Application of high-resolution metabolomics, data science, and integrative omics to evaluate health effects of environmental exposures

Karan Uppal, PhD

Assistant Professor of Medicine

Computational Systems Medicine & Metabolomics Lab (CSM2L)

Division of Pulmonary, Allergy, Critical Care and Sleep Medicine

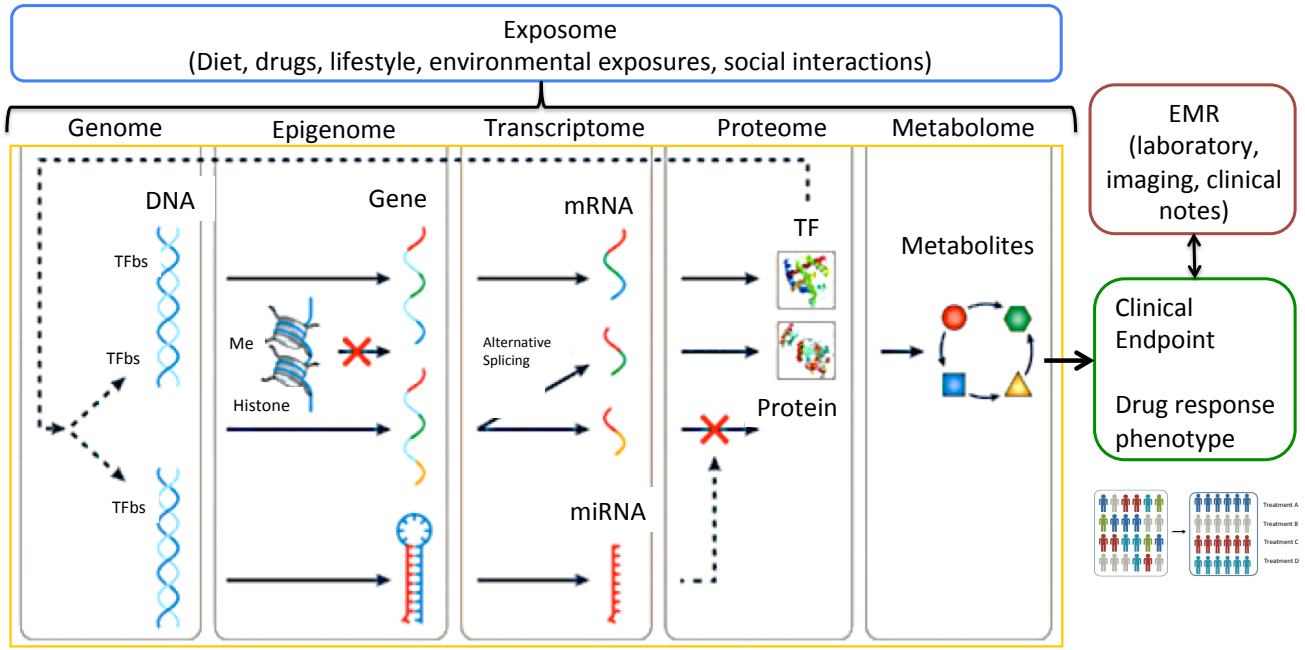
Emory University, Atlanta

Overview

- High-resolution metabolomics for exposure monitoring
- Bioinformatics tools for computational metabolomics (xMSanalyzer, xMSannotator), biomarker discovery (MetaboAnalyst), pathway analysis (Mummichog), and integrative omics (xMWAS)
- Case studies: metabolome-wide association studies of benzo(a)pyrene exposure, trichloroethylene exposure, and integrative analysis of health outcomes, biomolecular data, and exposure data

High-resolution metabolomics (HRM)

- HRM uses liquid or gas chromatography with high-resolution mass spectrometry and advanced data extraction algorithms to measure a broad spectrum of low and high abundance endogenous and exogenous chemicals in biologic sample
- Metabolomics provides a functional readout of cellular activity and physiological status
- Closest relationship to phenotype



Ritchie MD et al. 2015. Nat Rev Genet *16:85-97.*
Beger RD et al. 2016. Metabolomics 12 (10): 149.

High-Resolution Metabolomics for Exposome Research

Analysis provides capability to link exposures to internal dose to molecular responses to biomarkers of risk and health outcomes

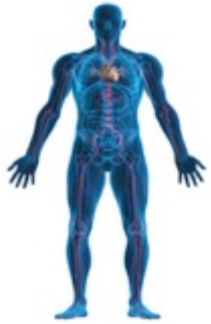
Exposome

Exposures

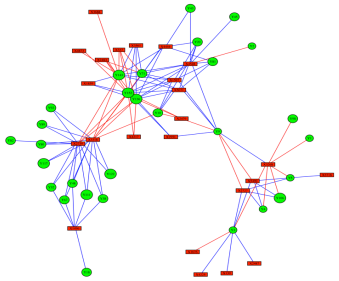
Food
Drugs
Exercise
Stress
Microbiome
Smoking
Alcohol
Supplements
Infectious agents
Environmental agents
Personal use products

High-Resolution Metabolomics

Internal Dose



Metabolic Response



Medicine

Biomarkers of Risk

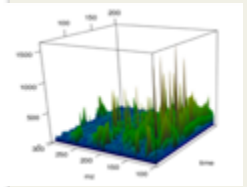
Health Outcomes

Walker et al 2016 in **Metabolic Phenotyping in Personalized and Public Healthcare**

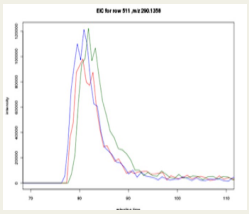
Clinical Biomarkers Laboratory
(Director: Dean P. Jones, Emory University)

Biofluids (e.g. plasma, serum, urine): 50 microliters
Tissues (e.g. liver, muscle, lung): 20mg

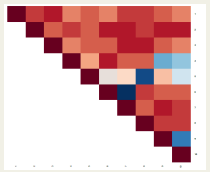
Ultra-high resolution MS
LTQ-FT
LTQ-Velos Orbitrap
Q-Exactive HF
Thermo Fusion
GC-Orbitrap



Raw data processing with built-in feature and sample quality assessment (apLCMS with **xMSanalyzer**)



Data Exploratory Analysis (Box plots, histograms, etc.)



Batch-effect evaluation and correction (Using ComBat); void volume filtering

Annotation of metabolites (**xMSannotator**)

1. Untargeted feature table
2. Targeted feature table
3. Annotated feature table

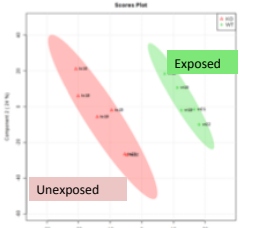
Confirmation of identity

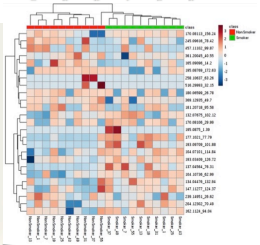
- MS/MS, authentic chemical standards
- **In-house library:** >1000 metabolites and environmental chemicals

Pathway analysis (Mummichog,MetaboAnalyst, MetaCore, MSEA)

Biomarker and network analysis (**xmsPANDA, MetabNet**, MetaboAnalyst)

