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

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

[a]Clinical Biomarkers Laboratory, Emory University, Atlanta, GA 30322

[b]Division of Pulmonary, Allergy and Critical Care Medicine, Emory University, Atlanta, GA 30322

[c]Division of Endocrinology, Metabolism and Lipids, Department of Medicine, Emory University, Atlanta, GA 30322

[d]Chemistry Department, Emory University, Atlanta, GA 30322

[e]Center for Health Discovery and Well-Being, Emory University, Atlanta, GA 30322

[f]Yerkes National Primate Research Center, Emory University, Atlanta, GA 30322

[g]Atlanta VA Medical Center, Decatur, GA 30033

[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

Address Correspondence to Dean P. Jones, Ph.D., Director, Clinical Biomarkers Laboratory, Division of Pulmonary, Allergy and Critical Care Medicine, 615 Michael Street, Suite 205P, Emory University, Atlanta, GA 30322 USA, Phone: 1-404-727-5970, Fax: 404-712-2974, dpjones@emory.edu.

Conflict of interest statement

The authors declare that they have no competing financial interests.

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-[^{13}C3]-caffeine (Cat#: CLM-514-0) and [^{15}N]-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 $13,000 \times g$ 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 phylogenic 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

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

(1)

where 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

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 × 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 × 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.

Highlights

Six mammal species had low-abundance environmental chemicals similar to humans

Of 3820 chemicals detected, 1485 common metabolites were present in the seven species

Metabolites with low interspecies variability identified for bioeffect monitoring

Analysis detected >2000 chemicals suitable for EWAS and G × E studies

---

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Bergen WG, Mersmann HJ. Comparative aspects of lipid metabolism: impact on contemporary research and use of animal models. The Journal of nutrition. 2005; 135:2499–2502. [PubMed: 16251600]

Boas M, Frederiksen H, Feldt-Rasmussen U, Skakkebaek NE, Hegedus L, Hilsted L, Juul A, Main KM. Childhood exposure to phthalates: associations with thyroid function, insulin-like growth factor I, and growth. Environ Health Perspect. 2010; 118:1458–1464. [PubMed: 20621847]

Bonkovsky HL, Kane RE, Jones DP, Galinsky RE, Banner B. Acute hepatic and renal toxicity from low doses of acetaminophen in the absence of alcohol abuse or malnutrition: evidence for increased susceptibility to drug toxicity due to cardiopulmonary and renal insufficiency. Hepatology. 1994; 19:1141–1148. [PubMed: 8175135]

Bunderson-Schelvan M, Pfau JC, Crouch R, Holian A. Nonpulmonary outcomes of asbestos exposure. J Toxicol Environ Health B Crit Rev. 2011; 14:122–152. [PubMed: 21534087]

Butt CM, Muir DC, Mabury SA. Elucidating the pathways of poly- and perfluorinated acid formation in rainbow trout. Environ Sci Technol. 2010; 44:4973–4980. [PubMed: 20518507]

CDC. Fourth National Report on Human Exposure to Environmental Chemicals 2011, Updated Tables. National Center for Environmental Health, Centers for Disease Control and Prevention; 2011.

Draper J, Enot DP, Parker D, Beckmann M, Snowdon S, Lin W, Zubair H. Metabolite signal identification in accurate mass metabolomics data with MZedDB, an interactive m/z annotation tool utilising predicted ionisation behaviour 'rules'. BMC Bioinformatics. 2009; 10:227. [PubMed: 19622150]

Faisandier L, Bonneterre V, De Gaudemaris R, Bicout DJ. Occupational exposome: A network-based approach for characterizing Occupational Health Problems. J Biomed Inform. 2011; 44:545–552. [PubMed: 21362498]

Frisbee SJ, Shankar A, Knox SS, Steenland K, Savitz DA, Fletcher T, Ducatman AM. Perfluorooctanoic acid, perfluorooctanesulfonate, and serum lipids in children and adolescents: results from the C8 Health Project. Arch Pediatr Adolesc Med. 2010; 164:860–869. [PubMed: 20819969]

Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106:9362–9367. [PubMed: 19474294]

Howard PH, Muir DC. Identifying New Persistent and Bioaccumulative Organics Among Chemicals in Commerce II: Pharmaceuticals. Environ Sci Technol. 2011; 45:6938–6946. [PubMed: 21740030]

