



HHS Public Access

Author manuscript

J Breath Res. Author manuscript; available in PMC 2018 December 06.

Published in final edited form as:

J Breath Res. ; 12(1): 016008. doi:10.1088/1752-7163/aa863c.

Metabolomic Profiles of Plasma, Exhaled Breath Condensate, and Saliva are Correlated with Potential for Air Toxics Detection

Chandresh Nanji Ladva¹, Rachel Golan², Roby Greenwald³, Tianwei Yu⁴, Stefanie Ebelt Sarnat¹, W. Dana Flanders^{4,5}, Karan Uppal⁶, Douglas I. Walker⁶, ViLinh Tran⁶, Donghai Liang¹, Dean P. Jones⁶, and Jeremy A. Sarnat¹

¹Department of Environmental Health, Rollins School of Public Health, Emory University 1518 Clifton Road, Atlanta, GA 30322

²Department of Public Health, Ben-Gurion University of the Negev Beer Sheva, 84105, Israel

³Department of Environmental Health, Georgia State University 1 Park Place, Suite 714, Atlanta, GA 30303

⁴Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University 1518 Clifton Road, Atlanta, GA 30322

⁵Department of Epidemiology, Rollins School of Public Health, Emory University 1518 Clifton Road, Atlanta, GA 30322

⁶Clinical Biomarkers Laboratory, Division of Pulmonary, Allergy, and Critical Care Medicine, School of Medicine, Emory University, 100 Woodruff Circle, Atlanta, GA 30322

Abstract

Introduction—Advances in the development of high-resolution metabolomics (HRM) have provided new opportunities for their use in characterizing exposures to environmental air pollutants and air pollution-related disease etiologies. Exposure assessment studies have considered blood, breath, and saliva as biological matrices suitable for measuring responses to air pollution exposures. The current study examines comparability among these three matrices using HRM and explored their potential for measuring mobile source air toxics.

Methods—Four participants provided saliva, exhaled breath concentrate (EBC), and plasma before and after a 2-hour road traffic exposure. Samples were analyzed on a Thermo Scientific QExactive MS system in positive electrospray ionization (ESI) mode and resolution of 70,000 FWHM with C18 chromatography. Data were processed using apLCMS and xMSanalyzer on the R statistical platform.

Results—The analysis yielded 7,110, 6,019, and 7,747 reproducible features in plasma, EBC, and saliva, respectively. Correlations were moderate-to-strong ($R = 0.41 - 0.80$) across all pairwise comparisons of feature intensity within profiles, with the strongest between EBC and saliva. The associations of mean intensities between matrix pairs were positive and significant, controlling for subject and sampling time effects. Six out of 20 features shared in all three matrices putatively matched a list of known mobile-source air toxics.

Conclusions—Plasma, saliva, and EBC have largely comparable metabolic profiles measurable through HRM. These matrices have the potential to be used in identification and measurement of exposures to mobile source air toxics, though further, targeted study is needed.

Keywords

exhaled breath condensate; saliva; plasma; high-resolution metabolomics; traffic pollutants; air toxics; exposure assessment; correlation

Introduction

Identifying causal agents associated with traffic-related health effects is made challenging given the complex chemical composition of traffic pollution and numerous biological pathways elicited upon exposure. Recently, some have proposed the study of the ‘exposome’ (1, 2), a theoretical framework which seeks to broaden the scope and capacity within environmental exposure science that aims to measure lifetime exposures to external chemicals and internal endpoints. Moreover, this framework is based upon approaches that link exposure assessment to mechanistic insights linking both exposure and response (3–5). A key aim of the burgeoning field of environmental metabolomics is the development and validation of analytical platforms to support exposome-related research.

High-resolution metabolomics (HRM) allows for the identification and quantitation of small molecules via high-throughput, high-yield analytical chemistry and bioinformatics. HRM may serve as an ideal tool for conducting exposure assessment of complex external and internal stressors. The use of HRM within traditional environmental epidemiologic settings has been limited, but growing. Previous studies have used metabolomic approaches to characterize high and low exposure to polycyclic aromatic hydrocarbons (PAHs) or metals, as well as differential perturbations in metabolism, between high and low exposure cohorts (6–10).

Traditional exposure assessment methods have experienced difficulty in identifying the specific compounds, compound classes, or non-chemical exposures of traffic-related pollution (TRP) that are primarily driving human health response. However, the toxic effects of exposure to airborne total particulate matter (PM) from primary TRP have been demonstrated (11–14). These toxic effects range from increased short-term morbidity in individuals with preexisting respiratory and cardiovascular disease to excess mortality rates over life-long exposure durations (15, 16). Given the extreme chemical and physical heterogeneity of TRP, however, it is not clear which components or mixtures are most responsible for driving the observed adverse responses (17–20), nor which specific biological pathways contribute to PM-specific disease etiology (20–25). Metabolomics may provide much needed insight into these gaps that traditional exposure assessments conducted to date have been unable to address.

As an initial step, understanding how metabolomic profiles vary by biological matrices in the same individuals may inform future study designs in this exposure science subfield. Non-invasive sampling of exhaled breath (EBC) (12, 26) and saliva (27) may serve as viable alternatives to more invasive blood sampling. Comparative metabolomic profiling across the

plasma, EBC, and saliva matrices may assist with the identification of the most relevant components of PM exposure for health. To date, studies of HRM and human health have mainly considered human blood and urine. Relatively few studies have been conducted on saliva or EBC using HRM, with even fewer examining metabolomic profiles across multiple biological matrices simultaneously (28–33). Another question of longitudinal changes of profiles remains as few studies have examined daily changes in metabolomic profiles for humans (34–39). The paucity in the literature HRM profile comparability for these three matrices is an opportunity to examine their utility for air pollution epidemiology.

To address these questions, a pilot commuter study was conducted in which plasma, EBC, and saliva were collected for a small set of participants. The overarching aims of this pilot study were to a) assess feasibility of conducting metabolomics analyses on these three matrices using a novel, powerful HRM platform; b) compare metabolomic profiles in plasma samples with those in the two, non-invasively collected matrices, saliva and EBC, and; c) assess whether the metabolomics platform we used, has the potential to identify features expected to be associated with exposure to TRP, specifically.

Methods

The current analysis used data collected as a pilot field study, nested within two larger panel studies of daily car commuters in Atlanta, GA (40–42). To conduct this pilot study, we recruited a convenience sample of four participants, who provided repeated samples of venous blood, EBC, and saliva over a multi-day period. These participants lived within 5-miles of the study site, and refrained from highway traffic exposures over the weekend preceding sample collection. To mirror the larger commuter study designs, the pilot sampling was conducted before and after a 2-hr commute during morning rush hour over prescribed, predominantly highway routes. This study was approved by the Emory Institutional Review Board as part of the larger Atlanta Commuters Exposures Study.

The pilot study protocol comprised four repeated measures of biological samples for four participants. Two sample collections were conducted on the first day (Day 0), a Monday; one pre-commute session at 7AM, and one immediately following the commute at 10AM. The subsequent sampling times occurred at 11AM on both +1 and +2 days. The design permitted careful control of exposure and follow-up with a small panel of participants with the expectation of producing thousands of measurements of chemicals within samples at every time point.

Biological Sampling

For each subject, sampling sessions included whole blood collection, passive saliva and exhaled breath condensate collection (41). Briefly, venous blood was collected in 3mL, ethylenediaminetetraacetic acid-treated (EDTA) tubes (BD Vacutainer) from either the left or right median cubital vein and centrifuged immediately for plasma retrieval. Saliva was collected passively, meaning it was not spit or expectorated but allowed to pool, into sterile 15mL conical centrifuge tubes (Falcon) held against the lips. Rtubes (Respiratory Research), disposable exhaled breath condensers, were used to collect roughly ~2mL of exhaled breath condensate. Biological samples from each matrix were aliquoted into microcentrifuge tubes

and stored at -80°C within 20 minutes of collection. Metabolomic analyses were conducted within four months of collection.

