

Analytical and cheminformatic framework for studying drugs and their metabolites in human plasma using high resolution mass spectrometry

Randolph R. Singh^{1,*}, PhD, Carolina Duarte-Hospital¹, Yuki Mizuno¹, Gabriela Jackson¹, Saurabh Dubey¹, Haotian Wu¹, Christina M. Eckhardt^{1,2,3}

¹Department of Environmental Health Sciences, Columbia University Mailman School of Public Health, New York, NY, United States

²Department of Medicine, Columbia University Vagelos College of Physicians and Surgeons, New York, NY, United States

³Present address: Christina M. Eckhardt, Merck & Co., Inc, Rahway, NJ, United States.

*Corresponding author: Randolph R. Singh, PhD, Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, NY 10032, United States (rrs2196@cumc.columbia.edu).

Abstract

Pharmacological interventions are required to manage a wide array of chronic diseases. Drug metabolism is heavily influenced by individual-level genomic and environmental factors, and there is substantial variability in drug efficacy for many medications. Novel strategies are needed to evaluate drug performance and improve therapeutic effectiveness. Recent technological advancements in high-resolution mass spectrometry (HRMS) have enabled broad exposomic analyses that can evaluate how an individual metabolizes drugs. Here, we present a new analytical framework to annotate drug metabolites in human plasma samples. We performed nontargeted analyses using liquid-chromatography (LC)-HRMS and developed an analytical workflow that leverages tandem mass spectra for structural annotation of unknown molecular features. In addition to database matching, we implemented molecular networking to link parent pharmaceutical compounds to potential metabolite features using shared fragmentation patterns followed by *in silico* chemical formula and structure prediction. In doing so, we successfully annotated downstream metabolites of an array of pharmaceutical agents and created a new investigational pipeline towards measuring drug metabolites in human samples for which reference standards or spectra are rarely available. The analytical framework presented in this study can be used to develop new pharmacokinetic models where both the parent drugs and the metabolites are simultaneously measured and monitored. Similarly, this workflow can be applied for annotating xenobiotic chemicals in various matrices. Our results contribute to efforts in developing a personalized approach to measuring drug metabolites and may provide new opportunities to maximize therapeutic response and improve drug efficacy in humans.

Key words: HRMS; nontargeted analysis; pharmacoexposomics; drug metabolites; unknown MS feature annotation.

Introduction

Pharmacological interventions are a cornerstone of modern medicine and are required to manage a wide array of chronic diseases. Unfortunately, there is substantial variability in drug efficacy and drug tolerance for many medications, which can jeopardize the health of individuals who depend on medications to prevent disease progression and improve their quality of life.¹ Emerging research has shown the therapeutic response to specific medications is determined in part by an individual's metabolic state.² An individual's biochemical profile and resulting metabolic phenotype greatly affect drug metabolism and are shaped by a combination of genomic and environmental factors. For example, genetic polymorphisms influence cellular metabolism and can substantially alter drug metabolism and efficacy.³⁻⁵ Similarly, environmental factors including diet, lifestyle, and exposure to environmental pollutants can affect an individual's metabolic phenotype and the ability to metabolize certain drugs.⁶ Insight into an individual's ability to metabolize drugs

and analyses of the resulting metabolite levels may elucidate the expected long-term therapeutic response. For example, individuals who rapidly metabolize drugs into a form that is easily eliminated (eg, glucuronidation) may experience subtherapeutic effects. Thus, it is critically important to develop novel strategies to analyze drug metabolite levels and predict the therapeutic response.

Liquid chromatography high-resolution mass spectrometry (LC-HRMS) data is typically composed of full scan data (MS1), which typically measures the mass of the intact molecules, and a corresponding fragmentation spectrum (MS2), which provides structural information.⁷ MS1 and MS2 data can be used to match metabolites with spectral databases, producing putative chemical annotations.⁷ In cases where there is no database match, *in silico* structural prediction can be used to extend the capability of data annotation pipelines by predicting both the chemical formula and probable structure of an unknown chemical feature using empirical data. Recent developments in the field have

Received: October 18, 2025; Revised: January 10, 2026; Accepted: January 24, 2026

© The Author(s) 2026. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

paved the way for tools like SIRIUS, which is a java-based software framework for the analysis of LC-HRMS data of small molecules. SIRIUS integrates a collection of tools including CSI: FingerID, COSMIC,⁸ ZODIAC,⁹ and CANOPUS.¹⁰ Usage of these tools is facilitated by other tools like MZmine,^{11,12} which performs large scale data processing and visualization, performs molecular networking,¹³⁻¹⁵ and allows export of data to the SIRIUS¹⁶ suite of tools. Peak picking and feature alignment during data processing are some of the most essential steps in data analysis.¹⁷ Both MZmine and SIRIUS have extensive and user friendly documentation for a more detailed insight on what these tools can do. Molecular networking enables the identification of molecules that may be structurally related using similarities in their fragmentation spectra. In this work, we demonstrate application of these tools to identify drug metabolites in human plasma samples, which may facilitate the development of precision-guided methods to predict drug efficacy and improve the therapeutic response.

Methods

Study design

The Fibrotic Biomarkers among Subjects with Interstitial Scarring (FIBROSIS) Study is a prospective cohort study that enrolled men and women with clinically diagnosed idiopathic pulmonary fibrosis (IPF) as well as age- and sex-matched healthy control subjects with no known lung disease from 2022-2023. Exclusion criteria included inability to provide informed consent, current smoking, and IPF exacerbation in the month preceding enrollment. The study was approved by the Columbia University institutional review board (AAAU7264). All study participants provided written informed consent. Among 68 study participants, the mean age was 69.8 ± 7.7 years (Supporting Information Table S1). There were 49 (72.1%) male participants. Overall, 36 participants (52.3%) reported former use of tobacco products. With respect to medical comorbidities, 8 (11.8%) participants had diabetes, 36 (52.9%) had hyperlipidemia, 23 (33.8%) had hypertension, and 47 (69.1%) had IPF.