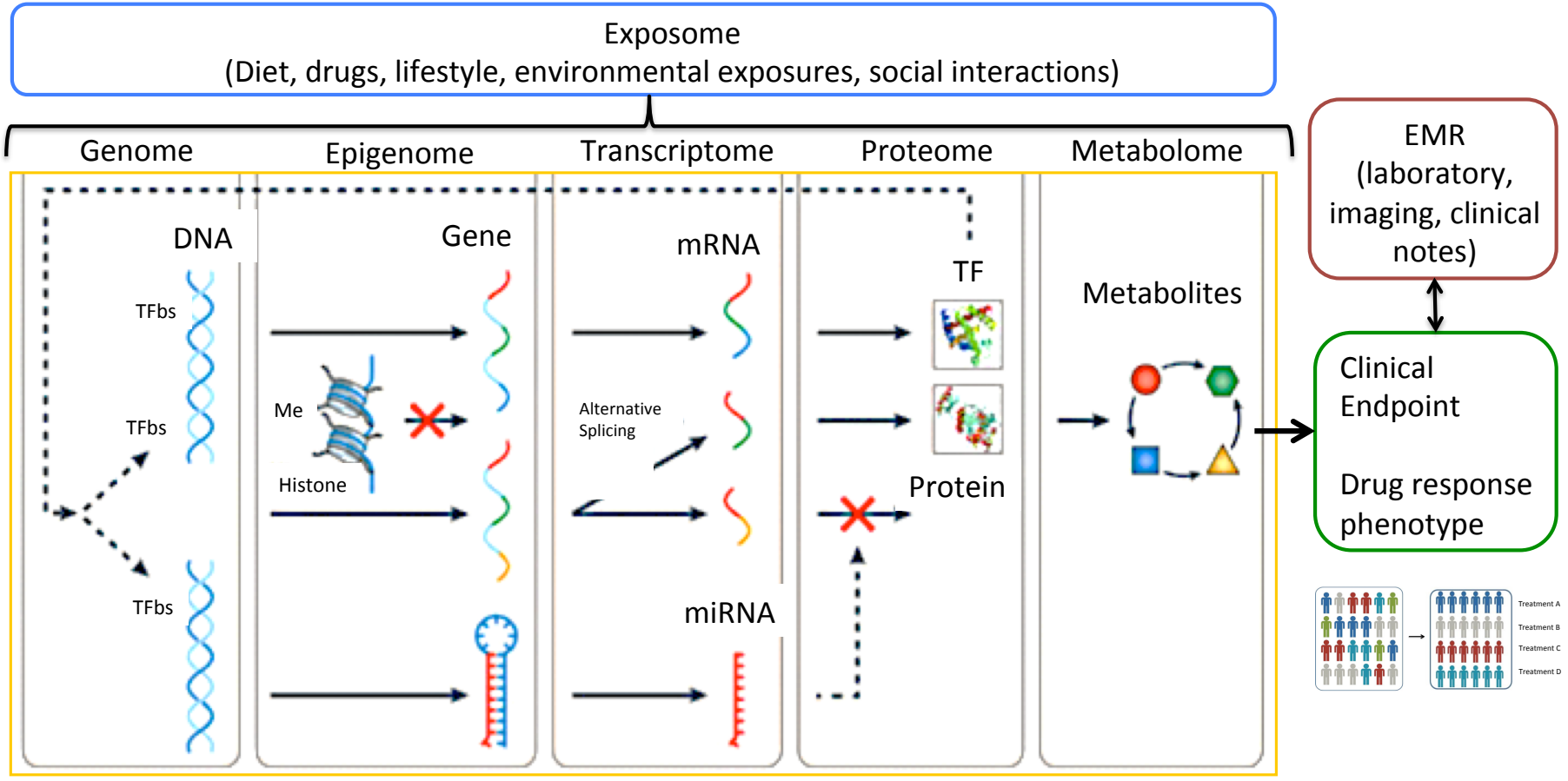
- **Univariate**: Limma t-test, ANOVA
- **Multivariate** and predictive analysis:
  Support vector machine, Random forest, PLSDA
- **Clustering**: Two-way Hierarchical clustering analysis
- Targeted and untargeted MWAS





Computational Systems Medicine:

Computational methods for multi-scale aggregation of integrating clinical, biomolecular, and environmental data for improved diagnosis, prognosis, and targeted therapies



Main approaches for data integration

- Pathway or knowledge-based integration
  - Datasets are analyzed individually (differentially expressed genes, metabolites, proteins) and integration is performed at the pathway level
  - Examples: MetaboAnalyst, iPEAP, MetScape
- Using literature-derived associations for integration
  - Using co-occurrence criteria for establishing relationship
  - Examples: CoPub, ArrowSmith, PolySearch2.0
- Data-driven integration using meta-dimensional analysis
  - Integration is performed globally such that data from multiple omics layers are combined simultaneously
  - Examples: 3Omics, mixOmics, **xMWAS**

xMWAS: a data-driven integration and differential network analysis (Uppal 2018, Bioinformatics)

URL: https://kuppal.shinyapps.io/xmwas/

R package: https://github.com/kuppal2/xMWAS

Transcriptomics

Gene	Sample 1	Sample 2	Sample 3
akap9	-0.24	-0.6	-0.47
macf1	-0.3	-0.3	0.48
RNPEP	0.24	0.85	0.15
SDHA	0.1	0.37	0.18

Metabolomics

Metabolite	Sample 1	Sample 2	Sample 3
4277439	-0.3109937	-0.2792995	-0.2548517
441	-0.2967872	-0.2895908	-0.2674823
69362	-0.3183692	-0.2828533	-0.272917
10258	-0.0614116	-0.1180467	-0.1231662

Proteomics

Protein	Sample 1	Sample 2	Sample 3
P14060	2.06	1.2	1.61
P26439	1.8	3.57	2.04
P29372	-0.64	-0.71	-0.21
Q96J02	-0.52	-1.34	-0.15

↓

Pairwise (sparse) Partial Least Squares regression for data integration (Cao 2009)

↓

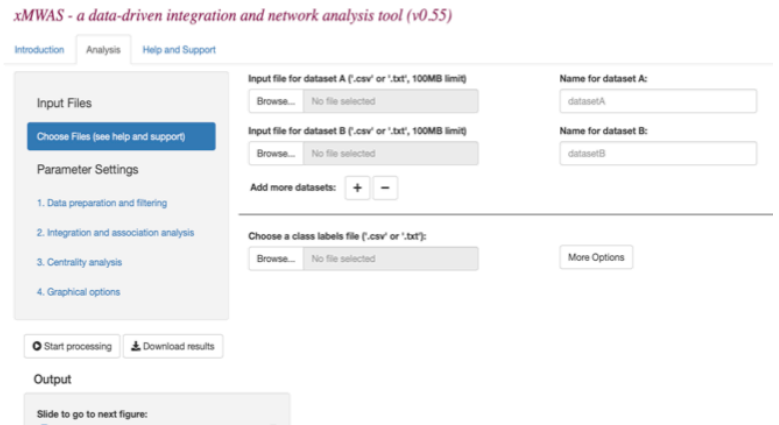
Approximation of Pearson correlation using PLS components

