- **Exposomics**

Structural similarity guided design of a set of internal standards for exposure science enabling class based normalization in untargeted metabolomics studies.

Kruttika Dabke[1], Jennifer Kyle[2], Tao Liu[2], Jiajun Lei[1], Jerry S.H. Lee[1,3,4,5,6], Jonathan E. Katz[1,3,5,6].

[1]Ellison Medical Institute, Los Angeles, CA
[2]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA
[3]Murtha Cancer Center/Research Program, Uniformed Services University of the Health Sciences/Walter Reed National Military Medical Center, Bethesda, MD
[4]Department of Chemical Engineering and Material Sciences, Viterbi School of Engineering, University of Southern California, Los Angeles, CA
[5]Dornsife College of Letters, Arts and Sciences, USC
[6]Department of Medicine, Keck School of Medicine, University of Southern
California, Los Angeles, California
Correspondence: jkatz@emila.org
Reprints: https://reprints.katzlab.org/

For Poster Reprints and Supplemental Information:
https://reprints.katzlab.org/asms-2026-exposurestandards/

Introduction

Stable isotopically labeled internal standards (SIL-IS) are essential for assessing data quality and improving quantitative reliability in metabolomics and lipidomics experiments. Commercially available SIL-IS panels are typically designed for machine validation or for broad profiling or targeted pathways and may not optimally support exposure-focused workflows. To address this gap, we curated an Exposure Compound Library comprising environmentally relevant contaminants, carcinogens, and representative endogenous metabolites.

Structural redundancy was reduced using chemical similarity scoring, yielding a diverse set of candidate standards appropriate for electrospray ionization based metabolomics. This collection is intended to enhance robustness and enable class based normalization in exposure science applications. Here we report the initial use and validation of our panel against various cohorts of human samples.

Methods

To establish our set of internal standards, we used compounds sourced from publications and databases including known and potential carcinogens that deployed military personnel may have been exposed to. Common human metabolites representing a range of metabolic pathways were also included in the broader list (**Table 2**). With the help of CHEMPASS (CHEMical Profiling and Analysis for Similarity Scoring) and manual review, we selected structurally non-redundant compounds to end up with a subset of exemplars from various compound classes which resulted in 62 structurally unique internal standards. Of these, 38 were expected to be measurable in electrospray ionization (**Table 1**). Each standard underwent characterization via dilution series (**Figure 2.A**), retention-time assessment (**Figure 2.C**), and preliminary recovery experiments (**Figure 2.D**). For comparing structural similarity between IS and annotated features identified in a metabolomics study, annotated features from a Hilic ESI- run were obtained with MSDIAL V-4.9.221218 against MSMS-Public_experimentspectra-neg-VS19.msp library and minimum identification score cutoff of 70. The two list compare tab of CHEMPASS was used with FCFP4 fingerprints to generate the top 3 structurally matched IS to the annotated features (**Figure 3.B**). Pairwise Tanimoto similarity scores were computed (best IS match retained per feature and identity match excluded) and Fruchterman-Reingold layout weighted by similarity was used in constructing the bipartite igraph network (**Figure 3.C**).

Conflict of Interest

The authors declare no competing financial interests.

Results

The curated SIL-IS were purchased at physiological concentrations based on their observed concentrations from literature either in whole blood, serum, plasma or urine. Since this resulted in their concentrations spanning several orders of magnitude, we evaluated these standards using our typical workflow for compound visibility and utility in physiological samples. Internal standards were incorporated at two steps in the sample preparation process for untargeted metabolomics. A set of 22 compounds were added at the extraction step (quench step) while an additional 16 were added at the resuspension step (**Table 1**). Based on the dilution series, 33 of the 38 compounds reliably gave a strong signal in at least one of our metabolomics measurement configurations (ionization polarity and chromatography system) (**Figure 2.A**). Five compounds that were not detected were present at very low concentrations in the mixture, likely below the detection limit of the instrument as configured (**Figure 2.A**). Retention times were highly stable across replicates in dilution series (**Figure 2.C**).