Metabolomics and Data Processing

Analytical chemistry was performed following procedures developed by Soltow et al. (2013) (43). For each sample, 65 μL were diluted into 130 μL of acetonitrile spiked with 14 stable isotopic standards (43). After a 30-minute equilibration on ice and protein precipitation, triplicate 10 μL aliquots were placed into separate autoinjector vials and run across two batches through a C18 reverse phase liquid chromatography coupled with high resolution mass spectrometry (HR-MS) on a QExactive (Thermo Scientific) (43). The analytical platform allows both positive and negative electrospray ionization (ESI), providing two, distinct runs per replicate. For this analysis, we focused only on positive phase features from positive ESI, since a greater number of features for this phase have been validated and annotated using spectral information available in the Human Metabolome Database (HMDB).

Raw mass spectra output from the analytical platform are processed to extract features, which represent individual ions detected from the mass spectrometer. Each feature is indicated by a unique combination of its mass-to-charge ratio (m/z) and retention time (RT). The relative concentration of a given feature within a replicate is indicated by the integrated ion intensity. Notably, we use the term ‘feature’ as distinct from the term “metabolite,” which would be the amalgamation of isotope and adduct peaks that correspond to a single, identifiable compound.

Feature extraction was performed using two, integrated algorithms developed by the Emory University Clinical Biomarkers Laboratory (clinicalmetabolomics.org), a research center integrating expertise in metabolomics and informatics. The two algorithms were the adaptive processing of liquid chromatography mass spectrometry (apLCMS, v. 6.0.1) and the xMSanalyzer (v. 2.0.6.1) packages, which were developed for use in R (v. 3.0.1) (44, 45). The apLCMS program was run in ‘hybrid’ mode, which allowed us to input information related to known metabolite masses and, thus, improve extraction efficiency of those potential metabolites from chromatograms (46). The list of known metabolites was comprised of monoisotopic masses from HMDB (v. 3.6) and the US Environmental Protection Agency’s Master List of Compounds Emitted by Mobile Sources. The xMSanalyzer package then modifies the resulting data from apLCMS for quality and reproducibility, including cut-offs for triplicate coefficients of variation, triplicate correlations, prevalence across biological samples (Supplemental Table 1). Complete feature extraction and data processing was done separately for each biological matrix in order to reduce noise introduced by mixing sample types. This resulted in three distinct feature tables for each sample type: plasma, EBC, and saliva.

The feature tables were pre-processed following Soltow et al. (2013) (43). Briefly, the ion intensities in each of the feature tables were \log_2 transformed after the addition of 1 to all intensities, including zeroes. The zeros resulting from \log_2 transformation were set to missing for downstream analysis. Triplicate intensity values for a feature within a biological sample were averaged, with missing values ignored, to produce the final feature tables for

statistical analysis. Metabolite levels of select features were generated using reference standardization and compared to previously published ranges (47).

Statistical Analysis

Statistical analyses were conducted in R (v.3.1.3). To evaluate the overall associations among the biological matrices, three key data subsets were sequentially examined: 1) All features detected in a biological matrix ('All'); 2) a subset of features measured with low triplicate variability in intensity (i.e., < 10% median coefficient of variation) ('<10% CoV'); and 3) a subset of features with masses that matched compounds in the USEPA Master List of Compounds Emitted by Mobile Sources ('EPA-matched'). Feature counts across matrices were compared, descriptively in Venn diagrams using functions provided in 'xMSanalyzer.'

For this analysis, a shared feature across matrix pairs was defined as a feature with a m/z within 10 ppm and a retention time within 300 seconds of a feature in another matrix. Only shared features between matrix pairs were used for subsequent analyses, similar to the interbiofluid analysis performed by Pinto et al. (2015) (32).

Scatterplots of 'All' shared features across matrix pairs provided a visual assessment of the joint distributions of mean feature intensities across subjects at the pre-exposure sampling. Correlations between matrix pairs of mean feature intensities at each sampling time were examined using Spearman's rank correlation analysis (r_s), with confidence intervals estimated using the Fisher transformation (48).

To examine the impacts of subjects and sampling time on the association of average intensities across matrix pairs, linear regression was used on the EPA-matched subset. Bivariate analyses were conducted to assess the association between subject or sampling time variables and feature intensity in a given matrix. The analysis focused on this subset because its members were features potentially representing traffic-related pollution. The model constructed examined associations of a feature intensity in one biological matrix (X_{ijk}) to the m/z and RT-matched feature intensity in another (Y_{ijk}). For subject, i , at sampling time, j , the association between features, k , can be expressed in the regression:

$$Y_{ijk} = \mu + \alpha_{ijk} X_{ijk} + \sum_{i=1}^{n-1} \beta_i Z_i + \sum_{j=1}^{t-1} \gamma_j W_j + \varepsilon_{ijk}, \quad (1)$$

$$\text{where } Z_i = \begin{cases} 0, & \text{subject 1} \\ 1, & \text{subject 2} \\ 2, & \text{subject 3} \\ 3, & \text{subject 4} \end{cases} \text{ and } W_j = \begin{cases} 0, & \text{pre-exposure} \\ 1, & \text{post-exposure} \\ 2, & +1 \text{ day} \\ 3, & +2 \text{ day} \end{cases} .$$

The variables, Z_i and W_j , were subject and sampling time identifiers. Here, μ is the mean intensity of the referent group (subject 1 at pre-exposure) and ε_{ijk} , the residual error, is normally distributed around 0. The association between the feature intensity in matrix X_{ijk} and feature intensity in Y_{ijk} is denoted by α . The coefficients β and γ represent the offsets to μ of the referent group. The model was run separately for each pair of matrices: 'EBC = plasma', 'saliva = plasma', and 'saliva = EBC'. Both the predictor and dependent variables

were log₂-transformed. A secondary analysis was conducted to explore the range of intensities in the sample population. The variability of specific features in the EPA-matched subset may have differential influence on observed associations and were examined using scatterplots (Supplementary Figure 1).

Annotation of shared, EPA-matched compounds

Feature annotation is the process in which a putative identification of a feature can be made through the mass querying online chemical databases like the HMDB. In this pilot study, the EPA-matched subset provided an opportunity to compare identifications with HMDB and ChemSpider, a comprehensive, chemical structure database curated from roughly 500 sources, to evaluate the feature annotation process and needs for use in environmental epidemiological applications.

Since full validation, using tandem MS approaches, was beyond the scope of this pilot analysis, we conducted a manual annotation of the metabolomics output, on those features from the 'EPA-matched' subset shared across all three matrices without multiple-matches (N=20) using HMDB and ChemSpider (49). First, the m/z ratios from the 20 features were used to calculate computed masses using only proton adducts and queried in HMDB. If a mass-match was successful, a comparison was made between the compound entry or entries in HMDB and compound identification provided in the USEPA Master List. The number of compound matches that overlapped, if any, was recorded. Secondly, entries from HMDB and the USEPA Master List were compared qualitatively. For example, if an entry in HMDB was detected, quantified, and validated as an endogenous metabolite, then this was determined to be a better identification than a dissimilar compound match with the Master List. However, if no match was found in HMDB, the putative identification using the Master List was used to search ChemSpider. The purpose of this step was to examine other potential sources for exposures. If the putative identification is for a compound shared in other potential exposures, such as consumer products or foods, it was no longer considered a putative TRP-related feature.

Results

We successfully retrieved 14 of 16 plasma, 15 of 16 EBC, and 16 of 16 saliva samples across a three-day sampling period across all participants. Participant metabolic characteristics, derived from the metabolomics analysis of plasma, at baseline, were comparable to or within a range of previously reported values for selected features (Table 1) (50). In total, raw feature extractions yielded unique 9,377–10,583 features in both plasma and saliva, and 7,688–8,565 features in EBC. After summarization and data cleaning, 7,110, 6,019, and 7,747 features were measured in plasma, EBC, and saliva, respectively. The median coefficients of variation across triplicates of samples within each biological matrix was less than 15% (Plasma: 13.5%, EBC: 14.3%, and Saliva: 13.2%). Saliva had the largest number of unique features (N=4,227) that were not present in either EBC or plasma; compared to the number of unique features in both plasma (N = 3,718) and EBC (N = 2,381). Most identified features in EBC (60.5%) were also found in the other matrices (Figure 1). Generally,

associations between metabolomic profiles from matrix pairs were consistent across repeated measures and across subjects.