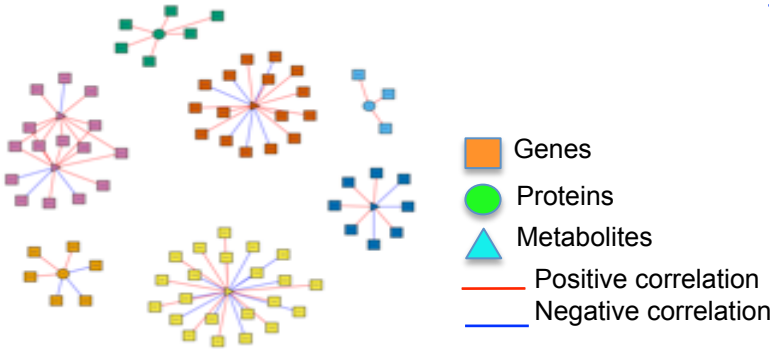
↓

Filtering based on |r|>threshold and p-value<alpha criteria

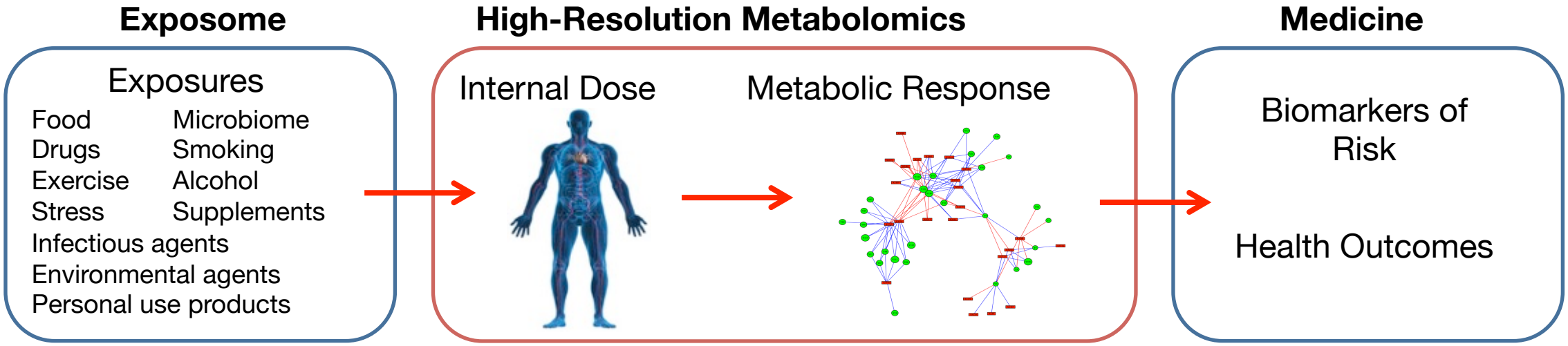
↓

Community (clusters) detection and centrality (importance) analysis





Case Studies



Case study 1: Pilot Metabolome-Wide Association Study of Benzo(a)pyrene in Serum from Military Personnel (n=30)

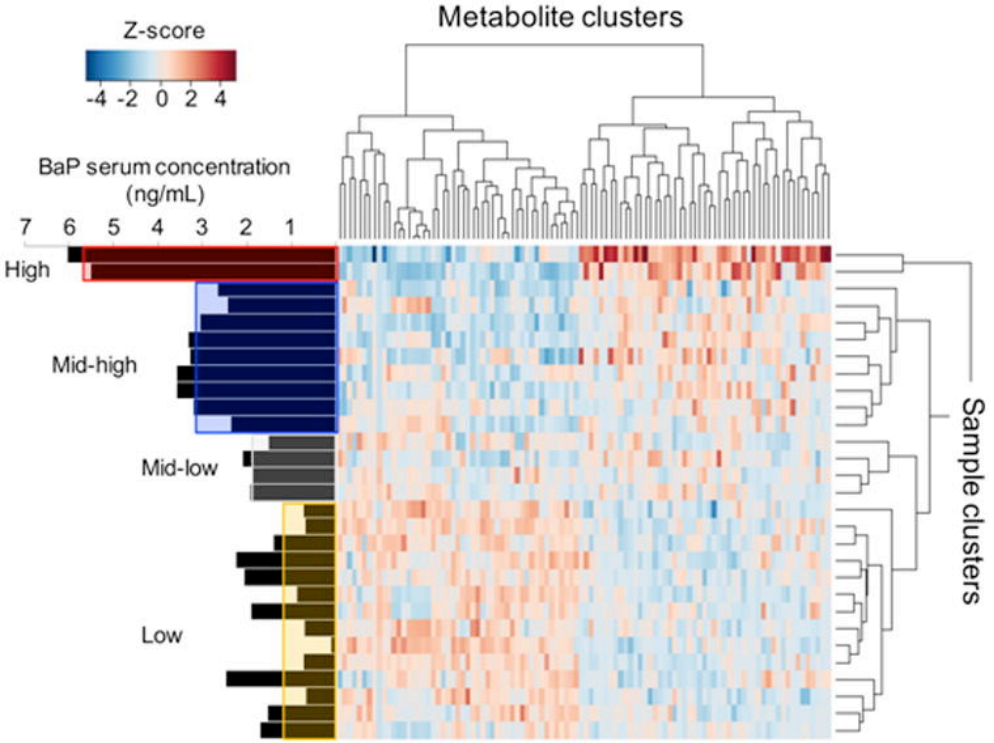
Walker DI, Pennell KD, Uppal K, Xia X, Hopke PK, Utell MJ, Phipps RP, Sime PJ, Rohrbeck P, Mallon CT, Jones DP. Pilot Metabolome-Wide Association Study of Benzo(a)pyrene in Serum From Military Personnel. J Occup Environ Med. 2016 Aug;58(8 Suppl 1):S44-52

Untargeted metabolomics: LC-MS
Benzo(a)pyrene measurement: GC-MS

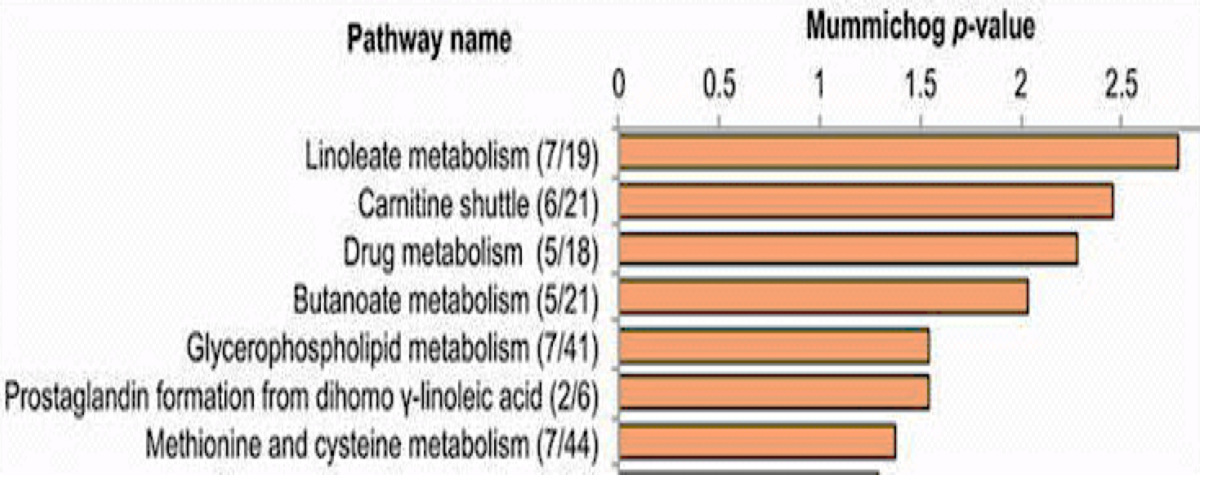
Study overview

- Thirty unidentified human serum samples were obtained from the Department of Defense Serum Repository (DoDSR)
  - Samples originally collected for mandatory HIV testing in armed forces personnel
  - Following collection, samples were stored at -30°C
- Samples analyzed using high-resolution mass spectrometry for metabolomics
- Serum levels of free benzo(a)pyrene (BaP), a carcinogenic polycyclic aromatic hydrocarbon (PAH) arising from the combustion of organic material, were measured using gas chromatography-mass spectrometry
- Metabolic associations with BaP were determined using a metabolome-wide association study (MWAS) and metabolic pathway enrichment