Performance and utility of our panel of internal standards were evaluated in cohorts of human samples. Recoveries in plasma of 16 IS indicated class/channel-dependent but generally manageable matrix effects (**Figure 2.D**). Endogenous metabolites corresponding to several of the selected internal standards such as hypoxanthine, xanthine, taurine, as well as ibuprofen, acetamidophenol, were also observed in cohorts of human samples. For 9/16 of the compounds, IS signal differed from endogenous levels by roughly 10 orders of magnitude, while for 5/16 compounds the levels matched and for 2 compounds (acetamidophenol and caffeine) the IS had several magnitudes higher signal compared to endogenous (**Figure 3.A**). 166 annotated features from a Hilic ESI- run were matched to their structurally most similar IS using the two-list compare tab of CHEMPASS. BHB-d4 and lactic acid-d3 emerged as structural nearest neighbors for the largest number of features with retained Tanimoto scores ranging from > 0.3 and < 1 (**Figure 3.B-C**). These exploratory findings indicate that the IS panel covers a sufficiently wide structural neighborhood to enable class based normalization of annotated features. The practical benefit of this approach over other normalization strategies remains to be assessed in downstream analysis.

References:

1. Anderson RH, Thompson T, Stroo HF, Leeson A. US Department of Defense-Funded Fate and Transport Research on Per- and Polyfluoroalkyl Substances at Aqueous Film-Forming Foam-Impacted Sites. Environ Toxicol Chem. 2021 Jan;40(1):37-43. doi: 10.1002/etc.4694. Epub 2020 Jun 2. PMID: 32077141; PMCID: PMC7984261.
2. Phillips KA, Chao A, Church RL, Favela K, Garantziotis S, Isaacs KK, Meyer B, Rice A, Sayre R, Wetmore BA, Yau A, Wambaugh JF. Suspect Screening Analysis of Pooled Human Serum Samples Using GC × GC/TOF-MS. Environ Sci Technol. 2024 Jan 30;58(4):1802-1812. doi: 10.1021/acs.est.3c05092. Epub 2024 Jan 13. PMID: 38217501; PMCID: PMC11459241.
3. Smith MR, Uppal K, Walker DI, Utell MJ, Hopke PK, Mallon TM, Krahl PL, Rohrbeck P, Go YM, Jones DP. Environmental Chemicals Altered in Association With Deployment for High Risk Areas. J Occup Environ Med. 2019 Dec;61 Suppl 12(Suppl 12):S15-S24. doi: 10.1097/JOM.0000000000001647. PMID: 31800447; PMCID: PMC7299071.
4. Xia X, Carroll-Haddad A, Brown N, Utell MJ, Mallon CT, Hopke PK. Polycyclic Aromatic Hydrocarbons and Polychlorinated Dibenzo-p-Dioxins/Dibenzofurans in Microliter Samples of Human Serum as Exposure Indicators. J Occup Environ Med. 2016 Aug;58(8 Suppl 1):S72-9. doi: 10.1097/JOM.0000000000000743. PMID: 27501107; PMCID: PMC4977992.
5. Moschet C, Anumol T, Lew BM, Bennett DH, Young TM. Household Dust as a Repository of Chemical Accumulation: New Insights from a Comprehensive High-Resolution Mass Spectrometric Study. Environ Sci Technol. 2018 Mar 6;52(5):2878-2887. doi: 10.1021/acs.est.7b05767. Epub 2018 Feb 13. PMID: 29437387; PMCID: PMC7239036.
6. Liu KH, Lee CM, Singer G, Bais P, Castellanos F, Woodworth MH, Ziegler TR, Kraft CS, Miller GW, Li S, Go YM, Morgan ET, Jones DP. Large scale enzyme based xenobiotic identification for exposomics. Nat Commun. 2021 Sep 14;12(1):5418. doi: 10.1038/s41467-021-25698-x. PMID: 34521839; PMCID: PMC8440538.
7. Djoumbou Feunang Y, Eisner R, Knox C, Chepelev L, Hastings J, Owen G, Fahy E, Steinbeck C, Subramanian S, Bolton E, Greiner R, Wishart DS. ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. J Cheminform. 2016 Nov 4;8:61. doi: 10.1186/s13321-016-0174-y. PMID: 27867422; PMCID: PMC5096306.
8. Kim HW, Wang M, Leber CA, Nothias LF, Reher R, Kang KB, van der Hooft JJJ, Dorrestein PC, Gerwick WH, Cottrell GW. NPClassifier: A Deep Neural Network-Based Structural Classification Tool for Natural Products. J Nat Prod. 2021 Nov 26;84(11):2795-2807. doi: 10.1021/acs.jnatprod.1c00399. Epub 2021 Oct 18. PMID: 34662515; PMCID: PMC8631337.

