



Sequencing the exposome: A call to action[☆]



Dean P. Jones^{*}

Department of Medicine, Emory University, Whitehead Biomedical Research Building, Suite 205P, 615 Michael Street, NE Atlanta, GA 30322, USA

ARTICLE INFO

Article history:

Received 25 November 2015

Accepted 25 November 2015

Available online 28 November 2015

Keywords:

Mass spectrometry

Biomonitoring

Analytical chemistry

Metabolomics

Environmental surveillance

ABSTRACT

The exposome is a complement to the genome that includes non-genetic causes of disease. Multiple definitions are available, with salient points being global inclusion of exposures and behaviors, and cumulative integration of associated biologic responses. As such, the concept is both refreshingly simple and dauntingly complex. This article reviews high-resolution metabolomics (HRM) as an affordable approach to routinely analyze samples for a broad spectrum of environmental chemicals and biologic responses. HRM has been successfully used in multiple exposome research paradigms and is suitable to implement in a prototype universal exposure surveillance system. Development of such a structure for systematic monitoring of environmental exposures is an important step toward sequencing the exposome because it builds upon successes of exposure science, naturally connects external exposure to body burden and partitions the exposome into workable components. Practical results would be repositories of quantitative data on chemicals according to geography and biology. This would support new opportunities for environmental health analysis and predictive modeling. Complementary approaches to hasten development of exposome theory and associated biologic response networks could include experimental studies with model systems, analysis of archival samples from longitudinal studies with outcome data and study of relatively short-lived animals, such as household pets (dogs and cats) and non-human primates (common marmoset). International investment and cooperation to sequence the human exposome will advance scientific knowledge and also provide an important foundation to control adverse environmental exposures to sustain healthy living spaces and improve prediction and management of disease.

© 2015 The Author. Published by Elsevier Ireland Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction and definitions

The exposome is the cumulative measure of environmental influences and biological responses throughout the lifespan [28]. The definition is inclusive, consistent with Christopher Wild's concept [44] to broadly complement the genome with non-genetic factors impacting human health. Measurements of all environmental, dietary, microbiome, behavioral, therapeutic and endogenous processes present a daunting challenge for systematic study, especially when considered cumulatively throughout life.

Less than a half-century ago, sequencing the human genome was unimaginable by few scientists. Yet visionary leaders championed this goal, societies and industries invested in new technologies, and success was attained. Despite monumental barriers, earlier generations succeeded in quests to control communicable diseases,

eradicate smallpox, prevent polio, cure childhood leukemia, etc. Sequencing the human exposome is a formidable challenge for contemporary science and technology, but with vision and commitment, this is attainable.

Wild [44] provided a thoughtful outline to address the challenge, and others have emphasized the need for critical thought and design to pursue this goal [34,1,45]. In the present article, I provide a brief review on environmental applications of high-resolution metabolomics (HRM; see Table 1 for acronym definitions) and comment on the potential to use this as a central framework to initiate sequencing the human exposome. I use the term “sequencing” to emphasize the time domain of the exposome; the cumulative measure of exposures for an individual cannot be sequenced in the same way as the genome can be sequenced. But there are many types of exposure memory systems [17], and there is little doubt that improved understanding of epigenetics and other exposure memory systems will allow certain aspects of an individual's exposome to be measured retrospectively. So the present review of HRM is not intended as a final chapter on how to sequence the exposome, but rather as a place to begin.

I first describe the use of high-resolution mass spectrometry for clinical metabolic profiling and discuss advantages and

[☆] This article was presented in a session entitled “Genome, Epigenome and Exposome” at the 45th Annual Symposium of the Society of Toxicology of Canada, Ottawa, December 4, 2013.

^{*} Corresponding author.

E-mail address: dpjones@emory.edu

limitations. This is followed by a discussion of computational workflows, which support both targeted and discovery analyses using commonly accessible biological samples and advanced informatics methods. I then consider use of this analytical platform as a universal surveillance tool to monitor environmental exposures and evaluate biological effects. Finally, I discuss HRM as a possible central element of a “Human Exposome Project” to sequence the cumulative environmental influences and biological responses throughout lifespan. This is presented as a call to the international environmental health research community to champion this effort and work together in this common goal. At Emory University, in collaboration with Georgia Institute of Technology and with support of the National Institute of Environmental Health Sciences, we are building toward this goal through the HERCULES Exposome Research Center directed by Gary W. Miller, Ph.D. (<http://humanexposomeproject.com>).

2. Rationale for development of high-resolution metabolomics

Analytical traditions, as well as regulatory and policy needs of government, have limited flexibility to address important challenges in health and environmental research. During the period from 1958, when automated amino acid analysis was introduced [29], and 2007, when we began experimental development of advanced blood chemistry analyses with high-resolution mass spectrometry [16], analytical chemistry provided merely a ten-fold improvement in the number of chemicals that could be measured in a routine analysis of plasma or serum, i.e., from about 30 to 300. During this time period, DNA sequencing progressed from a complete inability to sequence the human genome to ability to accomplish the task within a few days.

While considerable financial investment through the Human Genome Project contributed importantly to this success, practical differences from analytical chemistry also existed in the scope of the human genome initiative and in the tolerance of the project for errors. Genome sequencing was intended to be more or less complete and was pursued even for genes without known function. Additionally, whole genome sequencing was developed with an expectation for errors in sequencing and assembly. Instead of abandonment because methods were inadequate or errors were common, progress was made by embracing new methods and

developing new approaches to address errors. Although imperfect in many ways, this resulted in overall success.

In contrast, progress in analytical coverage of small molecules in biologic systems appears to have languished more due to analytical traditions than to limitations in technology. With a focus on matching proteomic capabilities to those of genomics, tremendous advances in technology were achieved in mass spectrometry (MS). A byproduct of the successes of proteomics was the transformation of MS capabilities for detection and measurement of small molecules.

2.1. Mass spectrometry for chemical profiling

Mass spectrometry (MS) has a special role in analytical chemistry because the mass of a chemical is an absolute property. Thus, if the measured mass does not match that of the purported chemical, then the chemical identification is incorrect. MS involves measurement of chemicals or derived fragments of chemicals as ions (m/z , mass-to-charge ratio) in the gas phase [13]. The ions can be formed by interaction of a neutral chemical with H^+ , Na^+ or other cation, by loss of H^+ as occurs with ionization of carboxylic acids, or by dissociation of a chemical into product ions. With introduction of electrospray ionization (ESI) [47], routine measurement of a very broad range of small molecules in biological materials without extensive fragmentation became practical.

Mass spectrometers have three components, an ion source, which generates ions; a mass analyzer, which separates ions according to m/z ; and a mass detector, which measures ion intensity, i.e., the amount of ions with the respective m/z (Fig 1A). Many physical and chemical principles are used for MS instruments, and a recent overview is available [25]. Mass spectrometers differ in sensitivity to detect low abundance ions, resolve ions with very similar m/z , and provide accurate estimates of m/z . For instruments with poorer mass resolution and mass accuracy, chromatographic separation prior to MS is necessary to allow measurement of chemicals with the same unit mass (Fig 1B). Both gas chromatography (GC) and liquid chromatography (LC) are often coupled to MS to improve measurement of chemicals according to characteristic retention time and m/z (Fig 1A and B).

Quadrupole mass analyzers, often designated simply by “Q”, selectively stabilize or destabilize the paths of ions passing through oscillating electrical fields between 4 parallel rods. Common

Table 1
Abbreviations and acronym definitions [For additional MS details, see [30]].

Term	Definition	Term	Definition
AMU	Atomic mass unit	m/z	Mass to charge ratio for an ion measured by mass spectrometry
apLCMS	Adaptive processing algorithms for extraction of MS data	MB	Maneb, a fungicide
Asp	Aspartate	MS	Mass spectrometry
Cd	Cadmium	MS/MS	First product ion spectrum from ion dissociation MS; same as MS ²
CV	Coefficient of variation	MS ¹	Spectrum of a precursor ion
ESI	Electrospray ionization	MS ²	First product ion spectrum from ion dissociation MS; same as MS/MS
FDR	False discovery rate	MS ⁿ	Product ion spectra from a sequence of ion dissociations
FT	Fourier transform	MWAS	Metabolome-wide association study
FT-ICR	Type of mass analyzer	PBDE	Polybrominated diphenyl ether,
GC	Gas chromatography	PD	Parkinson disease
GC-MS	Combined gas chromatography and mass spectrometry	Phe	Phenylalanine
GWAS	Genome-wide association study	PLS-DA	Partial least squares-discriminant analysis
HIV	Human immunodeficiency virus	PQ	Paraquat, an herbicide
HMDB	Human metabolome database	Q	Quadrupole, a type of mass analyzer
HPLC	High performance liquid chromatography	Q-TOF	Tandem mass spectrometer combining quadrupole and TOF
HRM	High-resolution metabolomics	SOP	Standard operating procedures
ICR	Ion cyclotron resonance, a type of mass analyzer	SRM1950	National institute of standards pooled reference human plasma
Ile	Isoleucine	TMWAS	Transcriptome metabolome-wide association study
KEGG	Kyoto encyclopedia of genes and genomes	TOF	Time-Of-Flight, a type of mass analyzer
LC	Liquid chromatography	Tyr	Tyrosine
LC-MS	Combined liquid chromatography and mass spectrometry	UHRAM	Ultra-high resolution accurate mass, used in reference to FT MS
Leu	Leucine	XCMS	Algorithms for MS data extraction
LIMMA	Linear models for microarray, software for differential expression	xMSanalyzer	Algorithms to improve data extraction by apLCMS or XCMS

