pcb138_cadj_Log2
hs_pcb153_cadj_Log2
hs_pcb170_cadj_Log2
hs_pcb170_madj_Log2
hs_pcb180_cadj_Log2
hs_pfhxs_c_Log2
hs_pfna_c_Log2
hs_pm10_yr_hs_h_None
hs_sumPCBs5_cadj_Log2

PartCorr
- Cohort
- Age
- Sex

BioDataVars
MultiOmics: 244
Proteome: 7
Serum: 37
Urine: 7
genExpr: 5
Methylation: 0

GLM

BioDataVars

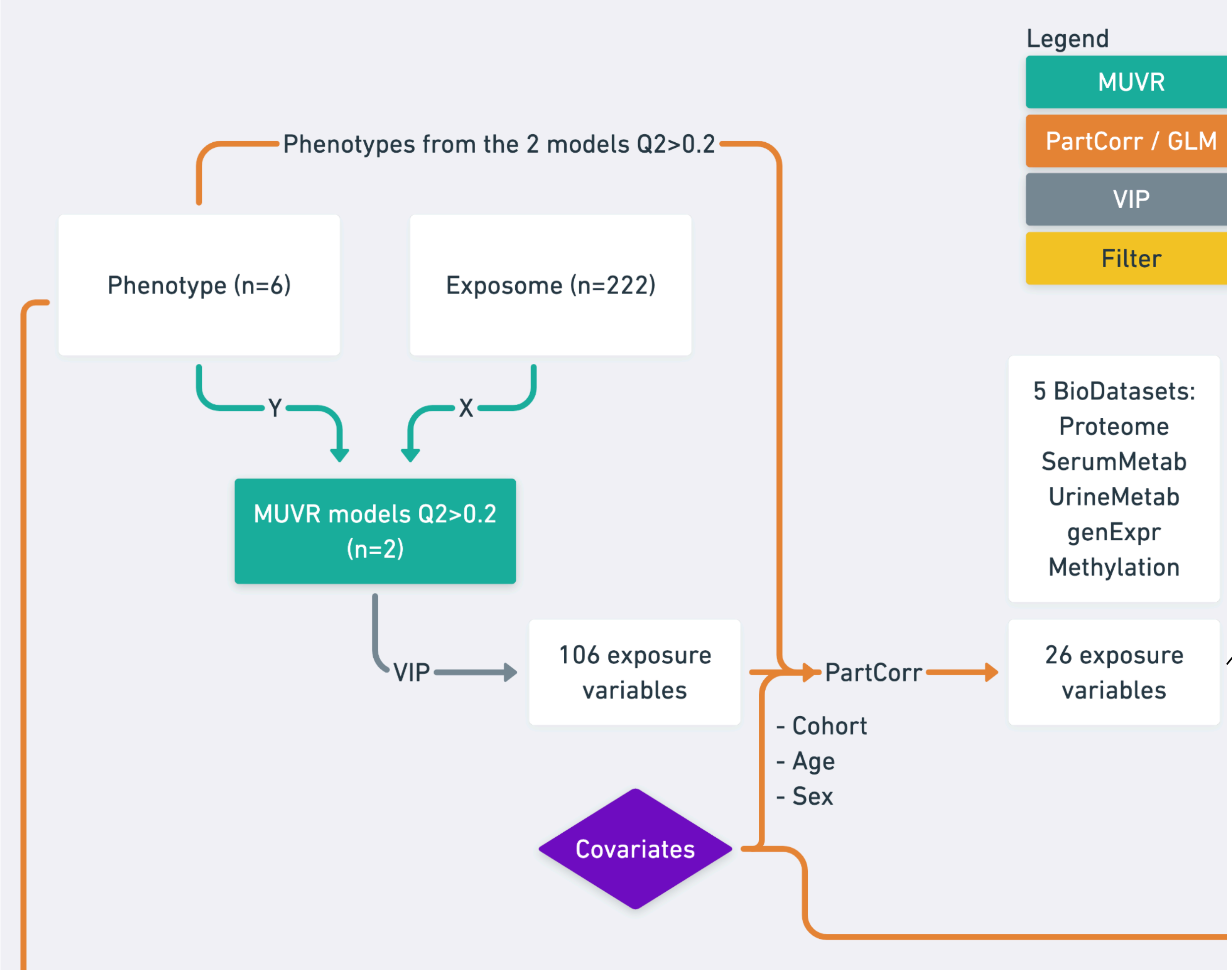
	BMI	Raven
Total	48	16
Proteome	4	0
Serum	15	3
Urine	2	1
GenExpr	5	1
Methylation	22	11

