High-performance metabolic profiling of plasma from seven mammalian species for simultaneous environmental chemical surveillance and bioeffect monitoring

Youngja Park, Emory University
Kichun Lee, Emory University
Quinlyn A. Soltow, Emory University
Fred Strobel, Emory University
Kenneth L Brigham, Emory University
Richard E. Parker, Emory University
Mark Wilson, Emory University
Roy Sutliff, Emory University
Keith G. Mansfield, Harvard University
Lynn M. Wachtman, Harvard University

Only first 10 authors above; see publication for full author list.

Journal Title: Toxicology
Volume: Volume 295, Number 1-3
Publisher: Elsevier: 12 months | 2012-05-16, Pages 47-55
Type of Work: Article | Post-print: After Peer Review
Publisher DOI: 10.1016/j.tox.2012.02.007
Permanent URL: http://pid.emory.edu/ark:/25593/f4gtg

Final published version: http://dx.doi.org/10.1016/j.tox.2012.02.007

Copyright information:
© 2012 Elsevier Ireland Ltd. All rights reserved.
This is an Open Access work distributed under the terms of the Creative Commons Attribution-NonCommerical-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

Accessed January 5, 2020 4:36 AM EST
High-performance metabolic profiling of plasma from seven mammalian species for simultaneous environmental chemical surveillance and bioeffect monitoring

Youngja H. Park\textsuperscript{a,b}, Kichun Lee\textsuperscript{a,b}, Quinlyn A. Soltow\textsuperscript{a,b}, Frederick H. Strobel\textsuperscript{d}, Kenneth L. Brigham\textsuperscript{b,e}, Richard E. Parker\textsuperscript{b,e}, Mark E. Wilson\textsuperscript{f}, Roy L. Sutliff\textsuperscript{g,h}, Keith G. Mansfield\textsuperscript{h}, Lynn M. Wachtman\textsuperscript{b}, Thomas R. Ziegler\textsuperscript{c,e}, and Dean P. Jones\textsuperscript{a,b,e}

\textsuperscript{a}Clinical Biomarkers Laboratory, Emory University, Atlanta, GA 30322
\textsuperscript{b}Division of Pulmonary, Allergy and Critical Care Medicine, Emory University, Atlanta, GA 30322
\textsuperscript{c}Division of Endocrinology, Metabolism and Lipids, Department of Medicine, Emory University, Atlanta, GA 30322
\textsuperscript{d}Chemistry Department, Emory University, Atlanta, GA 30322
\textsuperscript{e}Center for Health Discovery and Well-Being, Emory University, Atlanta, GA 30322
\textsuperscript{f}Yerkes National Primate Research Center, Emory University, Atlanta, GA 30322
\textsuperscript{g}Atlanta VA Medical Center, Decatur, GA 30033
\textsuperscript{h}New England Primate Research Center, Harvard University, Southborough, MA 01772

Abstract

High-performance metabolic profiling (HPMP) by Fourier-transform mass spectrometry coupled to liquid chromatography gives relative quantification of thousands of chemicals in biologic samples but has had little development for use in toxicology research. In principle, the approach could be useful to detect complex metabolic response patterns to toxicologic exposures and to detect unusual abundances or patterns of potentially toxic chemicals. As an initial study to develop these possible uses, we applied HPMP and bioinformatics analysis to plasma of humans, rhesus macaques, marmosets, pigs, sheep, rats and mice to determine: 1) whether more chemicals are detected in humans living in a less controlled environment than captive species, and 2) whether a subset of plasma chemicals with similar inter-species and intra-species variation could be identified for use in comparative toxicology. Results show that the number of chemicals detected was similar in humans (3221) and other species (range 2537 to 3373). Metabolite patterns were most similar within species and separated samples according to family and order. A total of 1485 chemicals were common to all species; 37% of these matched chemicals in human metabolomic databases and included chemicals in 137 out of 146 human metabolic pathways. Probability-based modularity clustering separated 644 chemicals, including many endogenous metabolites, with...
inter-species variation similar to intra-species variation. The remaining chemicals had greater inter-species variation and included environmental chemicals as well as GSH and methionine. Together, the data suggest that HPMP provides a platform that can be useful within human populations and controlled animal studies to simultaneously evaluate environmental exposures and biological responses to such exposures.