Liquid chromatography Mass spectrometry workflow with Internal Standard (IS) set

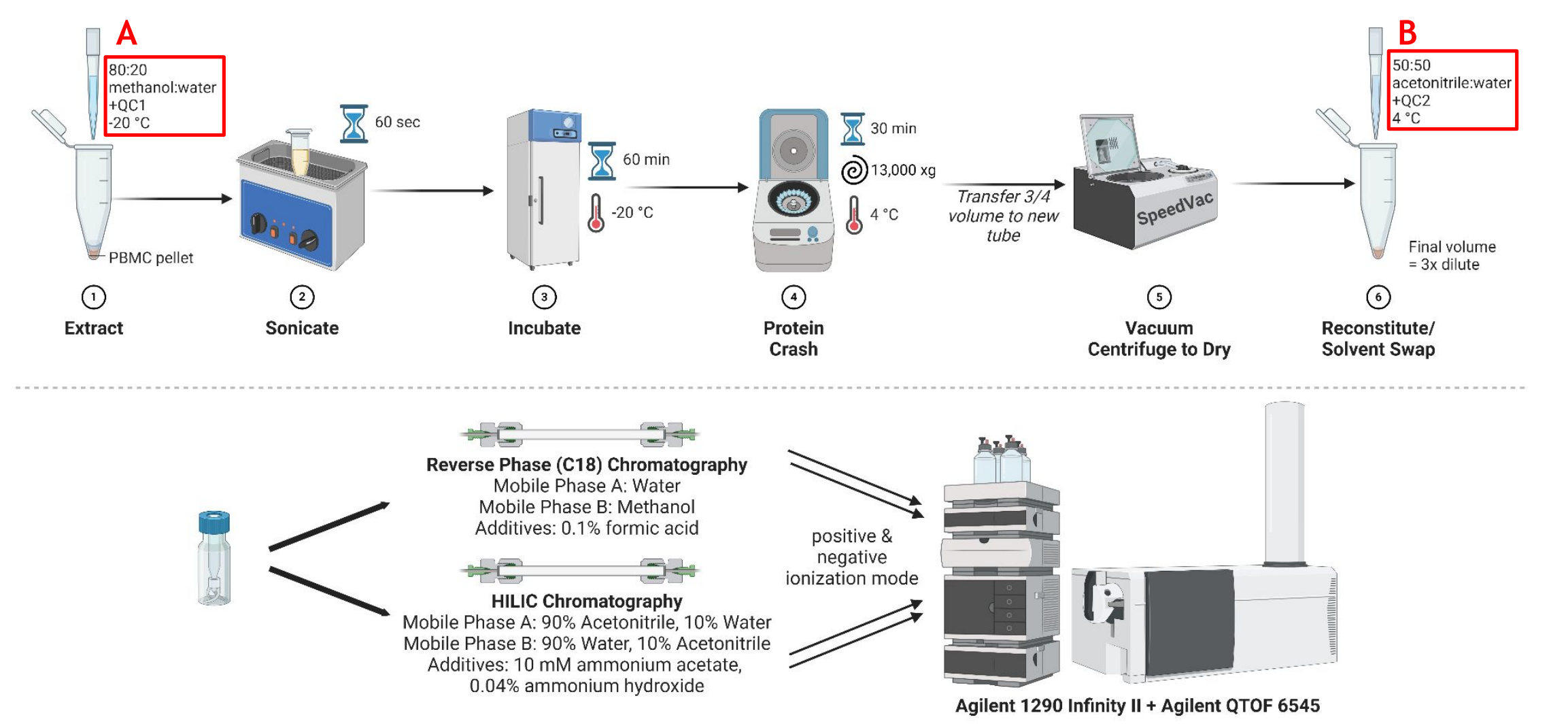

A
80:20 methanol:water +QC1 -20 °C
PBMC pellet
60 sec
60 min
-20 °C
30 min
13,000 xg
4 °C
Transfer 3/4 volume to new tube
SpeedVac
B
50:50 acetonitrile:water +QC2 4 °C
Final volume = 3x dilute
① Extract
② Sonicate
③ Incubate
④ Protein Crash
⑤ Vacuum Centrifuge to Dry
⑥ Reconstitute/ Solvent Swap
Reverse Phase (C18) Chromatography
Mobile Phase A: Water
Mobile Phase B: Methanol
Additives: 0.1% formic acid
HILIC Chromatography
Mobile Phase A: 90% Acetonitrile, 10% Water
Mobile Phase B: 90% Water, 10% Acetonitrile
Additives: 10 mM ammonium acetate, 0.04% ammonium hydroxide
positive & negative ionization mode
Agilent 1290 Infinity II + Agilent QTOF 6545

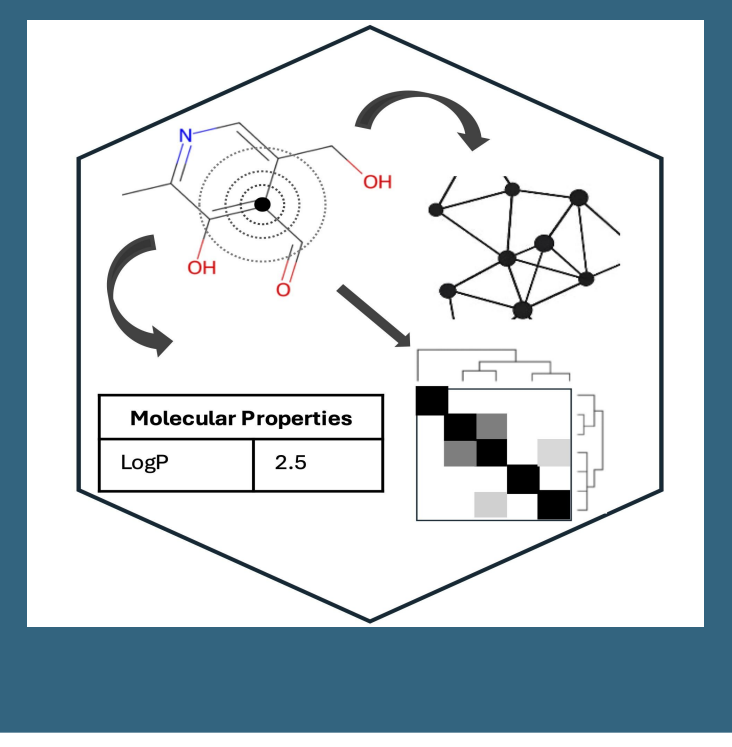
Figure 1: Sample prep workflow with internal standards

- LC-MS workflow incorporates two sets of internally developed internal standards: one prepared in methanol (Set A) and one in water (Set B).
- Set A (methanol) is introduced at the **extraction step**; Set B (water) is introduced at the **reconstitution step**.
- This two-step design enables monitoring of both **sample preparation** and **instrument variability**.
