ATHLETE
The Exposome from Evidence to Translation

Open science and causality in the Exposome era

Lessons learned from the Exposome Data Challenge

Léa Maitre, PhD

Barcelona Institute for Global Health (**ISGlobal**)

ISGlobal Barcelona Institute for Global Health

EXCELENCIA SEVERO OCHOA

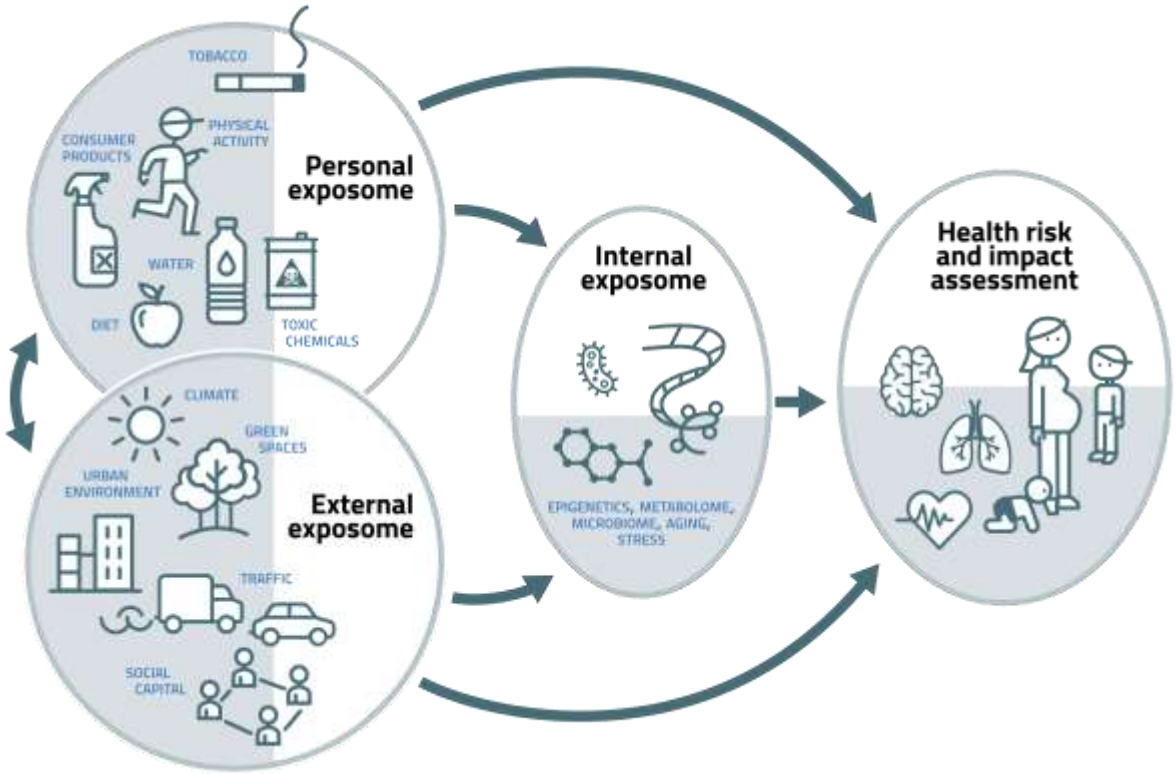
June 23rd 2022 – One Health Conference

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 874583

The Exposome concept

The totality of environmental exposures (meaning all non-genetic factors) that a person experiences, from conception onwards.

Chris Wild, 2005.



Classical approach in environmental epidemiology: single-exposure association

- Selective reporting of associations → Publication bias
- No correction for multiple testing (separate papers for each exposure)
- Cannot take into account confounding by co-exposures
- Lack of consideration of “mixture effect”

Exposome approach

calls for a holistic view of the effects of environmental exposures on human health

A collaborative, open-science approach to promote and accelerate innovation in Exposome research