Inouye M, Kettunen J, Soininen P, Silander K, Ripatti S, Kumpula LS, Hamalainen E, Jousilahti P, Kangas AJ, Mannisto S, Savolainen MJ, Jula A, Leiviska J, Palotie A, Salomaa V, Perola M, Ala-Korpela M, Peltonen L. Metabonomic, transcriptomic, and genomic variation of a population cohort. Mol Syst Biol. 2010; 6:441. [PubMed: 21179014]

Johnson JM, Yu T, Strobel FH, Jones DP. A practical approach to detect unique metabolic patterns for personalized medicine. Analyst. 2010; 135:2864–2870. [PubMed: 20838665]

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

prioritization: the ToxCast project. Environ Health Perspect. 2010; 118:485–492. [PubMed: 20368123]

Judson R, Richard A, Dix D, Houck K, Elloumi F, Martin M, Cathey T, Transue TR, Spencer R, Wolf M. ACToR--Aggregated Computational Toxicology Resource. Toxicology and applied pharmacology. 2008; 233:7–13. [PubMed: 18671997]

Judson R, Richard A, Dix DJ, Houck K, Martin M, Kavlock R, Dellarco V, Henry T, Holderman T, Sayre P, Tan S, Carpenter T, Smith E. The toxicity data landscape for environmental chemicals. Environ Health Perspect. 2009; 117:685–695. [PubMed: 19479008]

Kanehisa M. The KEGG database. Novartis Found Symp. 2002; 247:91–101. discussion 101–103, 119–128, 244–152. [PubMed: 12539951]

Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000; 28:27–30. [PubMed: 10592173]

Kraly JR, Holcomb RE, Guan Q, Henry CS. Review: Microfluidic applications in metabolomics and metabolic profiling. Analytica chimica acta. 2009; 653:23–35. [PubMed: 19800473]

Kwan D, Bartle WR, Walker SE. Abnormal serum transaminases following therapeutic doses of acetaminophen in the absence of known risk factors. Digestive diseases and sciences. 1995; 40:1951–1955. [PubMed: 7555449]

Markley JL, Anderson ME, Cui Q, Eghbalnia HR, Lewis IA, Hegeman AD, Li J, Schulte CF, Sussman MR, Westler WM, Ulrich EL, Zolnai Z. New bioinformatics resources for metabolomics. Pac Symp Biocomput. 2007:157–168. [PubMed: 17990489]

Marshall AG, Hendrickson CL. High-resolution mass spectrometers. Annu Rev Anal Chem (Palo Alto Calif). 2008; 1:579–599. [PubMed: 20636090]

Miura D, Fujimura Y, Yamato M, Hyodo F, Utsumi H, Tachibana H, Wariishi H. Ultrahighly sensitive in situ metabolomic imaging for visualizing spatiotemporal metabolic behaviors. Analytical chemistry. 2010; 82:9789–9796. [PubMed: 21043438]

Muir DC, Howard PH. Are there other persistent organic pollutants? A challenge for environmental chemists. Environ Sci Technol. 2006; 40:7157–7166. [PubMed: 17180962]

McGaw EA, Phinney KW, Lowenthal MS. Comparison of orthogonal liquid and gas chromatography-mass spectrometry platforms for the determination of amino acid concentrations in human plasma. J Chromatogr A. 2010; 1217:5822–5831. [PubMed: 20696434]

Newman ME. Modularity and community structure in networks. Proceedings of the National Academy of Sciences of the United States of America. 2006; 103:8577–8582. [PubMed: 16723398]

Nordstrom A, O'Maille G, Qin C, Siuzdak G. Nonlinear data alignment for UPLC-MS and HPLC-MS based metabolomics: quantitative analysis of endogenous and exogenous metabolites in human serum. Analytical Chemistry. 2006; 78:3289–3295. [PubMed: 16689529]

Pant N, Kumar R, Mathur N, Srivastava SP, Saxena DK, Gujrati VR. Chlorinated pesticide concentration in semen of fertile and infertile men and correlation with sperm quality. Environ Toxicol Pharmacol. 2007; 23:135–139. [PubMed: 21783749]

Pant N, Shukla M, Kumar Patel D, Shukla Y, Mathur N, Kumar Gupta Y, Saxena DK. Correlation of phthalate exposures with semen quality. Toxicology and applied pharmacology. 2008; 231:112–116. [PubMed: 18513777]

Patel CJ, Bhattacharya J, Butte AJ. An Environment-Wide Association Study (EWAS) on type 2 diabetes mellitus. PloS one. 2010; 5:e10746. [PubMed: 20505766]

Peng J, Oo ML, Andersen JK. Synergistic effects of environmental risk factors and gene mutations in Parkinson's disease accelerate age-related neurodegeneration. Journal of neurochemistry. 2010; 115:1363–1373. [PubMed: 21039522]