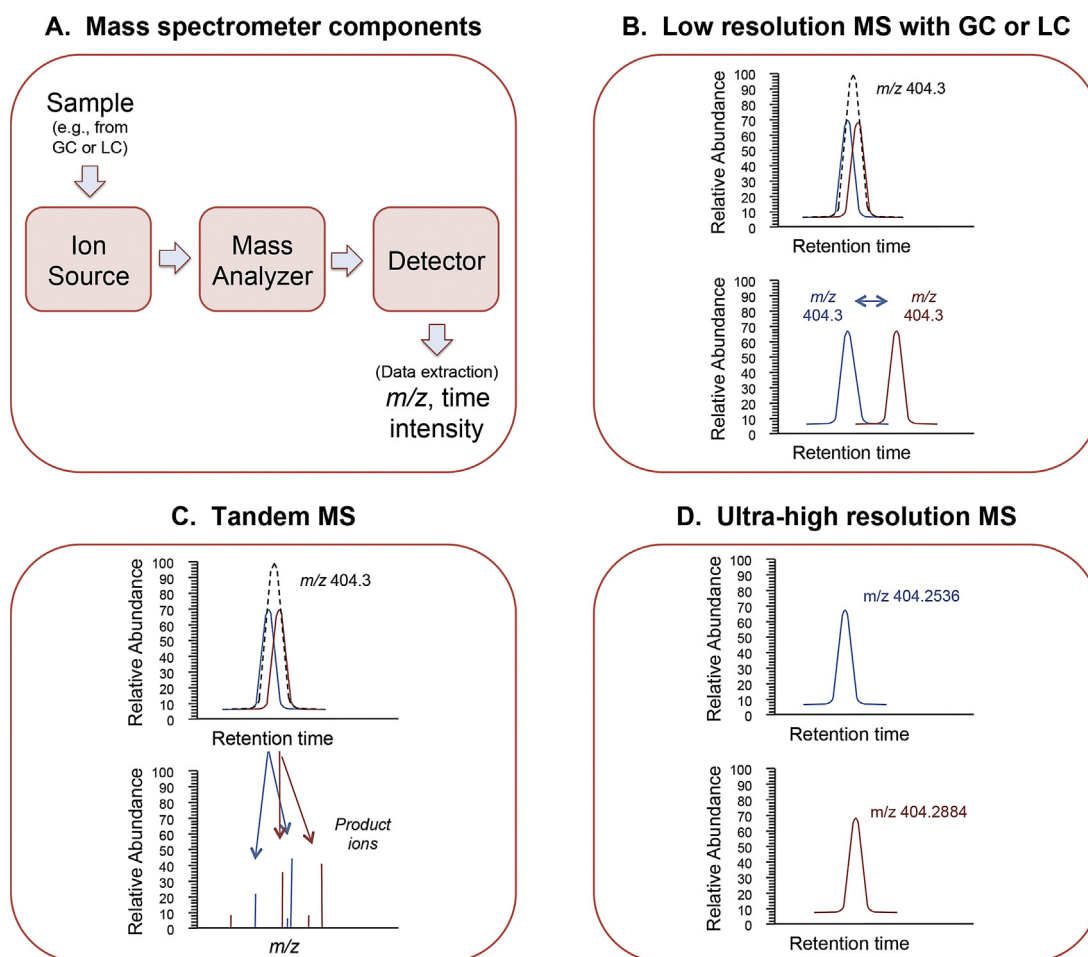


Fig. 1. Mass spectrometry (MS) for metabolomics. A. Mass is a fundamental characteristic of a chemical, and mass spectrometers measure mass by converting chemicals to ions in the gas phase and measuring the movement of the ions in an electromagnetic field. Mass spectrometers have an ion source to generate ions, a mass analyzer that separates ions according to mass to charge ratio (m/z) and a detector to quantify the intensity of respective ions with corresponding m/z . For metabolomics, samples are usually fractionated prior to delivery to the ion source by gas chromatography (GC) or liquid chromatography (LC). B. Many mass spectrometers measure m/z with ± 0.1 atomic mass unit (AMU), which is not sufficient to distinguish chemicals with very similar mass. Consequently, these instruments require separation by GC, LC or other methods prior to mass spectral analysis. Such configurations are designated by hybrid terms: GC-MS or LC-MS. C. Tandem mass spectrometry (e.g., triple quadrupole and Q-TOF instruments) involves use of combinations of mass spectrometry components to obtain m/z measurements on an ion and then subsequent measurement of m/z for product ions generated following ion dissociation. In some instruments, this process can be repeated multiple times (MS^n) to gain additional structural information; for quantification, the first ion dissociation (MS/MS or MS^2) is often used for targeted chemical analysis because it allows quantification of specific chemicals based upon product ions even when the precursor ion is not separated from chemicals with very similar mass. D. Ultra-high resolution accurate mass (UHRAM) mass spectrometers resolve ions and measure m/z much more precisely than other mass spectrometers. This mass resolution and mass accuracy simplifies separation requirements and provides improved capability to measure low abundance chemicals in complex matrices such as human plasma. Panels B–D were modified from [18] with permission.

variations include tandem mass spectrometers where sequential elements allow selection and dissociation of precursor ions into product ions. Simplified terminology for ion dissociation mass spectrometry is to refer to the spectrum of the precursor ion as MS or MS^1 , the spectrum of the products of ion dissociation as MS/MS or MS^2 , and subsequent dissociation spectra as MS^n , where n represents the corresponding fragmentation sequence. Tandem mass spectrometers include triple quadrupole mass spectrometers (“triple quad”) with three quadrupole stages, and Q-TOF mass spectrometers, which combine selection and fragmentation using quadrupoles and time-of-flight (TOF) analyzers. The TOF analyzers provide greater mass accuracy than quadrupoles and are very popular for metabolomics analyses (Fig 1C).

Ion traps use oscillating electric fields within a variety of two- and three-dimensional geometries to trap ions for detection and ion dissociation. Ion traps are very powerful for ion dissociation studies and also to accumulate selected ions for subsequent high-resolution analysis in combination with ultra-high resolution mass detectors. Ultra-high resolution mass detectors measure mass

through Fourier transformation of the image current of ions cycling within an electromagnetic field. These include ion-cyclotron resonance (ICR) mass spectrometers and Orbitraps (Thermo-Fisher). The ultra-high resolution and mass accuracy provide an advantage for high-resolution metabolomics by decreasing the separation requirements of GC-MS and LC-MS, and increasing the range of low abundance chemicals that can be measured compared to MS/MS (Fig 1D). Some of the Q-TOF instruments provide similar accurate mass for high abundance ions but are not as effective in resolving ions with very similar mass at low abundance. An advantage of Q-TOF instruments lies in fast scan speeds, and there is likely benefit from a union of technologies to sequence the exposome.

Many factors influence ability to accurately measure chemicals of interest, and most are not unique to MS, e.g., methods of sample collection and storage, avoidance of contamination and dynamic range of detection. For MS, additional factors include ionization, ability to resolve chemicals prior to or during analysis, and ability to measure m/z accurately. With advances in all aspects of separation sciences, spectrometry and physical and chemical analysis,