- The internal standard set composition is tailored for **exposure science applications** and supports **class-based normalization**.

Come chat if you are working on:

Internal standards tailored to your study design

Structural similarity analysis- try CHEMPASS, our new tool



https://chempass.emilabs.org/

Code: https://github.com/eitm-org/chempass_similarityApp

Internal standards list

Mix Solution in Methanol		
Compound	**Conc (ug/ml)**	**Molecular Formula**
4-Fluoro-3-phenoxy-benzoic acid - 13C6	1	13C6C7H9FO3
Triclosan-d3	1	C12H4Cl3D3O2
o-Cresol-d7	2.5	C7HD7O
Acrylamide-13C3	1	13C3H5NO
trans3-HydroxyCotinine-d3	2.5	C10H9D3N2O2
Cholic Acid-d4	20.06	C24H36D4O5
Glycocholic Acid-d4	22.5	C26H39D4NO6
Caffeine-13C3	748.25	13C3C5H10N4O2
Mefenacet-d3	1	C16D3H11N2O2S
Prometon-d14	1	C10H5D14N5O
Ibuprofen-d3	1003	C13H15D3O2
Diclofenac acid-d4	249.93	C14H7Cl2D4NO2
Dinoseb-13C6	498.35	13C6C4H12N2O5
4-Chlorobenzoic acid-d4	1	C7HClD4O2
Pirimicarb-d6	1	C11D6H12N4O2
Bisphenol A (BPA)-d16	1	C15D16O2
Acephate-d6	50.3	C4H4D6NO3PS
Dinotefuran-d3	250.75	C7H11D3N4O3
Warfarin-d5	247.75	C19H11D5O4
Nicotine-d4	7.47	C10H10D4N2
4-Acetamidophenol-d4	250.5	C8D4H5NO2
D-Sphingosine-d7	0.75	C18H30D7NO2