Plasma extraction protocol for pharmacometabolomic analyses

Venous blood was collected in EDTA tubes and centrifuged to separate plasma. Plasma samples were frozen at -80°C until they were thawed for use in this study. Prior to analyte extraction, 200 μL aliquots of plasma were transferred to clean Eppendorf tubes and mixed with 600 μL of isotopically labeled internal standards fortified acetonitrile (Optima LC/MS grade, Fisher Scientific). The final concentration per sample of each internal standard are 2500 nM [13C6]-D-glucose, 5 nM [15N]-indole, 75 nM L-Lysine-2HCl (α -¹⁵N), 4 nM [15N]-choline chloride, 12.5 nM [13C5]-L-glutamic acid, 5 nM [13C7]-benzoic acid, 25 nM [15N]-L-tyrosine, 0.0669 nM Uracil (1,3-¹⁵N₂), 5 nM [trimethyl-¹³C3]-caffeine and 12.5 nM [U-13C5, U-15N2]-L-glutamine (all purchased from Cambridge Isotope Laboratories, MA, USA, purity 98 or 99%). Internal standards were added to help monitor inter-injection variability in instrument performance and retention time shifts. The addition of acetonitrile is done to precipitate proteins and minimize spectral interferences. The samples equilibrated on ice for 30min and were centrifuged (4°C) for 20min at 14,000 revolutions per minute (RPM). A 500 μL aliquot of the supernatant was transferred to a fresh Eppendorf tube and concentrated to 50 μL using a nitrogen gas concentrator. The concentration step was employed as it helped increase the signal

for the metabolites which in turn allowed acquisition of experimental fragmentation spectra. The resulting sample was reconstituted to 100 μL using 20mM ammonium formate (pH 3.0) to match the starting mobile phase condition, which helps improve chromatographic peak shapes. In addition to study samples, a pooled sample extract and internal standard spiked solvent blanks were injected at the beginning and end of the worklist. Solvent blanks were injected after every five samples.

Liquid-chromatography high resolution mass spectrometry analysis

Analyte separation of plasma samples was performed with a Waters CORTECS Premier C₁₈ column (2.1 \times 150mm) in an Acquity I class HPLC system coupled to an Orbitrap IQ-X (Thermo Fisher Scientific, Waltham, MA, USA) equipped with a heated electrospray ionization probe (Figure 1). The injection volume was 10 μL and the flow rate was set to 0.4 mL/min. The auto-sampler was maintained at 4°C while the column temperature was set to 30°C . Mobile phase A consisted of 20 mM ammonium formate in water while mobile phase B (MP B) consisted of 20 mM ammonium formate (LiChropur, MilliporeSigma) in acetonitrile. The percentage of the organic modifier (MP B) started at 10%B and was maintained for 1 min followed by a linear ramp to 100% at 7.5 min. This was maintained for 2.5 min before going back to the starting mobile phase condition at 10.1 min. The column was re-equilibrated for 4.5min before injection of the next sample, giving a total run time of 15 min. All samples were analyzed in positive ionization using the data-dependent acquisition mode (top 10), mass range (m/z 50-1000), maximum injection time (100 ms), at a resolution of 120000 in full-scan mode, and 30000 for MS2, normalized HCD collision energy (25%). We used the following source parameters: spray voltage (3500), vaporizer temperature 350°C , ion transfer tube temperature 325°C , the sheath gas flow rate 29 au, the auxiliary gas flow rate 11 au, the sweep gas at 1 au, the RF lens level 60%. Dynamic exclusion was employed after 1 spectra or an exclusion duration of 5 secs and a mass tolerance of 10 ppm. Isotopes were also excluded. Instrument performance during instrumental analysis was monitored by ensuring reproducible LC pressure profiles followed by ensuring the inter-sample intensity of internal standards did not vary by more than 10%.

Data processing workflow

Raw files from participant samples, blanks, and quality control samples were analyzed using MZmine (version 4.5.0), which is an open-source software program for mass spectrometry data processing.^{17,18} Screenshots of the specific settings used for data processing within MZmine including molecular networking are provided in the Figure S1. Molecular features including full scan (MS1) and fragmentation (MS2) spectra were extracted, aligned, and matched against an open spectral database downloaded from the website of Computational Mass Spectrometry (CompMS)/MS-DIAL (version 19).¹⁹ MS-DIAL is an open source software program for processing non-targeted metabolomics data. Features that were detected in the blank samples were retained in the analyses if the average peak area of the feature was three times greater than the corresponding signals in the blank samples.

Database matches (similarity score > 0.85) for pharmaceutical compounds were prioritized for molecular networking analyses within MZmine. Features that were connected to known pharmaceutical compounds were identified via shared MS2 fragments. The MS2 spectra for features connected to known

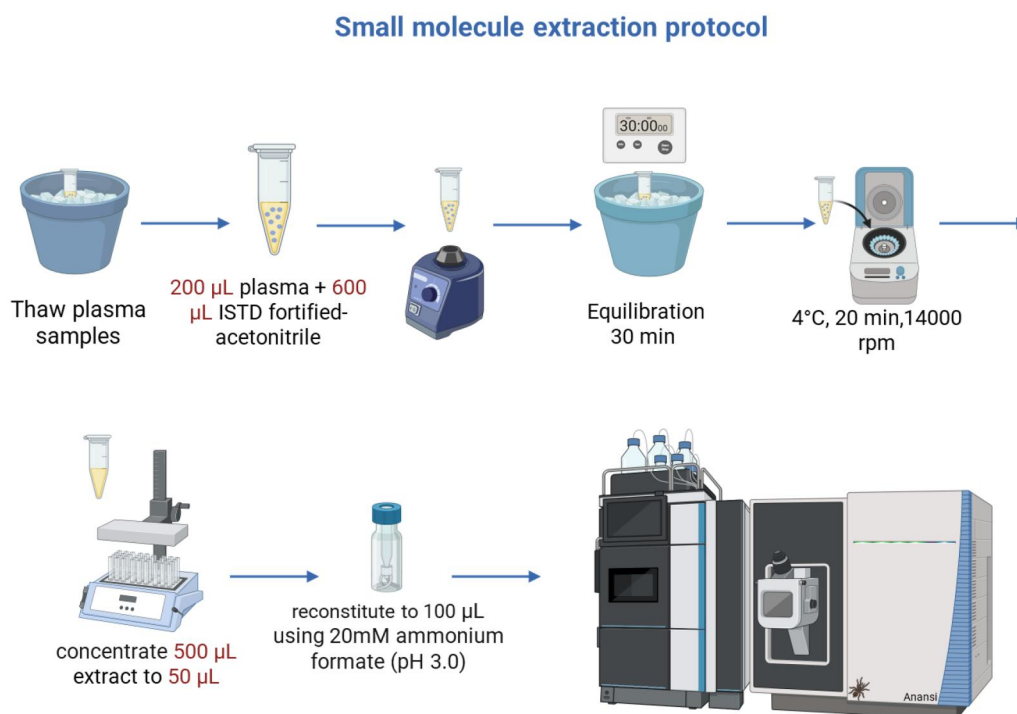


Figure 1. Protocol for extracting and analyzing small molecules from human plasma. Abbreviations: RPM = Revolutions per minute. µL = Microliter. Analyte separation of plasma samples was performed using an Orbitrap IQ-X mass spectrometer equipped with a Waters CORTECS Premier C18 column (2.1 × 150 mm) for reversed phase chromatography. All samples were analyzed in positive ionization using the data-dependent acquisition mode. Figure made using BioRender.

pharmaceutical compounds were exported for further analyses in SIRIUS (version 5.8.6), which is a software program that combines the analysis of isotope patterns in MS spectra with the analysis of fragmentation patterns in tandem mass spectra.^{20,21} Features were exported using the COSMIC workflow, which provides *in silico* structure database generation and annotation.

