Multi-Omics Data in Environmental Health

David Conti, PhD
Norris Comprehensive Cancer Center
Keck School of Medicine
University of Southern California
How can we advance mechanistic insight linking the human exposome to health across the life-course?

- Phthalates and Phenols
- Persistent pollutants (PFAS)
- Trace metals
- Air pollution

**Biological Mechanisms**

Can we identify the risk factors?

Can we identify the important biological pathways or patterns?
Measuring multiple exposures and omics layers

- Phthalates and Phenols
- Trace metals
- Persistent pollutants (PFAS)
- Air pollution

- Epigenome
- miRNA
- Transcriptome
- Proteome
- Metabolome
Application

Maternal biomarker-based EDC exposures (DDE, PCB-153, PFOA, PFOS, BPA, PBDE, Phthalates, heavy metals)

Pregnancy trimester-specific air pollution (PM, NO₂, traffic load)

Omics Data
- Urinary metabolomics
- Serum metabolomics

Proteomics
- Methylation
- Transcriptomics

MicroRNA (miRNA)
- Mitochondrial DNA content

CHILD LIVER INJURY
- Liver Enzymes
- CK-18
- Cytokines and Adipokines

n= 1200 mother-child pairs

Baseline

Birth/Infancy

Years
Liver Injury Risk in the HELIX cohort

In Utero Exposure

PFAS Exposure

Epigenome
- Methylation at 3,586 CpG sites

microRNA
- 1,117 mRNA

Transcriptome
- 58,254 mRNA

Proteome
- 36 proteins

Metabolome
- 177 Metabolites

Childhood non-alcoholic fatty liver disease

Liver Injury
Analysis frameworks for multiple exposures, multiple omics layers, and an outcome.
Multi-omics Integration

Legend:
E: Environmental exposure
Y: Health outcome
X: Latent Factors

High Dimensional
Molecular level traits

Mediation with Latent Factors
Step 1: Estimate X
Step 2: Mediation

Integrated, quasi-mediation
Estimate X using joint environmental and ‘omics data

Early
Concatenate ‘omics into single matrix

Intermediate
Combine ‘omics through inference on joint model

Late
Individually model each ‘omics layer

Analytical Framework
Liver Injury Risk in the HELIX cohort
Future Directions

Analytic Considerations

- **Omic features:**
  - High dimensional features within each omic layer.
    - Currently use machine learning for feature selection.
    - Omic features often highly correlated.
  - Balance estimation and inference within and across omic layers.

- **Need to adjust for study design covariates.**

- **Temporal or biological relation to data:**
  - Exposures -> Omics -> Outcome

- **Potential to incorporate external biological info:**
  - From experiments, ontologies, etc.

**Overall goals:**
- Identify causal features.
- Identify relevant biological patterns.
- Predict outcomes.