**Keywords**

metabolomics; plasma; mass spectrometry; probability-based modularity clustering; exposome

1. Introduction

A recent biomonitoring survey of Americans by the Centers for Disease Control and Prevention (CDC) found traces of 212 environmental chemicals, including chemicals from plastics, pesticides and flame retardants (CDC 2011). Such exposures appear to mostly reflect common commercial uses, do not exceed relevant safety thresholds and have little if any evidence for associated adverse health effects. Furthermore, the chemicals in the CDC study were selected because of postulated high health concerns (e.g., bioaccumulative/persistent, plasticizers, and chemicals with ongoing exposure) so that the findings suggest that chemicals with short environmental or biological half-lives are likely to have even lower internal-dose exposure profiles and even lower potential for adverse health outcomes.

Despite the extensive scientific foundation of the methods and the re-assuring nature of such findings, contemporary risk assessment and surveillance strategies were devised as an affordable approach to minimize population risks and have several limitations. For instance, detection of an exposure of concern is possible only if it occurs at a frequency sufficient to be detected within the number of samples measured; if 20,000 people are studied in the US and they are representatively distributed among the 50 states and across the age and sex distributions of the population, then significant exposures to a single chemical may have to affect >1% of a relevant subpopulation, such as middle-age men in a single geographical region, e.g. the Mississippi Delta, to be detected. Second, because of cost limitations, chemicals are prioritized for targeted analysis based upon evidence of hazard and potential risk. Unknown or unidentified risks are excluded by this targeted approach. For instance, effects of low-level (0.001%) contaminants of commercial chemicals may not be readily detected in toxicology studies but be environmentally relevant if they are stable and substantial amounts of the agent are produced. Similarly, chemicals with a short environmental half-life are converted to other chemicals; it is very difficult to know whether these include minor but persistent products that bioaccumulate. Third, the selection of chemicals for analysis depends upon in vitro toxicity models, animal testing and extrapolation. While sound in principle, this approach cannot include the full range of possible adverse health effects due to complex genetic, epigenetic and co-exposure interactions in humans. Fourth, the approach assumes that individuals have similar sensitivities to environmental exposures, i.e., that inclusion of a safety factor is adequate to account for deviation of individuals from the norm. Knowledge from therapeutic drug use shows that even an extensively consumed drug like acetaminophen, used safely by hundreds of millions of people, causes toxicity at therapeutic doses in some individuals (Bonkovsky et al. 1994; Kwan et al. 1995). Thus, even though environmental health risk assessment and surveillance are effective in providing cost-effective means to minimize risk, there are continuing needs to improve this process.

New chemical analysis and profiling technologies can potentially provide cost-effective approaches to: a) more thoroughly evaluate biological responses of cell and animal studies.
for extrapolation to humans; b) more broadly monitor exposures on a personal basis; and c) more effectively detect adverse exposure effects within the context of other pathophysiologic processes under real-life conditions. Such technologies and innovative applications include advanced biosensors (Rea et al. 2011), microfluidics (Kraly et al. 2009), ultra-high resolution chemical separations (Nordstrom et al. 2006), and improved sensitivity mass spectrometry (Soltow et al. 2011), nuclear magnetic resonance spectrometry (Inouye et al. 2010) and high-resolution imaging (Miura et al. 2010). Development of such approaches has the promise to provide more comprehensive, rapid and less expensive methods to identify existing and newly emerging exposures of concern (Want et al. 2010).

High-performance metabolic profiling (HPMP) is a high-throughput chemical analysis developed as a practical approach for use in personalized medicine (Johnson et al. 2010). The method uses the high resolution and mass accuracy of Fourier-transform mass spectrometry (Marshall and Hendrickson 2008) to support measurement of up to 7000 chemicals in 20 µl samples in 20 min (Soltow et al. 2011), thereby making this method potentially affordable for routine measurement of endogenous metabolites and metabolic patterns for disease diagnosis and health management. In application of this technology to biological samples, there is an ambiguity in terminology because “metabolomics” is used as a general term for all chemicals, yet all chemicals in biological samples are not biologically relevant “metabolites”. In the present report, we use “chemical” in a non-specific way to refer to any chemical detected. HPMP was developed as a high-throughput approach to study metabolism and disease, and we use the term “endogenous metabolite” in the present study to refer to these biochemicals, as represented by KEGG (Kyoto Encyclopedia of Genes and Genomes) (Kanehisa 2002; Kanehisa and Goto 2000) human metabolic pathways. Other chemicals include those derived from diet, enteric flora, pharmaceuticals and environmental sources, many of which are currently unidentified. The present study is based upon the observation that HPMP analyses include environmental chemicals, e.g., insecticides, fungicides and plasticizers, raising the possibility that HPMP could be adopted as part of a universal exposure surveillance strategy for health and environmental exposures (Soltow et al. 2011). In principle, combinations of chemical separation and ionization strategies could provide an approach to survey the environmental chemical space (Howard and Muir 2011) measured as the abundance or patterns of chemicals in routine blood or urine samples from individuals during normal healthcare visits (Soltow et al. 2011; Weis et al. 2005).