Risk-estimate = β-coeff sign. diff. from 0

Triplots

https://whimsical.com/exposomechallenge-824ga2eBEyAjvuxGybN75p

Selecting Exposure Variables



Legend

- MUVR
- PartCorr / GLM
- VIP
- Filter

Phenotypes from the 2 models Q2>0.2

Phenotype (n=6)

Exposome (n=222)

Y

X

MUVR models Q2>0.2 (n=2)

VIP

106 exposure variables

PartCorr

26 exposure variables

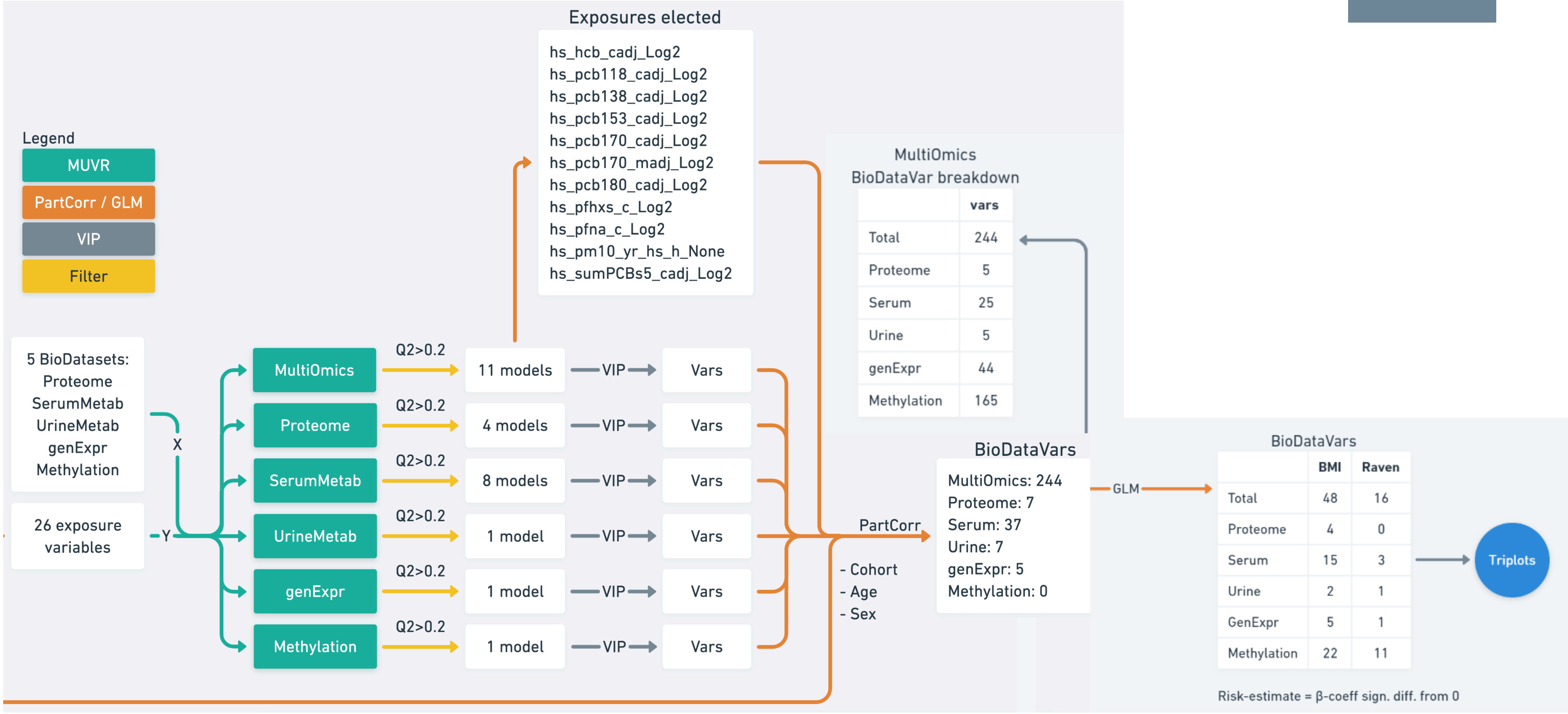
5 BioDatasets:
Proteome
SerumMetab
UrineMetab
genExpr
Methylation