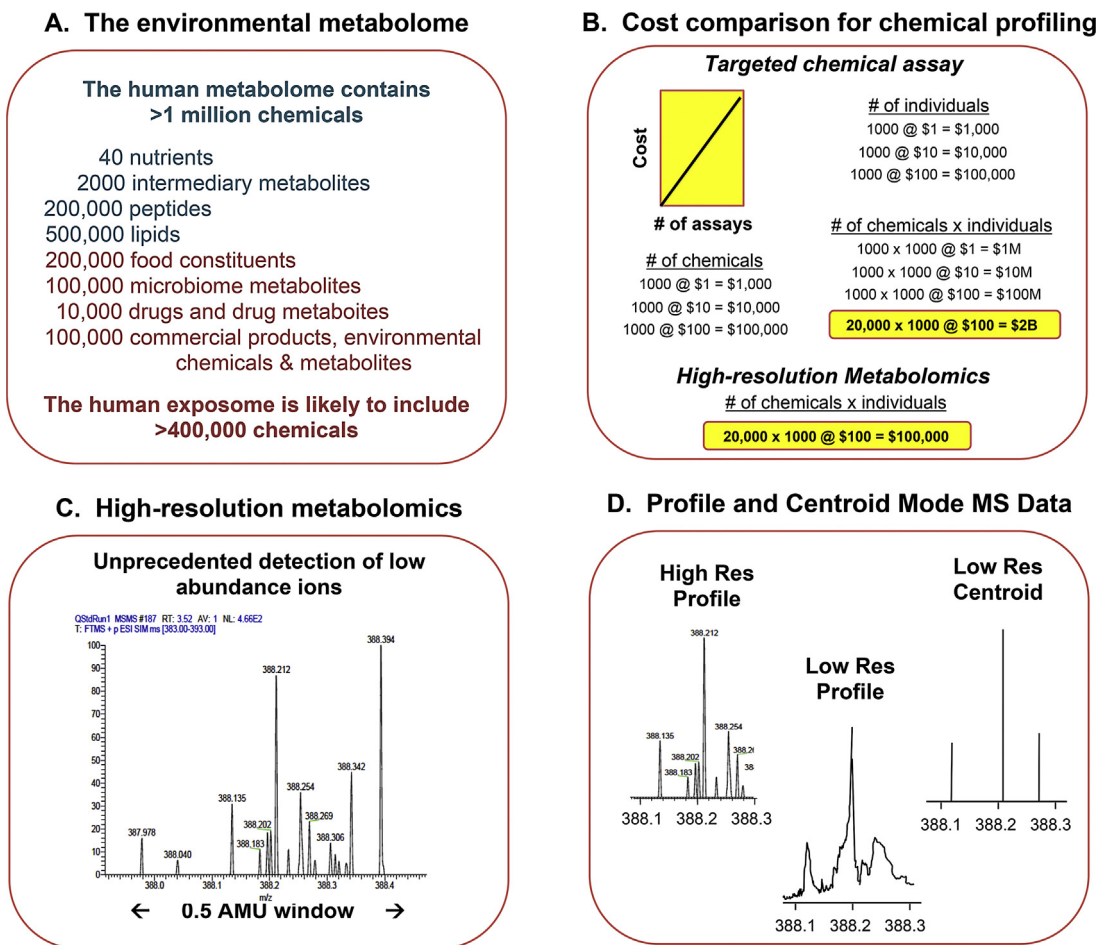


Fig. 2. High-resolution metabolomics (HRM) for advanced chemical profiling. A. The human metabolome is complex and likely to include >400,000 environmental chemicals. B. The cost for measurement of large numbers of environmental chemicals is impractical with targeted analytical methods but can be affordable if large numbers of chemicals are measured in a single analysis. C. High-resolution metabolomics uses high-resolution mass spectrometry with liquid chromatography to measure >20,000 chemicals based upon high mass resolution and high mass accuracy. D. An important advantage of high-resolution mass spectrometry is that data in profile mode contains more information than commonly used centroid mode. The centroid mode decreases the m/z window to zero, thereby being much more efficient for data storage. As this simulation shows, ions resolved by a high-resolution (High Res) mass spectrometer are not resolved by a low-resolution (low Res) instrument. When the latter is expressed in centroid mode, only 3 of 9 ions are detectable.

analytical techniques are available to address almost any complexity concerning targeted analyses of chemicals of interest. Despite this, there is no available strategy to provide comprehensive coverage of chemicals in biological space, i.e., to accomplish chemical measurement in the same way as one can obtain comprehensive coverage of base-pair sequences in genomic space.

In fairness, the scope of analytical chemistry never included intent to provide platforms to measure everything; such a thought is absurd given that chemical space appears to be infinite [21]. Perhaps more relevant, half a century ago, public concern over health effects of chemical pollutants led to government policies to limit exposures. This occurred long before current analytical technologies were available, and policies were developed based upon technologies available at the time. And even with more limited technology, an effective strategy was developed to identify hazards, evaluate sources and risks of exposure, and develop means to monitor and limit hazardous exposures. This approach involved prioritization of hazards and risks, focusing on known hazards and developing rigorous analytical methods for these. This approach is sound and proved effective, using resources wisely for targeted analysis of chemicals of greatest concern.

This targeted strategy is limited by cost to address possible risks from large numbers of hazardous chemicals. Limitations are also evident to discover chemicals that are hazardous at lower expo-

sure levels, and also those that are toxic due to chemical–chemical or gene–environment interactions. Small molecules present in biologic systems are cumulatively described as the metabolome, and include essential nutrients, gene-directed products of these nutrients, dietary constituents, products of intestinal microbes, drugs and related metabolites and commercial and environmental chemicals (Fig 2A) [18]. About 100,000 agents are registered for commercial use with the US Environmental Protection Agency, and these plus drugs, supplements and related microbial and geothermal products, suggest that the environmental metabolome is likely to consist of 400,000 or more chemicals. Although most of these pose no environmental health risk, increases in prevalence of many diseases over the past half-century indicate that unidentified hazards are present [36].

Analysis of costs shows that targeted analysis of large numbers of chemicals broadly in populations is unaffordable (Fig 2B). For instance, if analyses are done individually, total cost increases as a function of the number of chemicals measured (Fig 2B). This restricts the number of chemicals measured to the most hazardous. For difficult chemicals, like dioxins, cost can be >US\$4000 per sample, so global sampling of populations is impossible. Consequently, routine environmental chemical surveillance by targeted methods is affordable only for a relatively small number of the most hazardous chemicals, and measurement of these is limited to rep-

representative sampling for populations. There is no opportunity to routinely monitor lesser hazards, evaluate mixtures for which no compelling evidence for hazard is present, or address unrecognized hazards.

A lesson from the Human Genome Project is that focus on a small number of hazards is not good enough. Prior to sequencing the human genome, many believed that knowledge of genetic variations would rapidly lead to understanding human disease because mutations like those causing phenylketonuria or sickle cell disease produced very evident disease consequences. As genome-wide association studies (GWAS) became available for common diseases, however, results showed that most genetic variations associated with disease have a small effect size. Individual variations often account for less than 1% of the overall risk of disease, and genetic associations of disease are not detectable unless studies are conducted with 20,000 or 30,000 individuals. If environmental exposures similarly have small effect size, where individual chemicals each contribute to less than 1% of overall risk of disease, then targeted analysis of high-risk chemicals in small populations provides no way to detect or monitor this type of exposure-related risk. Changes in prevalence of obesity, diabetes, autism, childhood brain cancer, breast cancer, Alzheimer's disease, parathyroid disease, and other disease processes could be linked to changes in multiple unidentified chemical exposures, each with small effect size. Hence, there is need to approach environmental causes of disease in the same way that geneticists found necessary to understand genetic factors: develop a comprehensive framework which can be applied to large populations.

3. High-resolution metabolomics

An advanced but demanding (liquid helium temperature) mass spectrometry method was developed in 1974 that provided a foundation to overcome critical limitations to measurement of thousands of small molecules in an affordable manner. Comisaro and Marshall [3] developed Fourier-transform ion cyclotron resonance (FT-ICR) MS, an approach in which ions were cycled within a high-field electromagnet, to enable separation of ions with m/z differing by <1 ppm. This means that instead of measuring an m/z with the accuracy 200.0 ± 0.1 , the ion would be measured with the accuracy 200.0000 ± 0.0002 . For many low-molecular mass chemicals, mass accuracy can be sufficient to predict elemental composition. Measurement of ions with this ultra-high mass resolution and mass accuracy provides a way to bridge between targeted analyses of chemicals of known importance and capture of information about uncharacterized chemicals present in biologic systems (Fig 2C). About 10 years ago a new form of ion trap (Orbitrap; ThermoFisher) was introduced which also provides ultra-high resolution and accurate mass detection but is less demanding for operations (does not require liquid helium). The detection and conversion to m/z also involves Fourier transformation so these can be generically described as FT instruments to distinguish them from time-of-flight (TOF) methods, some of which are also high-resolution instruments [26]. ThermoFisher has several FT instruments and several manufacturers (AB Sciex, Agilent, Bruker, Shimadzu) market high-resolution TOF instruments; comparison of the types of high-resolution instruments is available [26].

3.1. Advantages of ultra-high resolution MS

The importance of the mass accuracy and mass resolution is apparent when considering measurement of metabolites in the central human pathways of the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [19]. Of approximately 2000 metabolites, >95% have unique m/z when measured with ± 5 ppm accuracy.

This substantially simplifies analysis because most of these can be measured with minimal chromatographic separation. With known chemicals in common biologic samples, such as plasma, this often eliminates need for ion dissociation (MS^n) mass spectrometry for identification and measurement. In development of procedures, we studied five aminothiols differing by more than 3 orders of magnitude in abundance in human plasma [15]. We found that these could be resolved and identified without ion dissociation MS and that quantification was comparable to other analytical methods [15].

Building upon this quantitative framework, which showed that the intensity of accurate mass m/z features could be directly calibrated to yield reliable concentration measurements, we developed data extraction algorithms to enhance measurement of other chemicals [48]. New data extraction algorithms were needed to maximize information capture on low abundance metabolites. For extraction of low abundance m/z features, many ions are detected within 0.1 AMU of each other (Fig 2C). Data can be extracted in different ways, with a commonly used data extraction producing centroid mode data in which m/z for a feature is collapsed to zero width to decrease size of data files for processing and storage. In contrast, extraction of data in profile mode yields better data recovery than centroid mode because more information is retained for signals as a function of m/z (Fig 2D). Use of centroid mode for data extraction can result in merger of different ions into a single m/z , easily mistaken for a single, relatively abundant chemical; in some cases, signals can be merged to create the m/z of a chemical that is not present. For analyses of samples with unidentified and/or unknown low abundance environmental chemicals, profile mode provides a more reliable starting point for data extraction [48].

3.2. Improved data extraction algorithms

Starting with profile mode data, adaptive processing algorithms for liquid chromatography-mass spectrometry (apLCMS) were developed to improve the number and reproducibility of detected metabolites [48]. The improvement in chemical detection was apparent immediately, with an increase from about 300 chemicals to more than 1500 (Fig 3A). apLCMS uses the data characteristics, such as noise level and peak shapes, to guide data extraction. Improvements have included xMSAnalyzer, a routine to use different extraction parameters and statistical filters to merge results to enhance coverage and reproducibility [42]. A hybrid version of apLCMS has also been developed to support targeted data extraction in combination with untargeted extraction [49].

3.3. Applications for precision medicine

Application of apLCMS to human plasma samples measured on an FT-ICR MS with a 10 min separation on anion exchange chromatography showed that >1500 reproducible metabolic features were measured, where a feature is defined as an accurate mass m/z with associated retention time (RT) and ion intensity [16]. This definition of a metabolic feature is sufficiently robust to allow reproducibility of measurement at different times and on different instruments. By inclusion of internal standards to allow calibration of RT, unidentified chemicals can be retrospectively identified based upon the accurate mass m/z and RT. Because the >1500 features include amino acids, energy metabolites, vitamins and coenzymes, lipids and many other metabolites relevant to human health, the analytical platform provides a practical approach for personalized medicine [16].