Perrone-Capano C, Di Porzio U. Genetic and epigenetic control of midbrain dopaminergic neuron development. Int J Dev Biol. 2000; 44:679–687. [PubMed: 11061432]

Petronis A. Epigenetics as a unifying principle in the aetiology of complex traits and diseases. Nature. 2010; 465:721–727. [PubMed: 20535201]

Rappaport SM. Implications of the exposome for exposure science. J Expo Sci Environ Epidemiol. 2011; 21:5–9. [PubMed: 21081972]

Rappaport SM, Smith MT. Epidemiology. Environment and disease risks. Science. 2010; 330:460–461. [PubMed: 20966241]

Rea, G.; Polticelli, F.; Antonacci, A.; Lambreva, M.; Pastorelli, S.; Scognamiglio, V.; Zobnina, V.; Giardi, MT. Computational Biology, Protein Engineering, and Biosensor Technology: a Close Cooperation for Herbicides Monitoring. In: Larramendy, ML.; Solloneski, S., editors. Herbicides, Theory and Applications, InTech. 2011. p. 93-120.

Reif DM, Martin MT, Tan SW, Houck KA, Judson RS, Richard AM, Knudsen TB, Dix DJ, Kavlock RJ. Endocrine profiling and prioritization of environmental chemicals using ToxCast data. Environ Health Perspect. 2010; 118:1714–1720. [PubMed: 20826373]

Roos R, Andersson PL, Halldin K, Hakansson H, Westerholm E, Hamers T, Hamscher G, Heikkinen P, Korkalainen M, Leslie HA, Niittynen M, Sankari S, Schmitz HJ, van der Ven LT, Viluksela M, Schrenk D. Hepatic effects of a highly purified 2,2',3,4,4',5,5'-heptachlorbiphenyl (PCB 180) in male and female rats. Toxicology. 2011; 284:42–53. [PubMed: 21458519]

Sherer TB, Betarbet R, Greenamyre JT. Pesticides and Parkinson's disease. Scientific World Journal. 2001; 1:207–208. [PubMed: 12805673]

Smith CA, O'Maille G, Want EJ, Qin C, Trauger SA, Brandon TR, Custodio DE, Abagyan R, Siuzdak G. METLIN: a metabolite mass spectral database. Ther Drug Monit. 2005; 27:747–751. [PubMed: 16404815]

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. 2011

Soucek P, Gut I. Cytochromes P-450 in rats: structures, functions, properties and relevant human forms. Xenobiotica. 1992; 22:83–103. [PubMed: 1615711]

Soucek P, Gut I, Trneny M, Skovlund E, Grenaker Alnaes G, Kristensen T, Borresen-Dale AL, Kristensen VN. Multiplex single-tube screening for mutations in the Nijmegen Breakage Syndrome (NBS1) gene in Hodgkin's and non-Hodgkin's lymphoma patients of Slavic origin. Eur J Hum Genet. 2003; 11:416–419. [PubMed: 12734548]

Soutar CA, Robertson A, Miller BG, Searl A, Bignon J. Epidemiological evidence on the carcinogenicity of silica: factors in scientific judgement. Ann Occup Hyg. 2000; 44:3–14. [PubMed: 10689755]

Steenland K, Fletcher T, Savitz DA. Epidemiologic evidence on the health effects of perfluorooctanoic acid (PFOA). Environ Health Perspect. 2010; 118:1100–1108. [PubMed: 20423814]

Stone EA, Ayroles JF. Modulated modularity clustering as an exploratory tool for functional genomic inference. PLoS Genet. 2009; 5:e1000479. [PubMed: 19424432]

Travlos GS, Hard GC, Betz LJ, Kissling GE. Chronic progressive nephropathy in male F344 rats in 90-day toxicity studies: its occurrence and association with renal tubule tumors in subsequent 2-year bioassays. Toxicol Pathol. 2011; 39:381–389. [PubMed: 21422264]

Vis DJ, Westerhuis JA, Hoefsloot HC, Pijl H, Roelfsema F, van der Greef J, Smilde AK. Endocrine pulse identification using penalized methods and a minimum set of assumptions. Am J Physiol Endocrinol Metab. 2010; 298:E146–E155. [PubMed: 19861586]

Want EJ, Wilson ID, Gika H, Theodoridis G, Plumb RS, Shockcor J, Holmes E, Nicholson JK. Global metabolic profiling procedures for urine using UPLC-MS. Nat Protoc. 2010; 5:1005–1018. [PubMed: 20448546]