Between-matrix scatterplots among 'All' features showed positive linear relationships between paired matrices with at least one subject having measurable concentrations in both matrices ($N = 1,862 - 2,039$) (Figure 2). When examining 'All' shared features identified at each of the sampling periods, correlations between EBC and plasma ranged from 0.51 to 0.58 across all four sampling time points (p -value < 0.0001 , for all pairwise correlations) (Figure 3). EBC-plasma correlations were comparable, albeit slightly stronger ($r_s = 0.56 - 0.65$; $p < 0.0001$), for those features measured with greater confidence (i.e., ' $< 10\%$ CoV') ($N = 674 - 861$). EBC-plasma correlations involving the subset of features within the 'EPA-matched' subset, ranged between 0.41 and 0.52 ($p < 0.0001$), across the various sampling time points. Saliva-plasma and saliva-EBC pairwise correlations were also positive and moderate-to-strong across all data subsets ($r_s = 0.51 - 0.69$; $p < 0.0001$ and $0.60 - 0.80$; $p < 0.0001$, respectively). The estimates of the correlation for EBC vs. saliva were stronger across all subsets ('All': $\rho=0.60-0.66$; ' $<10\%$ CoV': $0.72-0.80$; 'EPA-matched': $0.66-0.79$).

Generally, correlation patterns across all repeated measure sampling time points were consistent within each pairwise matrix comparison and data subset and did not vary by more than 13%. In the examination of the features matching the USEPA Master List ('EPA-matched'), the correlations varied from 16% in saliva vs. EBC to 23% in plasma vs. saliva. The greater variation in correlation estimates may be explained by the relatively small number of features being compared ($N = 36 - 54$).

Linear regression for the 'EPA-matched' subset demonstrated that the mean feature intensities did not vary by subject or sampling time when examining bivariate associations (Subject: $\beta = -0.55 - 0.08$, $p < 0.07 - 0.92$). This finding was also true across different sampling time points (Sampling Time: $\beta = -0.37 - 0.15$, $p < 0.09 - 0.86$). Multiple linear regression echoed this as well, with the exception of +1 Day sampling time in the EBC = plasma model. The feature intensities in EBC were significantly reduced ($\beta(\text{SE})$: $-0.66 (0.30)$) at the +1 Day time point relative to the pre-exposure measurement holding plasma intensity and subject constant. Strengths of associations between matrix pairs across all three models were positive and statistically significant ($p < 0.0001$) (Table 2). For example, a 10% increase in the geometric mean of feature intensities in plasma corresponded to a 6.0% increase in the geometric mean intensities in EBC. The same increase (10%) in plasma or EBC was associated with a corresponding 7.7% and 7.4% increase in saliva for models 2 and 3, respectively.

Highlighting the EPA-matched subset permitted an examination of feature annotation. Of features that were found in all three matrices ($N = 20$), only eight uniquely matched the curated Master List of Compounds Emitted from Mobile Sources and were not found in HMDB (Table 3). These features present new, putative identifications found in plasma, EBC, and saliva and are less likely features of endogenous metabolism. A ninth feature, matching the USEPA Master List as *n*-nitrosomorpholine, had two matches to drug metabolites in HMDB which are unlikely in our healthy population. Three of these features, putatively identified as *n*-nitrosodiethylamine, *n*-nitrosomorpholine, and C29H50 (i.e., 30-

normoretane), had high relative concentrations across all matrices. Further examination of the usage of the eight compounds, using ChemSpider as the source, suggested that three of them may be from exposure to consumer hygiene or food products.

Discussion

Comparison of feature information across multiple, biological matrices is a critical initial step towards evaluating HRM's usefulness as a sensitive, less invasive method for exposure assessment. A number of studies have previously compared metabolite features measured across matrices sampled from the same persons (30, 32, 33, 51–56). To our knowledge, however, the results from this pilot study are among the first to examine matrix comparability between plasma (relatively invasive) and saliva or EBC (non-invasive) using this measurement method. Our data showed moderate-to-strong associations among extracted features in each of the three matrices, with the strongest association between features in EBC and saliva. These results provide general support for the use of HRM to measure metabolic profiles in matrices collected non-invasively, specifically in EBC and saliva. The number of shared features as well as the between-matrix associations among these shared features suggest that either saliva or EBC may also be comparable to plasma.

Importantly, using our bioinformatics platform for metabolomics, we were able to extract a greater number of features than previous investigations (45). In addition, the use of xMSanalyzer and apLCMS yielded thousands of reproducible features in each matrix, with a majority of features shared across matrices. Although saliva had the greatest absolute number of individual features extracted, EBC shared a larger proportion of detected features than either plasma or saliva.

Generally, we observed strong associations between the mean feature intensities in plasma and corresponding intensities in both EBC and saliva. As indicated by the results from the linear models, these associations were not driven by individual subject effects nor sampling time points. The lack of significant subject effect is contrary to a previous repeated measures study of metabolomics (57). A possible alternative explanation may be that the sample size of our study does not reflect the true heterogeneity of metabolomes in larger populations. However, intensities of shared features between matrix pairs showed pronounced linearity in their relationships to each other in the scatterplots. Do et al. reported that 93.5% of total between-fluid correlations of plasma and saliva matched to the same biochemically-valid networks conferring confidence in annotation, but the average of the significant, partial correlations reported ($N = 14$ metabolites) was very weak between the same metabolites ($r = 0.06$) (2015) (30). Our results, with respect to plasma and saliva, show a moderate correlation among feature intensities between the matrices. The difference in interpretation of the association, aside from differing statistical methods, is likely attributable to the authors exclusion of unknown, unannotated features.

In our results, the relative strength of association was similar regardless of the data subset analyzed, suggesting that the cross-matrix correlation patterns were not merely driven by features that were more reliably measured or identifiable (Figure 3). Our regression modeling results using the 'EPA-matched' subset supported these findings, indicating

positive and significant associations of mean intensities between matrices that did not vary by subject or time, appreciably. These associations echoed the patterns of correlations between biological matrices in the 'All' and '<10% CoV' subsets. These results provide promising initial support for the use of our environmental metabolomics platform for the robust detection of compounds that may be associated with potential environmental sources.

Previous studies that have examined within- and between-person variability in metabolomic profiles across a variety of analytical platforms and matrices have found that the measured metabolites had much less within-person than between-person sources of variability (58–60). In contrast, our profiling indicated negligible differences when comparing the mean feature intensities between persons. This finding is consistent with Crews et al. (2009), and suggests that a large set of features from metabolomic profiles with greater intensities do not exhibit substantial daily variability within a multi-day sampling protocol (61). It is worth emphasizing that, although metabolomic profiles overall may exhibit low between-person variability in healthy persons, certain features do have greater between-person variability. In Crews et al., features with high integrated intensities ($>10^6$) clustered with CoVs less than 20% across biological samples in plasma. Our secondary analysis shows that some features within the EPA-matched subset have very little variation while others exhibit broad ranges of intensity values. Metabolomic analyses of low intensity features may be liable to greater inter-individual variability. This may be true, for example, in an assessment targeting features indicative of an exposure to an environmental contaminant, which would likely result in relatively low intensity values compared to endogenous metabolism (62).

The moderate correlations ($r_S \approx 0.5$ to 0.6) between EBC and saliva intensities from the 'All' subset were somewhat unexpected. de Laurentiis et al. reported dissimilar spectral profiles between EBC and saliva in samples of healthy subjects and patients with airway disease, suggesting the composition of measurable compounds between the two matrices are different when using nuclear magnetic resonance spectrometer (NMR) (29). Our results, using mass spectrometry, may support this as many more features were found in saliva than in EBC. Also, our correlation findings show that EBC samples may reflect a large portion of features in both saliva and plasma, and that the metabolomic profiles of EBC were more strongly correlated with saliva. For this protocol, we collected unstimulated, passive saliva, known to mostly be generated from the submandibular glands, but still comprising a mixture of saliva produced from the parotid and sublingual glands (63). The marginally greater number of features measured in saliva may reflect unique microbial signatures. Previous literature on the salivary metabolome have reported the presence of both endogenous and microbial signatures within this matrix (64). This interpretation should be viewed cautiously, however, since we were not able to discern metabolites generated from human or microbial metabolism for lack of specificity in putative identification. While definitive validation methods, such as tandem MS approaches (MS/MS), were beyond the scope of this pilot analysis, future targeted analyses including formal feature validation may help discern specific contribution from microbial as well as other sources.