- Cohort
- Age
- Sex

Covariates

h_Benzene_Log
h_PM_Log
hs_as_c_Log2
hs_bpa_madj_Log2
hs_bupa_madj_Log2
hs_cu_c_Log2
hs_dde_cadj_Log2
hs_ddt_cadj_Log2
hs_dep_cadj_Log2
hs_hcb_cadj_Log2
hs_mo_c_Log2
hs_pbde153_cadj_Log2
hs_pcb118_cadj_Log2
hs_pcb138_cadj_Log2
hs_pcb153_cadj_Log2
hs_pcb170_cadj_Log2
hs_pcb170_madj_Log2
hs_pcb180_cadj_Log2
hs_pfhxs_c_Log2
hs_pfhxs_m_Log2
hs_pfna_c_Log2
hs_pfoa_c_Log2
hs_pfoa_m_Log2
hs_pfos_c_Log2
hs_pm10_yr_hs_h_None
hs_sumPCBs5_cadj_Log2

Selecting Omics variabels



Legend
- MUVR
- PartCorr / GLM
- VIP
- Filter

5 BioDatasets:
Proteome
SerumMetab
UrineMetab
genExpr
Methylation

26 exposure variables

X

Y

	Q2>0.2			
MultiOmics	Q2>0.2	11 models	VIP	Vars
Proteome	Q2>0.2	4 models	VIP	Vars
SerumMetab	Q2>0.2	8 models	VIP	Vars
UrineMetab	Q2>0.2	1 model	VIP	Vars
genExpr	Q2>0.2	1 model	VIP	Vars
Methylation	Q2>0.2	1 model	VIP	Vars

Exposures elected

hs_hcb_cadj_Log2
hs_pcb118_cadj_Log2
hs_pcb138_cadj_Log2
hs_pcb153_cadj_Log2
hs_pcb170_cadj_Log2
hs_pcb170_madj_Log2
hs_pcb180_cadj_Log2
hs_pfhxs_c_Log2
hs_pfna_c_Log2
hs_pm10_yr_hs_h_None
hs_sumPCBs5_cadj_Log2

PartCorr

- Cohort
- Age
- Sex

MultiOmics
BioDataVar breakdown

	vars
Total	244
Proteome	5
Serum	25
Urine	5
genExpr	44
Methylation	165

BioDataVars

MultiOmics: 244
Proteome: 7
Serum: 37
Urine: 7
genExpr: 5
Methylation: 0

GLM

BioDataVars

	BMI	Raven
Total	48	16
Proteome	4	0
Serum	15	3
Urine	2	1
GenExpr	5	1
Methylation	22	11