Weis BK, Balshaw D, Barr JR, Brown D, Ellisman M, Lioy P, Omenn G, Potter JD, Smith MT, Sohn L, Suk WA, Sumner S, Swenberg J, Walt DR, Watkins S, Thompson C, Wilson SH. Personalized exposure assessment: promising approaches for human environmental health research. Environ Health Perspect. 2005; 113:840–848. [PubMed: 16002370]

Whitehead T, Metayer C, Buffler P, Rappaport SM. Estimating exposures to indoor contaminants using residential dust. J Expo Sci Environ Epidemiol. 2011

Whitrow MJ, Smith BJ, Pilotto LS, Pisaniello D, Nitschke M. Environmental exposure to carcinogens causing lung cancer: epidemiological evidence from the medical literature. Respirology. 2003; 8:513–521. [PubMed: 14629658]

Willett WC. Dietary fat plays a major role in obesity: no. Obes Rev. 2002; 3:59–68. [PubMed: 12120421]

Willett WC, Colditz GA, Hiatt RA. Combating environmental causes of cancer. N Engl J Med. 2011; 364:2266. author reply 2267–2268. [PubMed: 21651410]

Wright AL, Holberg CJ, Taussig LM, Martinez FD. Factors influencing the relation of infant feeding to asthma and recurrent wheeze in childhood. Thorax. 2001; 56:192–197. [PubMed: 11182011]

Yu T, Park Y, Johnson JM, Jones DP. apLCMS--adaptive processing of high-resolution LC/MS data. Bioinformatics. 2009; 25:1930–1936. [PubMed: 19414529]

Ziech D, Franco R, Pappa A, Malamou-Mitsi V, Georgakila S, Georgakilas AG, Panayiotidis MI. The role of epigenetics in environmental and occupational carcinogenesis. Chem Biol Interact. 2010; 188:340–349. [PubMed: 20599843]
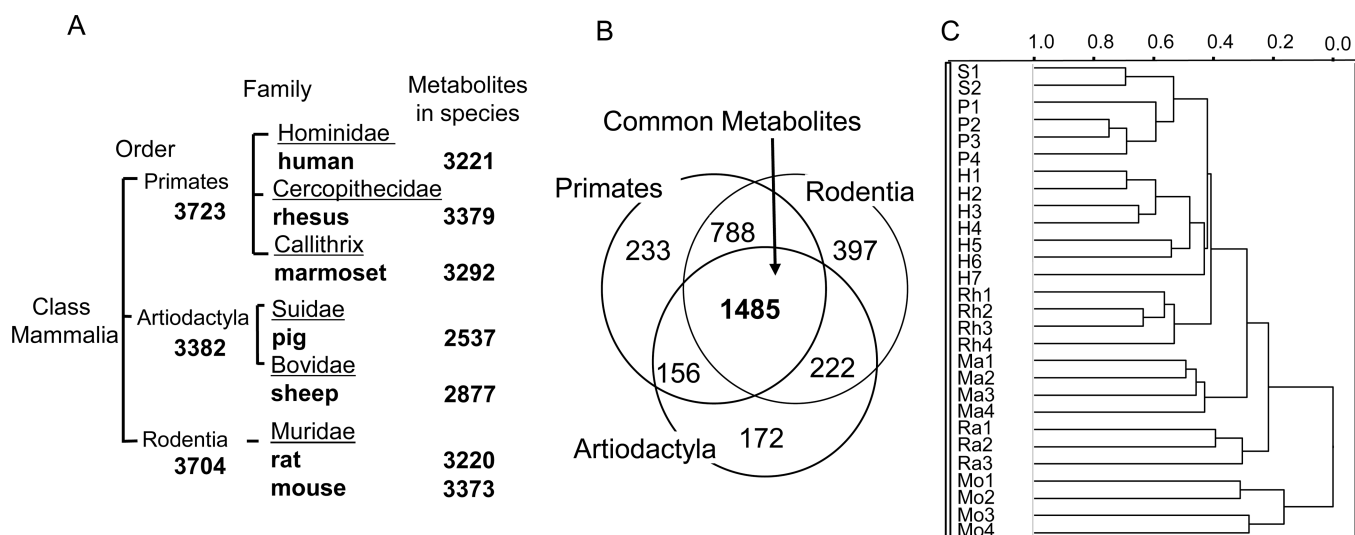
A