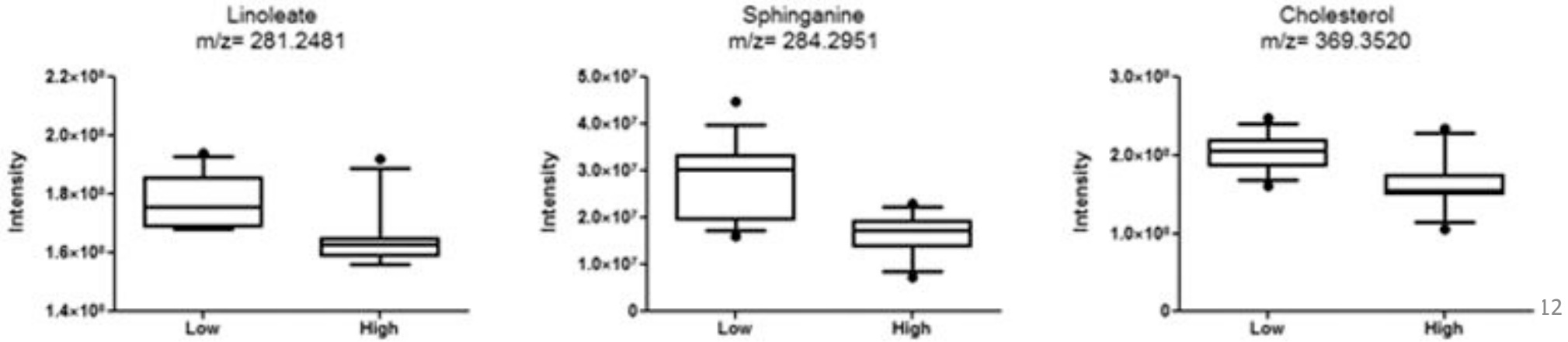
A. Top 100 metabolic features (sparse PLS) associated with Benzopyrene levels



B. Top enriched pathways using Mummichog



C. Boxplots for select metabolites associated with glycerophospholipid metabolism pathway



Case study 2: High-resolution metabolomics of occupational exposure to trichloroethylene

	Number of subjects
Controls (matched)	95
Exposed (0.4 to 230 ppm)	80

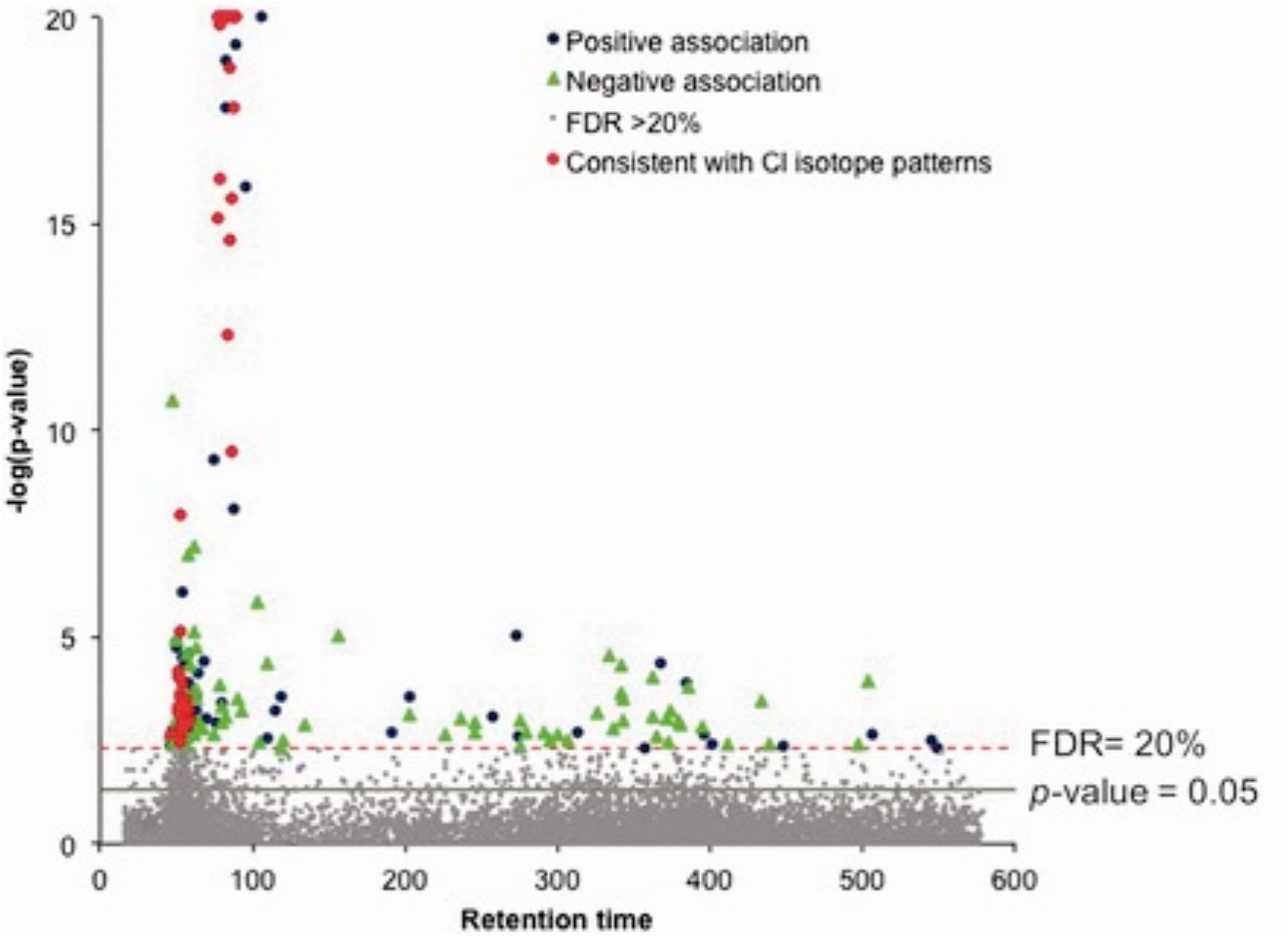
Walker DI, Uppal K, Zhang L, Vermeulen R, Smith M, Hu W, Purdue MP, Tang X, Reiss B, Kim S, Li L, Huang H, Pennell KD, Jones DP, Rothman N, Lan Q. High-resolution metabolomics of occupational exposure to trichloroethylene. Int J Epidemiol. 2016 Oct;45(5):1517-1527.