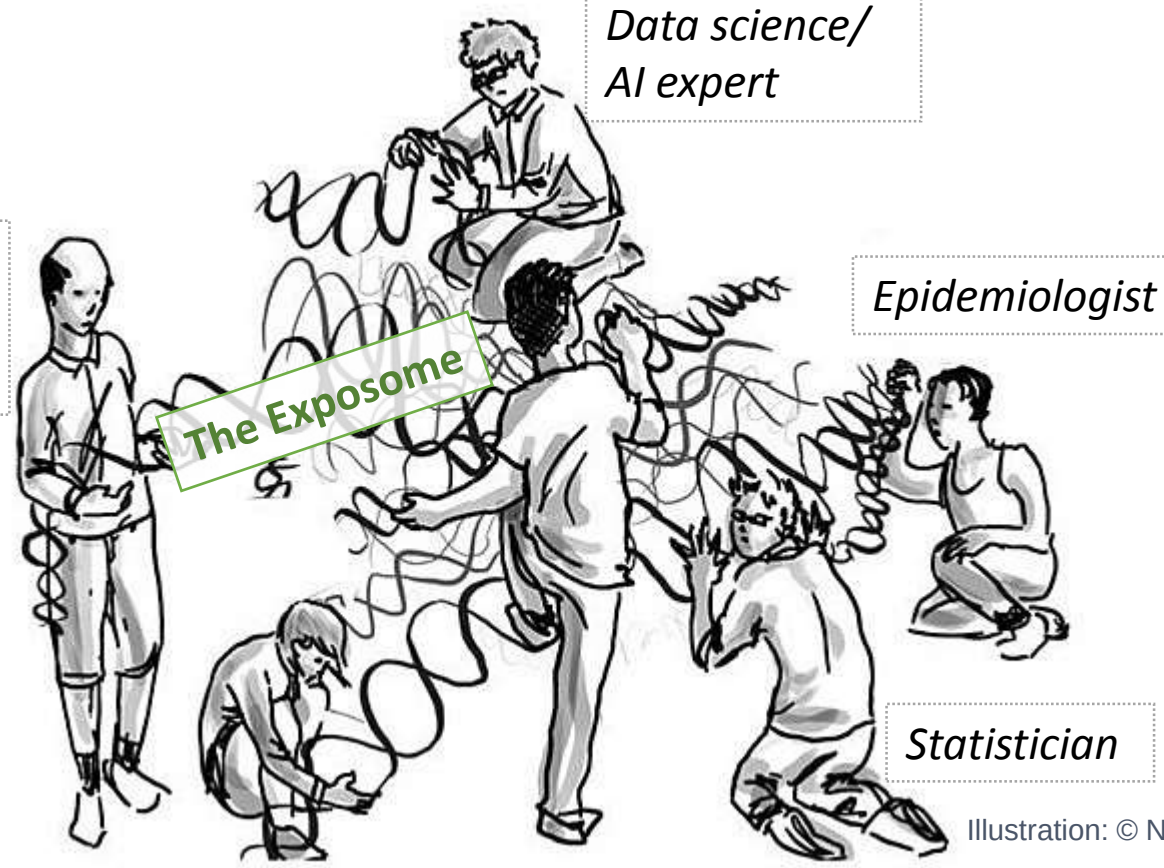


Illustration: © Natasha Stolovitzky-Brunner

The parable of the blind men and the elephant

Data Challenges to promote:

- Accelerated innovation
- Interdisciplinary collaboration
- Participation from around the world in COVID time
- Training material for educational purposes

The Exposome data challenge

- Event created in the framework of the ISGlobal Exposome hub and H2020 ATHLETE project

- Scientific publication pre-print https://arxiv.org/abs/2202.01680 (under review)

- Simulated data (based on the HELIX project) publicly available to challenge researchers on statistical tools to study exposome-health associations