SIRIUS settings were modified for orbitrap data analysis and directed to search through all databases available in the graphical user interface (Figure S2). The CSI: FingerID scores were obtained for all features of interest. CSI: FingerID is a tool embedded within SIRIUS software that predicts a compound's molecular fingerprint based on tandem mass spectrum (MS/MS) data. To identify molecules of interest, SIRIUS generated 100 predicted matching structures that were ranked in order of descending CSI: FingerID Score.

To identify the optimal CSI: FingerID scores, CSI: FingerID scores were benchmarked against mixture 506 from the Environmental Protection Agency's non-targeted analysis collaborative trial (ENTACT).²²⁻²⁴ The ENTACT trial was designed to evaluate non-targeted laboratory methods and assess their ability to identify unknown chemicals in human samples.²² The ENTACT mixtures, numbered 499-508, comprise high purity standards of known toxicologically relevant chemicals with varying degrees of mixture components. Access to datafiles was possible through previous participation in the trial.^{23,24} We compiled the CSI: FingerID scores, octanol/water partition coefficients (XLogP), mass-to-charge ratios (m/z), retention times, and ranks of correct annotation for both the ENTACT 506 components and the probable drugs and drug metabolites detected in the present study. Features that were detected using positive electrospray ionization were retained for analyses.

Performance was evaluated by classifying the Rank 1 CSI: FingerID predictions against the known ENTACT mixture

composition. A true positive was defined as a rank 1 prediction matching a known standard, while a false positive was defined as a rank 1 prediction to a structure not present in the mixture. The receiver operating characteristic (ROC) curve was constructed using the score distributions of these two classes, allowing for the direct calculation of sensitivity and specificity. Additionally, to establish a practical and high-confidence cutoff for this study, a score threshold was determined by controlling the False Discovery Rate (FDR) at 0.05. All analyses were conducted in Python (v. 3.12.2) using the scikit-learn, pandas, numpy, and matplotlib libraries. The identity of the annotations were confirmed (Level 1) by comparing the m/z , retention time, and fragmentation pattern against data obtained from analyzing the ApexBio DiscoveryProbe Drug Library (Apexbio, Catalog Number L1021). Manual curation was performed to verify whether the annotated metabolites were detected in the same sample as the parent drug. The annotation confidence level system proposed by Schymanski et al. was used.²⁰

Results

SIRIUS benchmarking using ENTACT mixture 506

Using SIRIUS and positive electrospray ionization, a total of 196 features were annotated from the ENTACT 506 mixture (Supporting Information Table S2). In the present study, 146 of the 196 identified features were characterized by matching MS-DIAL's open mass spectral database with similarity scores that ranged from 0.718 to 0.895 (19 features) and from 0.900 to 0.994 (127 features). When the fragmentation spectra (MS/MS) of the matched features were exported to SIRIUS, 130 features had correctly predicted annotations that received a ranking of 1 when SIRIUS generated predicted matching structures that were

ranked in order of descending CSI: FingerID Score. A total of 138 features received a correct annotation ranked within the top 5 of predicted matching structures, and 142 features received a ranking within the top 10. Only four features (propham [rank 12], 2-methyl-4'-(methylthio)-2-morpholinopropiophenone [rank 15], CP-728663 [rank 22], and dodecylamine [rank 96]) had predicted annotations that were not within the top 10 of predicted matching structures in SIRIUS. Close inspection of the remaining candidates showed they were positional isomers of the correct structure. These results suggest that the predictions obtained from SIRIUS have a high level of accuracy and precision except in cases where there exists many positional isomers.

In addition to identifying features using MS/MS matching, we applied SIRIUS to predict the structure of 50 compounds that did not have a database MS/MS match but that matched the precursor ion of known ENTACT 506 components within 5 ppm. In total, 44 compounds had annotations that received a ranking of 1 when SIRIUS ranked predicted matching structures in order of descending CSI: FingerID Score. A total of 47 compounds received a ranking within the top 5 of predicted matching structures, and 48 compounds were ranked within the top 10. Only two compounds (Basic Blue 7 [rank 22] and dipentylphthalate [rank 94]) had predicted structures that were not ranked within the top 10 of predicted matching structures. Using ENTACT as a validation set, SIRIUS was able to predict the correct structure as the top 1, within the top 5, and within the top 10, 89%, 94%, and 97% of the time, respectively.

As an illustrative example, an in-depth SIRIUS analysis is presented below for an unknown feature with m/z 453.1438 (Figure 2). The isotope distribution of the MS1 spectrum suggested the feature was singly chlorinated which is supported by the predicted chemical $C_{22}H_{21}ClN_6O_3$. The mirror plot (Figure 2A) showed strong agreement between the theoretical and experimental full scan spectra including the proper isotope pattern where the $A + 1$ ion was expected to be approximately 35% of the base peak. The formulas for the fragments were also calculated and constructed like a tree to show neutral losses (Figure 2B). The fragmentation tree confirmed the predicted formulas supported the putative formula of the intact molecule. The fragment peaks and the corresponding predicted structure of the fragments were consistent with the structure of the top-ranked candidate (Figure 2C). In addition to the predicted structure, SIRIUS also provides a link to the specific online databases (experimenter specified during analysis) where the molecule was found. A PubChem search of the molecule using the structural identifier (eg, chemical name, InChIKey, etc.) as a search term, or through the provided database link in the software, generated the annotation of the unknown molecule. In this instance, the unknown molecule was losartan carboxylic acid, which is a metabolite of the antihypertensive drug losartan. Aside from the ranking of the predicted structures, CSI: FingerID also provides a score, which is -74.227 for losartan carboxylic acid.

To determine a practical way of using CSI: FingerID scores to gauge the accuracy of feature predictions, we compiled the scores for the 196 features that had annotated matches in the ENTACT 506 mixture and performed a Receiver Operating Characteristic (ROC) analysis. The analysis resulted in an Area Under the Curve (AUC) of 0.90, indicating high discriminatory performance (Figure 3). Subsequently, we established a practical CSI: FingerID score threshold of -81.18 by controlling the False Discovery Rate (FDR) at 5%.

Manual inspection of the predicted structures revealed that most of the top-ranked candidates had fragment peaks that were

consistent with the proposed structure. These results suggest SIRIUS CSI: FingerID can accurately classify the identity of most unknown metabolomic features. We also showed operationally that a score of greater than -81 is a reliable indicator of a molecule's structural match.

Molecular networking facilitates identification of pharmaceutical agents and their metabolites

To identify pharmaceutical agents and their respective metabolites, we employed database matching and molecular networking functions within the MZmine data analysis platform. We applied an arbitrary but conservative score threshold of >0.900 to indicate a good match. Using this approach, we identified 67 drugs from multiple different therapeutic areas including antihypertensive drugs, antidiabetic drugs, psychoactive drugs, antimicrobial drugs, anti-fibrotic drugs, anti-histamines, and anti-inflammatory drugs. A list of the identified pharmaceutical compounds is provided in the Supporting Information (Table S3).