Mix Solution in water (w/ 0.1N HCl)		
Compound	**Conc (ug/ml)**	**Molecular Formula**
L-Leucine-13C6	1005.8	13C6H13NO2
L-Lysine-13C6	1509.3	13C6H14N2O2
Perfluorooctanoic acid (PFOA)-13C8	1	13C8HF15O2
Citric Acid-d4	1999	C6H4D4O7
Adenosine-1',2',3',4',5'-13C5	751.5	13C5C5H13N5O4
Hypoxanthine-d4	25	C5D4N4O
Taurine-13C2	499.5	13C2H7NO3S
Lactic acid-d3 sodium salt	5001.7	C3H3D3O3
L-Aspartic Acid-13C4,15N	248.4	13C4H715NO4
Muconic acid (trans,trans)-d4	1	C6H2D4O4
Creatinine-d3	252.2	C4D3H4N3O
L-S-Phenylmercapturic acid (S-PMA) -d5	1	C11H8D5NO3S
L-Kynurenine-d4	24.8	C10H8D4N2O3
beta-hydroxybutyric acid (BHB) d4 sodium salt	250	C4H4D4O3
Guanosine-13C10	12.5	13C10H13N5O5
Xanthine13C,15N2	7.5	13CC4H415N2N2O2

Table 1: Exposure Internal standards
Listed are the Internal standards with their concentrations and formulas. The concentrations used in the internal standard mix were designed to be physiological (per literature) when used at a 50:1 dilution. The two sets of IS in methanol and water were added at two steps in the sample preparation process as described in Figure 1.

Acknowledgement

The authors would like to thank Warren Casey, Jennifer Kyle, John Wambaugh and Michelle Embry for their contributions in developing the exposure compound library and exposure reference standards, which led to the development of this web application.

Results:

Building custom internal standard set

Xenobiotics (Table 2.A)		
Reference	**No of compounds**	**Rationale**
PMCID: PMC7984261[1]	8	PFAS in AFFFs (aqueous film forming foam)
EPA NPDWR; https://www.epa.gov/sdwa/and-polyfluoroalkyl-substances-pfas	5	PFAS in drinking water
PMCID: PMC11459241[2]	10	Suspect Screening Analysis of Pooled Human Serum Samples Using GC × GC/TOF-MS
NASEM Reports (2011, 2017, 2020, 2022)	14	Compiled data on airborne hazards associated with Gulf War (oil well fires/ chemical agents/ diesel etc..) and post 9/11 deployments (Burn pits)
PMCID: PMC7299071[3]	30	Identified 153 environmental chemicals detected in DoDSR samples. ~45 of these could be used to differentiate high- and low-risk deployments and pre- and post- deployment samples. Chemical classes include: PFAS, Phthalates, Pesticides (organonitrogen/phosphorous/chloro acid/carbamate), Phosphodiester flame retardants, Brominated flame retardants, Phenols. No PAHs were detected.
PMCID: PMC4977992[4]	24	Most frequently detected PAHs (n=9) and PCDDs/PCDFs (n=15)
PMCID: PMC7239036[5] PMCID: PMC11459241[2] PMCID: PMC8440538[6]	16	Exposure compounds associated with cosmetics, diet, plastics, herbicide, medication (common OTC and prescription)

Metabolites (Table 2.B)	
Reference	**No of compounds**
Amino acids and derivatives	10
Nucleosides, Nucleotides & Purines	4
Lipids & Lipid-Like Molecules	7
Carbohydrates & Conjugates	3
Organic Acids	3
Other	2
Total (A+B)	**136**

A list of compounds for internal standards was generated from an exposure compound library. These compounds were sourced from publications and databases (like EPA and NASEM) including known and potential carcinogens that deployed military personnel may have been exposed to. Common human metabolites representing a range of metabolic pathways were also included in the list, resulting in 136 compounds (**Table 2**).