	Family	Metabolites in species
Order	Hominidae	
Primates 3723	**human**	**3221**
	Cercopithecidae	
	rhesus	**3379**
	Callithrix	
Class Mammalia	**marmoset**	**3292**
Artiodactyla 3382	Suidae	
	pig	**2537**
	Bovidae	
	sheep	**2877**
Rodentia 3704	Muridae	
	rat	**3220**
	mouse	**3373**

B

Common Metabolites

Primates   Rodentia

233   788   397

1485

156   222

Artiodactyla   172

C

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.
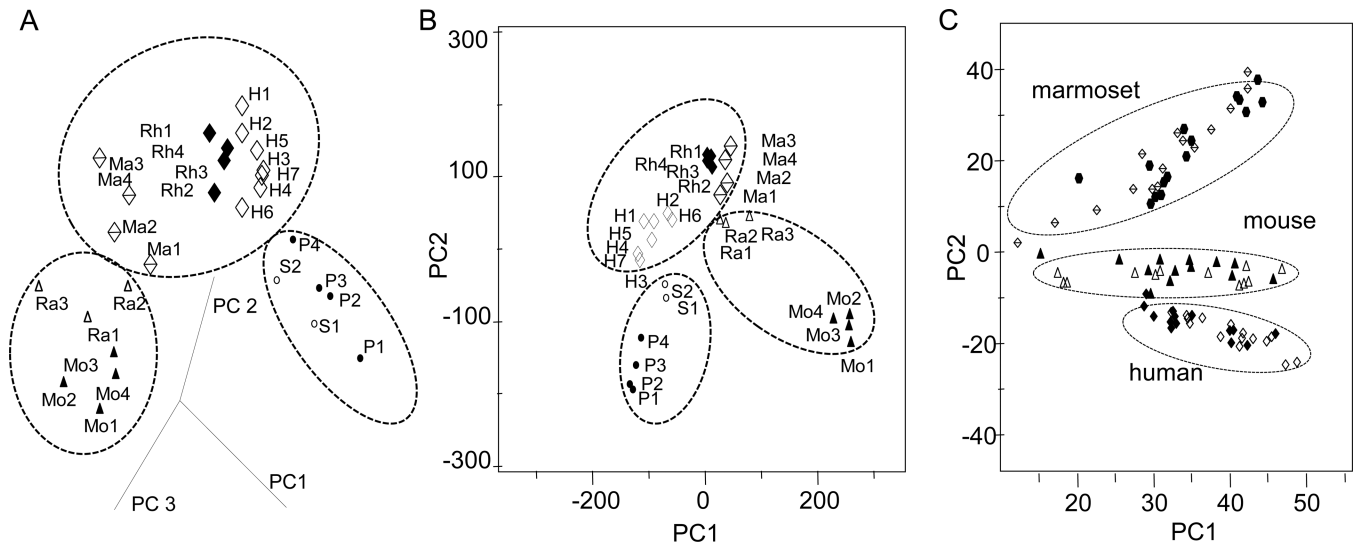
Figure 2.
A. Three-dimensional PCA score plot, in which the first three principal components (PCs) explained 