We then aimed to develop a framework for identifying drug metabolites using established molecular networking tools. Annotated features in MZmine have associated molecular networks, which are determined by identifying similar fragments between two or more features. The analytical platform created molecular networks by grouping compounds with similar molecular scaffolds. Representative fragmentation patterns of features that are related to the drug losartan are shown in Figure 4. Inspection of the core structure of losartan compared to irbesartan and valsartan shows great similarity which makes it unsurprising that molecular networking would suggest that these molecules belong to the same network. The MS1 spectra of some of these features (except the feature with m/z 459.2140) showed evidence of chlorination (MS2 spectra shown in Figure 4B), which supported their structural similarity to losartan.

The candidate structures were calculated using the same protocol that was implemented to benchmark ENTACT chemicals (SIRIUS settings are reported in the SI). The spectra were annotated as losartan (retention time [t_r] = 5.51), losartan carboxylic acid (t_r = 5.78), 3-hydroxylosartan (t_r = 4.50), 1-hydroxylosartan (t_r = 4.91), hydroxy-losartan carboxylic acid (t_r = 4.60), and desbutyl-(3-carboxypropyl)irbesartan (t_r = 4.4, 4.79). The CSI: FingerID scores for losartan and its metabolites ranged from -78 to -46 and were within the range of the acceptable scores obtained for the ENTACT chemicals. In addition, with the exception of losartan carboxylic acid, the early elution of the putative metabolites also aligned with the expectation that hydroxylation renders a molecule to be more polar than its parent drug. Losartan carboxylic acid may represent a neutral form that had later elution compared to the parent drug, which is plausible as the samples were in acidic media (pH 3.0). Using the same molecular network, we successfully annotated irbesartan and eight of its metabolites, valsartan and three of its metabolites, and olmesartan, which did not have a database match.

Using the aforementioned approach, we successfully annotated 127 drug metabolites derived from 44 (of 67) parent drugs (Table S3). The CSI: FingerID scores for the parent drugs and most of the metabolites were within the acceptable range based on the benchmarking results obtained from ENTACT 506 (Figure 3). Venlafaxine (14 metabolites), irbesartan (8 metabolites), nortriptyline (8 metabolites), and carvedilol (7 metabolites) had the highest number of annotated drug metabolites. Multiple different forms of metabolism-induced drug transformation were observed including carboxylation, dealkylation (desmethyl, desacetyl, desbutyl, and didemethyl), dehydroxylation,

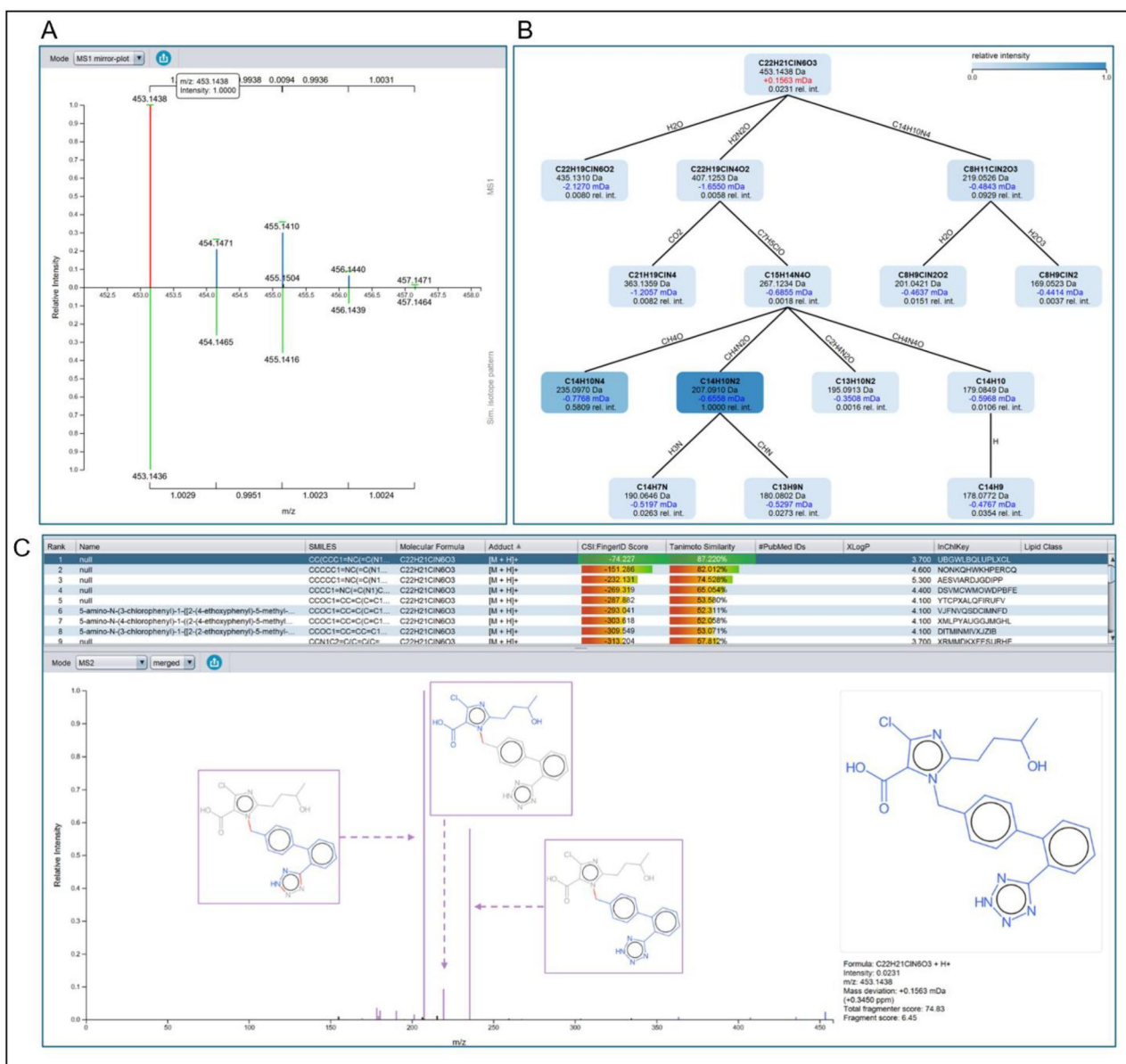


Figure 2. SIRIUS analysis identified a downstream metabolite of the anti-hypertensive drug losartan. (A) The MS1 mirror plot showed strong agreement between the experimental formula and the simulated formula. (B) The formulas for the fragments were calculated and constructed like a tree to show neutral losses. (C) The experimental fragments were matched to the candidate molecular structure: losartan carboxylic acid.

glucuronidation, hydroxylation, N-oxidation, and sulfation. A summary of the parent drugs and their downstream metabolites is presented in Table S3.