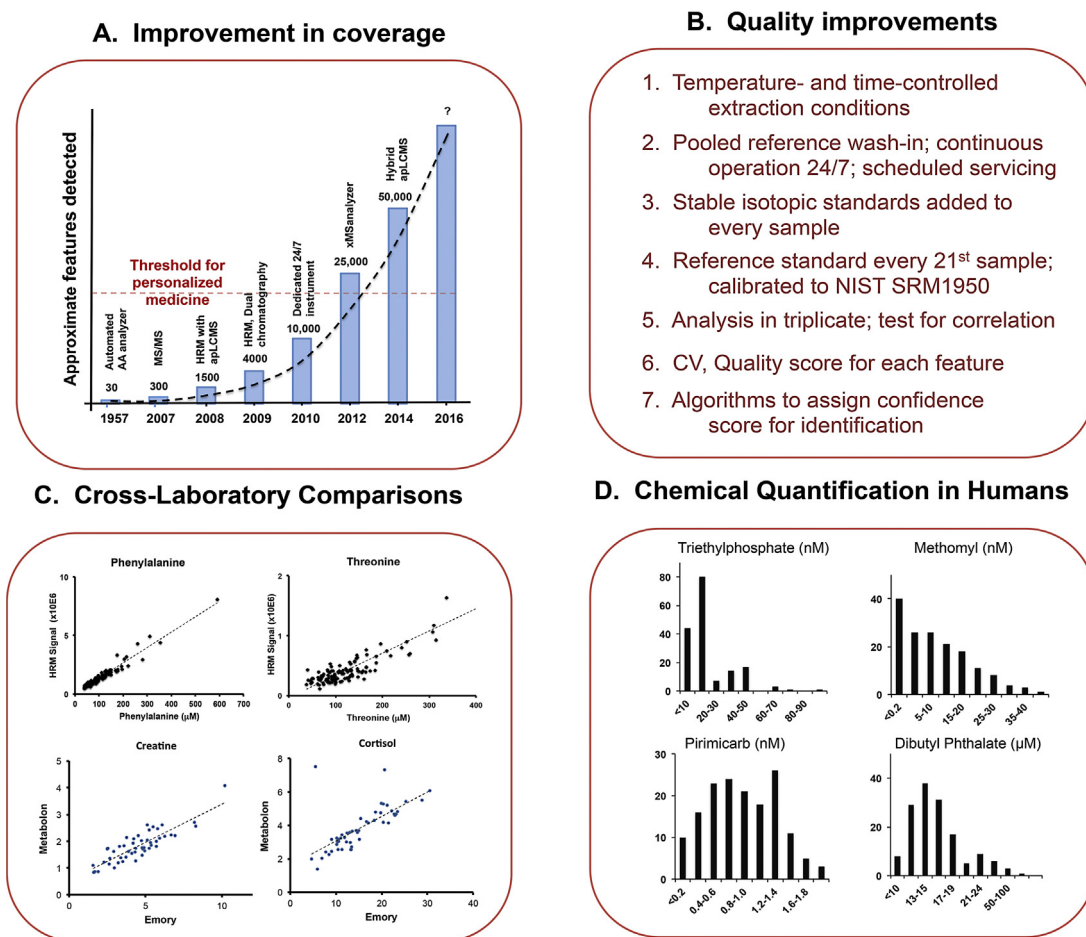


Fig. 3. Recent improvements in high-resolution metabolomics. A. Improved coverage of chemicals in human plasma have occurred over the past several years due to improved data extraction algorithms, improved instrumentation and improved standard operating procedures (graphic based upon figure prepared by DI Walker). B. Multiple steps have been introduced to improve consistency and data quality. C. Cross laboratory comparisons provide an effective approach to verify correct identification and evaluate quantification of metabolites. Phenylalanine (Phe) and threonine (Thr) comparisons were between HRM and amino acid measurements by the Emory Human Genetics Laboratory. Creatine and cortisol comparisons were between HRM and measurements by Metabolon. D. Quantification of environmental chemicals in human plasma by HRM. Chemical identities in human plasma were confirmed by co-elution with authentic standards and matched product ions in ion dissociation spectra. Quantification was performed in plasma from 163 healthy individuals by reference standardization with method of additions. Figures from [11] were used with permission conveyed through Copyright Clearance Center.

3.4. Improved affordability of chemical profiling with HRM

Cost considerations are highly relevant to analytical methods, and the aminothiols method described above is a useful example. The high-resolution mass spectrometry method required only 10 min for analysis, plus 10 min wash time, to obtain the same measurement as obtained in one hour by HPLC with fluorescence detection. However, the high-resolution mass spectrometry analysis required averaging of duplicates to obtain the same data quality, and the cost was US\$25/chemical, approximately five-fold the cost of the HPLC method (US\$4.40/chemical). However, if 2000 chemicals were measured, the cost would be only US\$0.06/chemical. Thus, value is derived from simultaneous measurement of many chemicals.

3.5. Limitations and strategies to improve HRM

During recent years, we have focused on costs, coverage and data quality as criteria to improve HRM. Although details are beyond the scope of the present summary, some points warrant comment. Cost per analysis is approximately cut in half by using a dual chromatography setup in which one LC column is washed and re-equilibrated while the mass spectrometer is used to analyze a sample on another

column [40]. With this approach, we were able to use two types of columns to increase detection to more than 4000 chemical features [40]. This was further increased to >10,000 by acquisition of an LTQ-Velos Orbitrap and dedicating use to HRM (Fig 3A).

With this configuration, data extraction using xMSanalyzer [42] with apLCMS [48] resulted in >20,000 chemical features. This number can be increased to >100,000 by allowing extraction of chemical features found in only a small fraction of the samples. This can be important, for instance, for chemicals found in only 1% or less of the population. The actual number of chemicals measured is difficult to determine because many chemicals are detected in different ion forms, e.g., H^+ and Na^+ . Most ions detected in plasma by HRM are of relatively low abundance, however, and these appear to be present only as one form. Consequently, >100,000 ions detected probably reflects >50,000 chemicals in the sample. By varying extraction parameters, results indicate that >300,000 ions are detected (K Uppal, unpublished), and by combining data from analyses using different chromatography and ionization methods, results indicate that >800,000 ions are detected (DI Walker, unpublished). Thus, the analytical capabilities with HRM are good enough to set a milestone for analytical chemistry: capability to analyze one million chemicals in small microliter volumes of plasma and other biologic materials.

3.6. Development of standard operating procedures (SOP)

With development of effective methods for measurement of such a large number of chemicals, focus must switch to quality of data. For this, we adopted LC–MS and data extraction procedures to routinely measure 20,000 ions [42]. For about 1000 ions, including most of the amino acids and many other intermediary metabolites, reproducibility of HRM is good enough ($CV < 10\%$) so that technical replicates are not needed. Our tests showed, however, that most m/z features had greater CV so that replicate analyses were needed to improve data quality. We adopted an SOP (Fig 3B) with three technical replicate analyses of each sample. With xMSanalyzer, approximately 20,000 m/z features are obtained with median CV of about 25%. Because triplicate analyses are performed, the accuracy of measurement is better expressed by the standard error of the mean, which is $< 15\%$. Under these conditions, the cost for each chemical feature is less than US\$0.005. Newer high-field Orbitrap instruments have improved scan speeds, and our initial studies indicate that lower cost can be obtained with similar quality using smaller columns, shorter run time and five replicate analyses of the same sample. Additionally, our comparison of different combinations of liquid chromatography columns and ion sources show that dual chromatography using positive ESI with HILIC and negative ESI with C18 provides the most extensive coverage within a single analysis (DI Walker et al., unpublished).

I view the routine measurement of 20,000 metabolites in a human sample as a critical threshold for personalized medicine. Already within the sequences of 20,000 genes, information is available to predict risk of many diseases. Similarly, knowledge of environmental chemical exposures can be used to predict disease risk so that, in principle, an analytical platform providing internal doses of environmental chemicals advances capabilities for risk prediction. Furthermore, specific gene-environment interactions have been identified so that the combination of detailed metabolite profiles with detailed genetic information can be expected to provide an enhanced analytic structure for risk evaluation and prediction.

Genetic information alone can provide unambiguous identification of an individual and is useful for risk prediction, but in the case of identical twins, where genetic information is identical, considerable discordance in health outcome can still occur. These differences appear to be due to differences in individual exposures. In terms of health prediction, combination of information on 20,000 human genes with information on 20,000 metabolites from HRM provides a genome \times metabolome interaction network of 400 million interactions, without considering interaction effects. In our unpublished analyses of this type, data appear to be sufficient to define individual genome-metabolome interaction structure in an unambiguous manner. Because the number of interactions (400 million) is greater than the number of Americans (310 million), this combination appears sufficient to develop precision medicine programs that are truly personalized. Thus, the practical implication of an affordable means to routinely measure 20,000 chemicals in a blood or urine sample is that with a commitment to obtain genetic information, a sufficient foundation is available to create a powerful precision medicine system in which environmental exposures are included within predictive, diagnostic and management models.

3.7. Focus on data quality to obtain affordable high-throughput analysis

With this prospect of affordable high-throughput chemical profiling now approaching reality, data quality must be critically evaluated and substantiated. Although traditional analytical chemistry methods are expensive, the quality standards are very high. To consider adoption of a new approach, key questions must

be addressed concerning accuracy and reproducibility of chemical detection, identification and quantification. Most of these can be satisfactorily addressed but require adoption of different, and somewhat less flexible laboratory practices.