62% of total variation, is arbitrarily rotated to visualize discrimination according to species, family and order. B. Two-dimensional PCA score plot with PC1 and PC2 corresponding to Panel D. 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.
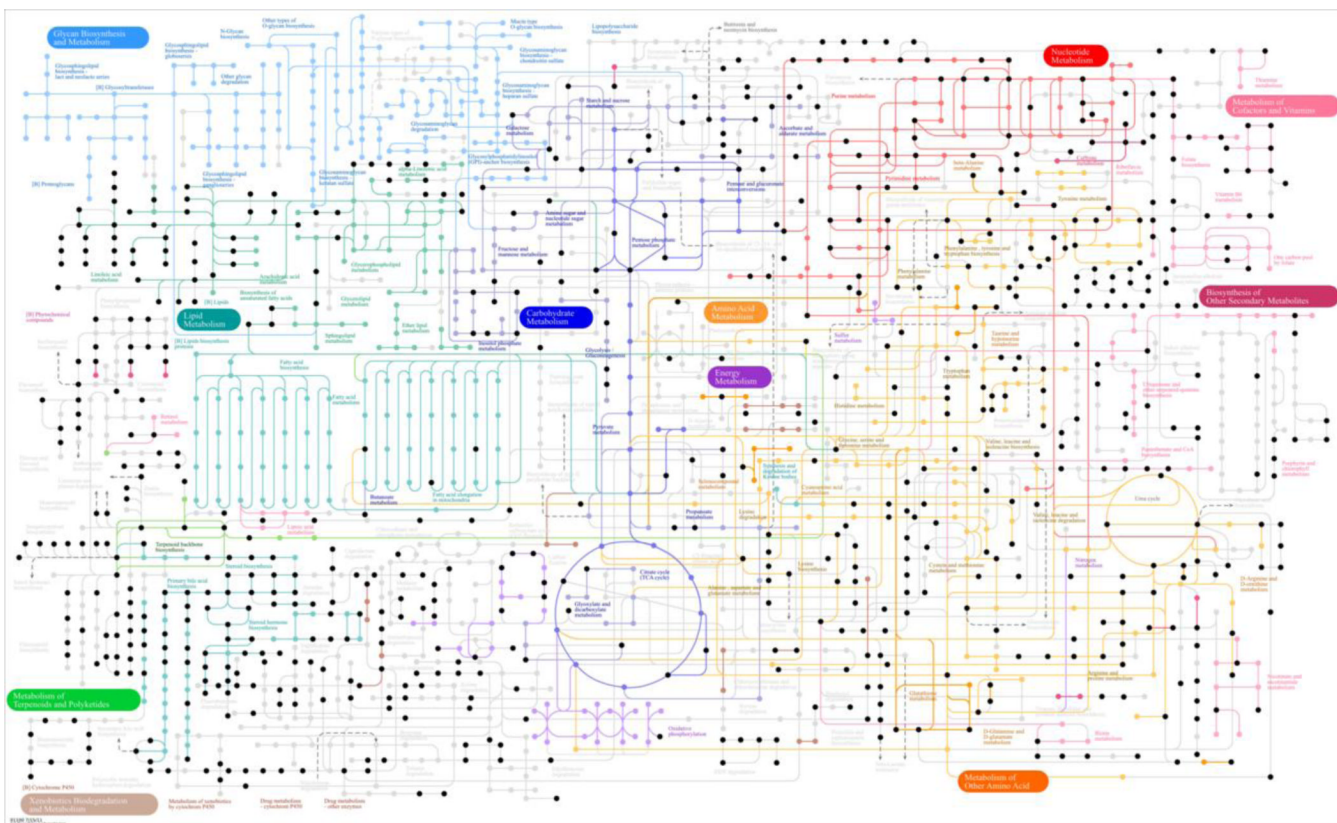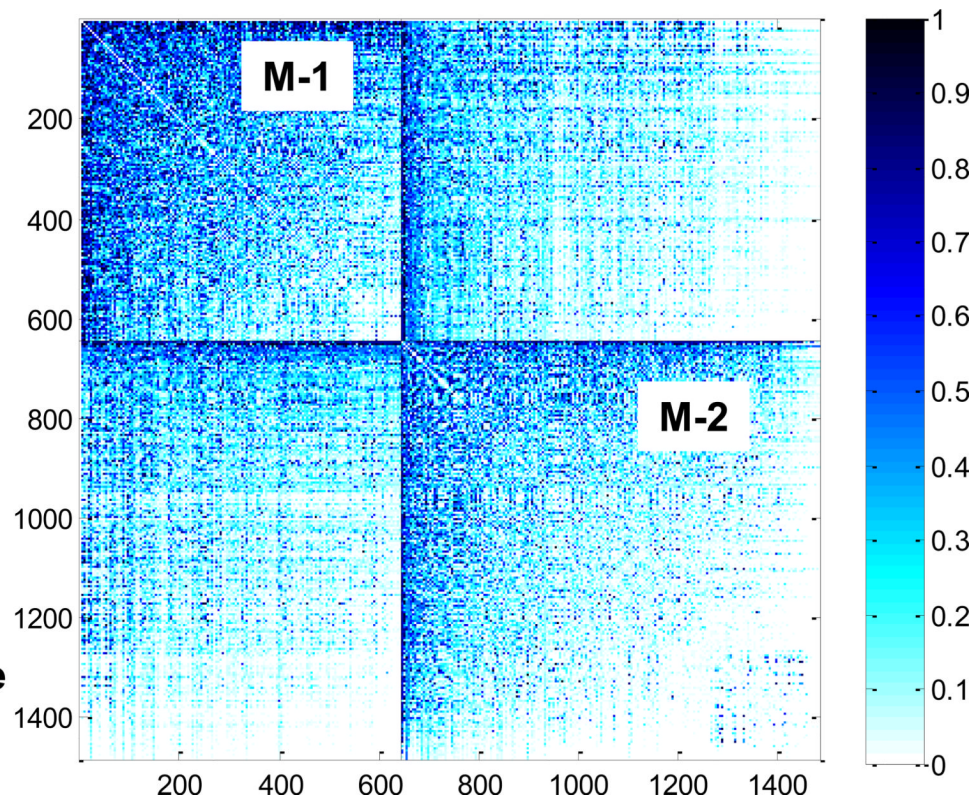
Figure 3.
KEGG metabolic pathway matches for *m/z* features that are common to mammalian species. High mass accuracy *m/z* for 1485 common chemicals revealed 666 matches as shown in black, including matches for metabolites in 137 of the total 146 KEGG