It is possible that the strength of the saliva-EBC association may be explained, in part, by salivary contamination of EBC. When examining immunological markers, however, previous studies have found a negligible influence of salivary concentrations on EBC

concentrations (65). Despite this, our metabolomic analysis captures a much larger chemical space than the more limited immunological markers examined previously. In two recent studies of oral and lung microbiota, the taxonomic composition of oral washes and bronchoalveolar lavage were reported to be mostly similar, though still different in specific, lower-abundance taxa by sampling location (66, 67). These samples, at least partially, were collected from similar physiological locations we sampled through our passive saliva and exhaled breath. The microbiota occupying these similar locations may produce shared metabolites contributing to the observed correlations between EBC and saliva, relative to correlations between either of these matrices with plasma.

Another key finding from this analysis was the consistent strength of association of the four sampling time points over the span of three consecutive sampling days. This finding suggests observed associations between the metabolomic profiles in this pilot study were generally not time-varying over this duration. The striking consistency in the strengths of associations across time, also support our conclusion that these results were not spurious and driven by artifacts in the data analysis plan. However, the result should not be interpreted to mean that the profiles themselves are time-invariant. Ang et al. (2012) found both time-invariant and time-variant metabolites in pooled samples (57). Their study suggests that the selection of metabolites from a total metabolic profile may warrant special considerations depending on their respective temporal characteristics. Our study had repeated samples at nearly the same time of day from healthy individuals, which may explain the robustness of the associations over time.

Although correlations were typically strong among a broad suite of metabolites, without conducting intensive validation, it is unclear whether these were primarily abundant metabolites associated with common metabolic processes, which would not be expected to exhibit daily fluctuation or variability. In a secondary analysis, we examined the range of intensities across all subjects and sampling times of the EPA-matched features. Some metabolites had consistent measures across the study period, while others were far more variable. From our feature annotation, we found that the 20 metabolites identified from the EPA-matched subset were more likely to be features associated with endogenous metabolism, a food chemical, or consumer product exposure. However, six features, detected in all samples and matrices, were exclusively identified using the USEPA mobile source metabolite list. This may indicate an environmental origin and suggested, albeit non-validated, that the current metabolomics platform detected features that are potentially from exposure to TRP. This conclusion should be viewed cautiously as the utility of mass-matching of high-resolution masses is limited by the comprehensive capture of the annotation databases used. It may be possible that these features are truly unrelated to TRP exposures, and that HMDB has yet to identify these with as great reliability as other compounds. Our comparative approach for identifying metabolites of interest does not replace formal validation procedures, such as tandem MS; it does provide a relatively quick and inexpensive means of leveraging available information about the identity and sources of several thousand features.

One limitation of the analysis is that the matches were identified strictly through a search of proton adducts to a parent compound ion. For example, our putative annotation of tin as a

feature in all three matrices is an unlikely candidate using our LC ESI+ chemical analysis. As our samples were diluted in acetonitrile, there is a high likelihood of acetonitrile, sodium, and potassium adducts, which complicate database annotation by providing many more matches to a single mass-to-charge ratio. Until tools are available to utilize data structure, correlations, and expectations of presence in various biological matrices, simple mass matching will continue to be a challenge for environmental metabolomics when environmental data is still poorly represented in available public datasets.

There are numerous caveats in interpreting the findings from this analysis. Most notably is the sample size, which was conducted on a non-random convenience sample of four subjects. Similar sample sizes, however, are not unprecedented in metabolomics analyses (61). To offset noise and potential confounding from uncontrolled factors associated with field-based observational designs involving multiple subjects, we elected to enroll four subjects, who were carefully monitored and followed during their participation in the study protocol. This degree of monitoring would not have been possible given a larger number of participants in a pilot study. Additionally, the purpose of the pilot study was to compare the metabolomic profiling of three distinct matrices, which requires, primarily, a sufficient number of features measured in the participants. The use of apLCMS and xMSAnalyzer produced measurements of 6,000 – 8,000 features from HRM, providing more than adequate power to detect differences among the matrices. Despite this, although the model results showed that the strength of the associations did not vary by subject, given the limited number of observations, it is not possible to generalize these findings to a broader population.

Relatedly, although not a traditional limitation, there may be a concern of noise retained from untargeted metabolomics approaches. Standard, targeted metabolomics practices often subtract signals from laboratory blanks to remove background signals. This approach is a gold standard when the targets of the analysis, the research question, are known and quantifiable. In untargeted or hybrid metabolomics approaches, such as those used in this study, a blank correction is inappropriate, especially when measuring low intensity features representing ubiquitous environmental exposures. This is compensated for through the use of advanced bioinformatics methods, quality controls, isotopic standards, and a high-resolution metabolomics platform in tandem (3, 68). Thus, noise is dramatically reduced in the final feature tables used in the analyses here without the explicit and overly conservative removal of signals.

To further address the issues of sample size and generalizability to other study populations, we repeated a similar analysis, using an identical sample collection and metabolic analysis protocol, examining linear correlations between plasma and saliva metabolites in a cohort of 54 college-aged students. EBC was not collected as part of this specific protocol. In this analysis (results not shown), we found comparable correlations in the subset of highly-reproducible features (<10% CoV) ($r_S = 0.59$ vs. 0.64). While our results are still limited in generalizability to the participants in the study, the general concordance of the current pilot results to those shown in this larger cohort provides some external validation that the pilot study findings were not spurious.

Future work should contribute to the capture of spectral data of those compounds in the Master List of Mobile Source Air Toxics to public annotation databases. The degree of curation in resources such as the Human Metabolome Database provide an invaluable resource for metabolomic analysis, but still lack in compounds associated with or direct metabolites from environmental exposures. Nevertheless, the access to monoisotopic masses of the mobile source air toxics list enabled putative matching using a curated list of compounds known to be emitted from mobile sources. The detection of EPA-matched features in all matrices is encouraging for air pollution exposure science, but cannot be determined with computation alone. Only after definitive identification, through targeted MS/MS analysis, of compounds associated with traffic pollution exposure in each matrix can a determination be made on which of plasma, EBC, or saliva is best suited for exposure assessment while minimizing participant burden.

Conclusions

Collectively, we believe this pilot analysis efficiently demonstrated matrix comparability between plasma, saliva, and EBC and the putative detection of several features that may be associated with TRP. Our methodology was successful in retrieving thousands of features and was able to characterize robust, consistent metabolomic profiles across plasma, saliva and exhaled breath condensate. As metabolomic databases become better curated for environmental health applications, analyses in studies such as this will presumably improve. Today, there is much ambiguity in putative matching by accurate mass with respect to mobile source air toxics. Detection and validation of these compounds in all three matrices will help towards understanding exposure and relevant biological pathways associated with traffic pollution toxicity.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to acknowledge the support provided by J Kancherla and R Judson from USEPA with procuring monoisotopic mass from the USEPA Master List of Compounds Emitted by Mobile Sources. The presented research was supported by a Clean Air Research Center grant to Emory University and the Georgia Institute of Technology from the US Environmental Protection Agency (USEPA, RD834799) and a pilot award from the Human Exposome Research Center (HERCULES) at Emory University. Shared instrumentation for metabolomics analysis was support through the National Institutes of Health (NIH, OD018006). C Ladva acknowledges additional pre-doctoral support from the Burroughs Wellcome Fund through the Molecules the Mankind Doctoral Pathway at Emory University. R Golan gratefully acknowledges support by a post-doctoral fellowship from the Environment and Health Fund, Jerusalem, Israel. The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the USEPA. Further, USEPA does not endorse the purchase of any commercial products or services mentioned in the publication.