The current study was designed to gain information about potential use of HPMP in exposure surveillance through study of plasma from seven mammalian species (human, rhesus macaque, common marmoset, pig, sheep, rat, mouse). We performed HPMP on plasma and applied bioinformatic approaches to determine the fraction of plasma chemicals that are common among the mammalian species, characterize these in terms of matches to metabolic databases and identify ones with low inter-species variation that could be suitable to support biological response monitoring along with environmental chemical surveillance.

2. Materials and Methods

2.1. Materials

Acetonitrile (HPLC grade), formic acid (puriss. p.a. 98%), water (HPLC grade) and caffeine were obtained from Sigma–Aldrich (St. Louis) and the reverse phase test mix (Cat#: 47641-U) was from Supelco Analytical (Bellefonte, PA). Trimethyl-[13C3]-caffeine (Cat#: CLM-514-0) and [15N]-L-tyrosine (Cat#: NLM-590-0) were obtained from Cambridge Isotope Laboratories, Inc (Andover, PA).
2.2. Plasma Samples

For the purpose of characterization of species and/or sex on phylogenetic differences in metabolic pathways, 2 sample sets (Sample Set 1 and 2) were analyzed at different times. In Sample Set 1, archival samples from seven mammalian species were studied (Supplemental Material, Table 1). Samples from animal species were obtained under standard experimental conditions, respectively differing in diet and environment according to husbandry practices for the different species (Supplemental Material, Table 1). Samples were stored at −80°C until analysis, in all cases for <2 y; at present there is no systematic knowledge concerning the long-term stability of the thousands of chemicals in the samples. Sample Set 2 (Supplemental Material, Table 1) was designed to determine whether metabolic differences according to sex could be discriminated from phylogenetic differences by HPMP.

2.3. High-Performance Metabolic Profiling (HPMP)

Details have been previously published (Johnson et al. 2010; Soltow et al. 2011). Briefly, plasma (50 µl) was treated with acetonitrile (2:1) and an internal standard mix (Soltow et al. 2011) and then centrifuged at 13,000 × g for 5 min to remove protein. Extracts were placed in a refrigerated autosampler and 10 µl volumes were analyzed in duplicate with DC-FTMS platforms, one using an anion exchange (AE) column (Hamilton PRPX-110S, 2.1 mm ×10 cm, Reno) and the other using a C18 column (17 mm × 2.1 mm, Pompton Plains), both with aC18 precolumn (Higgins Analytical Targa column, Mtn View). Samples were fractionated with a formate gradient, ionized with electrospray ionization in the positive mode, and detected with an LTQ-FT spectrometer (Thermo, San Jose) with m/z from 85–850. Data were extracted using apLCMS (Yu et al. 2009) as m/z features, where an m/z feature is defined by m/z (mass/charge), RT (retention time) and ion intensity (integrated ion intensity for the peak). Endogenous metabolites were annotated using Madison Metabolomics Consortium Database (MMCD) (Markley et al. 2007), Metlin Mass Spectrometry Database (Smith et al. 2005) and MS/MS.