ATHLETE
The Exposome from Evidence to Translation

ISGlobal
Barcelona Institute for Global Health

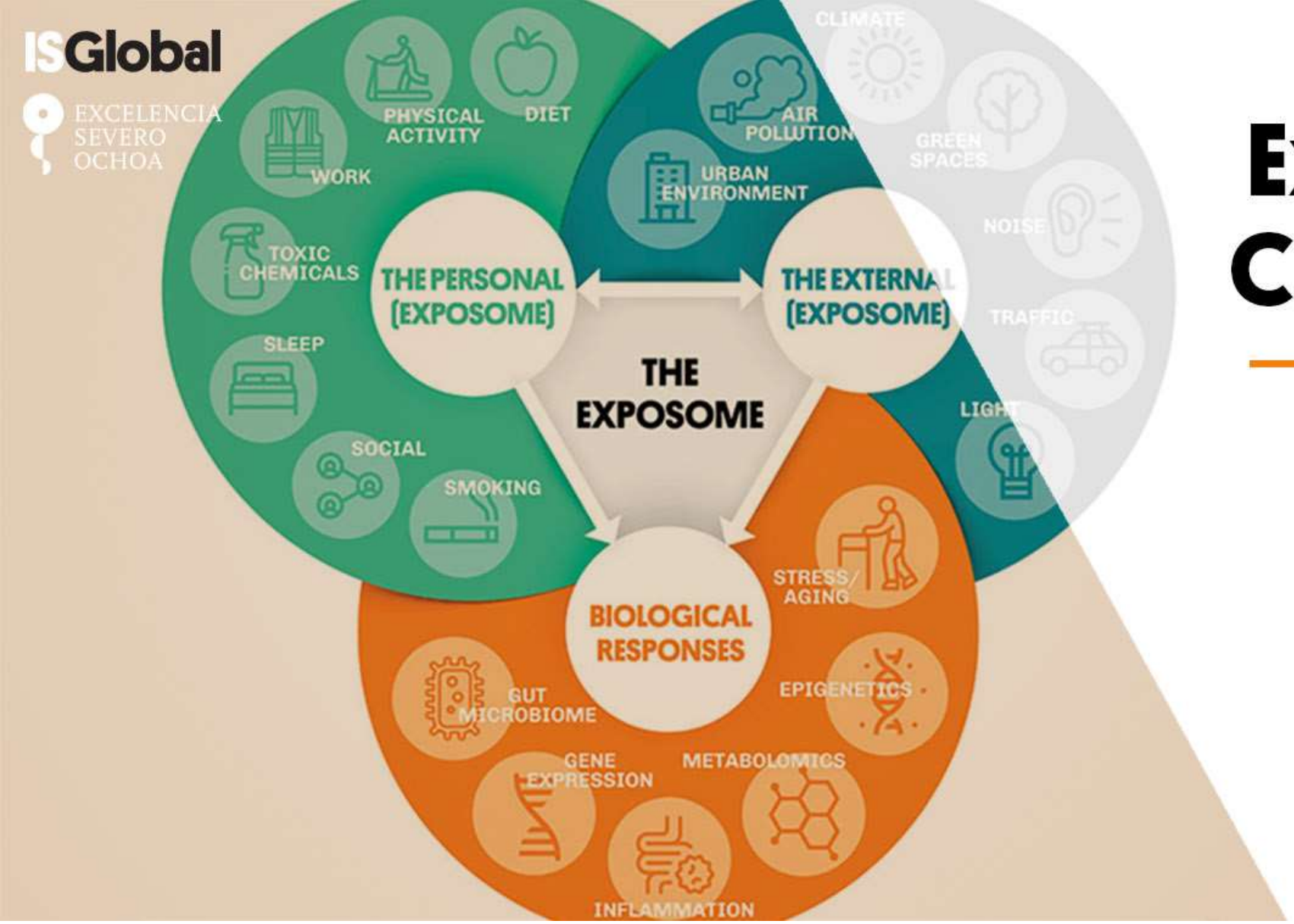
ISGlobal

EXCELENCIA SEVERO OCHOA

#ISGlobalExposome

Exposome Data Challenge Event

28-30 April 2021



- 25 selected teams out of 39 abstracts sent
- 307 participants online:
  - 101 North America
  - 186 Europe
  - 9 Asia
  - 8 Latin America
  - 2 Africa
  - 1 Australia

- Awards attributed from the public and from the comittee

The Human early-life exposome project (HELIX)

Study design

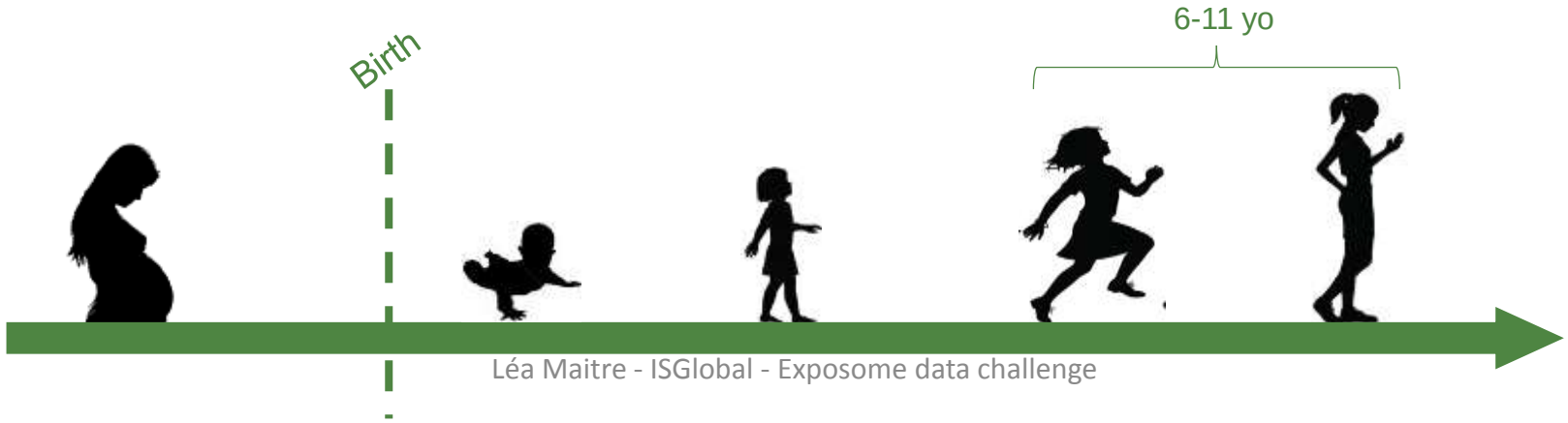
Six mother-child cohorts in Europe (n=1301)

Pregnant women enrolled at the beginning of their pregnancy

Follow-up of the children with a standardized clinical examination at 6-11 years old

Aim

To study the association between multiple exposures, molecular signatures, and child health outcomes.



Exposure assessment

>100 environmental factors

Assessed during **pregnancy** and **childhood** (at the time of the children follow-up, 6-11yo)

Outdoor exposures
(Geographic Information System)

Air pollution*
Noise†
Built environment†
Natural spaces†
Traffic
Meteorology*
Water DBP
Indoor air

Chemicals
(blood or urine biomarkers)

Organochlorines
PBDE
PFAS
Metals
Phthalates
Phenols
Organophosphate pesticides

Lifestyles
(questionnaires)

Smoking

Diet

Physical activity

Social and economic capital

Sleep

* Postnatal exposures available within different time window

† Postnatal exposures available at different location: home and school

Health outcomes

6 health outcomes

At birth or at the time of the children follow-up (6-11yo)

Continuous variables

Birth weight

Body mass index at 6-11yo

Categorical variables

Asthma at 6-11yo (binary)