Sixty-two structurally unique non-redundant internal standards were selected from the list of 136 compounds with the help of CHEMPASS (CHEMical Profiling and Analysis for Similarity Scoring) and manual review, to end up with a subset of exemplars from various compound classes. Of the 62 compounds, 38 were purchased from AChemtek as two separate mixes (**Table 1**).

Characterizing internal standard set

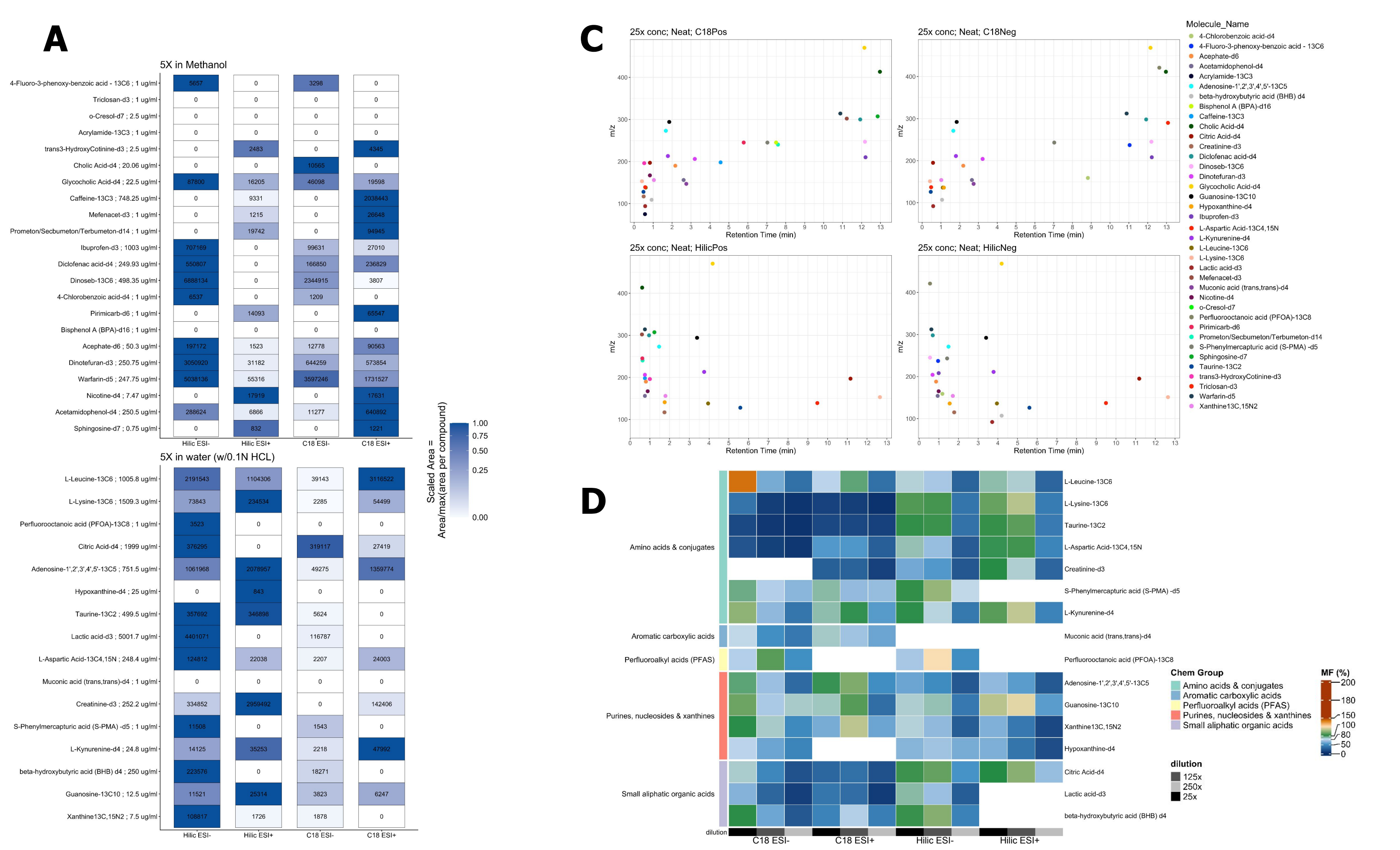


Figure 2: Characterization of exposure internal standards

A. Heatmap of internal standards spiked in plasma, colored by scaled area across 4 channels, to visualize appropriate channel per IS and flag missing IS.
B. Pie chart of the different classes of the 38 compounds in exposure internal standard set generated with classyfire[7] and manual review.
C. Retention time distribution of the exposure IS across 4 channels in neat solution spanning 0.5-13 minutes of the 22 minute run.