References

1. Miller GW, Jones DP. The Nature of Nurture: Refining the Definition of the Exposome. *Toxicological Sciences*. 2014; 137(1):1–+. [PubMed: 24213143]

2. Wild CP. Complementing the genome with an “exposome”: The outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiology Biomarkers & Prevention*. 2005; 14(8):1847–50.
3. Walker, DI., Go, Y-M., Liu, K., Pennell, KD., Jones, DP. *Metabolic Phenotyping in Personalized and Public Healthcare*. Boston: Academic Press; 2016. Chapter 7 – Population Screening for Biological and Environmental Properties of the Human Metabolic Phenotype: Implications for Personalized Medicine; p. 167-211.
4. Rappaport SM. Implications of the exposome for exposure science. *Journal of Exposure Science and Environmental Epidemiology*. 2011; 21(1):5–9. [PubMed: 21081972]
5. Park YH, Lee K, Soltow QA, Strobel FH, Brigham KL, Parker RE, et al. High-performance metabolic profiling of plasma from seven mammalian species for simultaneous environmental chemical surveillance and bioeffect monitoring. *Toxicology*. 2012; 295(1–3):47–55. [PubMed: 22387982]
6. Ellis JK, Athersuch TJ, Thomas LDK, Teichert F, Perez-Trujillo M, Svendsen C, et al. Metabolic profiling detects early effects of environmental and lifestyle exposure to cadmium in a human population. *Bmc Medicine*. 2012; 10
7. Wei Y, Wang Z, Chang CY, Fan T, Su L, Chen F, et al. Global metabolomic profiling reveals an association of metal fume exposure and plasma unsaturated fatty acids. *PLoS One*. 2013; 8(10):e77413. [PubMed: 24143234]
8. Wang KC, Kuo CH, Tian TF, Tsai MH, Chiung YM, Hsieh CM, et al. Metabolomic characterization of laborers exposed to welding fumes. *Chem Res Toxicol*. 2012; 25(3):676–86. [PubMed: 22292500]
9. Romano A, Aiani MR, Brambilla M, Brambilla P, Carcano M, Galbiati E, et al. Metabolomic profiles of exhaled breath condensate of 39 nickel exposed workers. *G Ital Med Lav Ergon*. 2012; 34(3 Suppl):682–6. [PubMed: 23405751]
10. Wang Z, Zheng Y, Zhao B, Zhang Y, Liu Z, Xu J, et al. Human metabolic responses to chronic environmental polycyclic aromatic hydrocarbon exposure by a metabolomic approach. *J Proteome Res*. 2015; 14(6):2583–93. [PubMed: 25990285]
11. Paffett ML, Zychowski KE, Sheppard L, Robertson S, Weaver JM, Lucas SN, et al. Ozone Inhalation Impairs Coronary Artery Dilatation via Intracellular Oxidative Stress: Evidence for Serum-Borne Factors as Drivers of Systemic Toxicity. *Toxicological Sciences*. 2015; 146(2):244–53. [PubMed: 25962394]
12. Benor S, Alcalay Y, Domany KA, Gut G, Soferman R, Kivity S, et al. Ultrafine particle content in exhaled breath condensate in airways of asthmatic children. *J Breath Res*. 2015; 9(2):026001. [PubMed: 25830607]
13. Baxter LK, Duvall RM, Sacks J. Examining the effects of air pollution composition on within region differences in PM2.5 mortality risk estimates. *Journal of Exposure Science and Environmental Epidemiology*. 2013; 23(5):457–65. [PubMed: 23250195]
14. Tainio M, Tuomisto JT, Pekkanen J, Karvosenoja N, Kupiainen K, Porvari P, et al. Uncertainty in health risks due to anthropogenic primary fine particulate matter from different source types in Finland. *Atmospheric Environment*. 2010; 44(17):2125–32.
15. Hoek G, Krishnan RM, Beelen R, Peters A, Ostro B, Brunekreef B, et al. Long-term air pollution exposure and cardio-respiratory mortality: a review. *Environmental Health*. 2013; 12
16. Tetreault LF, Perron S, Smargiassi A. Cardiovascular health, traffic-related air pollution and noise: are associations mutually confounded? A systematic review. *International Journal of Public Health*. 2013; 58(5):649–66. [PubMed: 23887610]
17. Eeftens M, Hoek G, Gruzieva O, Moelter A, Agius R, Beelen R, et al. Elemental Composition of Particulate Matter and the Association with Lung Function. *Epidemiology*. 2014; 25(5):648–57. [PubMed: 25061921]
18. Fang T, Verma V, Guo H, King LE, Edgerton ES, Weber RJ. A semi-automated system for quantifying the oxidative potential of ambient particles in aqueous extracts using the dithiothreitol (DTT) assay: results from the Southeastern Center for Air Pollution and Epidemiology (SCAPE). *Atmospheric Measurement Techniques*. 2015; 8(1):471–82.

19. Pardo M, Shafer MM, Rudich A, Schauer JJ, Rudich Y. Single Exposure to near Roadway Particulate Matter Leads to Confined Inflammatory and Defense Responses: Possible Role of Metals. *Environmental Science & Technology*. 2015; 49(14):8777–85. [PubMed: 26121492]
20. Recio A, Linares C, Ramon Banegas J, Diaz J. Road traffic noise effects on cardiovascular, respiratory, and metabolic health: An integrative model of biological mechanisms. *Environmental Research*. 2016; 146:359–70. [PubMed: 26803214]
21. Schisler JC, Ronnebaum SM, Madden M, Channell M, Campen M, Willis MS. Endothelial inflammatory transcriptional responses to an altered plasma exposome following inhalation of diesel emissions. *Inhalation Toxicology*. 2015; 27(5):272–80. [PubMed: 25942053]
22. Uski O, Jalava PI, Happonen MS, Leskinen J, Sippula O, Tissari J, et al. Different toxic mechanisms are activated by emission PM depending on combustion efficiency. *Atmospheric Environment*. 2014; 89:623–32.
23. Vattanasit U, Navasumrit P, Khadka MB, Kanitwithayanun J, Promvijit J, Autrup H, et al. Oxidative DNA damage and inflammatory responses in cultured human cells and in humans exposed to traffic-related particles. *International Journal of Hygiene and Environmental Health*. 2014; 217(1):23–33. [PubMed: 23567252]
24. Vella RE, Pillon NJ, Zarrouki B, Croze ML, Koppe L, Guichardant M, et al. Ozone Exposure Triggers Insulin Resistance Through Muscle c-Jun N-Terminal Kinase Activation. *Diabetes*. 2015; 64(3):1011–24. [PubMed: 25277399]
25. Xia M, Viera-Hutchins L, Garcia-Lloret M, Rivas MN, Wise P, McGhee SA, et al. Vehicular exhaust particles promote allergic airway inflammation through an aryl hydrocarbon receptor-notch signaling cascade. *Journal of Allergy and Clinical Immunology*. 2015; 136(2):441–53. [PubMed: 25825216]
26. Romieu I, Barraza-Villarreal A, Escamilla-Nunez C, Almstrand AC, Diaz-Sanchez D, Sly PD, et al. Exhaled breath malondialdehyde as a marker of effect of exposure to air pollution in children with asthma. *J Allergy Clin Immunol*. 2008; 121(4):903–9 e6. [PubMed: 18234317]
27. Ising H, Lange-Asschenfeldt H, Moriske HJ, Born J, Eilts M. Low frequency noise and stress: bronchitis and cortisol in children exposed chronically to traffic noise and exhaust fumes. *Noise Health*. 2004; 6(23):21–8.
28. Breier M, Wahl S, Prehn C, Fugmann M, Ferrari U, Weise M, et al. Targeted metabolomics identifies reliable and stable metabolites in human serum and plasma samples. *PLoS One*. 2014; 9(2):e89728. [PubMed: 24586991]
29. de Laurentiis G, Paris D, Melck D, Maniscalco M, Marsico S, Corso G, et al. Metabonomic analysis of exhaled breath condensate in adults by nuclear magnetic resonance spectroscopy. *Eur Respir J*. 2008; 32(5):1175–83. [PubMed: 18653649]
30. Do KT, Kastenmuller G, Mook-Kanamori DO, Yousri NA, Theis FJ, Suhre K, et al. Network-based approach for analyzing intra- and interfluid metabolite associations in human blood, urine, and saliva. *J Proteome Res*. 2015; 14(2):1183–94. [PubMed: 25434815]
31. Maher AD, Cysique LA, Brew BJ, Rae CD. Statistical integration of 1H NMR and MRS data from different biofluids and tissues enhances recovery of biological information from individuals with HIV-1 infection. *J Proteome Res*. 2011; 10(4):1737–45. [PubMed: 21244037]
32. Pinto J, Barros AS, Domingues MR, Goodfellow BJ, Galhano E, Pita C, et al. Following healthy pregnancy by NMR metabolomics of plasma and correlation to urine. *J Proteome Res*. 2015; 14(2):1263–74. [PubMed: 25529102]
33. Taylor SL, Ruhaak LR, Kelly K, Weiss RH, Kim K. Effects of imputation on correlation: implications for analysis of mass spectrometry data from multiple biological matrices. *Brief Bioinform*. 2016
34. Chen R, Mias GI, Li-Pook-Than J, Jiang L, Lam HYK, Chen R, et al. Personal Omics Profiling Reveals Dynamic Molecular and Medical Phenotypes. *Cell*. 2012; 148(6):1293–307. [PubMed: 22424236]
35. Fidalgo TKS, Freitas-Fernandes LB, Almeida FCL, Valente AP, Souza IPR. Longitudinal evaluation of salivary profile from children with dental caries before and after treatment. *Metabolomics*. 2015; 11(3):583–93.