Body mass index at 6-11yo (4 categories)

Count variables

Intelligence quotient at 6-11yo
Total correct answers (RAVEN test)

Neuro behavior at 6-11yo
Internalizing and externalizing problems (CBCL scale)

Covariates, potential confounders

Maternal and child data

Maternal data

Cohort of inclusion

Age

Education

Pre-pregnancy body mass index

Weight gain during pregnancy

Parity

Child data at birth

Sex

Gestational age

Year of birth

Native origin

Child data at 6-11yo

Age

Weight

Height

Omics data

>400,000 features

Assessed during **childhood** only (at the time of the children follow-up, 6-11yo)

White blood cells

DNA methylation
(386,518 CpGs)

Transcription

Gene expression
(58,254 transcripts)

Plasma & serum

Proteins
(36 proteins)

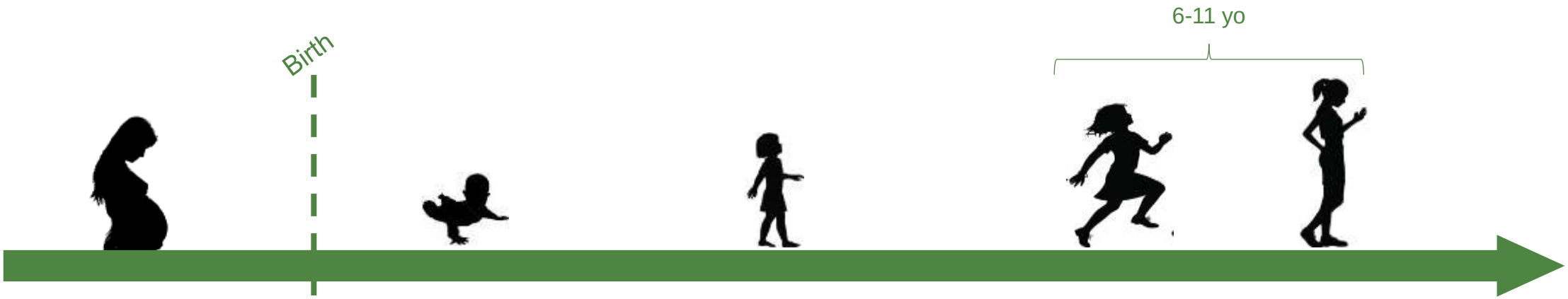
Metabolites
(177 metabolites)

Urine

Metabolites

(44 metabolites)