Untargeted metabolomics: LC-MS
Trichloroethylene measurement: 3M vapor-monitoring badges

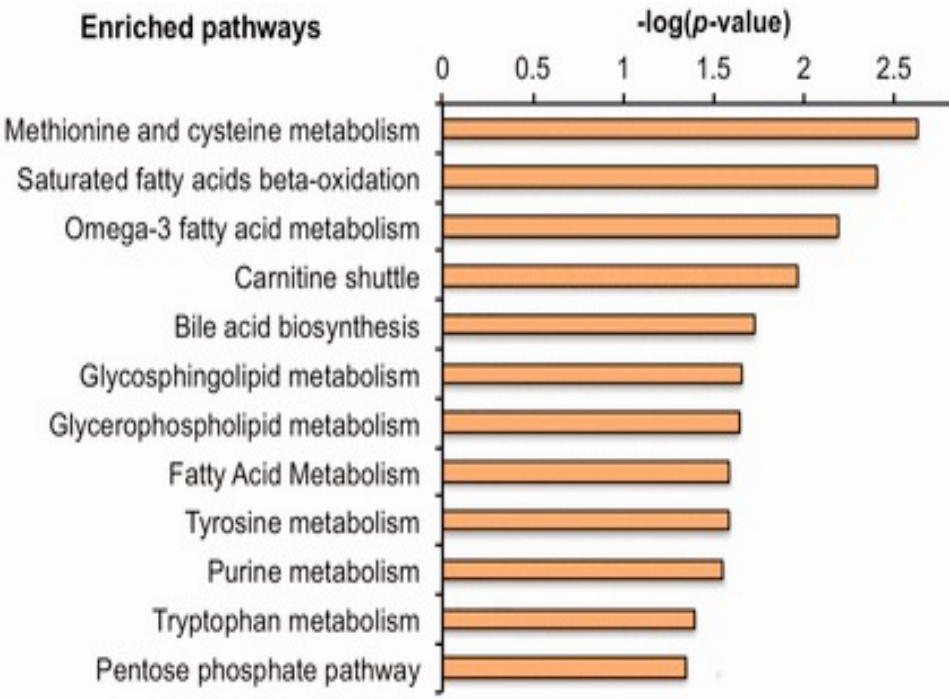
Study Overview

- Trichloroethylene measurement: 3M vapor-monitoring badges
- Non-targeted metabolomics analysis of plasma obtained from 80 TCE-exposed workers [full shift exposure range of 0.4 to 230 parts-per-million of air (ppm_a)] and 95 matched controls were completed by ultra-high resolution mass spectrometry
- Biological response to TCE exposure was determined using a metabolome-wide association study (MWAS) framework, with metabolic changes and plasma TCE metabolites evaluated by dose-response and pathway enrichment
- Biological perturbations were then linked to immunological, renal and exposure molecular markers measured in the same population

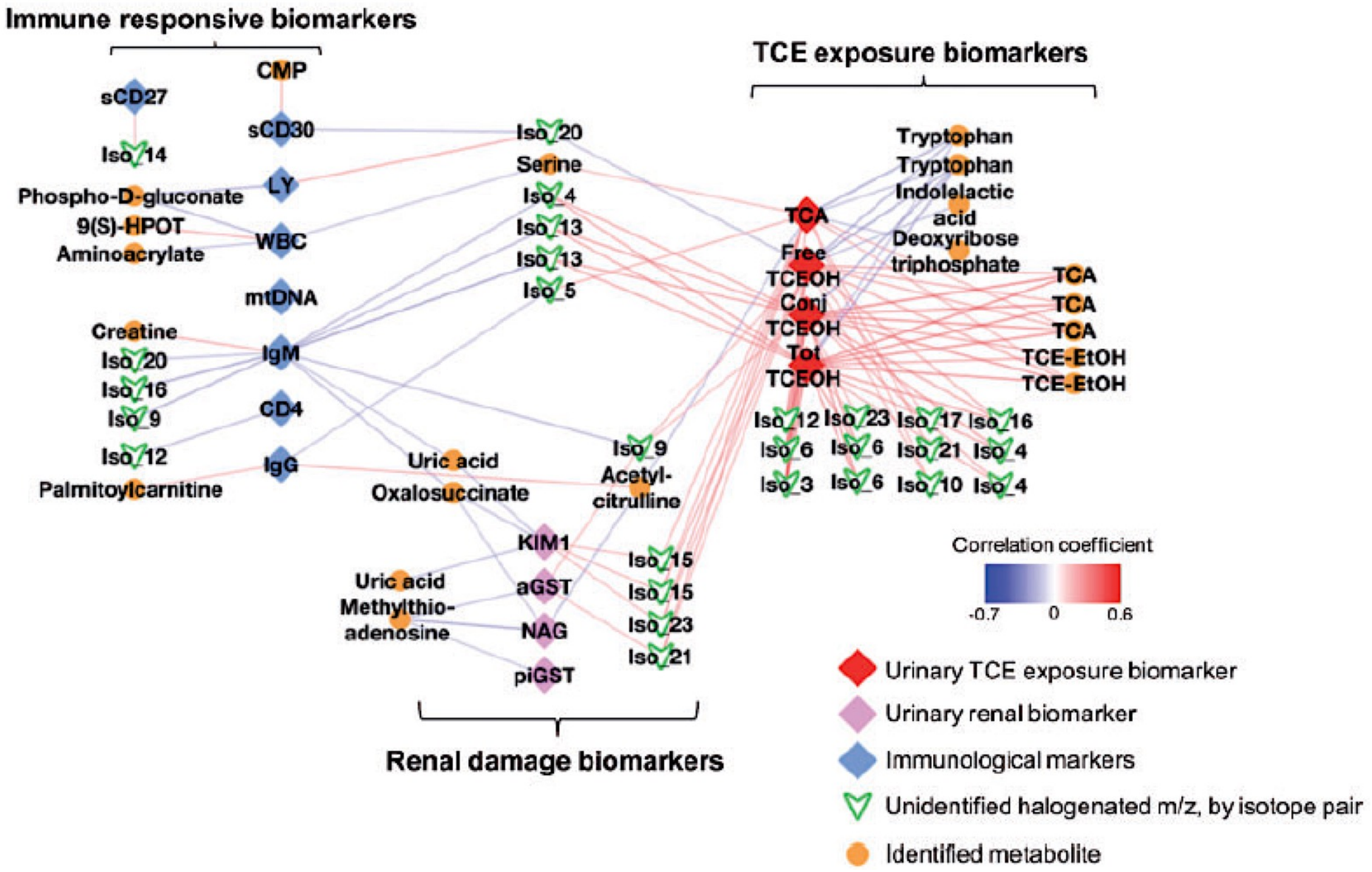
A. 188 metabolic features associated with TCE exposure using linear regression (model adjusted for age, sex, and BMI)



B. Pathway analysis using Mummichog



C. Correlation-based network analysis of 188 metabolic features associated with TCE exposure and molecular markers previously tested for association with TCE exposure



Case study 3: Integrative network analysis of clinical, biomolecular, and environmental exposure data from a dataset of 66 service personnel post-deployment

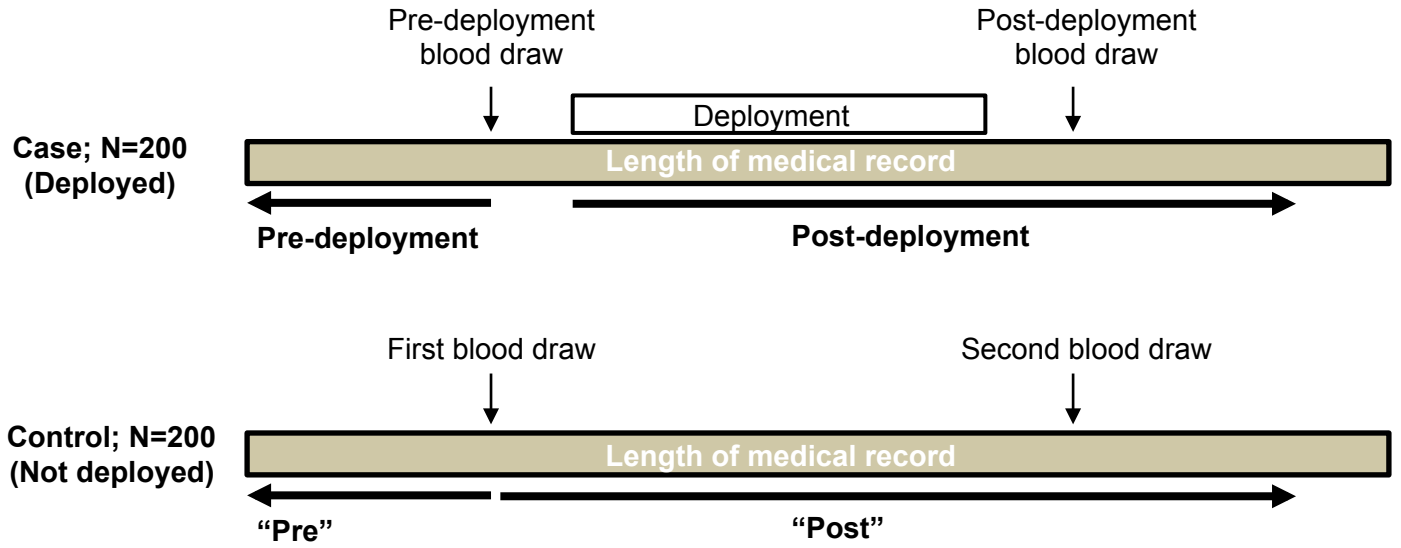
	Number of subjects
Cardiopulmonary symptoms post-deployment (Yes)	41
Cardiopulmonary symptoms post-deployment (No)	25