36. Lindsay KL, Hellmuth C, Uhl O, Buss C, Wadhwa PD, Koletzko B, et al. Longitudinal Metabolomic Profiling of Amino Acids and Lipids across Healthy Pregnancy. *Plos One*. 2015; 10(12)
37. Paige LA, Mitchell MW, Krishnan KRR, Kaddurah-Daouk R, Steffens DC. A preliminary metabolomic analysis of older adults with and without depression. *International Journal of Geriatric Psychiatry*. 2007; 22(5):418–23. [PubMed: 17048218]
38. Xu T, Holzapfel C, Dong X, Bader E, Yu Z, Prehn C, et al. Effects of smoking and smoking cessation on human serum metabolite profile: results from the KORA cohort study. *Bmc Medicine*. 2013; 11
39. Yousri NA, Kastenmueller G, Gieger C, Shin S-Y, Erte I, Menni C, et al. Long term conservation of human metabolic phenotypes and link to heritability. *Metabolomics*. 2014; 10(5):1005–17. [PubMed: 25177233]
40. Mirabelli MC, Golan R, Greenwald R, Raysoni AU, Holguin F, Kewada P, et al. Modification of Traffic-related Respiratory Response by Asthma Control in a Population of Car Commuters. *Epidemiology*. 2015; 26(4):546–55. [PubMed: 25901844]
41. Sarnat JA, Golan R, Greenwald R, Raysoni AU, Kewada P, Winquist A, et al. Exposure to traffic pollution, acute inflammation and autonomic response in a panel of car commuters. *Environ Res*. 2014; 133:66–76. [PubMed: 24906070]
42. Golan, R.Ladva, CN.Greenwald, R.Raysoni, AU.Liang, D.Winquist, A., et al., editors. Associations Between In-Vehicle Pollutant Mixtures And Acute Respiratory Response In Panels Of Car Commuters. 2015 Conference of the International Society of Environmental Epidemiology (ISEE); 2015; Las Vegas, NV. 2015
43. Soltow QA, Strobel FH, Mansfield KG, Wachtman L, Park Y, Jones DP. High-performance metabolic profiling with dual chromatography-Fourier-transform mass spectrometry (DC-FTMS) for study of the exposome. *Metabolomics*. 2013; 9(1 Suppl):S132–S43. [PubMed: 26229523]
44. Yu T, Park Y, Johnson JM, Jones DP. apLCMS–adaptive processing of high-resolution LC/MS data. *Bioinformatics*. 2009; 25(15):1930–6. [PubMed: 19414529]
45. Uppal K, Soltow QA, Strobel FH, Pittard WS, Gernert KM, Yu T, et al. xMSAnalyzer: automated pipeline for improved feature detection and downstream analysis of large-scale, non-targeted metabolomics data. *BMC Bioinformatics*. 2013; 14:15. [PubMed: 23323971]
46. Yu T, Park Y, Li S, Jones DP. Hybrid feature detection and information accumulation using high-resolution LC-MS metabolomics data. *J Proteome Res*. 2013; 12(3):1419–27. [PubMed: 23362826]
47. Go Y-M, Walker DI, Liang Y, Uppal K, Soltow QA, ViLinh T, et al. Reference Standardization for Mass Spectrometry and High-resolution Metabolomics Applications to Exposome Research. *Toxicological Sciences*. 2015; 148(2):531–43. [PubMed: 26358001]
48. Caruso JC, Cliff N. Empirical size, coverage, and power of confidence intervals for Spearman's rho. *Educational and Psychological Measurement*. 1997; 57(4):637–54.
49. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, et al. HMDB 3.0–The Human Metabolome Database in 2013. *Nucleic Acids Res*. 2013; 41:D801–7. (Database issue). [PubMed: 23161693]
50. Accardi CJ, Walker DI, Uppal K, Quyyumi AA, Rohrbeck P, Pennell KD, et al. High-Resolution Metabolomics for Nutrition and Health Assessment of Armed Forces Personnel. *J Occup Environ Med*. 2016; 58(8 Suppl 1):S80–8. [PubMed: 27501108]
51. Stiegel MA, Pleil JD, Sobus JR, Morgan MK, Madden MC. Analysis of inflammatory cytokines in human blood, breath condensate, and urine using a multiplex immunoassay platform. *Biomarkers*. 2015; 20(1):35–46. [PubMed: 25495125]
52. Yousri NA, Mook-Kanamori DO, Selim MM, Takiddin AH, Al-Homsi H, Al-Mahmoud KA, et al. A systems view of type 2 diabetes-associated metabolic perturbations in saliva, blood and urine at different timescales of glycaemic control. *Diabetologia*. 2015; 58(8):1855–67. [PubMed: 26049400]
53. Martel J, Wu CY, Hung CY, Wong TY, Cheng AJ, Cheng ML, et al. Fatty acids and small organic compounds bind to mineralo-organic nanoparticles derived from human body fluids as revealed by metabolomic analysis. *Nanoscale*. 2016

54. Barnes VM, Kennedy AD, Panagakos F, Devizio W, Trivedi HM, Jonsson T, et al. Global metabolomic analysis of human saliva and plasma from healthy and diabetic subjects, with and without periodontal disease. *PLoS One*. 2014; 9(8):e105181. [PubMed: 25133529]
55. Peralbo-Molina A, Calderon-Santiago M, Priego-Capote F, Jurado-Gamez B, Luque de Castro MD. Development of a method for metabolomic analysis of human exhaled breath condensate by gas chromatography-mass spectrometry in high resolution mode. *Anal Chim Acta*. 2015; 887:118–26. [PubMed: 26320793]
56. Mook-Kanamori DO, Selim MM, Takiddin AH, Al-Homsi H, Al-Mahmoud KA, Al-Obaidli A, et al. 1,5-Anhydroglucitol in saliva is a noninvasive marker of short-term glycemic control. *J Clin Endocrinol Metab*. 2014; 99(3):E479–83. [PubMed: 24423354]
57. Ang JE, Revell V, Mann A, Mantele S, Otway DT, Johnston JD, et al. Identification of human plasma metabolites exhibiting time-of-day variation using an untargeted liquid chromatography-mass spectrometry metabolomic approach. *Chronobiol Int*. 2012; 29(7):868–81. [PubMed: 22823870]
58. Gika HG, Theodoridis GA, Earll M, Wilson ID. A QC approach to the determination of day-to-day reproducibility and robustness of LC-MS methods for global metabolite profiling in metabonomics/metabolomics. *Bioanalysis*. 2012; 4(18):2239–47. [PubMed: 23046266]
59. Kim K, Mall C, Taylor SL, Hitchcock S, Zhang C, Wettersten HI, et al. Mealtime, temporal, and daily variability of the human urinary and plasma metabolomes in a tightly controlled environment. *PLoS One*. 2014; 9(1):e86223. [PubMed: 24475090]
60. Zivkovic AM, Wiest MM, Nguyen U, Nording ML, Watkins SM, German JB. Assessing individual metabolic responsiveness to a lipid challenge using a targeted metabolomic approach. *Metabolomics*. 2009; 5(2):209–18. [PubMed: 20442800]
61. Crews B, Wikoff WR, Patti GJ, Woo HK, Kalisiak E, Heideker J, et al. Variability analysis of human plasma and cerebral spinal fluid reveals statistical significance of changes in mass spectrometry-based metabolomics data. *Anal Chem*. 2009; 81(20):8538–44. [PubMed: 19764780]
62. Rappaport SM, Barupal DK, Wishart D, Vineis P, Scalbert A. The Blood Exposome and Its Role in Discovering Causes of Disease. *Environmental Health Perspectives*. 2014; 122(8):769–74. [PubMed: 24659601]
63. Lukacs JR, Largaespada LL. Explaining sex differences in dental caries prevalence: Saliva, hormones, and “life-history” etiologies. *American Journal of Human Biology*. 2006; 18(4):540–55. [PubMed: 16788889]
64. Takeda I, Stretch C, Barnaby P, Bhatnager K, Rankin K, Fu H, et al. Understanding the human salivary metabolome. *Nmr in Biomedicine*. 2009; 22(6):577–84. [PubMed: 19259987]
65. Tufvesson E, van Weele LJ, Ekedahl H, Bjermer L. Levels of cysteinyl-leukotrienes in exhaled breath condensate are not due to saliva contamination. *Clinical Respiratory Journal*. 2010; 4(2): 83–8. [PubMed: 20565481]
66. Beck JM, Schloss PD, Venkataraman A, Twigg H III, Jablonski KA, Bushman FD, et al. Multicenter Comparison of Lung and Oral Microbiomes of HIV-infected and HIV-uninfected Individuals. *American Journal of Respiratory and Critical Care Medicine*. 2015; 192(11):1335–44. [PubMed: 26247840]
67. Hare KM, Grimwood K, Leach AJ, Smith-Vaughan H, Torzillo PJ, Morris PS, et al. Respiratory Bacterial Pathogens in the Nasopharynx and Lower Airways of Australian Indigenous Children with Bronchiectasis. *Journal of Pediatrics*. 2010; 157(6):1001–5. [PubMed: 20656297]
68. Uppal K, Walker DI, Liu K, Li S, Go YM, Jones DP. Computational Metabolomics: A Framework for the Million Metabolome. *Chem Res Toxicol*. 2016; 29(12):1956–75. [PubMed: 27629808]