Data summary



Exposures ++

Health

Covariates

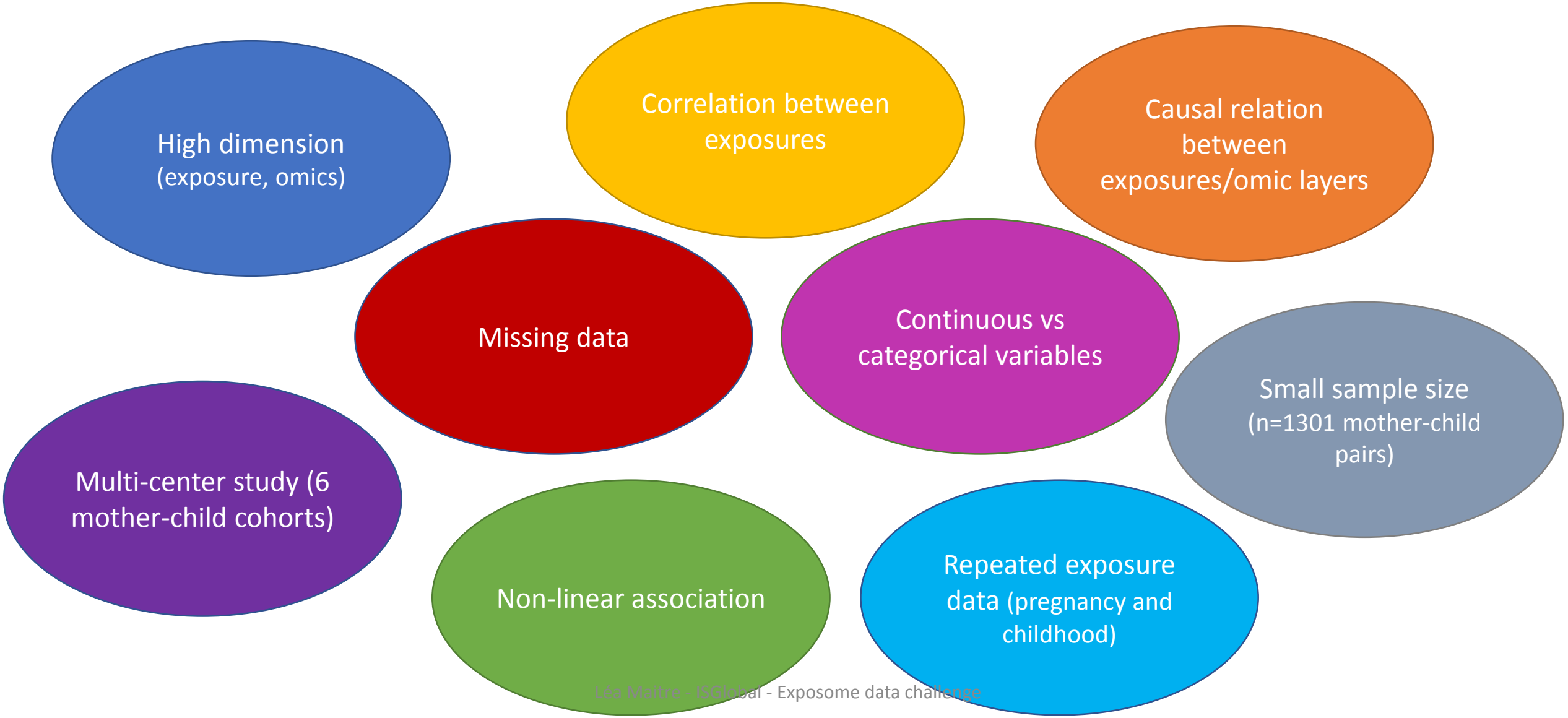
Exposures ++

Health

Covariates

Omics ++++

Particularity of the data



Challenge examples

Challenge 1: Combined effects of exposures

Identify combination of exposures, high-order interactions or exposure patterns that are particularly harmful or beneficial for one or several health outcomes.

Challenge 2: Using omics data to improve inference on the link between exposome and health.

Incorporate the different omics layers into the analysis linking the exposome and one or more health outcomes.

Challenge 3: Multi-omics analysis

Incorporate different layers of omics data (including exposome as one of the layers) to find patterns that can explain variations in one or more health outcomes.

Challenge 4: Causal structure in the exposome

Define hypothesized causal relationships between the different exposures and one health outcome, and incorporate this information into the analysis.

Challenge 5: Visualization techniques

Tools to visualize the complex relationships between the different components of the analysis, with the main aim of illustrating determinants of health effects.

/!\ Control for potential confounders and multicenter design.
Handle missing data.

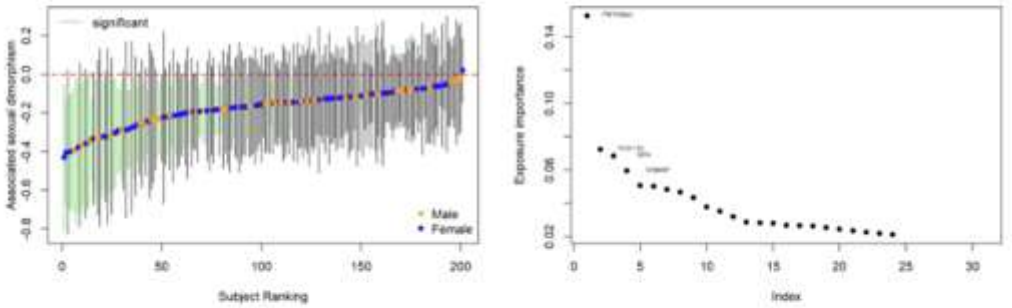
Popular vote: winner

Using causal random forest to determine exposure environments with high sexual dimorphisms

Alejandro Caceres, ISGlobal

Individuals with significant sex-effects on BMI

We found 46 individuals from 155 (test set) with significantly negative sexual dimorphism in BMI (M>F).



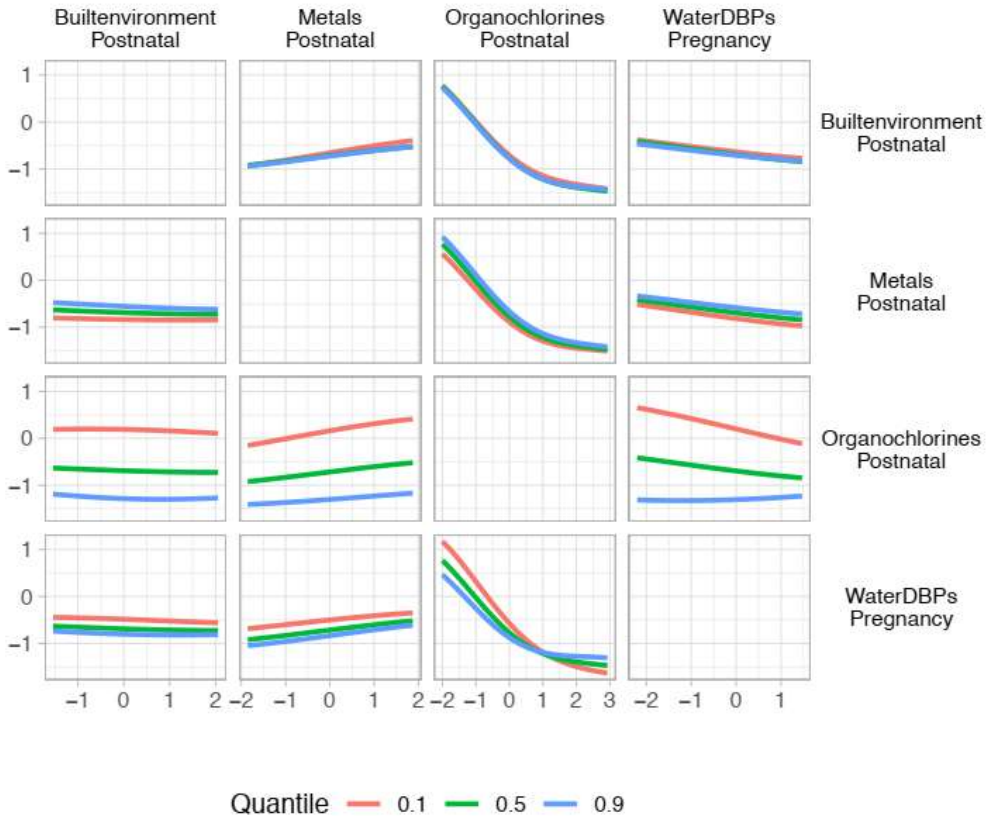
- We **selected** 31 informative exposures: Those whose interaction with sex significantly associated (nominal level) with zBMI. we adjusted by all covariates available in the exposome dataset.