Risk-estimate = β-coeff sign. diff. from 0

Triplots

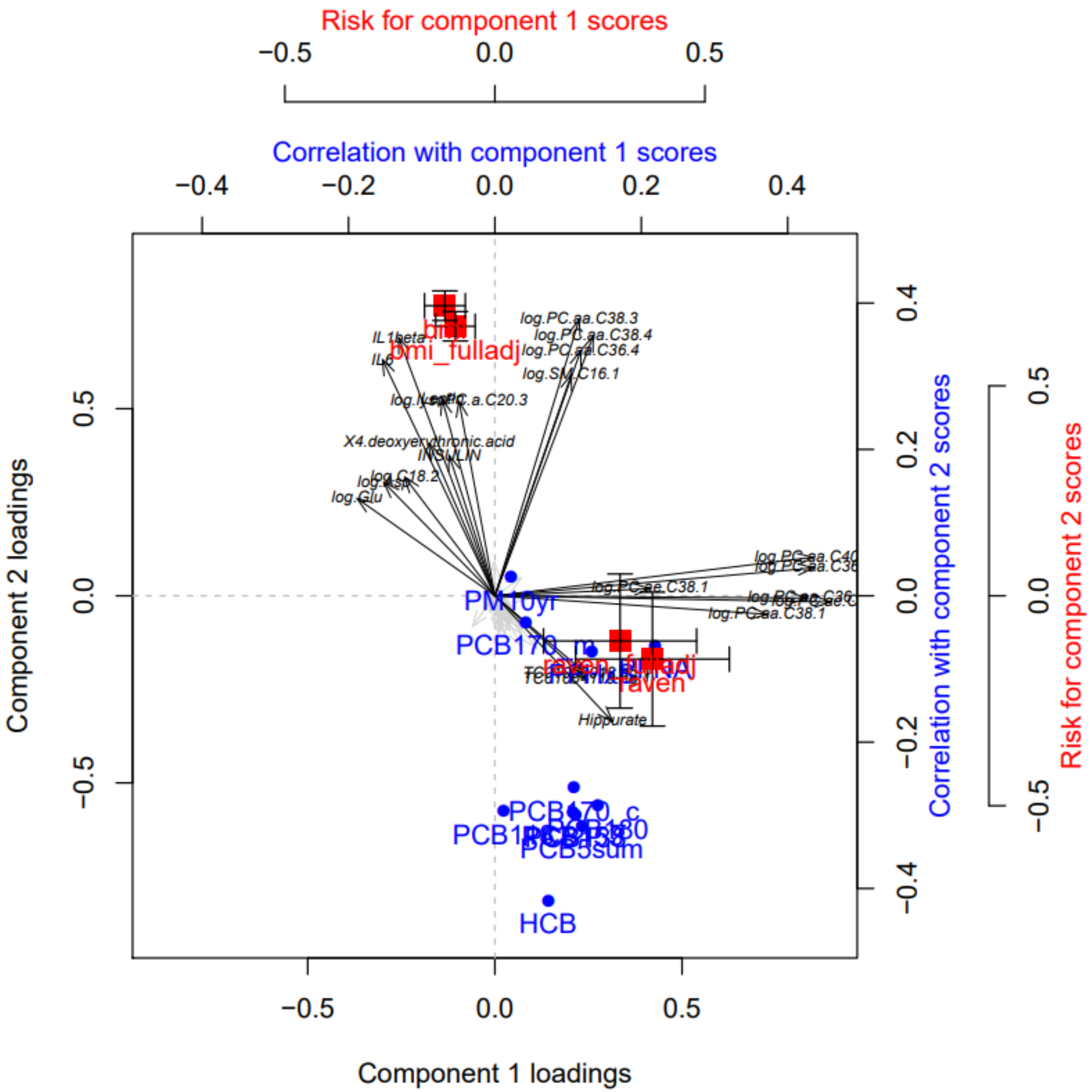
Selecting Omics variabels

Exposures	Multiomics	SerumMetab	UrineMetab	Proteomics	Genomics	Methylation
hs_pm10_yr_hs_h_None	0.62	0.59			0.22	0.20
hs_hcb_cadj_Log2	0.38	0.33	0.21	0.30		
hs_pcb118_cadj_Log2	0.23	0.36				
hs_pcb138_cadj_Log2	0.24			0.20		
hs_pcb153_cadj_Log2	0.35	0.24		0.26		
hs_pcb170_cadj_Log2	0.23					
hs_pcb170_madj_Log2	0.26	0.32				
hs_pcb180_cadj_Log2	0.24					
hs_sumPCBs5_cadj_Log2	0.36	0.25		0.27		
hs_pfhxs_c_Log2	0.28	0.28				
hs_pfna_c_Log2	0.23					
hs_pfos_c_Log2		0.20				