human reference metabolic pathways. Corresponding metabolite names are available at http://www.genome.jp/kegg-bin/show_pathway?13154256908831/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.
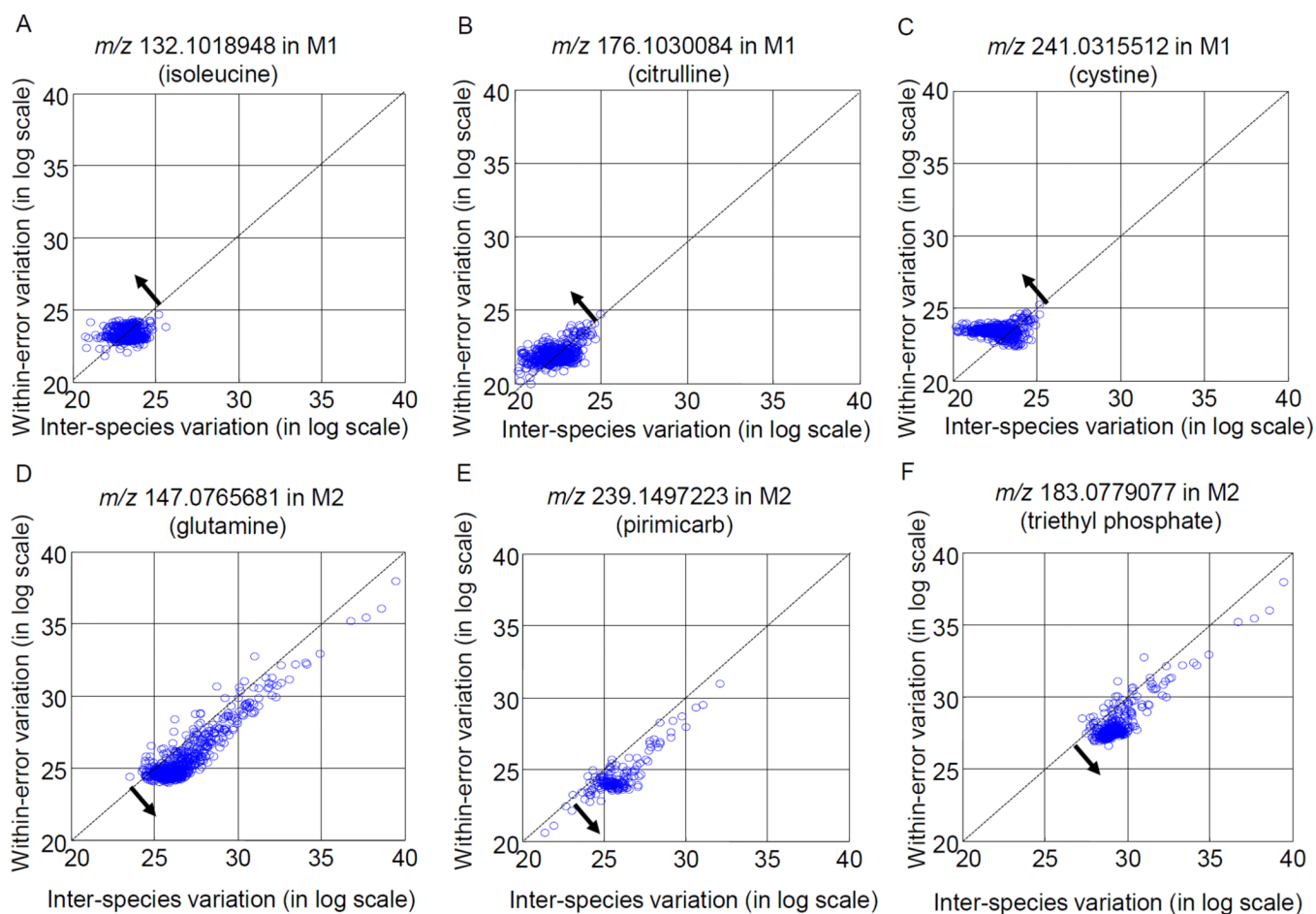
Figure 5.
The relationship between inter-species variation and within-error variation for selected chemicals in Module 1 (M1) and 2 (M2) of probability-based modularity clustering analysis shown in Fig 3. Trajectories above the diagonal line, represented by the arrows in A–C, indicate that within-species 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).