Committee vote: winner 1

Quantifying Exposome-Health Associations with Bayesian Multiple Index Models

Glen McGee, University of Waterloo

github.com/glenmcgee/bsmim2
glen.mcgee@uwaterloo.ca

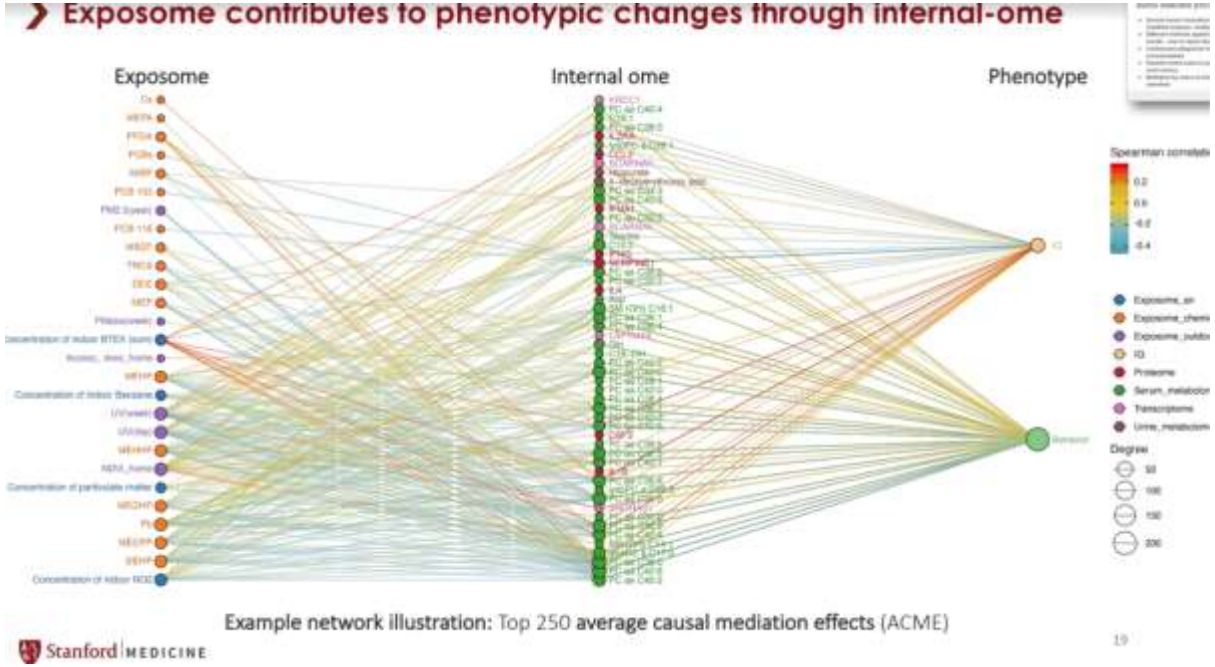


Interactions between multipollutant index and individual covariates

Committee vote: winner 2

Decoding unknown links between the exposome and health outcomes by multi-omics analysis

Xiaotao Shen, Stanford University



Ressources

Exposome data challenge

Dataset: https://github.com/isglobal-exposomeHub/ExposomeDataChallenge2021/blob/main/README.md

Data description: https://docs.google.com/document/d/1ul3v-sIniLuTjFB1F1CrFQIX8mrEXVnvSzOF7BCOnpQ/edit

Scientific publication https://arxiv.org/abs/2202.01680 (under review)

Slides and videos of the presentations https://www.isglobal.org/-/exposome-data-analysis-challenge
https://www.youtube.com/channel/UC0F3hR04UzUeKkcfAyiklta/featured

Code used shared on GitHub: https://github.com/isglobal-exposomeHub/ExposomeDataChallenge2021/tree/main/R_Code_Presentations

- ✓ Publicize methods for developers
- ✓ Code source for analysists

HELIX project

Data inventory: https://www.projecthelix.eu/index.php/es/data-inventory

Tamayo-Uria I, et al. The early-life exposome: Description and patterns in six European countries. Environ Int. 2019.
Maitre L, et al. Human Early Life Exposome (**HELIX**) study: a European population-based exposome cohort. BMJ Open. 2018.
Vrijheid M, et al. The human early-life exposome (**HELIX**): **project** rationale and design. Environ Health Perspect. 2014