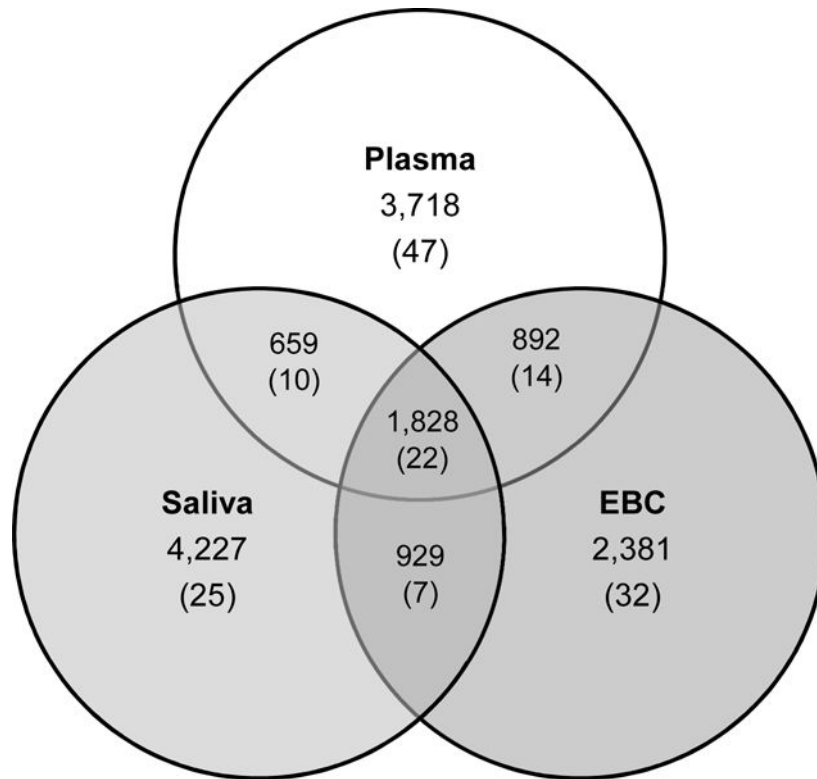


Figure 1. Venn Diagram of Feature Counts. Counts are from features detected at any sampling time. Features in one matrix were matched to features in another if they were within 10ppm of their respective m/z and within 30s of retention time. The numbers in the parentheses indicate the counts of features from the USEPA Master List of Mobile Source Air Toxics.

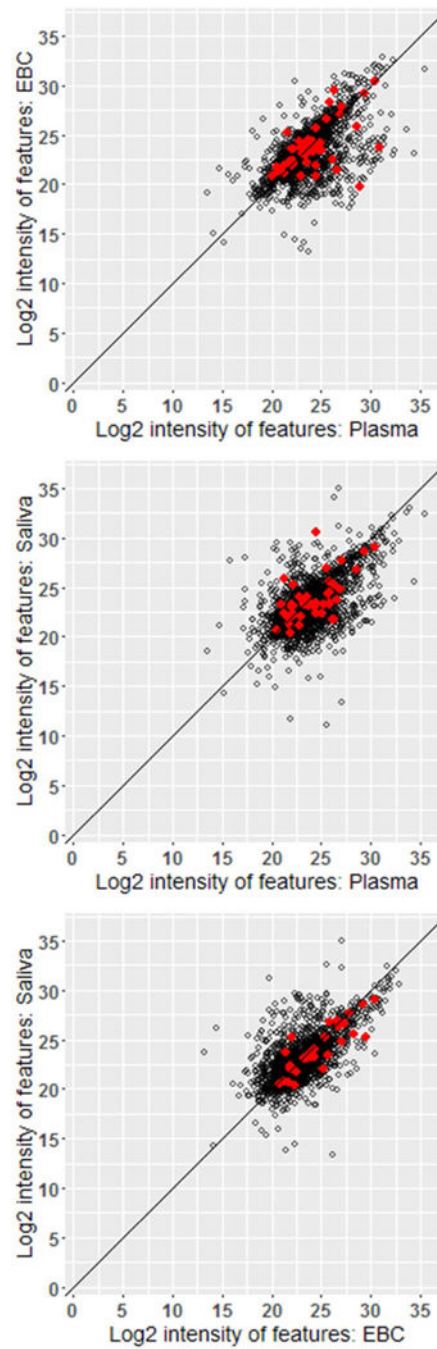


Figure 2.

Scatterplots of Mean Intensities of Shared Features across Subjects between Matrix Pairs at the Pre-Exposure Sampling Time. Red dots indicate those features that also match the Master List of Mobile Source Air Toxics. Sample sizes (from top to bottom) range from 1,862, 2,039, and 1,875 features for Plasma vs. EBC, Plasma vs. Saliva, and EBC vs. Saliva, respectively.