The most common drug metabolites annotated were glucuronidated pharmaceuticals. When curating the molecular networks, majority of the glucuronidated pharmaceutical metabolites clustered together. The fragmentation spectra showed these features were clustering because of a common neutral loss of 176.0321 (C₆H₅O₆) (Figure 5). Upon further investigation, we determined that when the major fragment produced was formed through the neutral loss of the sugar moiety, the glucuronidated metabolites were observed to cluster together. In contrast, when the major fragments were the same as the parent drug, the metabolites clustered with features that shared structural similarities (Figure 6).

In addition to identifying downstream metabolites, we also annotated features that appeared consistent with the structure

of known drug impurities. Nortriptyline impurity H had a CSI: FingerID score of -35 and all of the experimental fragments were consistent with the proposed structure (Figure S3). Nortriptyline impurity H (PubChem CID: 616060) has been previously reported to be produced through nortriptyline photodegradation²⁵ and decomposition in aqueous media.^{26,27} Similarly, valsartan impurity IV^{28,29} (PubChem CID: 53954839) had a CSI: FingerID score of -103 and all of the experimental fragments were consistent with the proposed structure (Figure S4). These impurities shared common MS2 fragments with the parent drugs and were measured in the same samples where the parent drugs were detected. The peak areas of the impurities were less than 1% of the peak areas observed for the parent drugs. Very little is known about the biological effects of drug impurities, much less their detection in actual human samples. Due to their structural similarity to the parent drug, they may or may not exhibit biological activity, which

raises questions on what their long term effects are considering that if these impurities are always present with the actual drug, then exposure to it can be considered chronic. By being able to detect and identify these unintended chemical components, we can start to ask further questions and work on understanding whether these may have long term effects on humans.

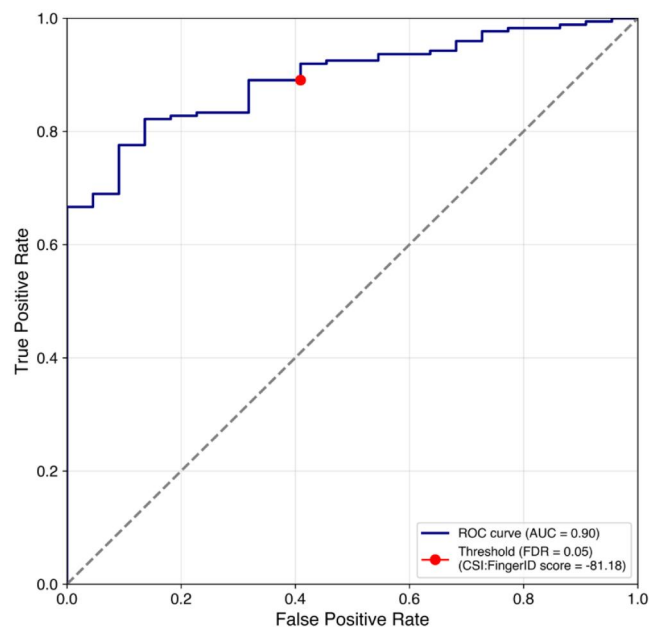


Figure 3. Receiver Operating Characteristics curve for CSI: FingerID score of compounds with annotated matches in ENTACT 506.

Discussion

Our study introduces a novel analytical framework to identify drug metabolites in human plasma samples. We performed non-targeted metabolomic analyses using LC-HRMS and developed an analytical workflow that classified the identities of unknown molecular features using tandem mass spectrum (MS/MS) data. We implemented molecular networking platforms to match features of interest with parent pharmaceutical compounds. In doing so, we successfully annotated downstream metabolites of an array of pharmaceutical agents and created a new investigational pipeline to relatively quantify drug metabolites in human samples. Our results contribute to the analytical toolbox to advance efforts to develop a personalized approach to measure drug metabolites and improve drug efficacy in humans.

In this work, we demonstrated how database matching coupled with molecular networking and *in silico* structure prediction expanded our ability to identify drug metabolites and drug-related impurities without *a priori* knowledge of first- and second-pass drug metabolism. This capability may help bridge knowledge gaps in analytical chemistry and toxicology, as knowledge related to first- and second-pass drug metabolism has historically been required to identify drug metabolites.²¹ In lieu of *a priori* knowledge of drug metabolism, the SIRIUS program implements mass spectrometry first principles to annotate unknown MS1 and MS2 fragmentation data sets. The SIRIUS program uses the precursor ion full scan spectrum to match the ions with corresponding isotopes. The program then calculates the formula of corresponding fragments and determines whether the experimentally acquired fragment peaks can explain the proposed structure. Our results showed the SIRIUS program can accurately identify unknown features that received a ranking of 1 with 88.8% consistency when SIRIUS ranked predicted matching structures in order of descending CSI: FingerID Score. Features that received a ranking within the top 5 of predicted matching