2.4. Bioinformatics

Hierarchical Clustering Analysis (HCA) (Pirouette software; Infometrix, Bothell, WA) was used as a data reduction method to visualize phylogenetic similarities of metabolic spectra among the 7 species. Principal Component Analysis (PCA) (Pirouette software; Infometrix, Bothell, WA) with autoscaling was used as a data reduction method to visualize discriminatory factors, shown in two-dimensional (2-D) and three-dimensional (3-D) PCA score plots. Metabolite pathway mapping was performed using KEGG (Kanehisa 2002; Kanehisa and Goto 2000) to match high-resolution m/z data to known endogenous metabolites and metabolic pathways. The large number of m/z features (>3000), most of which did not match endogenous metabolites in the metabolic databases, precluded confirmation of chemical identities of most features. However, over 90% of metabolites in the KEGG human metabolic database have unique m/z at the resolution used, and MS/MS and coelution studies confirmed identity of many features (e.g., amino acids, intermediary metabolites, environmental chemicals). One-way ANOVA and Tukey test were used to compare the relative intensities of chemicals in human plasma to those of 6 different species. Significance was expressed at p <0.05.

2.5. Probability-based modularity clustering

Probability-based modularity clustering was used to determine associations of chemicals within species and between species. This is a graph-based technique for automated clustering that provides numerical optimization across all possible clusterings. The method is based upon modularity clustering (Newman 2006), an approach which maximizes the "modularity function", a direct measure of the quality of a particular clustering of nodes in a
graph. A detailed discussion of maximization and formulation of modularity function is available (Stone and Ayroles 2009). Briefly, in an undirected graph with nodes and edges, modularity clustering determines the community structure of by constructing adjacency between two nodes and, to define the closeness of nodes. A modularity function is defined

\[ Q(P_k) = \sum_{c=1}^{k} \left[ \frac{A(V_c, V_c)}{A(V, V)} - \left( \frac{A(V_c, V)}{A(V, V)} \right)^2 \right] \]

where \( A \) is a partition of the nodes into groups and. Thus, measures the within-cluster sum of adjacency, measures the sum of adjacency over all edges attached to nodes in cluster, and measures the sum of all edge adjacency in the graph. The term \( / \) relates to the empirical probability that both ends of a randomly selected edge lie in cluster, and the term relates to the empirical probability that one specific end of a randomly selected edge lies in cluster. Hence, the modularity function of partition is the difference between the partition’s empirical within edge probability and the partition’s hypothetical randomly-placed edge probability. Stone and Ayroles (Stone and Ayroles 2009) used this correlation-based adjacency measure to cluster genes and, based upon this, we used a probability-based adjacency measure to cluster chemicals for phylogenetic groups. We used the adjacency between chemicals to measure contribution of individual chemicals to similarity among groups. Specifically, for (human, non-human primate, rodentia, artiodactyla), is the -th group effect. The adjacency between chemicals and, denoted by, is defined as the p-value of the test:

\[
\begin{cases} 
H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 \\
H_a: \text{not } H_0 
\end{cases}
\]

With this adjacency measure, high p-values lead to failure to reject the null hypothesis, which indicates that high p-values relate to the conclusion of the four species being similar. Adjacency close to 0 leads to rejection of the null hypothesis, meaning that metabolite and result in quite dissimilar, or divergent, effects of the relative intensities among species. By maximizing the modularity function as in equation (1), the adjacency constructed as above produces a partition of clusters that separates chemicals that are non-differentiating (have the highest average adjacency among groups) from those that are differentiating (have the lowest average adjacency among groups). A detailed discussion of iterative and automatic maximization of involving no cut-off value is available (Newman 2006). Within each cluster, chemicals are ordered according to adjacency within the cluster, with the upper left chemicals showing best associations among groups and the lower right chemicals showing greatest divergence among groups. For this analysis, chemicals with >30% zero values were excluded, leaving 1418 of the common chemicals for reported results. Results were similar using for all species individually, but the current analysis was used because of the small \( n \) for some species.

3. Results

3.1. HPMP of different species

A complete table of results for \( m/z \) features for individuals is available from the authors upon request. Results for each order, family and species are summarized in Figure 1A and B. Each order had a similar total number of chemicals ranging from 3382 to 3723. Species from two families, pig and sheep, had relatively fewer features, but the number of features/individual was similar so this may reflect the smaller number of individuals studied. Of interest, the number of chemicals detected in human plasma was similar to that in the
experimental animals even though the human subjects were not in a controlled research environment and expected to have more variable diet and environmental exposure.

The distribution of features among the orders is shown as a Venn diagram in Figure 1B. Of the total 3820 $m/z$ features, less than half (1485) was present in all species (Supplemental Material, Table 2). A greater number of common features was found between primates and rodents than between artiodactyls and either primates or rodents.