Analytical facilities differ in the types of samples analyzed, the chemicals measured and the demands for precision in identification and quantification. To enable capture of the large number of chemicals by HRM, some important compromises must be made. One is the dedication of use of powerful instruments to a single purpose. By restricting an instrument to one or a small number of very similar matrices (e.g., plasma, food extracts, liver biopsy, etc.), variations in response characteristics can be readily monitored. Additionally, if only one type of sample (e.g., serum or plasma) is analyzed on an instrument, then there is little chance for contamination by chemicals present in laboratory toxicology experiments, food extracts, etc. Although difficult to precisely verify, the sensitivity to detect low abundance chemicals with exactly the same method appeared to nearly double when we transitioned from a common use to a dedicated use instrument. The practical meaning is that analytical facilities using HRM probably need at least two high-resolution instruments, one for dedicated use to the primary matrix and a second for multiple application use. In this regard, newer instruments are replacing many FT-ICR MS and earlier LTQ-Orbitrap MS used for proteomics, and this has created an availability of high-resolution instruments that can be re-purposed for high-resolution metabolomics projects.

3.8. Chemical identification strategies

Traditionally, verification of chemical identity is obtained by direct comparison to analysis of authentic standard on the same instrument under the same conditions, usually on the same day within the same sample batch. We avoid this common approach for two reasons. Although authentic standards are frequently $>99\%$ pure, they invariably are contaminated with some level of impurities. While this is not of concern in a targeted analysis, this poses a risk of introduction of low-abundance contaminants in untargeted analyses. The advanced data extraction algorithms increase dynamic range of detection to nearly eight orders of magnitude, so contamination by very minor components in purified standards cannot easily be evaluated. Thus, to avoid contamination of samples with unknown impurities, a simple procedure is to follow the same rule as above, maintain one instrument for routine analysis of a common matrix and perform chemical identity studies on a different but comparable system. By doing such studies of chemical identity independently of the first analysis, one also confirms reproducibility of detection on a different instrument with different LC column and solvents. By doing this, reproducibility is confirmed in a way that can be translated to different instruments in different facilities and measured at different times.

Several tools are available to aid identification of chemicals and discussed elsewhere [7,20,24,39]. Common principles are used and some are briefly discussed here. A key foundation for HRM involves the mass resolution and mass accuracy of the instruments. Mass resolution and resolving power refer to the separation of mass spectral peaks and have been defined by IUPAC (see http://fiehnlab.ucdavis.edu/projects/Seven_Golden_Rules/Mass_Resolution) for discussion. Higher values refer to better resolution, with quadrupoles and lower resolution TOF instruments having 10,000 resolution, high resolution TOF having 60,000 resolution, Orbitraps having about 100,000 resolution, and Fourier-Transform Ion-cyclotron Resonance (FT) instruments having up to 1,000,000. The latter instruments (Orbitrap, FT) are sometimes referred to as “ultra-high resolution accurate mass (UHRAM) mass spectrometers.

For high-resolution FT instruments (LTQ-FT; LTQ-Velos Orbitrap) operating at 50,000–60,000 resolution, database match within 10 ppm for common ionic forms (e.g., H^+ and Na^+) is often a suitable place to begin. In direct comparisons of data collected from human plasma on an LTQ-FT comparing 50,000 resolution (10 min LC run) and 100,000 resolution (20 min LC run) and on an LTQ-Velos Orbitrap comparing 60,000 (10 min LC run) and 120,000 (20 min LC run), we found that the lower resolution runs for 10 min captured 90–95% of the ions detected at the higher resolution. Instrument specifications indicate that <5 ppm accuracy should be obtained, but we found that improved ion detection was obtained by increasing the ion count to higher values than recommended; this results in improved detection of low abundance ions but also results in deterioration in mass accuracy. Specifically, on the LTQ-FT and LTQ-Velos Orbitrap, m/z for internal standards often are 8 ppm from the exact mono-isotopic mass. For the Q-Exactive or Q-Exactive-HF operating at 70,000–140,000 resolution, 5-min LC runs with smaller injection volumes and smaller LC columns provide better data, typically with detected m/z within 5 ppm of the monoisotopic mass.

In practice, greater tolerance is sometimes needed because the complexity of instrumentation and data extraction do not always yield the same precision for all ions. Routines are now available which allow grouping of ions according to likelihood that they are derived from one chemical. If such ions correlate in intensity among samples, this provides confidence that the precursor ion is correctly selected. Detection of correlated isotopic forms, e.g., ^{12}C and ^{13}C forms, also provides such confidence. With 20,000 ions detected by high-resolution MS, however, most of the ions have relatively low intensity and do not have strong correlations with high abundance ions. Consequently, most low abundance ions appear to be derived from unique, low abundance chemicals rather than being low-intensity ions derived from high abundance chemicals. This characteristic is of considerable importance for exposome research because it suggests that humans are exposed to a large number of low abundance, uncharacterized chemicals. Indeed, every MWAS of disease that has been performed by our research group has found accurate mass m/z features that do not match any metabolite in Metlin, HMDB or KEGG databases.

3.9. Rigid 24/7 work schedule

In characterization of conditions to establish SOP for high-throughput metabolomics, we found much better consistency of data when using a continuous 24-h per day, seven day per week analysis rather than work schedules with interruptions of analysis. Thus, I feel that a commitment to a continuous 24 h per day, 7 days per week analysis is necessary, and this structure (Fig 3B) can limit flexibility in an analytical laboratory. Analytical variation occurs due to stopping and starting chromatography because physical interactions and chemical reactions are ongoing on chromatography columns. These variations are minimized by using a rigorously enforced analytical strategy, with a fixed number of wash-in analyses, a fixed number of total analyses, and scheduled shut-down for cleaning, servicing, replacement of columns and recalibration. A byproduct of the routine scheduling of cleaning and replacement of serviceable parts is that autosamplers, LC systems and columns rarely fail within a 30-day analytic sequence, supporting consistent analyses of up to 600 samples in triplicate on two columns within a single run.

3.10. Cross-laboratory comparisons

With a stable analytical structure operated under routine conditions, known metabolites elute reproducibly so that confirmation of identity within each sample is unnecessary. To support such routine

use of accurate mass and retention time as a basis for identification, however, rigorous identification and verification of selective detection is required. Several approaches are available, with no single approach universally acceptable. A very convenient approach is to analyze the same samples by HRM and another validated platform (Fig 3C). We have done this for different chemicals with six different analytical facilities, with some examples provided (Fig 3C) for amino acids measured by the Emory University Genetics Laboratory and creatine and cortisol measured by Metabolon (Research Triangle Park). Such comparisons are useful both for identification and to test reliability of quantification. We perform such analyses in a randomized and blinded manner to assure integrity of results. So far, the results of such comparisons are re-assuring in terms of data reliability; despite this, some inconsistencies are present, and these provide useful information concerning needs for further identify verification.

3.11. Honest interpretation, honest reporting

Many mass spectral analyses are problematic because of inability to separate or distinguish chemicals. For example, both *D*- and *L*-aspartic acid (Asp) occur in biologic systems; even though *L*-Asp is more common, one cannot unequivocally identify *L*-Asp without appropriate consideration of *D*-Asp. Most analytical laboratories adopt a practical approach to assume that the more common form is the one measured and/or do not specify stereospecificity or positional isomers when the methods do not support such distinction. Designations can be used such as Leu/Ile, for methods where leucine (Leu) and isoleucine (Ile) are known to be present but are not resolved. Many lipids and environmental chemicals are problematic because isobaric (same mass) species exist. The practical matter is that if one form is predominant, i.e., represents 95% or more of the signal, then the measurement is principally of that chemical regardless of the contribution of other lower abundance ions.

3.12. Chemicals of low abundance or infrequent occurrence

Challenges exist for chemical identification of ions that are not uniformly present in samples from different individuals. For instance, if a chemical is only found in 1% of samples, then parameter settings are needed to allow detection, and large population sizes are needed to evaluate whether health risks are associated with that chemical. MS/MS studies can only be performed on samples for which the ion is detected, and the volume of sample available can limit ability to pursue identification. For low intensity ions, identification using MS/MS can be difficult if multiple ions are present within the ion selection window for MS/MS (see Fig 2C, where 0.5 AMU isolation window contains many low-abundance ions). In such cases, deconvolution MS/MS is needed to allow reconstruction of ion dissociation spectra [27]. These issues illustrate need to establish a universal surveillance system for study of populations that are large enough to address low abundance chemicals with low prevalence.

3.13. Quantification

Quantification is the holy grail of analytical chemistry. With genomics, transcriptomics and proteomics, considerable advances in knowledge have been made without absolute reference measurements. Similarly, HRM has mostly relied upon relative intensities, with pooled reference samples analyzed regularly to evaluate consistency of response characteristics. This is not ideal, but standard curves and daily high and low quality control analyses are not feasible for thousands of chemicals. To address this challenge for exposome research, a reference standardization protocol

was developed for HRM analyses [11]. This relies upon commonly used methodology of surrogate standardization.

Surrogate standardization is a useful approach that involves quantitative calibration of a target chemical relative to a known amount of a reference chemical. For instance, if a stable isotopic form of palmitic acid were added as an internal standard to quantify palmitic acid, it could also be used as a surrogate standard to estimate myristic acid and stearic acid concentrations. Inclusion of such reference chemicals as internal standards is extremely valuable for quantification of chemicals when multiple extraction or processing steps are required. Although frequently not appreciated, this “gold standard” approach, as used for quantification of analyses done by tandem mass spectrometry, is a single-point calibration method. Linearity is assumed but not explicitly established during analysis, and no upper and lower quantification range is determined. With such analysis, the error of the measurement includes a multiplication of the error for each component.

Common use of stable isotopic internal standards for quantification has, therefore, established a *de facto* single-point calibration as an acceptable alternative to standard curves for MS-based quantification. This is important because with this principle, one can readily convert MS signal intensities to absolute concentrations by referencing to calibrated samples, such as National Institute of Standards (NIST) Standard Reference Material 1950 (SRM1950). Because SRM1950 has concentrations certified, use of this reference material for single point calibration provides a versatile approach for absolute quantification of any chemical present within this sample. In combination with cross-platform validation studies (Fig 3C), one can verify that such approaches are valid and yield useful quantitative data [11].