D. Matrix Factor % (area_plasma/area_neat) calculated for 16 internal standards (in water) across 3 dilutions in 4 channels with each IS assigned a chemical group with Classyfire[7]. Recovery varies across IS, with apparent class/channel dependent effects.

Potential class based normalization

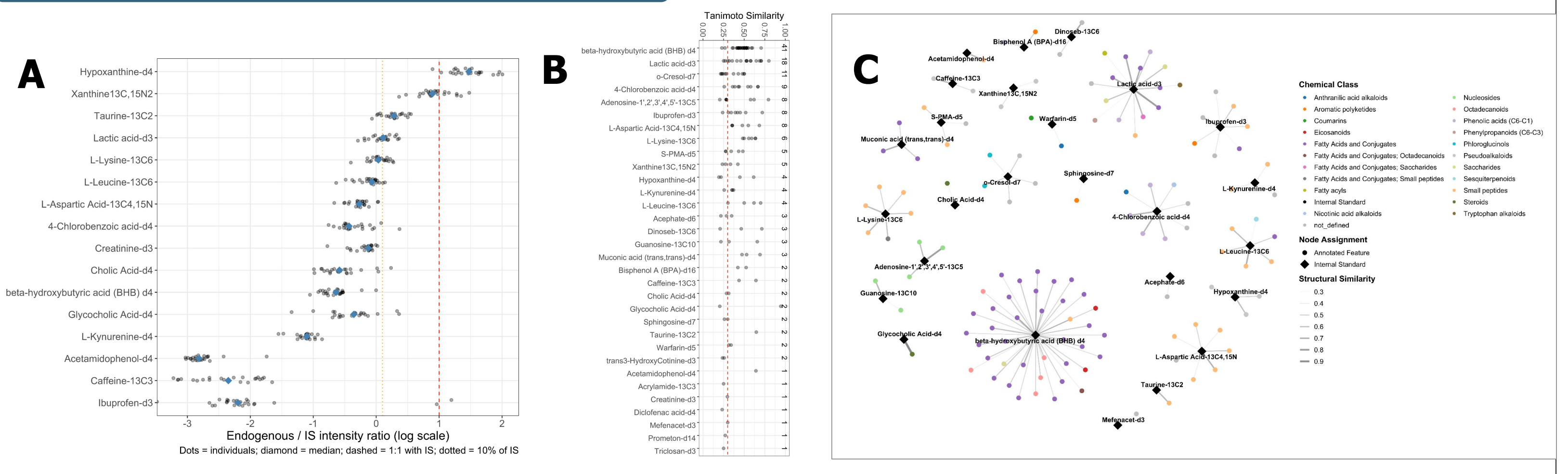


Figure 3: Characterization of exposure internal standards

A. Endogenous to IS peak height ratios across 24 individuals on a log10 scale showed varying orders of magnitude between endogenous and IS levels where 9/16 compounds were within -1 to 1 log10 ratios.
B. Tanimoto similarity scores for 166 annotated features (from Hilic ESI- channel) matched to their top structurally similar IS (CHEMPASS two list compare - FCFP4 fingerprints). BHB-d4 and Lactic acid-d3 captured the most matches.
C. Bipartite igraph network of each annotated feature linked to its best matched IS (Tanimoto score >= 0.3; 128/166 features retained). Line thickness and proximity scale with Tanimoto similarity score; FR layout; nodes colored by NPClassifier[8] superclass.

Ellison Medical Institute