### 3.2. Similarities of metabolic profiles using HCA

HCA was performed on the data for the 3820 chemicals in individual plasma samples to characterize similarity of metabolic spectra. The results (Figure 1C) showed that individuals within a species were more similar to other individuals within that species than to individuals from other species. However, species generally were not more similar within family than between family, e.g., humans and rhesus were more similar to pig and sheep than to marmoset. Thus, the similarities in metabolic character of plasma reflected in HCA only partially recapitulate phylogeny.

### 3.3. PCA of HPMP data for plasma from 7 mammalian species

Data reduction by PCA showed better classification according to taxonomic rank. A 3-D PCA score plot (Figure 2A), rotated to visualize maximal separation, showed that the 7 mammalian species separated by phylogenetic order (primata, actiodactyla, rodentia) using PC1, 2 and 3, which represents 63% of total variation. In Figure 2B, a 2-D PCA score plot representing 56% of total variation showed nearly complete separation according to species. This sample set did not have equal distribution according to sex, so a second set of samples was examined to determine whether individuals were separated in a PCA score plot according to sex. This set of samples contained 5 females and 5 males from human, marmoset and mouse. The 2-D PCA score plot (Figure 2C) showed that separation was largely due to species and not to sex. According to the results, the first three principal components of a PCA analysis of 3820 chemicals separates according to species with additional separation according taxonomic family and order but not according to sex.

### 3.4. Pathway analysis of 1485 chemicals common to 7 mammalian species

Metabolic databases provide the capability to search for chemicals that match the high-accuracy $m/z$ data. Searches against Metlin (Smith et al. 2005) and MMCD (Madison Metabolomics Consortium Database; (Markley et al. 2007) showed that less than half of the $m/z$ match known human endogenous metabolites (Draper et al. 2009; Soltow et al. 2011). KEGG pathway analysis similarly showed that only 673 of the $m/z$ matched known metabolites (Figure 3). Although these endogenous metabolites represent only about 20% of the total chemicals detected, they included metabolites in 137 pathways of the total 146 pathways in the KEGG human reference metabolic pathways. Examination of some selected chemicals (alanine, threonine, tyrosine, methionine, GSH, pirimicarb) showed significant differences in signal intensities between species (Supplemental Material, Figure 1). Thus, the results suggest that HPMP could provide a useful approach for comparative studies evaluating endogenous metabolites and other common chemicals.

### 3.5. Probability-Based Modularity Clustering of 1485 chemicals common to 7 mammalian species

We used probability-based modularity clustering as a means to analyze chemicals according to the significance of their association with other chemicals (Newman 2006). By setting criteria to classify chemicals according to significance of within-species correlations and significance of between-species correlations, a distribution was established in which
modules of chemicals that have similar variation among species (644 chemicals, Module 1, Figure 4) are discriminated from chemicals that have greater variation among species than within species (774 chemicals, Module 2, Figure 4). In each module, the heatmap provides a visualization of the p-value of correlations between chemicals so that chemicals are classified for how well they fit the classification; additional graphical discrimination between chemicals in Modules 1 and 2 is provided in Figure 5. Module 1 was enriched in endogenous metabolites, such as leucine/isoleucine, citrulline, cystine and other amino acids and metabolites for which we have previously confirmed identities by co-elution and MS/MS criteria (Johnson et al. 2010; Soltow et al. 2011). Module 2 was enriched in m/z matching environmental chemicals, such as pirimicarb and di-N-butyl phthalate, and also included some endogenous metabolites (GSH, methionine and glutamine). We have previously confirmed identities of the latter endogenous metabolites (Johnson et al. 2010; Soltow et al. 2011). To confirm identity of some of the presumed environmental chemicals, co-elution and MS/MS studies were performed on selected m/z that matched pirimicarb, triethyl phosphate, di-N-butyl phthalate and rotenone. Results confirmed these identifications except for rotenone (Supplemental Material, Figure 2). The MS/MS for the m/z corresponding to rotenone had an ion dissociation pattern that matched a minor contaminant present in the commercial rotenone preparation.