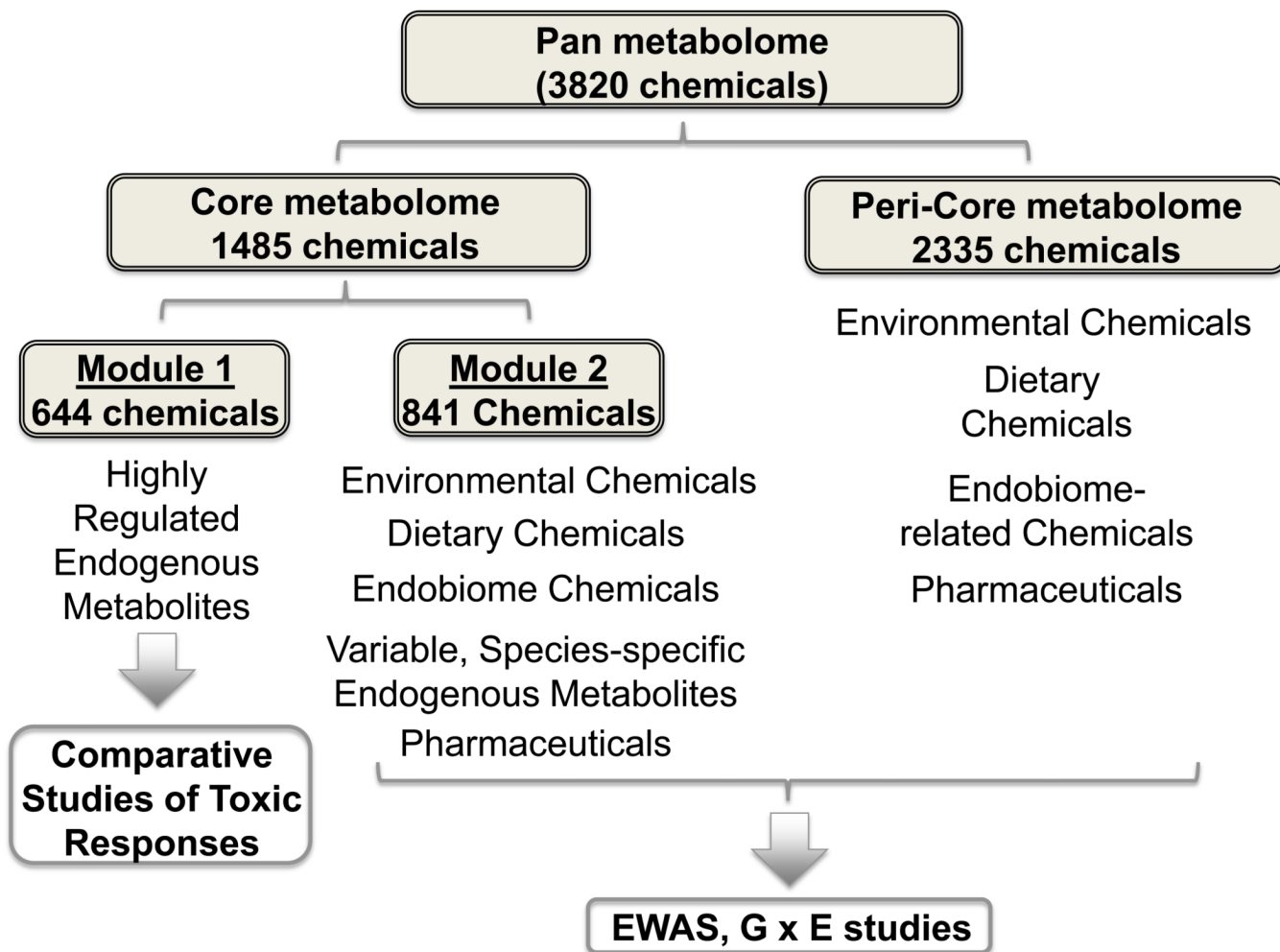
Figure 6.
Characterization of HPMP data for comparative toxicology and biomonitoring of exposures. HPMP analyses detected 3820 *m/z*, which reflects the total metabolome, designated here as the "pan" metabolome. The pan metabolome consists of a "core" metabolome that this is common to the 7 mammalian 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 × E) studies. Identification and classification of chemicals in Module 2 and the peri-core metabolome will greatly facilitate interpretation of EWAS and G × E study results.