Thakar J*, Thatcher TH*, Smith MR*, Woeller CF, Walker DI, Utell MJ, Hopke PK, Mallon TM, Krah PL, Rohrbeck P, Go YM, Jones DP, **Uppal K***, Integrative network analysis linking clinical outcomes with environmental exposures and molecular variations in service personnel deployed to Balad and Bagram. JOEM (in press);

Collaboration between Emory, U Rochester, USUHS, Mt Sinai Med Sch, and Armed Services Health Surveillance System.

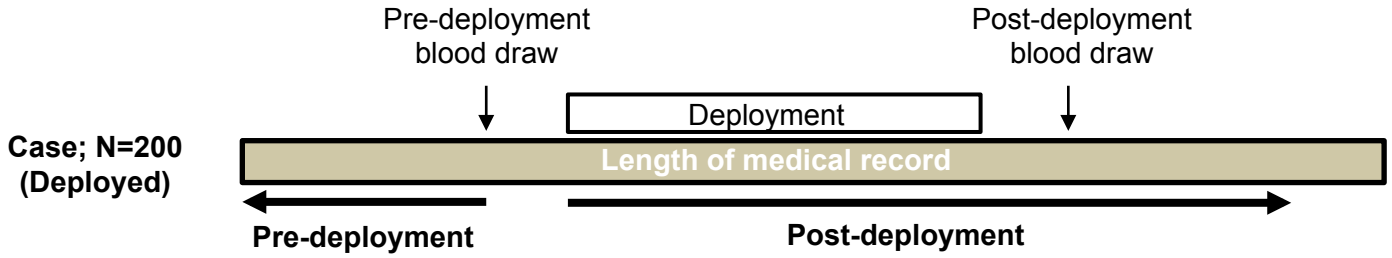
Study Overview

- Serum samples from 400 military personnel from the Department of Defense Serum Repository (DoDSR)
  - 200 deployed to Balad, Iraq and Bagram, Afghanistan (Case)
  - 200 non-deployed service personnel stationed domestically (Control)
  - Samples collected at two timepoints (Pre and Post)
- Samples originally collected for the mandatory HIV testing



Study Overview

- Serum samples from 400 military personnel from the Department of Defense Serum Repository (DoDSR)
  - 200 deployed to Balad, Iraq and Bagram, Afghanistan (Case)
  - 200 non-deployed service personnel stationed domestically (Control)
  - Samples collected at two timepoints (Pre and Post)
- Samples originally collected for the mandatory HIV testing



Case-onset: 41 (ICD-9 codes for cardiopulmonary symptoms during or after deployment)

Case-never: 25 (no coding for cardiopulmonary symptoms before or after deployment

Input data for integrative analysis using xMWAS

1. **Molecular data: metabolites, miRNAs, cytokines, and proteins**
(3,274 molecular variables x 66 subjects)

	Subject1	Subject2	-	Subject N
Metabolite 1	199	19	-	100
-	-	-		-
miRNA 1	50	30	-	20
-	-	-		-
Cytokine 1	33	12	-	39

ICD-9; (49 any cardiopulmonary ICD-9 codes x 66 subjects)

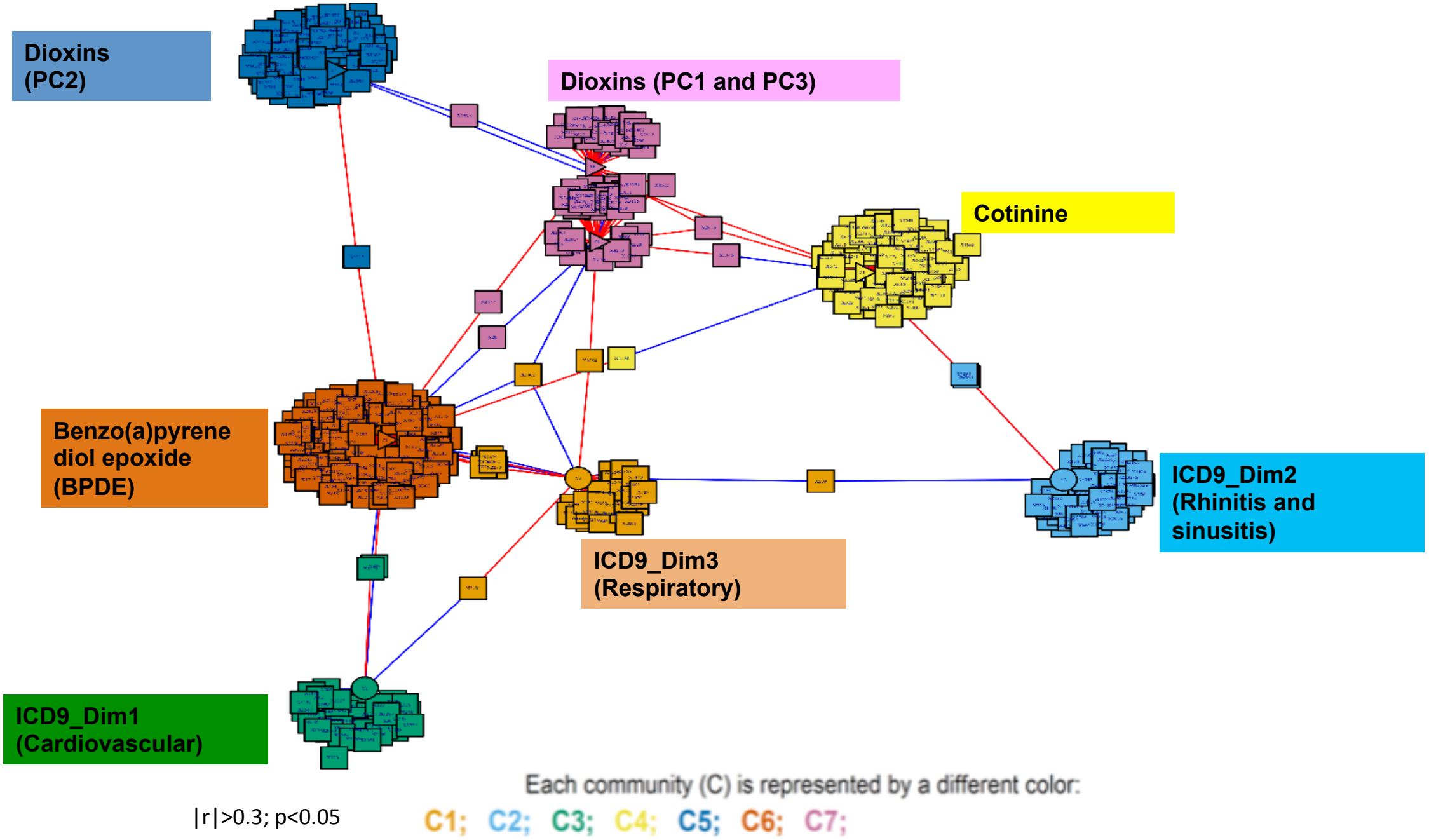
	Subject1	Subject2	-	Subject N
4019	0	1	-	0
4011	1	1	-	0
-	-	-	-	-
49301	1	0	-	0