4. Discussion

Toxicology testing is the primary approach to hazard identification (Toxicity testing in the 21st century, Natl Acad Sci Press) (Judson et al. 2008; Judson et al. 2009; Judson et al. 2010), and high-throughput in vitro screening capabilities now allow rapid evaluation of a range of adverse effects in cell and molecular systems. Coupling knowledge of such hazards to comparative research in animal models has contributed significantly to understanding of human risks by helping understand mechanisms and identifying potential exposures of concern (Bergen and Mersmann 2005; Soucek and Gut 1992). For example, sequencing of the genomes of many species (Soucek et al. 2003) along with detoxification and toxicokinetic studies (Butt et al. 2010; Soucek and Gut 1992) has facilitated extrapolations of human risks (Peng et al. 2010; Soucek and Gut 1992). There are, however, a number of limitations to this approach, and contemporary comparative biology increasingly relies upon new information-rich technologies and systems biology approaches (Bergen and Mersmann 2005; Roos et al. 2011; Travlos et al. 2011).

The present research shows that HPMP could potentially be used to improve comparisons between species in comparative toxicology research. The analyses of plasma from 7 species resulted in detection of 3820 total m/z, which can be considered a representation of the “pan” metabolome of mammals (Figure 6). These data confirm previous findings (Draper et al. 2009; Soltow et al. 2011) that humans and other mammalian species have thousands of chemicals in plasma, with over half not represented by chemicals in human metabolic databases such as Metlin (Smith et al. 2005) and MMCD (Markley et al. 2007). It should be noted that the full complement of the “pan” metabolome cannot be estimated from this analysis because plasma does not represent other biologic compartments and there is no way to evaluate the fraction of the total metabolome that is detected by the methods used. In the context of the current study, this pan metabolome can be considered to consist of a “core” metabolome of 1485 chemicals that are common to all species and a “peri-core” metabolome (i.e., chemicals surrounding the core metabolome) that includes the other 2335 chemicals (Figure 6). The core metabolome includes 644 chemicals with inter-species variation similar to intra-species variation, indicating that chemicals from this subset are likely to be useful to study toxic responses (Figure 6). Specifically, toxicity can occur as a consequence of exposure to chemicals that are relatively rapidly eliminated from an organism and, as a result, cannot be detected at later times. Comparative toxicity studies of
Module 1 chemicals could support identification of biomarkers and/or provide response patterns useful for related epidemiological research of such toxicities in humans.

An important priority to enhance use of HPMP data involves identification and annotation of the 1485 chemicals to allow subclassification according to endogenous metabolites, dietary chemicals, microbiome-related chemicals, therapeutic chemicals and environmental chemicals (Figure 6). Such identification and classification is labor-intensive but is critical to support environment-wide association studies (EWAS) (Patel et al. 2010) and gene-environment (G x E) associations (Petronis 2010) (Figure 6). Importantly, the m/z data also can be used to support metabolome-wide association studies (MWAS) of health and disease phenotypes and genome-metabolome (G x M) association studies even without identification and classification. In such cases, however, post-hoc identification will still be needed to fully understand mechanisms.

The presence of environmental chemicals in research animals highlights a limitation of low-dose exposure research. Environmental exposures from water and air may not substantially differ in research animals from that to which humans are exposed. Research animals also consume food formulated by agricultural industries that also produce human food, and the animal food is processed using commercial operations sharing many characteristics of human food production. Furthermore, research animals are exposed to plastics and other chemicals to which humans are exposed. Thus, despite the concept that research animals are maintained under highly controlled environmental conditions, the present data show that plasma of research animals contains a number and quantities of environmental chemicals similar to humans and therefore cannot be considered “unexposed”.