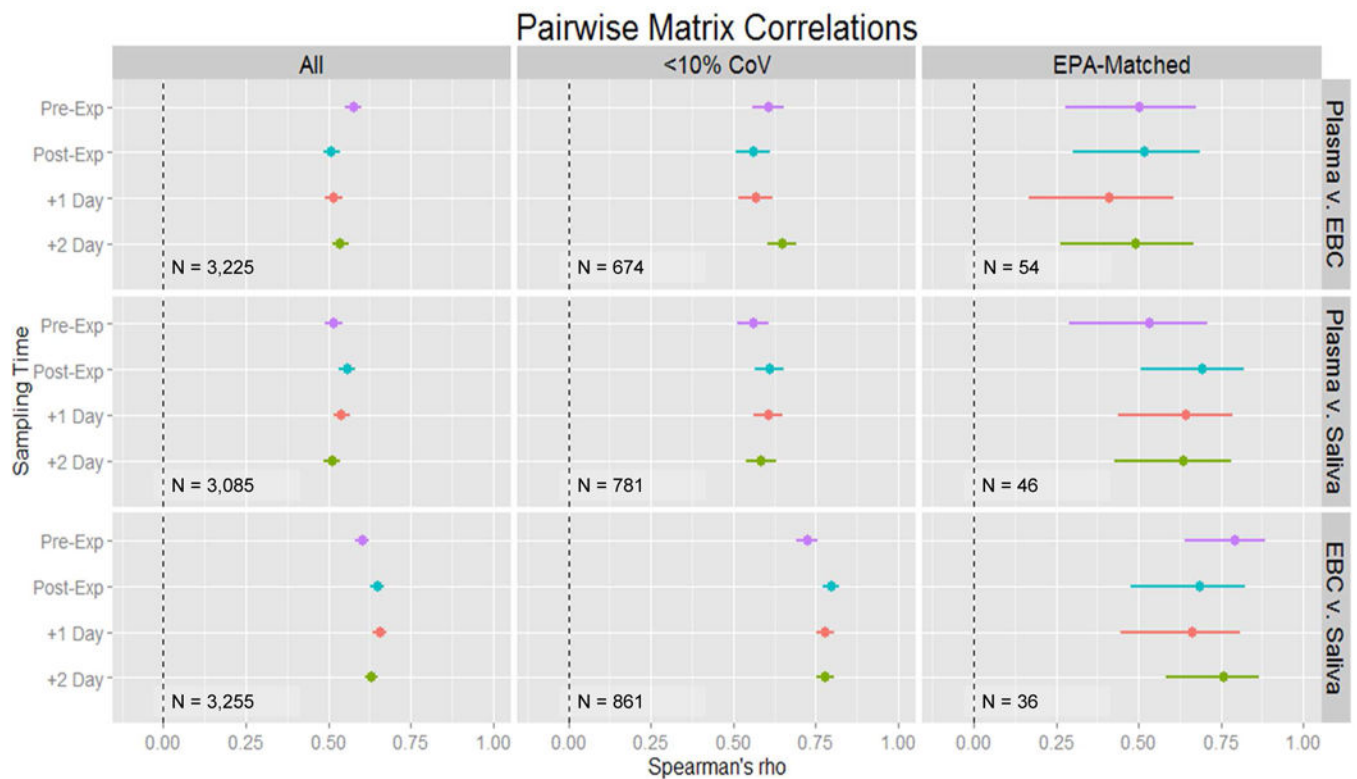


Figure 3.

Estimates of the Spearman's Rank Correlations between Matrices Stratified by Time. Three sets of the data were considered in turn. The All set (far left) is the complete list of shared, reproducible features. The <10% CoV set (middle) is a subset of All with a very high standard for chemical measurement. The EPA-matched (far right) is the subset of features that match the Master List of Mobile Source Air Toxics. All feature matching here was within m/z of 10ppm and RT of 300s.

Table 1

Participant Characteristics and Metabolite Levels* from Plasma Metabolomics

	Pre-Exp	Comb. Post-Exp	Accardi et al.	HMDB
Female (%)	25	–	–	–
Age (Range)	29 – 46	–	–	–
<i>Health Indicators</i>				
Glucose (mM)	2.6 ± 0.02	2.5 ± 0.06	6.3 ± 3	3.9 – 6.1
Creatinine	76 ± 0.8	76 ± 0.8	85 ± 15	56 – 108.8
<i>Amino Acids</i>				
Phenylalanine	81 ± 0.6	78 ± 0.6	142 ± 22	48 – 189
Methionine	30 ± 0.5	30 ± 0.4	21 ± 10	22 – 46
<i>Exogenous Chemical</i>				
Benzoic acid	18 ± 0.3	19 ± 0.4	4 ± 0.6	8.1 – 145.3

* Metabolite matching from reference standardization. Values are presented as mean ± SD or ranges. All metabolite levels are in μM concentrations, unless noted otherwise.

Table 2

Regression Estimates of Associations in the EPA-Matched Subset

X v. Y	Plasma		Plasma v EBC		EBC		Plasma v Saliva		Saliva		EBC v Saliva	
	Bivariate		Full		Bivariate		Full		Bivariate		Full	
N			413				318				240	
Coefficients (SE)												
α	-		0.60 (0.04)**		-		0.77 (0.04)**		-		0.74 (0.04)**	
β	-0.46 - 0.02 (0.18 - 0.33)		-0.28 - 0.15 (0.25 - 0.58)*		-0.52 - 0.08 (0.20 - 0.33)		-0.08 - 0.42 (0.24 - 0.25)		-0.55 - 0.13 (0.23 - 0.31)		-0.08 - 0.06 (0.27 - 0.27)	
γ	0.10 - 0.15 (0.20 - 0.20)		-0.66 - -0.43 (0.29 - 0.30)		-0.37 - -0.04 (0.23 - 0.23)		-0.53 - -0.34 (0.28 - 0.28)		-0.22 - -0.15 (0.25 - 0.25)		-0.19 - 0.22 (0.31 - 0.32)	

SE = standard error;

* = $p < 0.05$;** = $p < 0.001$; α is the coefficient for X_{ij} ; Factor coefficients, β and γ , are ranges of coefficients and SE across levels for subject and sampling time, respectively; Significance for range indicates at least one level was statistically significant

Table 3
Putative Annotation of EPA-Matched Features Shared in Plasma, EBC, and Saliva

#	Plasma		EBC		Saliva		RT	m/z	#HMDB/#EPA	Overlap	Putative EPA Compound Name/Formula [CAS RN]	Poss. TRP?
	m/z	RT	m/z	RT	m/z	RT						
1	101.0714	63.7508	101.0715	63.6356	101.0715	55.4801		2/1	100%	n-nitrosopyrrolidine [930-55-2]	N	
2	103.0871	40.9973	103.0871	65.5280	103.0871	41.0064		0/1	0%	n-nitrosodiethylamine [55-18-5]	Y	
3	111.1174	566.4460	111.1174	567.9528	111.1174	551.2129		5/4	0%	C8H14/C9H20 (i.e., octyne)	N	
4	115.0870	64.3534	115.0870	85.3724	115.0870	84.4289		2/1	0%	n-nitrosopiperidine [100-75-4]	N	
5	117.0663	50.9270	117.0662	58.1568	117.0662	59.8724		2/1	0%	n-nitrosomorpholine [59-89-2]	Y	
6	119.0495	92.0831	119.0495	74.8689	119.0495	92.1704		4/1	100%	2,3-benzofuran [271-89-6]	N	
7	121.0640	44.0034	121.0643	63.3738	121.0644	69.9346		8/7	100%	C8H8O (i.e., methylbenzaldehyde)	N	
8	122.9249	80.5810	122.9248	73.0977	122.9248	69.6379		0/1	0%	Tin [7440-31-5]	Y	
9	123.0795	75.2479	123.0794	70.5669	123.0794	70.3230		11/3	66%	C8H10O (i.e., 2,3-dimethylphenol)	N	
10	129.1277	490.6521	129.1277	462.2666	129.1277	489.3781		19/1	100%	octanal [124-13-0]	N	
11	149.1328	32.9848	149.1329	26.4474	149.1331	16.8234		3/12	17%	C11H16 (i.e., pentylbenzene)	N	
12	151.1118	525.9616	151.1119	534.2908	151.1119	516.0162		28/1	0%	2,3,5,6-tetramethylphenol [527-35-5]	N	
13	154.8806	83.0172	154.8806	69.3365	154.8806	66.7111		0/1	0%	carbon tetrachloride [56-23-5]	N	
14	185.0414	111.3406	185.0411	113.1992	185.0410	111.7597		0/1	0%	dibenzothiophene [132-65-0]	Y	
15	201.1855	550.7727	201.1856	543.2826	201.1854	544.6137		13/1	100%	lauric acid [143-07-7]	N	
16	223.0639	531.9464	223.0640	514.3256	223.0640	526.9002		4/1	0%	hexamethylcyclotrisiloxane [541-05-9]	N	
17	235.0572	116.1874	235.0585	101.9014	235.0571	111.0235		0/1	0%	benzonaphthothiophene [195-52-8]	Y	
18	266.8493	66.2241	266.8495	79.2979	266.8493	69.4841		0/1	0%	pentachlorophenol [87-86-5]	N	
19	297.0828	561.6014	297.0832	560.7830	297.0827	560.1216		0/1	0%	octamethylcyclotetrasiloxane [556-67-2]	N	
20	399.3951	33.7569	399.3953	37.7646	399.3954	35.3767		0/3	0%	C29H50 (i.e., 30-normoretane)	Y	