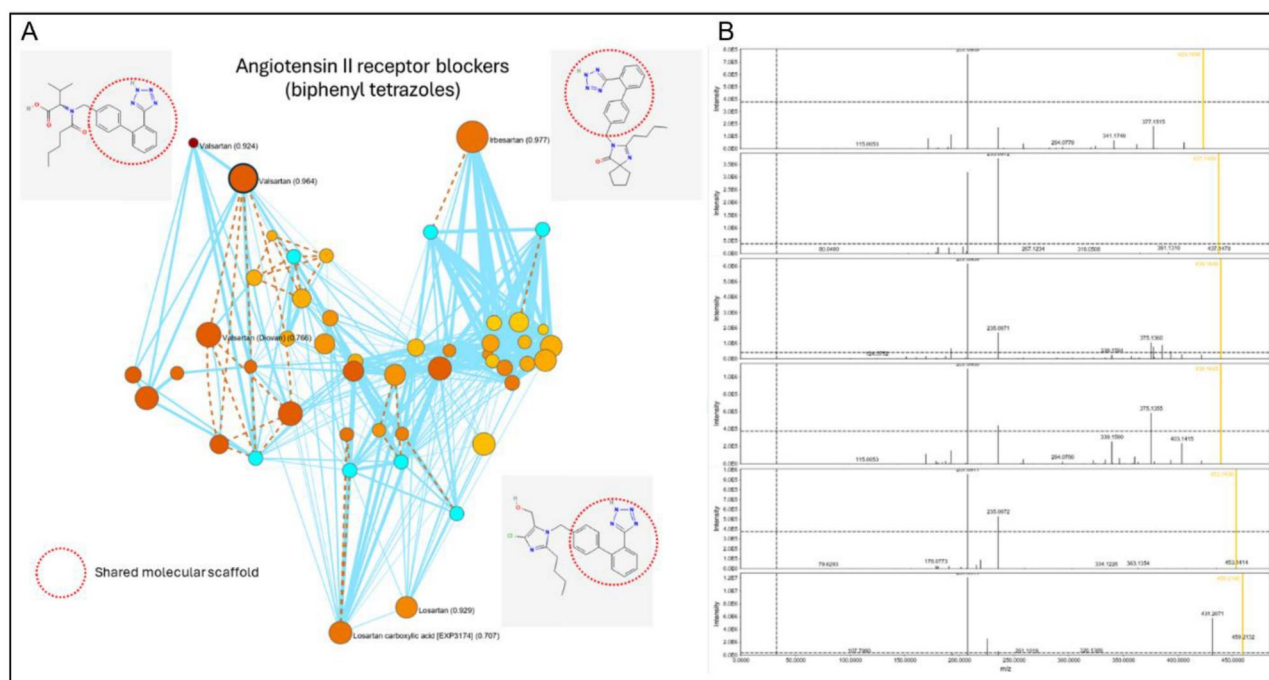


Figure 4. Representative fragmentation patterns of features that are related to the drug losartan. (A) shows the molecular network between structurally related compounds. (B) shows the comparison of some of the MS2 spectra of features that belong to the same network. Yellow lines represent the molecular ion.

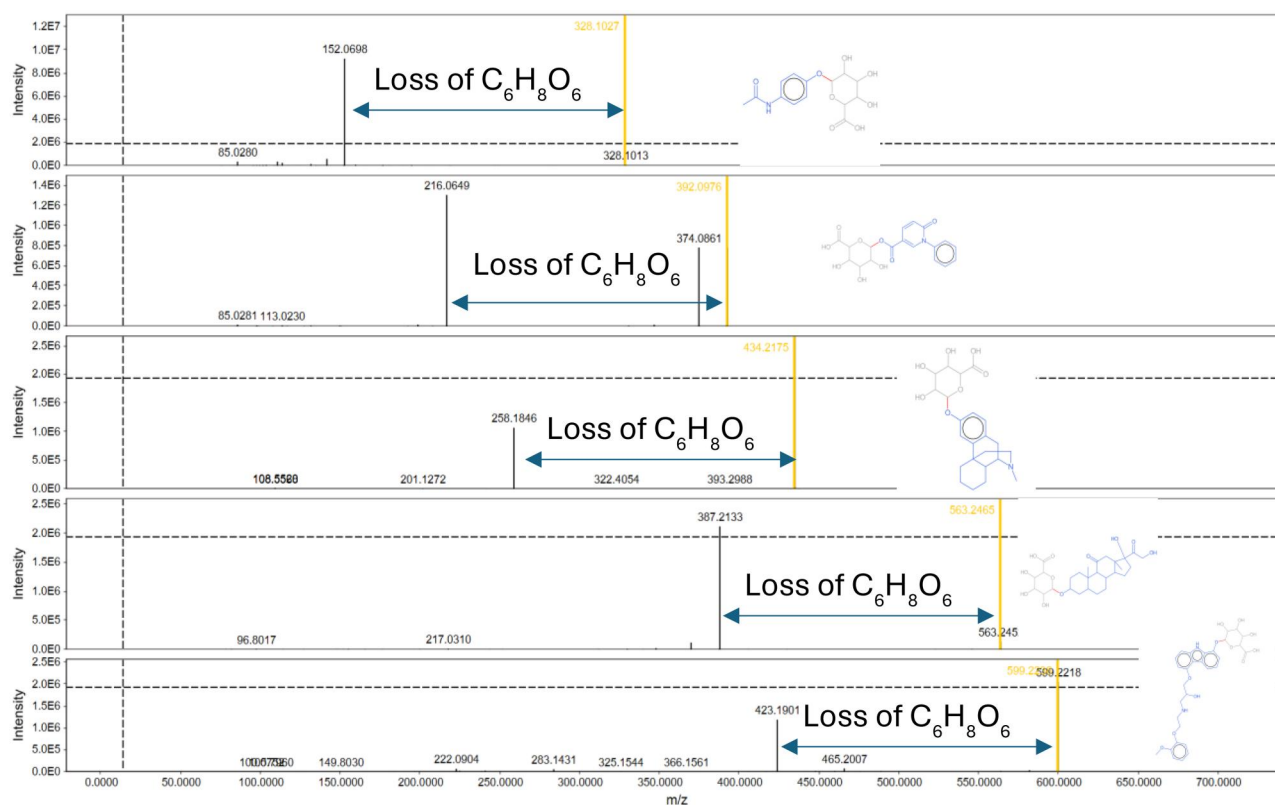


Figure 5. Fragmentation spectra of glucuronidated drug metabolites that cluster together in a molecular network. Top to bottom: acetaminophen glucuronide, pifenidone glucuronide, dextromethorphan glucuronide, aldosterone glucuronide, 8-hydroxycarvedilol O-glucuronide. The fragmentation spectra cluster together in a molecular network due to the major fragment in their respective MS2 spectrum coming from the neutral loss of C₆H₈O₆.

structures were annotated with 94.8% consistency, while features that received a ranking within the top 10 were accurately annotated with 97.4% consistency. The analytical framework presented in this study can be used to inform new pharmacokinetic models where both the parent drugs and the metabolites are simultaneously measured and monitored. Similarly, we envision that this has potential in clinical settings where unique individual drug metabolism and response can be determined immediately and thus help the primary care provider provide the patient with a drug dose that takes into account their personal metabolism.

Individual-level genomic and environmental factors can substantially alter drug metabolism,⁵ emphasizing the importance of novel approaches to measure drug metabolites in human samples. Analyses of the concentration of the parent drug and downstream drug metabolites have the potential to identify individuals with limited drug absorption or accelerated drug metabolism who may have a sub-optimal response to pharmacologic treatment. Measuring specific drug metabolites may also elucidate why some individuals experience sub-therapeutic effects or intolerable side effects that necessitate drug discontinuation. Developing a precision approach to quantifying drug metabolites may ultimately provide new opportunities to maximize the therapeutic response and improve drug efficacy. Future population-based studies should determine whether the concentration of specific metabolites in the blood can predict treatment response and facilitate interventions that improve drug efficacy.

The proposed analytical framework has several strengths. The SIRIUS program is easy to install, demonstrates high precision,

and can annotate MS1 and MS2 fragmentation data sets without requiring any *a priori* information about the features of interest. In addition, the analytical framework does not require an enzymatic deconjugation step,³⁰⁻³⁴ which improves the efficiency of the established workflow. However, the framework has several limitations. The algorithm uses fragmentation spectra to improve annotation. As a result, the accuracy of the algorithm depends on the availability of fragmentation spectra. This presents a challenge for mass spectrometers with slow data acquisition speeds, as they can only acquire a limited number of MS2 spectra per unit time. There is also a clear disadvantage for peaks of lower abundance when analyzed on instruments that use data-dependent acquisition based on the most abundant precursor ions. Nonetheless, the analytical framework provides a novel approach to measuring drug metabolites with high precision in human samples. It is important to note that these are probable annotations. Ultimately, it is recommended to confirm the identity of the molecules using authentic standards, whenever possible, especially the parent drugs. It is also important to ensure the parent drugs and their respective metabolites are detected in the same sample. Lastly, the study can benefit from being validated on samples coming from a more diverse group of people to account for potential matrix differences.