Accumulating evidence indicates that genetic factors represent only a fraction of total risk for chronic and age-related diseases (Rappaport 2011; Willett 2002; Willett et al. 2011). The majority of risk is linked to environmental or gene-environment interactions, emphasizing the need for systematic study of the exposome (Faisandier et al. 2011; Rappaport and Smith 2010). A pilot EWAS of type 2 diabetes (Patel et al. 2010) using CDC data on environmental chemicals suggested that environmental factors can be found with effect sizes comparable to loci found by GWAS (Hindorff et al. 2009). Other studies also link low-level environmental exposures to health risks, including cancer (Soutar et al. 2000; Whitrow et al. 2003), respiratory diseases (Whitehead et al. 2011; Wright et al. 2001), endocrine disruption (Reif et al. 2010; Vis et al. 2010), reproductive failure (Bunderson-Schelvan et al. 2011; Pant et al. 2007) and neurodegenerative diseases (Perrone-Capano and Di Porzio 2000; Sherer et al. 2001). While current approaches using analytically rigorous methods to measure subsets of known hazardous chemicals in populations at greatest risk is cost-effective in minimizing population risk to known hazards, such approaches can be inadequate to detect unanticipated chemicals, such as exemplified by the widespread distribution of perfluorooctanoic acid (PFOA), a chemical not recognized to be released into the environment (Frisbee et al. 2010; Steenland et al. 2010). Cumulative effects from exposure to multiple similar chemicals (Boas et al. 2010; Pant et al. 2008), and long-term and multigenerational effects due to genomic and epigenomic mechanisms (Perrone-Capano and Di Porzio 2000; Ziech et al. 2010), also highlight a need to characterize lifelong exposure histories (Faisandier et al. 2011; Rappaport 2011; Ziech et al. 2010). However, surveillance of all possible exposures is effectively impossible because greater than 80,000 agents are registered with EPA for commercial use. Furthermore, more than 2,800 chemicals have annual production volumes exceeding 454 tons/year (Muir and Howard 2006). If each such agent has 10 minor contaminants and each agent and contaminant is converted to 10 products in the biosphere, one would need to devise means to identify and perform toxicity screening for nearly a half-million additional chemicals to appropriately support associated risk assessment and surveillance.
The present data suggest an alternative approach could be developed using HPMP for MWAS studies that complements this risk assessment and surveillance approach and addresses some of its key limitations. The approach would start with analysis of representative samples from a large number of individuals (e.g. 10,000–20,000) for whom associated demographic and health phenotyping information is available. Such analyses could be accomplished within a single reference laboratory with routine, automated analysis using one to four high-resolution mass spectrometers (5000–10,000 samples per year per spectrometer). These samples would be collected, stored and analyzed in duplicate or triplicate, along with pooled reference plasma samples as available from the National Institute of Standards (McGaw et al. 2010), with standard analytic procedures and internal standards (Soltow et al. 2011). The high-resolution mass spectral data would be entered into a cumulative database structure with accurate mass m/z, chromatographic retention time (RT) and intensity information. This data would provide a resource for association studies, such as discovery of associations of chemicals with geographical regions, dietary patterns and phenotypic information present in the dataset. Known chemicals would be annotated in this database. For unknowns, the accurate mass m/z is sufficient to predict elemental composition in many cases, and along with RT, this m/z would provide information for post-hoc identification and validation studies. Associated MS/MS information could facilitate post-hoc identification and validation. Although initially limited by the sample size, extent of associated disease information and knowledge of chemical identities, this reference data would allow any newly designed study, whether from a specific geographical region or selected population, to be conducted within a framework allowing comparison to a more standardized reference population. Importantly, the data would include information on unknown and unidentified chemicals detected in human plasma, providing a basis to investigate exposure distributions by geography and other demographic factors. Addition of data from representative samples on an annual basis would allow detection of exposure trends, even for unidentified chemicals. Inclusion of disease outcome data would allow detailed EWAS and G × E studies, and repeat analyses on the same individuals would allow systematic study of lifelong exposures as conceptualized in the exposome.

5. Conclusions

This study shows that high-performance metabolic profiling (HPMP) detects a spectrum of low-level environmental chemicals in plasma of research animals that is comparable to that which is found in plasma of humans. Of 3820 chemicals detected, 1485 were common to all 7 mammalian species, suggesting a general utility of HPMP for comparative environmental health research. Analysis of the 1485 common chemicals by probability-based clustering identified a subset of 644 that included endogenous metabolites with characteristics useful for comparative studies of biologic responses to toxicologic exposure. The other >2000 chemicals can be used to provide a non-targeted surveillance of chemical exposures for MWAS of health and disease phenotypes and genome-metabolome (G × M) association studies. Together, the data suggest that HPMP provides a platform that can be useful within human populations and controlled animal studies to simultaneously evaluate environmental exposures and biological responses to such exposures.