4. HRM for environmental chemical analysis: standardized workflow

The step from clinical metabolomics (discussed above) to environmental chemical analysis is small yet challenging. Clinical metabolomics includes hundreds of metabolites present in the micromolar range. Many of these have been extensively studied. Many have metabolic precursors and products, which are structurally related, simultaneously measureable and also well characterized. Additionally, mammalian species have very similar central pathways of metabolism so the extensive knowledge of rodents and other species provides a resource for interpretation. In contrast, environmental chemical abundance in human samples is typically nanomolar or sub-nanomolar, often pushing limits of detection [35]. Further, xenobiotic metabolism is more highly variable among species and among individuals [33]. For analyses to measure exposures in individual human samples, this limits ability to rely upon data from other species, and sometimes limits ability to trust population-based human data.

Despite this, application of the principles described above shows that environmental chemicals can be identified in human samples [33] and that useful quantification can be obtained (Fig 3D). For clinical metabolomics, high-resolution metabolomics methods detect metabolites in most metabolic pathways [18], with accurate mass matches to more than half of the metabolites in KEGG human metabolic pathways. These central pathways include only 2000 metabolites, however, and this represents only 10% of the 20,000 ions routinely measured by HRM. Even accounting for multiple ions from individual chemicals, results show that large numbers of uncharacterized chemicals exist in plasma. Hence, the results suggest that application of HRM could provide useful new information on human exposures, and this has been substantiated by studies of Parkinson disease [38], age-related macular degenera-

tion [32], glaucoma [2], tuberculosis [8], HIV-1 infection [6] and alcohol abuse [31].

The standardized workflow developed during the past several years for use of HRM departs substantially from targeted environmental chemical analysis. Part of this is the result of the intent to provide a more affordable and powerful clinical metabolomics capability. Some of the procedures are briefly described here to expose details often not discussed in typical research manuscripts.

4.1. Sample collection and storage

This is the most uncontrolled variable for all aspects of analytical sciences. Samples collected under different work conditions by different individuals and with varying instrumentation, storage conditions and timing, are outside of the control of the analytical facility. There is no simple way to ascertain whether a sample was badly corrupted before delivery to the analytical laboratory. Thus, rather than potentially introducing bias by rejecting a sample, we use a simpler approach, to assume that those who collected the samples followed appropriate protocols, labeled samples correctly and discarded or noted samples thought to be corrupted. We analyze every sample and use quality control (total ion intensity; correlation of triplicate analyses) and downstream statistical criteria to evaluate data for exclusion. Such downstream criteria can include glucose (blood samples that are allowed stand at room temperature for long periods before centrifugation frequently have very low glucose concentration), heme (samples with extensive hemolysis have high heme content) or oxidized sulfur-containing chemicals (cystine oxidation to cystine sulfoxide and cystine sulfone occurs upon long-term storage).

4.2. Sample processing for metabolomics

Many small molecules bind to protein and cannot be separated and measured unless the protein is removed. Therefore the first step for most analyses of small molecules in biological samples is to remove protein. However, every sample manipulation has differential effects on recovery of chemicals so every effort must be made to perform this step consistently. Targeted methods include index chemicals to evaluate recovery at each step, but this is only effective for index chemicals with properties similar to the chemicals being analyzed. When thousands of chemicals are being analyzed, verification of recovery for each becomes impossible. Therefore, we adopted a simple operational method for preparation of 20 research samples and a pooled reference sample together for analysis, using very consistent processing.

To remove protein, 120 μ l of highest purity acetonitrile (or methanol) at 0 °C containing a mixture of stable isotopic standards is added to 60 μ l of each research or reference sample, followed by timed mixing. Acetonitrile and methanol are similar in terms of protein removal but acetonitrile is better in removal of oligonucleotides. Samples are allowed to stand on ice in the dark for 30 min prior to centrifugation at the top speed in a micro-centrifuge to remove protein. This largely removes protein and extracts a consistent amount of lipophilic compounds from the lipoproteins. However, it should be noted that this extraction is optimized to obtain a broad spectrum of polar and non-polar chemicals; an extensive literature exists for targeted lipidomics research, and other extraction procedures can improve extraction of different classes of lipids. After protein removal, samples are then loaded onto a refrigerated autosampler and analyzed as previously described in detail [10]. Validation studies for time in the autosampler showed parallel analyses of a sample set in opposite order resulted in comparable data [16]. Using this approach, the internal standards provide multiple internal standard references for quality control, retention time calibration and verification of instrument

performance. This minimal processing approach is limited in that there is no enrichment step for low abundance chemicals, and some chemicals are not well extracted. On the other hand, for the >20,000 ions that are measured, they are analyzed in a routine manner with little opportunity for failure. Integration with data obtained using other platforms, e.g., GC–MS, LC–MS/MS, will provide a robust core data structure as a reference for all types of exposure and phenotypic data.

4.3. Liquid chromatography-high resolution mass spectrometry

Only MS¹ (precursor ion spectra) data are captured during our routine analysis. This was a practical decision to simplify the programming requirements for data extraction. As indicated above, most common intermediary metabolites have unique elemental composition and can be unambiguously measured due to the mass resolution and mass accuracy of the instrument. As newer instruments and data extraction algorithms are developed, this analytical structure can be updated, especially with inclusion of MS/MS on targeted ions.

4.4. Data extraction

MS data are processed in profile mode to .cdf files and extracted with aplCMS [48] using xMSanalyzer [42]. A hybrid targeted/untargeted extraction routine is also available which allows input of targeted environmental chemical lists to enhance reproducibility of environmental chemical data extraction [49]. Data output from xMSanalyzer is provided in a form that can be used in popular spreadsheet and database formats. Data include quality control criteria; each sample is analyzed in triplicate and for each m/z feature in each sample, a coefficient of variation (CV) is calculated. Median CV, % missing values and a combined Quality Score is provided for each m/z (Fig 4A). Initial testing showed that use of five technical replicates further improves data quality and also allows useful information to be captured on lower abundance ions.

4.5. Bioinformatics and biostatistics

A wide array of bioinformatic and biostatistical tools are available for data analysis. We commonly use LIMMA in R programming, with false discovery weight adjustment [14] for initial statistical testing for group-wise differences, metabolome-wide associations with specific environmental chemical exposures, or metabolome-wide correlations with physiologic, clinical or other parameters of interest (see below). This generates Manhattan plots of $-\log p$ versus m/z to visualize significant features (Fig 4B and C). For many purposes, it is useful to see the significant features plotted both as a function of m/z and retention time, with color-coding to distinguished direction of effect [12]. Because one must protect against both type 2 as well as type 1 statistical error, a useful practice is to identify raw p with color and FDR with broken horizontal lines (Fig 4B and C). While statistical testing is useful to evaluate m/z that differ, partial least squares-discriminant analysis (PLS-DA) provides an approach to rank m/z in terms of their contribution to separation (see section IV.B). Features that are both significant following FDR correction and are among the top in contributing to PLS-DA separation are often of greatest interest for subsequent study.

4.6. Graphical presentation of data

After a list of features of interest is generated, multiple approaches are available for annotation and curation of chemical identity and linkage to pathways and functional networks. This is illustrated by figures from a recent metabolome-wide association

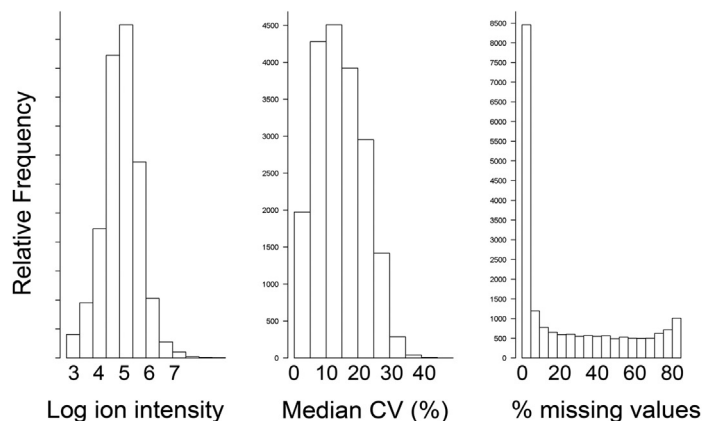
study (MWAS) of pulmonary tuberculosis (Fig 5). The MWAS data are plotted as a Manhattan plot (Fig 5A), and box and whiskers plots are generated for metabolites that differ (Fig 5B). These data are then used to perform a 2-way hierarchical cluster analysis of the distribution of significant features among subjects (Fig 5C). This is very useful to see how well the selected features classify the individuals and also see how similarly the features are distributed among the individuals. The clusters of metabolites are often closely related, including multiple ions from a single chemical (e.g., Na⁺ and H⁺ forms), multiple metabolites from common pathways [(e.g., phenylalanine (Phe) and tyrosine (Tyr)) and metabolites with very similar properties (e.g., different phosphatidylcholines). Among the quickest ways to identify pathways is to match the high-resolution m/z to predicted ions of chemicals in HMDB (<http://www.hmdb.ca>), Metlin (<https://metlin.scripps.edu/index.php>) or other online metabolomics database. Using KEGG compound identifiers, one can query the KEGG compound database for KEGG Pathway or KEGG Brite classification (Fig 5D). This is often inefficient, however, because there are many matches to some m/z and there is no statistical test to enhance confidence. Thus, a better approach is to use a pathway analysis program such as Mummichog [22], MetaboAnalyst [46] or MetaCore (thomsonreuters.com/metacore/) to perform pathway analyses with statistical testing routines. Pathway analysis can be challenging but is extremely valuable because this analytical structure allows linkage of exposures to biologic effects, i.e., perturbations of metabolism that are significantly associated with a chemical or chemical metabolite that is directly measured in the same biologic samples.