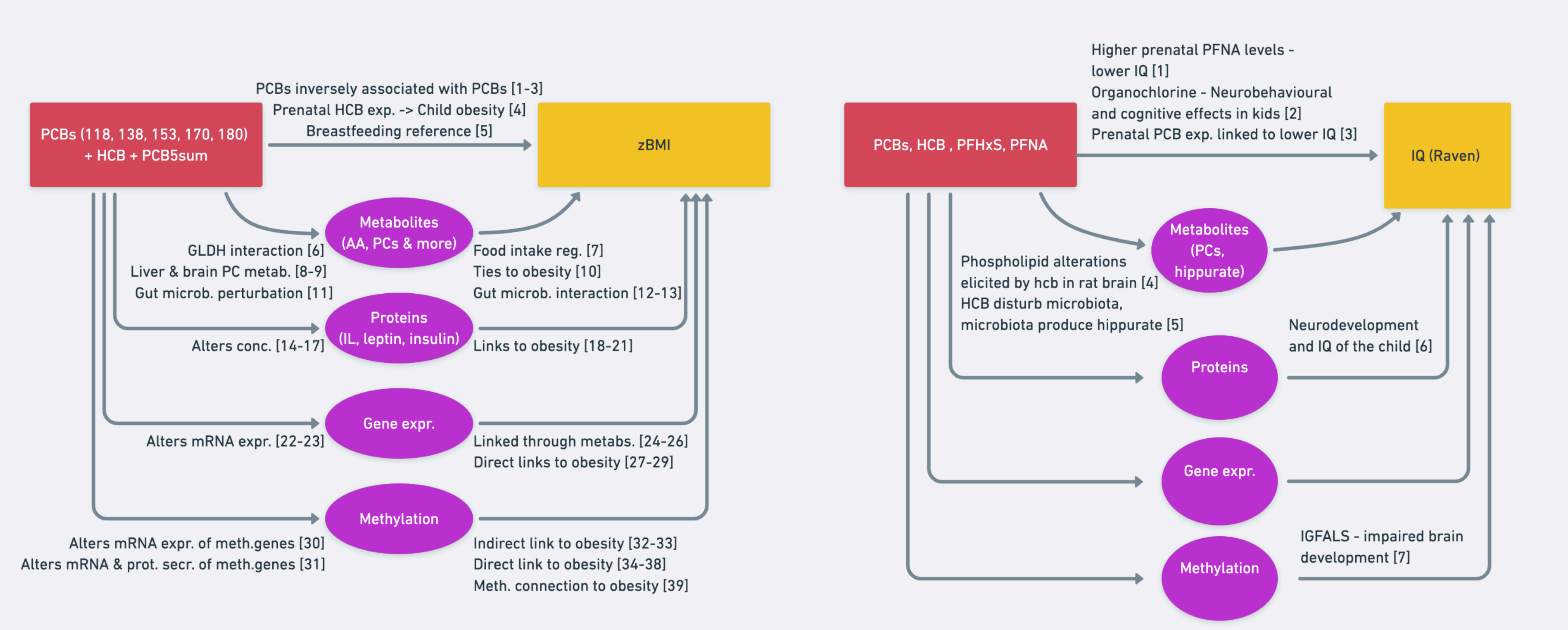
BMI RAVEN BOTH

Putting everything together: the Triplot

- Omics modules
- Partial correlation with exposures adjusted for cohort, sex and age
- Association with outcomes adjusted for cohort, sex and age
- Associations with outcomes additionally adjusted for maternal education/age/BMI, parity, year of birth, gestational age, weight gain



Diving deeper into BMI and IQ



PCBs (118, 138, 153, 170, 180) + HCB + PCB5sum

PCBs inversely associated with PCBs [1-3]
Prenatal HCB exp. -> Child obesity [4]
Breastfeeding reference [5]

zBMI

GLDH interaction [6]
Liver & brain PC metab. [8-9]
Gut microb. perturbation [11]

Metabolites (AA, PCs & more)

Food intake reg. [7]
Ties to obesity [10]
Gut microb. interaction [12-13]

Alters conc. [14-17]

Proteins (IL, leptin, insulin)

Links to obesity [18-21]

Alters mRNA expr. [22-23]

Gene expr.