Challenge challenges

- Making real-case, sensitive personal data, publicly available
  - Partial imputation of the data
- Finding a balance between well-defined research questions and generalizability of the results to a wide community
  - The first edition of the challenge was focused on general data analysis approach challenges, rather than specific research questions, which means methods presented could not be directly compared. The advantage was a broad catalogue of methods
- A short time frame, sometimes lack of adjustment for known biaises
- Appropriate recognition for the data-generating team

Challenge successes

- Knowledge transfer -> Training tools for students, already used by universities such as French Centrale sup elec or PhD students
- New collaborations formed
- Accelerated the rate of scientific discovery, out-of-the-box ideas created by multidisciplinary teams in a short time:
  - 25 teams who worked over a month, will never equal the work even of a consortium over 5 years (only 5 methods tested for the HELIX stat protocol, run mainly by one experienced statistician)

Thank you

ISGlobal
Barcelona
Institute for
Global Health

Contact me at
lea.maitre@isglobal.org

Extra slides

The early life period

- Vulnerable period of rapid organ development
- Many chronic diseases have part of their origin in early life (Barker hypothesis)
- Starting point for a life-course exposome

The Exposome data challenge

- Event created in the framework of the ISGlobal Exposome hub and H2020 ATHLETE project
- Organized by ISGlobal, Barcelona

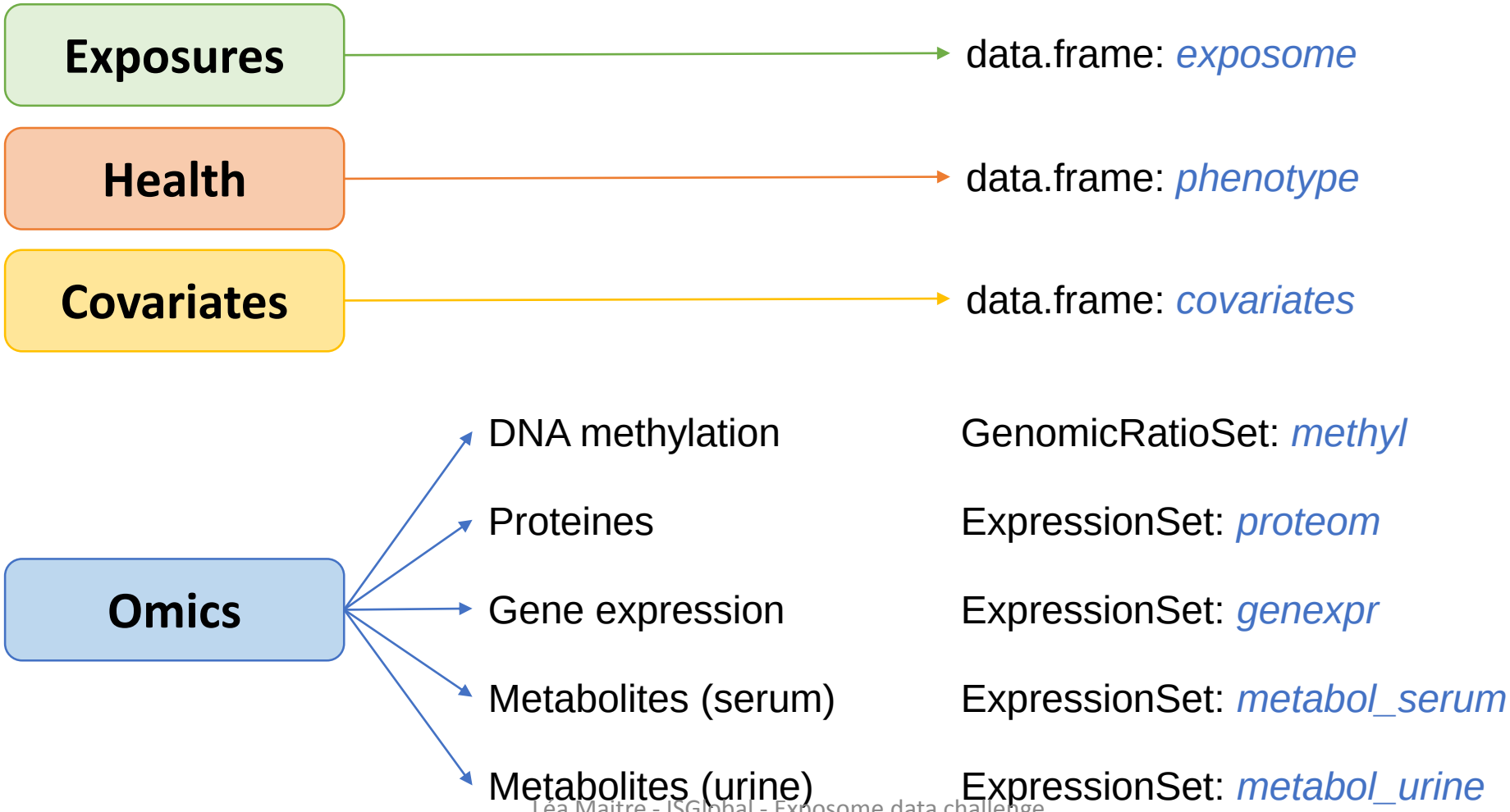
- Simulated data (based on the HELIX project) publicly available to challenge researchers on statistical tools to study exposome-health associations