2. **Environmental chemicals: first 3 PCs of dioxins, cotinine, benzo(a)pyrene diol epoxide (BPDE)**
(5 variables x 66 subjects)

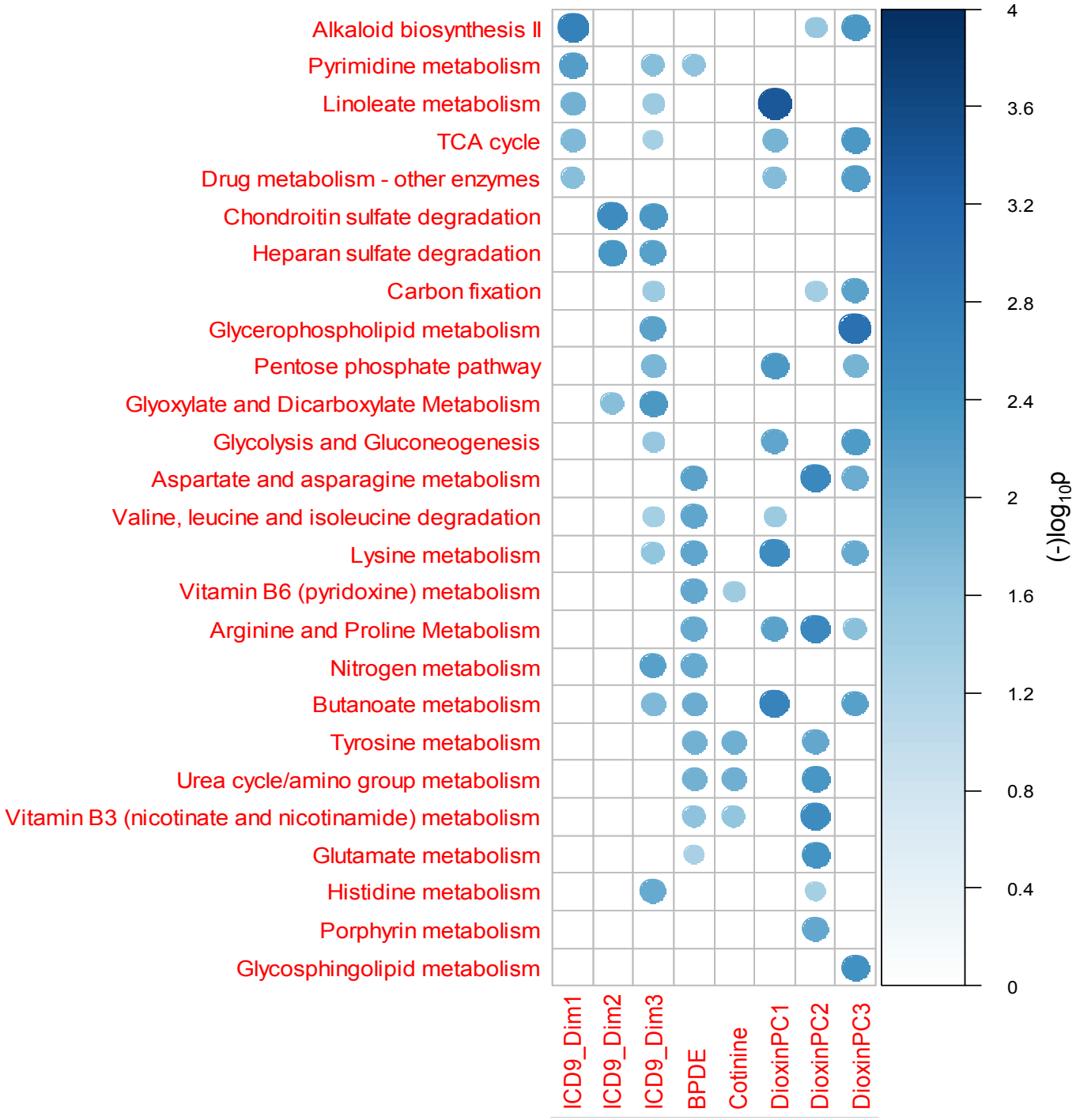
	Subject1	Subject2	-	Subject N
chemical 1	3	2.5	-	13
chemical 2	1	4	-	9
-	-	-		-
chemical s	5	3	-	2

3. Multiple correspondence analysis
(8 dimensions x 66 subjects)
(Only dimensions with >5% variance explained were included)

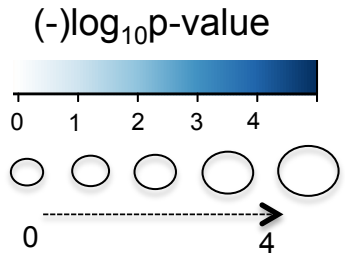
	Subject1	Subject2	-	Subject N
Dim 1	199	19	-	100
Dim 2	10	40		90
-	-	-		-
Dim r	50	30	-	20


Dioxins (PC2)
Dioxins (PC1 and PC3)
Cotinine
Benzo(a)pyrene diol epoxide (BPDE)
ICD9_Dim2 (Rhinitis and sinusitis)
ICD9_Dim3 (Respiratory)
ICD9_Dim1 (Cardiovascular)
Each community (C) is represented by a different color:
C1; C2; C3; C4; C5; C6; C7;

|r|>0.3; p<0.05



- Bubble plot showing metabolic pathways associated with clinical and environmental exposures data
- Metabolic pathway analysis performed using Mummichog



More examples

Environ Int. 2018 Nov;120:145-154. doi: 10.1016/j.envint.2018.07.044. Epub 2018 Aug 7.

Use of high-resolution metabolomics for the identification of metabolic signals associated with traffic-related air pollution.

Liang D[1], Moutinho JL[2], Golan R[3], Yu T[4], Ladva CN[5], Niedzwiecki M[6], Walker DI[7], Sarnat SE[7], Chang HH[4], Greenwald R[8], Jones DP[9], Russell AG[2], Sarnat JA[7].

Author information

Abstract

BACKGROUND: High-resolution metabolomics (HRM) is emerging as a sensitive tool for measuring environmental exposures and biological responses. The aim of this analysis is to assess traffic-related air pollution mixtures.

METHODS: We used untargeted HRM profiling t Vehicle Emission (DRIVE) study to identify meta related pollutants at multiple ambient and indoor students living in dormitories near (20 m) or far (was completed for both plasma and saliva samp association study (MWAS) framework with pathv

RESULTS: Weekly levels of traffic pollutants wer pollutants). In total, 20,766 metabolic features w detected and shared in both plasma and saliva s more traffic indicator, including black carbon, car features), after controlling for confounding and fa indicated elicitation of inflammatory and oxidativ chemical identities of 10 metabolites associated

CONCLUSIONS: Using HRM, we identified and with repeated measurement. Observed respons inflammation, and nucleic acid damage and repa development of metabolic biomarkers of traffic p

Environ Int. 2019 Jun;127:503-513. doi: 10.1016/j.envint.2019.04.003. Epub 2019 Apr 10.