In conclusion, we developed a novel LC-HRMS-based approach to measure pharmaceutical agents and downstream metabolites in humans. Our analytical workflow annotated downstream metabolites of an array of pharmaceutical agents and created a new investigational pipeline to identify drug metabolites in human samples. Given the wide variability in drug metabolism and

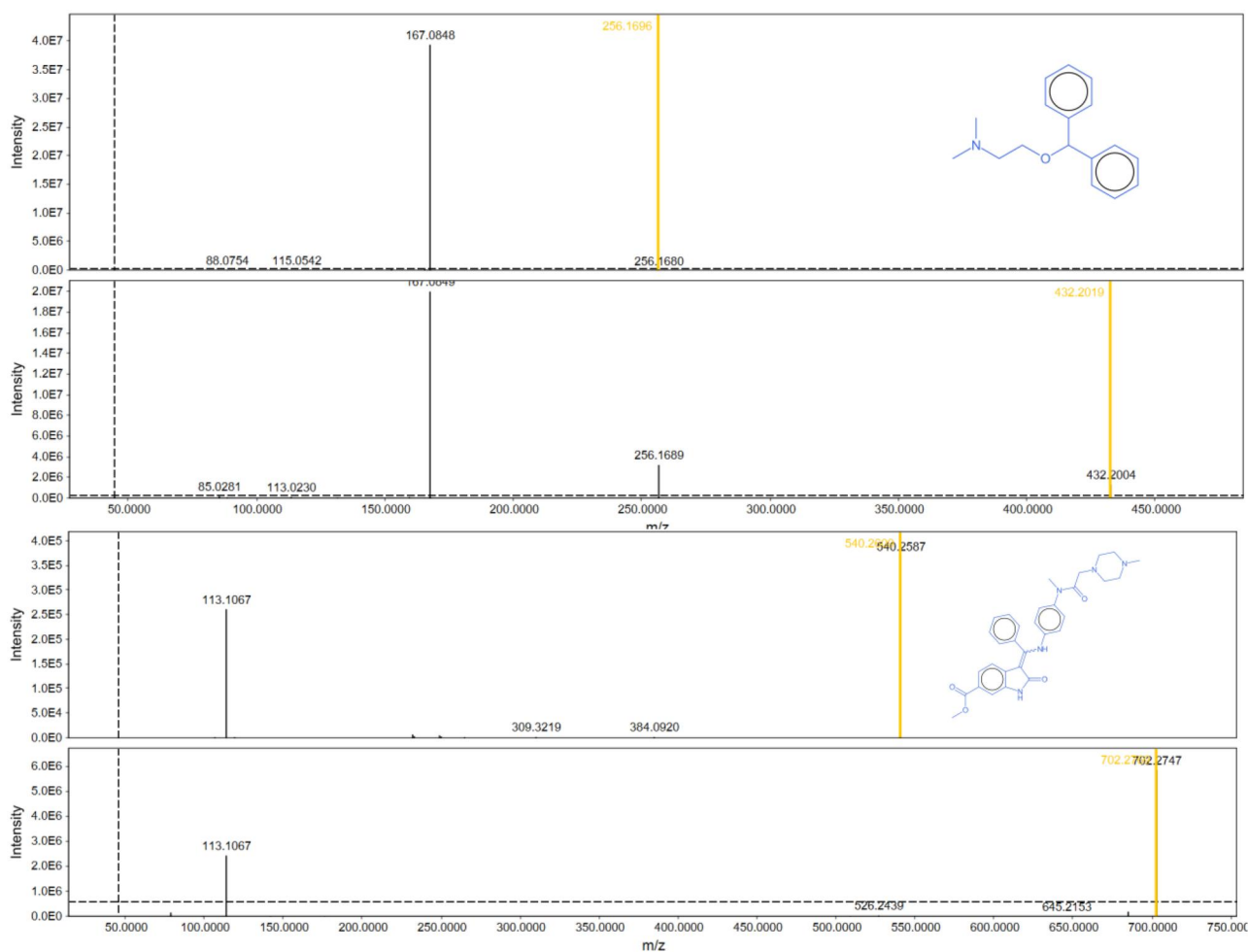


Figure 6. Fragmentation spectra of parent drug-metabolite pairs that cluster together in a molecular network. Top to bottom: diphenhydramine, diphenhydramine glucuronide, nintedanib, nintedanib glucuronide. Fragmentation spectra cluster together in a molecular network.

drug efficacy for many medications, our approach promises to enable advanced pharmacometabolomic analyses that can identify variations in drug metabolism and improve drug tolerance and drug efficacy on a large scale.

Funding

Christina Eckhardt was supported by the National Institutes of Health/National Heart, Lung, and Blood Institute (NHLBI) through grant no. K23HL171822 and the American Thoracic Society (ATS) through grant no. 23-24U4.

Author contributions

Randolph Singh (Conceptualization [equal], Data curation [lead], Formal analysis [lead], Investigation [lead], Methodology [lead], Resources [equal], Validation [lead], Visualization [lead], Writing—original draft [lead], Writing—review & editing [lead]), Carolina Duarte-Hospital (Data curation [supporting], Formal analysis [supporting], Investigation [supporting], Methodology [supporting], Writing—review & editing [supporting]), Yuki Mizuno (Data curation [supporting], Formal analysis [supporting], Visualization [supporting], Writing—review & editing [supporting]), Gabriela Jackson (Methodology [supporting], Writing—review & editing

[supporting]), Saurabh Dubey (Investigation [equal], Methodology [equal], Writing—review & editing [supporting]), and Haotian Wu (Conceptualization [equal], Data curation [supporting], Formal analysis [supporting], Funding acquisition [equal], Investigation [equal], Methodology [equal], Resources [equal], Writing—review & editing [supporting]), Christina M. Eckhardt (Conceptualization [lead], Formal analysis [equal], Funding acquisition [lead], Investigation [equal], Methodology [supporting], Project administration [lead], Resources [lead], Supervision [equal], Validation [equal], Writing—original draft [equal], Writing—review & editing [equal]).

Supplementary data

Supplementary data is available at *Exposome* online.

Conflicts of interest

Christina Eckhardt is an employee and shareholder in Merck & Co., Inc.

Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

References

- Rappaport SM, Barupal DK, Wishart D, Vineis P, Scalbert A. The blood exposome and its role in discovering causes of disease. *Environ Health Perspect*. 2014;122:769–774.
- Wild CP. The exposome: from concept to utility. *Int J Epidemiol*. 2012;41:24–32.
- Niedzwiecki MM, Walker DI, Vermeulen R, et al. The exposome: molecules to populations. *Annu Rev Pharmacol Toxicol*. 2019; 59:107–127.
- Vermeulen R, Schymanski EL, Barabási A-L, Miller GW. The exposome and health: where chemistry meets biology. *Science*. 2020;367:392–396.
- Pristner M, Warth B. Drug-exposome interactions: the next frontier in precision medicine. *Trends Pharmacol Sci*. 2020; 41:994–1005.
- Nemkov T, Stefanoni D, Bordbar A, et al. Blood donor exposome and impact of common drugs on red blood cell metabolism. *JCI Insight*. 2021;6:146175.
- Rehfeld A, Frederiksen H, Rasmussen RH, et al. Human sperm cells can form paracetamol metabolite AM404 that directly interferes with sperm calcium signalling and function through a CatSper-dependent mechanism. *Hum Reprod*. 2022; 37:922–935.
- David A, Chaker J, Léger T, et al. Acetaminophen metabolism revisited using non-targeted analyses: implications for human biomonitoring. *Environ Int*. 2021;149:106388.
- Ozdemir V, Gunes A, Dahl M-L, et al. Could endogenous substrates of drug-metabolizing enzymes influence constitutive physiology and drug target responsiveness? *Pharmacogenomics* 2006;7:1199–1210.
- Pant A, Maiti TK, Mahajan D, Das B. Human gut microbiota and drug metabolism. *Microb Ecol*. 2023;86:97–111.
- Lai Y, Koelmel JP, Walker DI, et al. High-resolution mass spectrometry for human exposomics: expanding chemical space coverage. *Environ. Sci. Technol* 2024;58:12 784–12 822. <https://doi.org/10.1021/acs.est.4c01156>.
- Sun J, Fang R, Wang H, et al. A review of environmental metabolism disrupting chemicals and effect biomarkers associating disease risks: where exposomics meets metabolomics. *Environ Int*. 2022;158:106941.
- González-Domínguez R, Jáuregui O, Queipo-Ortuño MI, Andrés-Lacueva C. Characterization of the human exposome by a comprehensive and quantitative large-scale multianalyte metabolomics platform. *Anal Chem*. 2020;92:13 767–13 775.
- Bloszies CS, Fiehn O. Using untargeted metabolomics for detecting exposome compounds. *Curr Opin Toxicol*. 2018;8:87–92.
- Bessonneau V, Geron RR, Trowbridge J, et al. Gaussian graphical modeling of the serum exposome and metabolome reveals interactions between environmental chemicals and endogenous metabolites. *Sci Rep*. 2021;11:7607.
- Raghu G, Remy-Jardin M, Richeldi L, et al. Idiopathic pulmonary fibrosis (an update) and progressive pulmonary fibrosis in adults: an official ATS/ERS/JRS/ALAT clinical practice guideline. *Am J Respir Crit Care Med*. 2022;205:e18–e47.
- Heuckeroth S, Damiani T, Smirnov A, et al. Reproducible mass spectrometry data processing and compound annotation in MZmine 3. *Nat Protoc*. 2024;19:2597–2641.
- Schmid R, Heuckeroth S, Korf A, et al. Integrative analysis of multimodal mass spectrometry data in MZmine 3. *Nat Biotechnol*. 2023;41:447–449.
- Takeda H, Matsuzawa Y, Takeuchi M, et al. MS-DIAL 5 multimodal mass spectrometry data mining unveils lipidome complexities. *Nat Commun*. 2024;15:9903.
- Dührkop K, Fleischauer M, Ludwig M, et al. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nat Methods*. 2019;16:299–302.
- Hoffmann MA, Nothias L-F, Ludwig M, et al. High-confidence structural annotation of metabolites absent from spectral libraries. *Nat Biotechnol*. 2022;40:411–421.
- Ulrich EM, Sobus JR, Grulke CM, et al. EPA's non-targeted analysis collaborative trial (ENTACT): genesis, design, and initial findings. *Anal Bioanal Chem*. 2019;411:853–866.
- Sobus JR, Grossman JN, Chao A, et al. Using prepared mixtures of ToxCast chemicals to evaluate non-targeted analysis (NTA) method performance. *Anal Bioanal Chem*. 2019; 411:835–851.
- Singh RR, Chao A, Phillips KA, et al. Expanded coverage of non-targeted LC-HRMS using atmospheric pressure chemical ionization: a case study with ENTACT mixtures. *Anal Bioanal Chem*. 2020;412:4931–4939.
- Epling GA, Sibley MT, Chou TT, Kumar A. Photofragmentation of phototoxic dibenzocycloheptadiene antidepressants. *Photochem Photobiol*. 1988;47:491–495.
- Enever RP, Po ALW, Millard BJ, Shotton E. Decomposition of amitriptyline hydrochloride in aqueous solution: identification of decomposition products. *J Pharm Sci*. 1975;64:1497–1499.
- Roman R, Cohen EM, Christy ME, Hagerman WB. Stability of amitriptyline hydrochloride in a commercial aqueous solution. *J Pharm Sci*. 1979;68:1329–1330.
- Ardiana F, Suciati, Indrayanto G. Chapter seven - valsartan. In: Brittain HG, ed. *Profiles of Drug Substances, Excipients and Related Methodology*. Vol 40. Academic Press; 2015:431–493.
- Sampath A, Reddy AR, Yakambaram B, et al. Identification and characterization of potential impurities of valsartan, AT1 receptor antagonist. *J Pharm Biomed Anal*. 2009;50:405–412.
- Fang N, Yu S, Adams SH, Ronis MJ, Badger TM. Profiling of urinary bile acids in piglets by a combination of enzymatic deconjugation and targeted LC-MRM-MS. *J Lipid Res*. 2016; 57:1917–1933.
- Li Y, Gu C, Gruenhagen J, et al. An enzymatic deconjugation method for the analysis of small molecule active drugs on antibody-drug conjugates. *MAbs*. 2016;8:698–705.
- Corbel T, Perdu E, Gayraud V, et al. Conjugation and deconjugation reactions within the fetoplacental compartment in a sheep model: a key factor determining bisphenol a fetal exposure. *Drug Metab Dispos*. 2015;43:467–476.
- Wan Y, Ma Q, Hayat W, Liu Z, Dang Z. Ten bisphenol analogues in Chinese fresh dairy milk: high contribution ratios of conjugated form, importance of enzyme hydrolysis and risk evaluation. *Environ Sci Pollut Res Int*. 2023;30:88 049–88 059.
- Feng Y-L, Singh R, Chao C, Li Y. Diagnostic fragmentation pathways for identification of phthalate metabolites in nontargeted analysis studies. *J Am Soc Mass Spectrom*. 2022;33:981–995. <https://doi.org/10.1021/jasms.2c00052>.

© The Author(s) 2026. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact permissions@oup.com.

Exposome, 2026, 6, 1–9

<https://doi.org/10.1093/exposome/osag004>

Research Article