5. Paradigms for use of HRM in exposome research

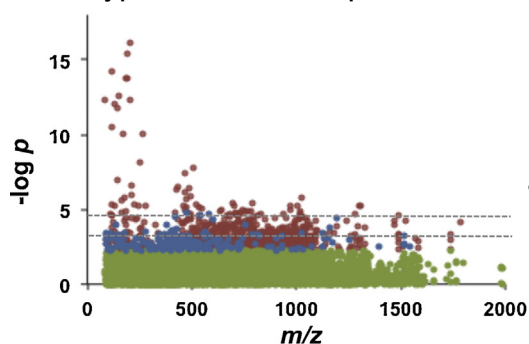
In an initial examination of combined exposure surveillance and bioeffect monitoring, we compared seven mammalian species with considerable differences in environment [33]. The hypothesis was that because humans have more varied exposures than animals housed in research facilities, humans would have more variable chemical content in plasma. The results showed, however, that research animals have exposures to some of the same insecticides, fire retardants and plasticizers, as humans. Additionally, the number of metabolites detected was similar, suggesting that dietary and microbiome chemicals also contribute to the metabolomes of other species in the same way as in humans. Finally, cluster analysis showed that some metabolites had greater interspecies variation than intraspecies variation; environmental chemicals and metabolites used for detoxification, such as glutathione, were present in the group that had greater variability between species. Thus, the data point to the importance of biomonitoring humans for exposure assessment rather than relying upon measurements performed in other species.

High-resolution metabolomics can be used in multiple ways to study the human exposome. These can be considered in linear cause-effect relationships, such as (1) linking environmental exposure data to respective plasma levels of a chemical or its metabolite, (2) linking plasma levels of an environmental chemical to associated metabolite levels and pathway effects, (3) linking dose-response effects in model systems with human dose response relationships, (4) integrating exposure associated metabolic effects with disease phenotypic markers and (5) using an integrated omics approach to identify central pathophysiological responses. Ongoing collaborative studies applying high-resolution metabolomics analyses to complement independently measured trichloroethylene exposure levels, polycyclic aromatic hydrocarbon concentrations, polyfluorinated hydrocarbon levels and cadmium body burden, show that these approaches are sound and broadly useful. Some examples are given below with comments on limitations and

A. xMSAnalyzer Data Characteristics



B. Type 1 Manhattan plot



C. Type 2 Manhattan plot

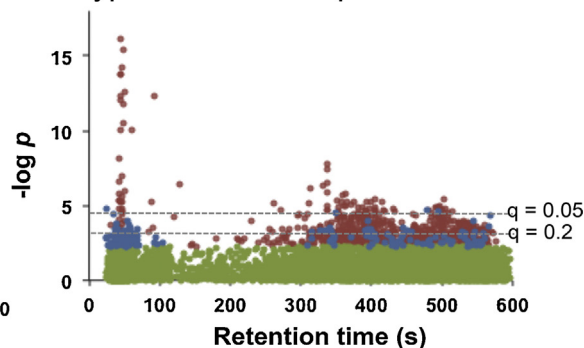


Fig. 4. Data characteristics for high-resolution metabolomics. A. xMSAnalyzer provides a summary of data characteristics including histograms showing the distribution of the log of ion intensity, the distribution of the median coefficient of variation for metabolites and the distribution of the percentage of missing values for the metabolites. B. A type 1 Manhattan plot is the negative log p as a function of m/z for a statistical analysis of each metabolite. This illustration is from a study to evaluate metabolites correlated with an amino acid. In the plot, ions with non-significant raw p are green, those positively associated are in red and those negatively associated are in blue. The broken lines are corresponding cutoffs for false discovery rates of 0.05 (top line) and 0.2 (bottom line). C. A type 2 Manhattan plot is the same data as in B plotted as a function of retention time. This plot is useful for separations obtained with C18 chromatography because more hydrophobic lipid-like chemicals elute at greater retention times. Thus, the retention time provides useful information concerning the properties of the chemical.

opportunities for development. In consideration of such data, one must remain vigilant to protect against interpretation of associations in metabolomics data as proof of causal relationships.

5.1. Metabolome-wide association study (MWAS) for environmental associations of disease

In a pilot study of neovascular age-related macular degeneration (NVAMD), we performed an MWAS of 26 patients and 19 controls and found 94 m/z features associated with NVAMD after false discovery rate correction at 0.05 (Fig 6A) [32]. Some unanticipated pathway associations were present, such as bile acid metabolism, but an observation of interest for possible environmental cause of disease was a feature with m/z 308.857, which had a high-resolution match to the K^+ adduct of beta-2,3,4,5,6-pentachlorocyclohexanol. Although there were no other matches in the database for this ion, a single match without confirmed identity may not be considered worthy of closer examination. In correlation analyses, however, this ion correlated with m/z 353.103, and both of these correlating ions also correlated with m/z 376.845. The latter ion had raw $p < 0.05$ for association with NVAMD and was excluded after FDR correction. m/z 376.845 matched the Na^+ adduct of 7-hydroxy-1,2,3,6,8-pentachlorodibenzofuran and 1,2,3,7,8-pentachlorodibenzodioxin. An associated ^{37}Cl form suggested correct identification as a chlorinated chemical, but the sample availability was too low to confirm identity for either.

Importantly, the results show that FDR protects against type 1 statistical error (false inclusion of incorrect chemicals), but does so by ignoring type 2 statistical error (exclusion of correct chemicals). The results showing one of two highly correlated pentachlorinated chemicals was excluded because of FDR filtering illustrates a weakness in use of FDR correction as being overly stringent for discovery-based research. A limitation in this discovery approach is that associations only provide circumstantial evidence that pentachlorochemicals could contribute to NVAMD. On the other hand, they illustrate the importance of interpreting data collectively. In this case, two ions matching pentachlorinated hydrocarbons were associated with NVAMD and indicated that follow-up is needed to determine whether these low-level environmental chemicals could be causally related to disease.

5.2. MWAS to discover co-exposures and possible underlying mechanisms

Demonstration of ability to detect co-exposures is provided by analysis of cotinine in 30 samples without information concerning smoking history. Partial least squares-discriminant analysis (PLS-DA) showed clear separation where the samples were classified as smokers and non-smokers based upon plasma cotinine levels (Fig 6B). MWAS of cotinine (Fig 6C) showed m/z features associated with cotinine; these included expected metabolites, hydroxycotinine (m/z 193.0966) and norcotinine (m/z 163.0863). In

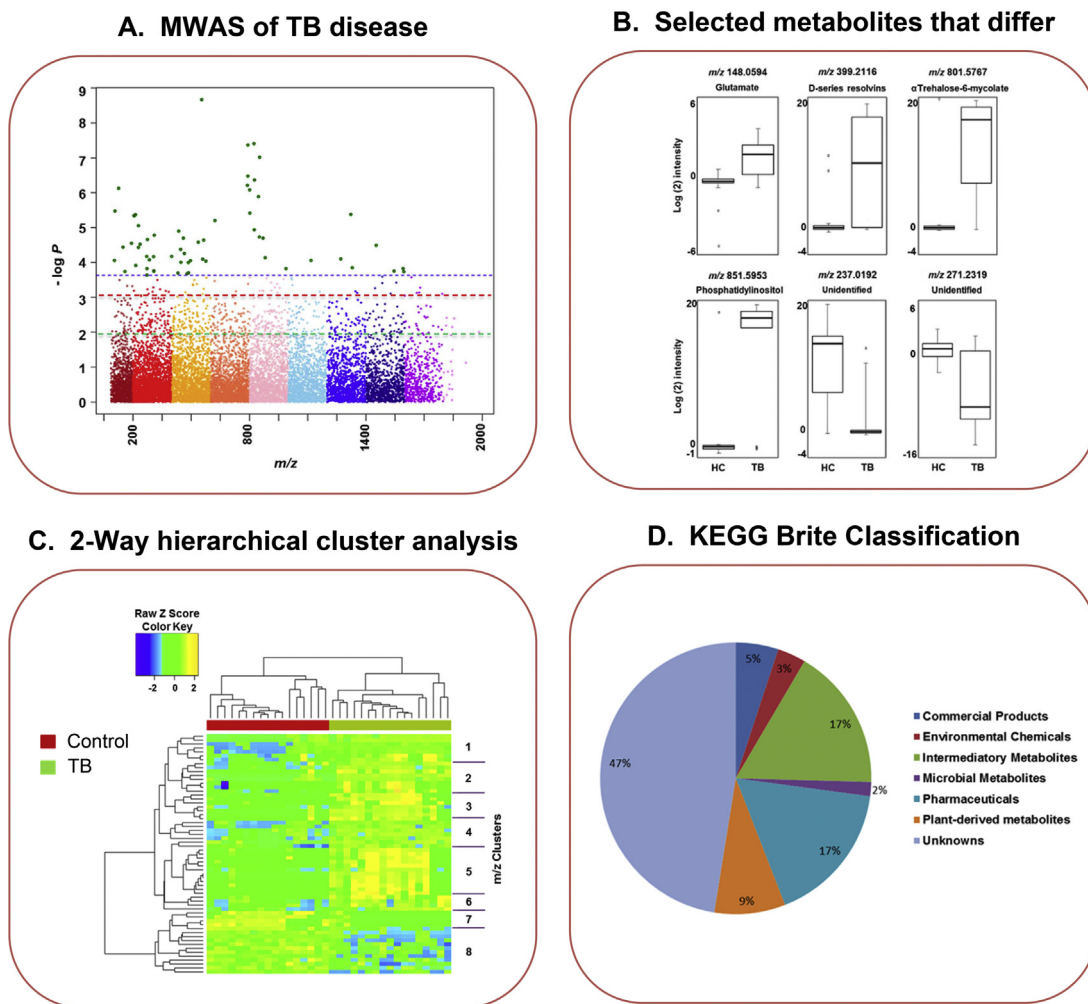


Fig. 5. High-resolution metabolomics workflow. A. An MWAS was performed for pulmonary tuberculosis (TB) patients compared to uninfected household controls. Respective broken lines, from bottom, raw $p = 0.05$, FDR = 0.2, FDR = 0.05. B. Selected metabolites that differ in A are plotted using box and whiskers plots. C. Two-way hierarchical cluster analysis of metabolites that differ in A shows that the metabolites separate the individuals and that the metabolites are associated into clusters. D. Use of Kyoto Encyclopedia of Genes and Genomes (KEGG) Brite Classification of metabolomics database matches give a depiction of the types of chemicals that differ between patients and controls. Data are from [8].