<table>
<thead>
<tr>
<th>Highlights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Six mammal species had low-abundance environmental chemicals similar to humans</td>
</tr>
<tr>
<td>Of 3820 chemicals detected, 1485 common metabolites were present in the seven species</td>
</tr>
<tr>
<td>Metabolites with low interspecies variability identified for bioeffect monitoring</td>
</tr>
<tr>
<td>Analysis detected &gt;2000 chemicals suitable for EWAS and G × E studies</td>
</tr>
</tbody>
</table>
Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported in part by NIH Grants ES016731, ES011195, AG038746, RR025008, HD46501; RR00165, HL083019, HL070892, RR023356

References


Judson RS, Houck KA, Kavlock RJ, Knudsen TB, Martin MT, Mortensen HM, Reif DM, Rotroff DM, Shah I, Richard AM, Dix DJ. In vitro screening of environmental chemicals for targeted testing

Toxicology. Author manuscript; available in PMC 2013 May 16.


Figure 1. 
A. Phylogenetic relationship of 7 mammalian species with associated number of chemicals detected by HPMP for each order, family and species. B. Venn diagram of the chemicals common among families. C. Hierarchical clustering analysis (HCA) of plasma metabolic profiles after variance scaling shows similarity of individuals of the same species.
Figure 2.
A. Three-dimensional PCA score plot, in which the first three principal components (PCs) explained 62% of total variation, is arbitrarily rotated to visualize discrimination according to species, family and order. B. Two-dimensional PCA score plot with PC1 and PC2 corresponding to Panel D. C. 2-D PCA score plot for Sample Set 2 to evaluate whether samples from different species (human, marmoset and mouse) were discriminated according to sex. In A and B, each symbol shows a different class: diamond, primate (human, H, filled; rhesus, Rh, open; marmoset, Ma, with line); circle, artiodactyla (pig, P, filled; sheep, S, open); triangle, rotentia (rat, Ra, filled; mouse, Mo, open). In C, males, filled symbols; females open symbols; human, diamond; marmoset, square; mouse, triangle.
Figure 3.
KEGG metabolic pathway matches for m/z features that are common to mammalian species. High mass accuracy m/z for 1485 common chemicals revealed 666 matches as shown in black, including matches for metabolites in 137 of the total 146 KEGG human reference metabolic pathways. Corresponding metabolite names are available at http://www.genome.jp/kegg-bin/show_pathway?1315425690831/hsa01100.args.
Figure 4.
Probability-based modularity clustering of 1485 common chemicals according to intra- and inter-species variation. Module 1 (M1) includes chemicals with similar characteristics in all 7 mammalian species while Module 2 (M2) contains chemicals that have different characteristics among the species. M1 showed a preponderance of endogenous metabolites while M2 showed a preponderance of other chemicals, including known environmental chemicals. These data suggest that HPMP can be used both for surveillance of environmental exposures by focusing on chemicals in M2 and for study of biologic responses to environmental exposures by focusing on chemicals in M1.
Figure 5.
The relationship between inter-species variation and within-error variation for selected chemicals in Module 1 (M1) and 2 (M2) of probability-based modularity clustering analysis shown in Fig 3. Trajectories above the diagonal line, represented by the arrows in A–C, indicate that within-species variation is large relative to interspecies variation. In contrast, trajectories below the diagonal line, represented by arrows in D–F, show that interspecies variation is large relative to within-species variation. This difference between chemicals in Module 1 and Module 2 provides a basis to use HPMP for two purposes, direct evaluation of environmental exposures (Module 2), and biological responses to environmental exposures (Module 1). A: leucine/isoleucine (M1), B: citrulline (M1), C: cystine (M1), D: glutamine (M2), E: pirimicarb (M2), F: triethylphosphate (M2).
Figure 6.
Characterization of HPMP data for comparative toxicology and biomonitoring of exposures. HPMP analyses detected 3820 m/z, which reflects the total metabolome, designated here as the “pan” metabolome. The pan metabolome consists of a “core” metabolome that is common to the 7 mammalian species and a “peri-core” metabolome that includes all of the chemicals surrounding the core metabolome. The core metabolome includes Module 1, consisting of chemicals that have inter-species variation similar to intra-species variation, and Module 2, consisting of chemicals that have greater inter-species variation than intra-species variation. Module 1 includes endogenous metabolites and has characteristics suitable for comparative studies of biologic responses to toxic exposure. Module 2 includes environmental chemicals, variable endogenous metabolites, chemicals derived from the diet, chemicals derived from the microbiome, and pharmaceuticals. Module 2 and peri-core metabolites can be used to support environment-wide association studies (EWAS) and gene-environment (G x E) studies. Identification and classification of chemicals in Module 2 and the peri-core metabolome will greatly facilitate interpretation of EWAS and G x E study results.