Perturbations of the arginine metabolome following exposures to traffic-related air pollution in a panel of commuters with and without asthma.

Liang D[1], Ladva CN[2], Golan R[3], Yu T[4], Walker DI[5], Sarnat SE[6], Greenwald R[7], Uppal K[8], Tran V[8], Jones DP[8], Russell AG[9], Sarnat JA[6].

Author information

Abstract

BACKGROUND: Mechanisms underlying the effects of traffic-related air pollution on people with asthma remain largely unknown, despite the abundance of observational and controlled studies reporting associations between traffic sources and asthma exacerbation and hospitalizations.

OBJECTIVES: To identify molecular pathways perturbed following traffic pollution exposures, we analyzed data as part of the Atlanta Commuters Exposure (ACE-2) study, a crossover panel of commuters with and without asthma.

METHODS: We measured 27 air pollutants and conducted high-resolution metabolomics profiling on blood samples from 45 commuters before and after each exposure session. We evaluated metabolite and metabolic pathway perturbations using an untargeted metabolome-wide association study framework with pathway analyses and chemical annotation.

RESULTS: Most of the measured pollutants were elevated in highway commutes ($p < 0.05$). From both negative and positive ionization modes, 17,586 and 9087 metabolic features were extracted from plasma, respectively. 494 and 220 unique features were associated with at least 3 of the 27 exposures, respectively ($p < 0.05$), after controlling confounders and false discovery rates. Pathway analysis indicated alteration of several inflammatory and oxidative stress related metabolic pathways, including leukotriene, vitamin E, cytochrome P450, and tryptophan metabolism. We identified and annotated 45 unique metabolites enriched in these pathways, including arginine, histidine, and methionine. Most of these metabolites were not only associated with multiple pollutants, but also differentially expressed between participants with and without asthma. The analysis indicated that these metabolites collectively participated in an interrelated molecular network centering on arginine metabolism, underlying the impact of traffic-related pollutants on individuals with asthma.

CONCLUSIONS: We detected numerous significant metabolic perturbations associated with in-vehicle exposures during commuting and validated metabolites that were closely linked to several inflammatory and redox pathways, elucidating the potential molecular mechanisms of traffic-related air pollution toxicity. These results support future studies of metabolic markers of traffic exposures and the corresponding molecular mechanisms.

JOEM 2016

JOEM 2019 (in press)

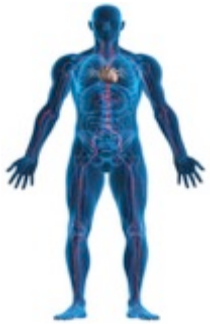
Exposome

Exposures

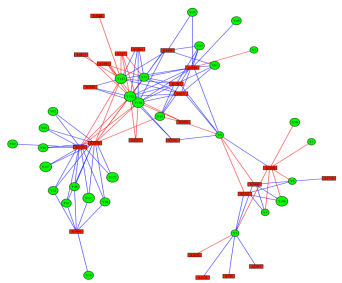
Food
Drugs
Exercise
Stress
Infectious agents
Environmental agents
Personal use products
Microbiome
Smoking
Alcohol
Supplements

High-Resolution Metabolomics

Internal Dose



Metabolic Response

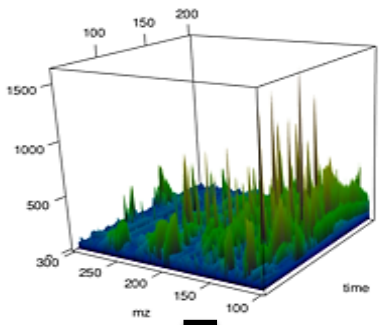


Medicine

Biomarkers of Risk

Health Outcomes

Raw data

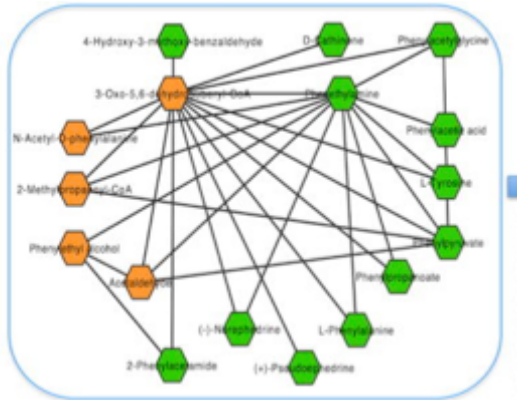


Data extraction (xMSanalyzer; Uppal 2013)

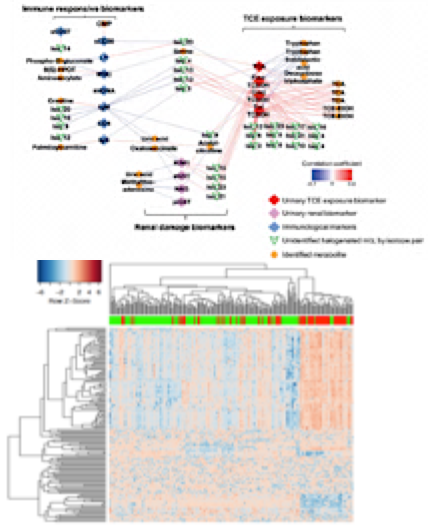
m/z	time	Smoker_11.raw Peak area	Smoker_15.raw Peak area	Smoker_13.raw Peak area
193.0970902	1.697928509	21590.09577	1465875.407	2921520.329
104.071007	1.914036922	1.68E+08	1.20E+08	1.18E+08
104.071007	1.914036922	1.68E+08	1.20E+08	1.18E+08
137.0456421	1.271814331	217380.9151	66352.25511	96353.93902
241.0307929	2.180590728	6.27E+07	8.42E+07	8.09E+07
134.044759	2.215654287	1.39E+07	2.77E+07	2.66E+07

Computational metabolomics, data mining, and integrative omics

Network-based annotation (xMSannotator; Uppal 2017)



Biomarker discovery and network analysis
xmsPANDA & optSelect
MetabNet (Uppal 2015)
xMWAS (Uppal 2018)



Acknowledgements

Clinical Biomarkers Lab; CSM^{2}L

Dean Jones, Young-Mi Go, Shuzhao Li, Karan Uppal, Chunyu Ma, Ken Liu, Kristine Dennis, ViLinh Tran, Michael Orr, Ryan Smith, Xin Hu, Jolyn Fernandes, Bill Liang, Yating Wang, Tiantian Zhang

Collaborators:
Mt. Sinai (Douglas Walker), Uniformed Services University of the Health Sciences (Timothy Mallon, Pamela Krahl), Armed Forces Health Surveillance Center (Ms Patricia Rohrbeck), University of Rochester (Mark Utell, Juilee Thakar, Collynn Woeller, Thomas Thatcher), National Cancer Institute (Nat Rothman)

Funding
Department of Defense award (306889-1.00-64239)
NIEHS, NIA, NCI, NHLBI, NIDDK, NIAAA, NIAID, Woodruff Foundation, Emory Dept of Medicine, Georgia Research Alliance Exposome Center Integrated Health Science and Facilities Core
NIEHS P30 ES019776
DK112341 (MoTrPAC)
AG057470
AA026928
EY022618
ES026071
DK117246-01

Questions/Comments?

Email: kuppal2@emory.edu

Website: www.csm2l.com