Linked through metabs. [24-26]
Direct links to obesity [27-29]

Alters mRNA expr. of meth.genes [30]
Alters mRNA & prot. secr. of meth.genes [31]

Methylation

Indirect link to obesity [32-33]
Direct link to obesity [34-38]
Meth. connection to obesity [39]

PCBs, HCB , PFHxS, PFNA

Higher prenatal PFNA levels - lower IQ [1]
Organochlorine - Neurobehavioural and cognitive effects in kids [2]
Prenatal PCB exp. linked to lower IQ [3]

IQ (Raven)

Phospholipid alterations elicited by hcb in rat brain [4]
HCB disturb microbiota, microbiota produce hippurate [5]

Metabolites (PCs, hippurate)

Proteins

Neurodevelopment and IQ of the child [6]

Gene expr.

Methylation

IGFALS - impaired brain development [7]

Strong support in literature for discovered associations!

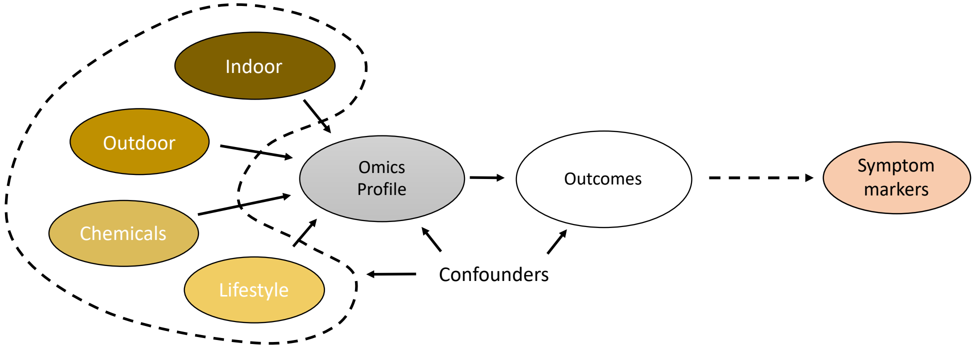
Potential targets fur future research!

(See weblink for references)

https://whimsical.com/7j3CLvmih62zPDMgAGeHYo

Limitations

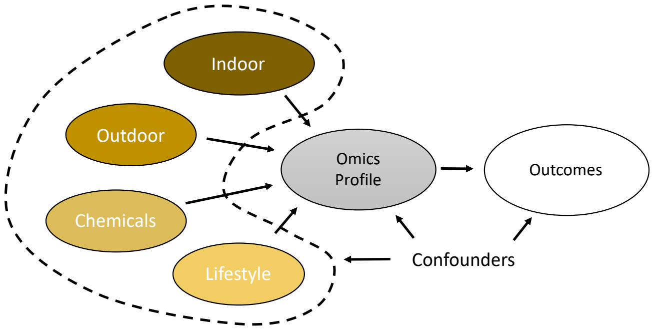
- Selection of exposures
  - Exposures with ”mixed associations” to outcomes may be ignored (example: PFAS – T2D; Schillemans et al 2020)
  - Sensitive to adjustment (example: PM10 – Raven; this data)
- Selection of omics variables
  - Mechanisms not captures by omics measurements (example: regulations in the mother before and during pregnancy; limited coverage of the metabolome and the proteome)
- Sensitivity to confounders
  - Several unmeasured confounders (example: breast feeding)
- We only investigate 2 Omics ”modules”
  - Details may be hidden in smaller components
- Does not investigate alternative causal hypotheses
  - Exposure -> Outcome -> Metabolism (Omics variables primarily as markers of symptoms)
- Does not investigate ”true” causality
  - Interventions
  - Medelian randomization
  - Mediation analysis



Take Home Message

- Data-driven approach
  - We don’t possess specific domain knowledge
  - Combines machine learning with GLM to manage both multi-collinearities, implicit and explicit modelling complexity
  - Relies on ”unbiased” variable selection
  - Meet-in-the-middle approach based on TriPlot tool
- Zooms in on ”most important” pathway from Exposure -> Metbolism -> Outcome
  - PCB -> BMI gives biological plausibility
  - Pollutants -> IQ provides potential targets for future research
- Meet-in-the-middle provides strong potential for generation of hypotheses

EXPOSURES → MOLECULAR MEDIATORS → CLINICAL PHENOTYPES



CHALMERS
UNIVERSITY OF TECHNOLOGY