ATHLETE
The Exposome from Evidence to Translation

ISGlobal
Barcelona
Institute for
Global Health

Organization of the datasets

Data available here: *https://github.com/isglobal-exposomeHub/ExposomeDataChallenge2021/blob/main/README.md*



Codebook

https://github.com/isglobal-brge/brgedata/blob/master/data/ExposomeDataChallenge2021/codebook.xlsx

Exposures **Health** **Covariates**

	B	C	D	E	F	G	H	I	J	K	L
1	variable_name	domain	family	subfamily	period	location	period_postnatal	description	var_type	transformation	labels
2	h_abs_ratio_preg_Log	Outdoor exposures	Air Pollution	PMAbsorb	Pregnancy	Home	NA	abs value (extrapolated back in time u	numeric	Natural Logarithm	PMabs
3	h_no2_ratio_preg_Log	Outdoor exposures	Air Pollution	NO2	Pregnancy	Home	NA	no2 value (extrapolated back in time u	numeric	Natural Logarithm	NO2
4	h_pm10_ratio_preg_No	Outdoor exposures	Air Pollution	PM10	Pregnancy	Home	NA	pm10 value (extrapolated back in time	numeric	None	PM10
5	h_pm25_ratio_preg_No	Outdoor exposures	Air Pollution	PM2.5	Pregnancy	Home	NA	pm25 value (extrapolated back in time	numeric	None	PM2.5
6	hs_no2_dy_hs_h_Log	Outdoor exposures	Air Pollution	NO2	Postnatal	Home	Day before examin	no2 value (extrapolated back in time u	numeric	Natural Logarithm	NO2(day)
7	hs_no2_wk_hs_h_Log	Outdoor exposures	Air Pollution	NO2	Postnatal	Home	Week before exam	no2 value (extrapolated back in time u	numeric	Natural Logarithm	NO2(week)
8	hs_no2_yr_hs_h_Log	Outdoor exposures	Air Pollution	NO2	Postnatal	Home	Year before exami	no2 value (extrapolated back in time u	numeric	Natural Logarithm	NO2(year)
9	hs_pm10_dy_hs_h_Non	Outdoor exposures	Air Pollution	PM10	Postnatal	Home	Day before examin	pm10 value (extrapolated back in time	numeric	None	PM10(day)
10	hs_pm10_wk_hs_h_Non	Outdoor exposures	Air Pollution	PM10	Postnatal	Home	Week before exam	pm10 value (extrapolated back in time	numeric	None	PM10(week)
11	hs_pm10_yr_hs_h_Non	Outdoor exposures	Air Pollution	PM10	Postnatal	Home	Year before exami	pm10 value (extrapolated back in time	numeric	None	PM10(year)
12	hs_pm25_dy_hs_h_Non	Outdoor exposures	Air Pollution	PM2.5	Postnatal	Home	Day before examin	pm25 value (extrapolated back in time	numeric	None	PM2.5(day)
13	hs_pm25_wk_hs_h_Non	Outdoor exposures	Air Pollution	PM2.5	Postnatal	Home	Week before exam	pm25 value (extrapolated back in time	numeric	None	PM2.5(week)
14	hs_pm25_yr_hs_h_Non	Outdoor exposures	Air Pollution	PM2.5	Postnatal	Home	Year before exami	pm25 value (extrapolated back in time	numeric	None	PM2.5(year)
15	hs_pm25abs_dy_hs_h_L	Outdoor exposures	Air Pollution	PMAbsorb	Postnatal	Home	Day before examin	pm25 absorbance value (extrapolated	numeric	Natural Logarithm	PMabs(day)
16	hs_pm25abs_wk_hs_h_	Outdoor exposures	Air Pollution	PMAbsorb	Postnatal	Home	Week before exam	pm25 absorbance value (extrapolated	numeric	Natural Logarithm	PMabs(week)
17	hs_pm25abs_yr_hs_h_L	Outdoor exposures	Air Pollution	PMAbsorb	Postnatal	Home	Year before exami	pm25 absorbance value (extrapolated	numeric	Natural Logarithm	PMabs(year)
18	h_accesslines300_preg_	Outdoor exposures	Built environr	Access	Pregnancy	Home	NA	Meters of public transport mode lines	numeric	Dichotomous	Access_ lines
19	h_accesspoints300_preg	Outdoor exposures	Built environr	Access	Pregnancy	Home	NA	Number of bus public transport mode	numeric	Natural Logarithm	Access_stops
20	h_builtdens300_preg_Sc	Outdoor exposures	Built environr	Building d	Pregnancy	Home	NA	Building density (m2 built/km2) within	numeric	Square root	Building
21	h_connind300_preg_Sqr	Outdoor exposures	Built environr	Connectiv	Pregnancy	Home	NA	Connectivity density (number of inters	numeric	Square root	Connectivity
22	h_fdensity300_preg_Log	Outdoor exposures	Built environr	Facility	Pregnancy	Home	NA	Number of facilities present divided by	numeric	Natural Logarithm	Facility_dens