addition, positive correlation was obtained with methoxy pyrazine (m/z 111.0555), a flavoring additive for cigarettes. Network associations (Fig 6D) performed by MetabNet [41] and metabolic pathway analysis revealed significant associations with methionine metabolism, with a negative association of cotinine with S-methylmethionine (m/z 208.0399). Although little studied, S-methylmethionine is a product of a mitochondrial enzyme that uses S-adenosylmethionine to methylate methionine; the function of S-methylmethionine is unknown. The results show that high-resolution metabolomics provides a platform for discovery of co-exposures and also to detect unanticipated association with altered methionine metabolism.

5.3. Use of translational models to discover exposure-associated metabolic effects

In a study of atrazine in mice, we found that short-term atrazine exposure (125 mg/kg) in male C57BL/6 mice disrupted tyrosine, tryptophan, linoleic acid, and α -linolenic acid metabolism [23]. In unpublished studies, two of the metabolites of atrazine found in mice, deisopropylatrazine (DIP; m/z 174.0536) and deethylatrazine (m/z 188.0700), had high-resolution matches in samples from healthy humans (m/z 174.0540 and m/z 188.0697, respec-

tively). The latter identification is ambiguous, however, because a tryptophan metabolite, indoleacrylic acid, also matches this m/z and has an overlapping retention time. The feature matching DIP had a significant but weak ($r = 0.26$) correlation with an ion matching hydroxyatrazine (m/z 220.1773) and both features correlated with Phe and other indole-containing metabolites. The data show that results from dose-response findings in model systems can be used to search HRM repositories of human metabolomics data to test for correlations of environmental chemicals and biologic effects. This use of established cause-effect relationships from animal models to test for the same associations in human populations can provide a cost-effective approach to improve understanding of relevant exposures in the general population. The data also show that caution will be required to avoid misinterpretation due to isobaric chemicals, which cannot be distinguished without further analysis.

5.4. Use in health prediction with case-control design

In a high-resolution metabolomics study of Parkinson Disease progression [38], significant differences between 80 patients and 20 controls included matches to a polybromodiphenyl ether (PBDE), tetrabromobisphenol A, octachlorostyrene and pentachloroethane.

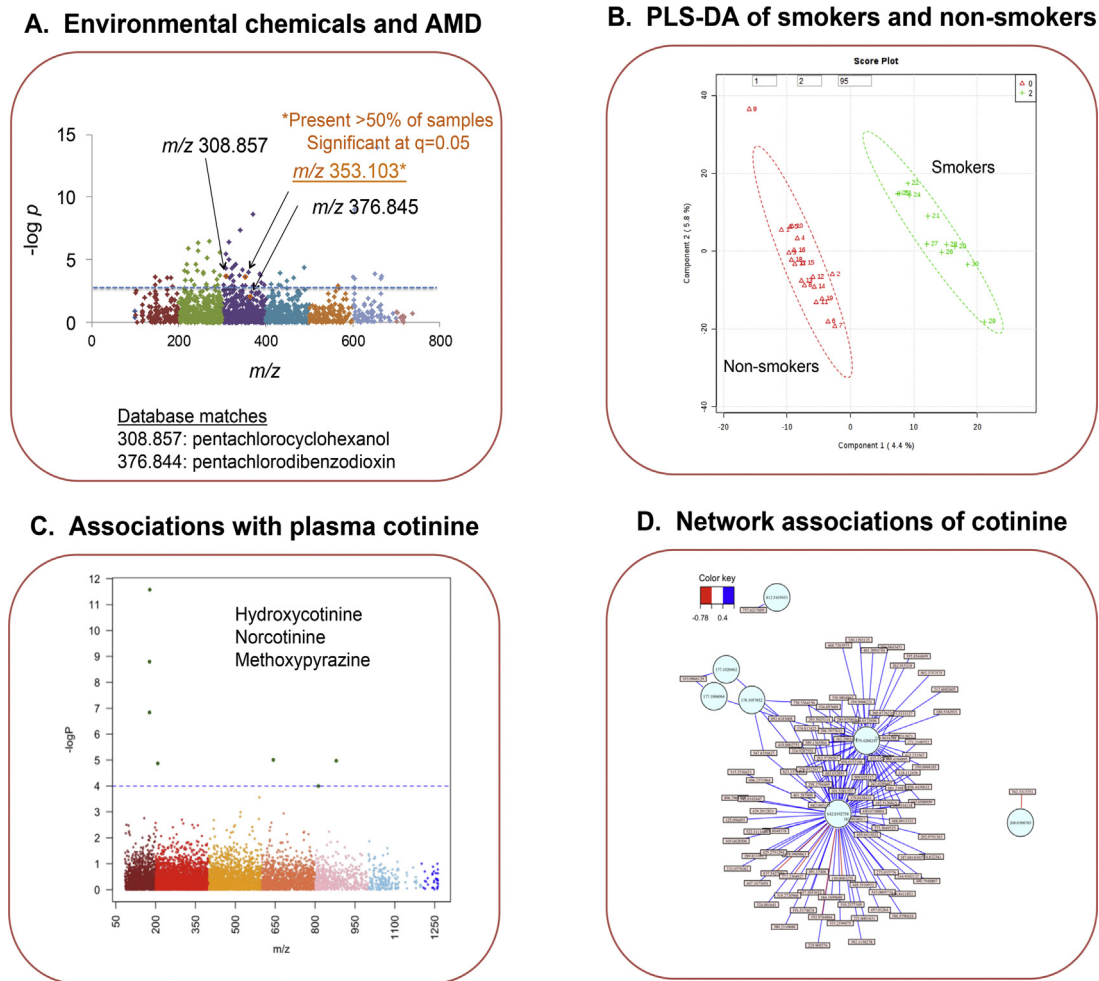


Fig. 6. High-resolution metabolomics provides an approach to detect environmental chemicals associated with disease and to discover co-exposures of chemicals. A. MWAS of age-related macular degeneration (AMD) patients and controls revealed associations of chemicals matching pentachlorochemicals with AMD [32]. B. PLS-DA of smokers and non-smokers, classified by plasma cotinine concentrations, showed complete separation. C. MWAS of the data in B showed that 7 metabolites were correlated with cotinine, including two cotinine metabolites and a cigarette flavoring agent, methoxy pyrazine. D. Network analyses using MetabNet [41] showed that cotinine was associated with disruption of methionine metabolism, specifically associated with decreased S-methylmethionine. B–D are unpublished data from mass spectral analyses performed by DI Walker and bioinformatics analyses performed by Karan Uppal.

Serum samples were from a case-control study that enrolled PD patients and population-based controls between 2001 and 2007 in Central California and followed cases until 2011, death or loss to follow-up [4]. Groups of subjects (80 from 146 PD and 20 from controls) were matched as best possible by age (+/– 5 years), gender (male, female), smoking status (never, former, current) and ambient pesticide exposure. The *m/z* matching PBDE had a mean intensity 50% above controls. *m/z* 282.046, matching 2-amino-1,2-bis(*p*-chlorophenyl) ethanol, was more than 50% higher in individuals with rapid disease progression. Although these population sizes are too small to make firm conclusions, the results indicate that high-resolution environmental metabolomics has the potential to provide a broadly applicable tool to screen for possible environmental contributions to disease and disease progression.

5.5. Integrated omics to support mechanistic understanding

HRM is powerful as a stand-alone platform but becomes much more powerful when used in combination with other omics approaches. A transcriptome–metabolome wide association study (TMWAS) was used to investigate complex cellular responses to a combination of a fungicide, maneb (MB), and an herbicide, paraquat (PQ) [10]. The combination of MB+PQ had been linked in epi-

demologic studies as a potential environmental cause of Parkinson Disease [37,43]. Using different concentrations and measuring gene expression changes and metabolomics under conditions without cell death, interactions of gene expression and metabolism were evaluated (Fig 7A). The results showed that four central clusters of gene-metabolite associations were present (Fig 7B), with one cluster representing a hub apparently linked to toxicity and consisting of an apoptosis gene and genes for two cation transporters and associated regulatory protein. Other clusters included two hubs of genes for adaptive systems and a final cluster representing a hub of stress response genes. The results show that central mechanistic hubs can be identified in terms of the strength of correlations of genes and metabolites under conditions of a toxicological challenge.

A second example of integrated omics involves use of redox proteomics in combination with HRM to study mechanisms of cadmium (Cd) toxicity. Cd is not redox active yet is known to preferentially cause oxidation of mitochondrial proteins. Mice were dosed with Cd and isolated liver mitochondria were analyzed for protein oxidation using MS-based redox proteomics [9]. The results showed that the most central mitochondrial pathways with proteins undergoing oxidation were those involved in β -oxidation of fatty acids and branched chain amino acid metabolism. HRM showed that the central mitochondrial pathways with metabo-

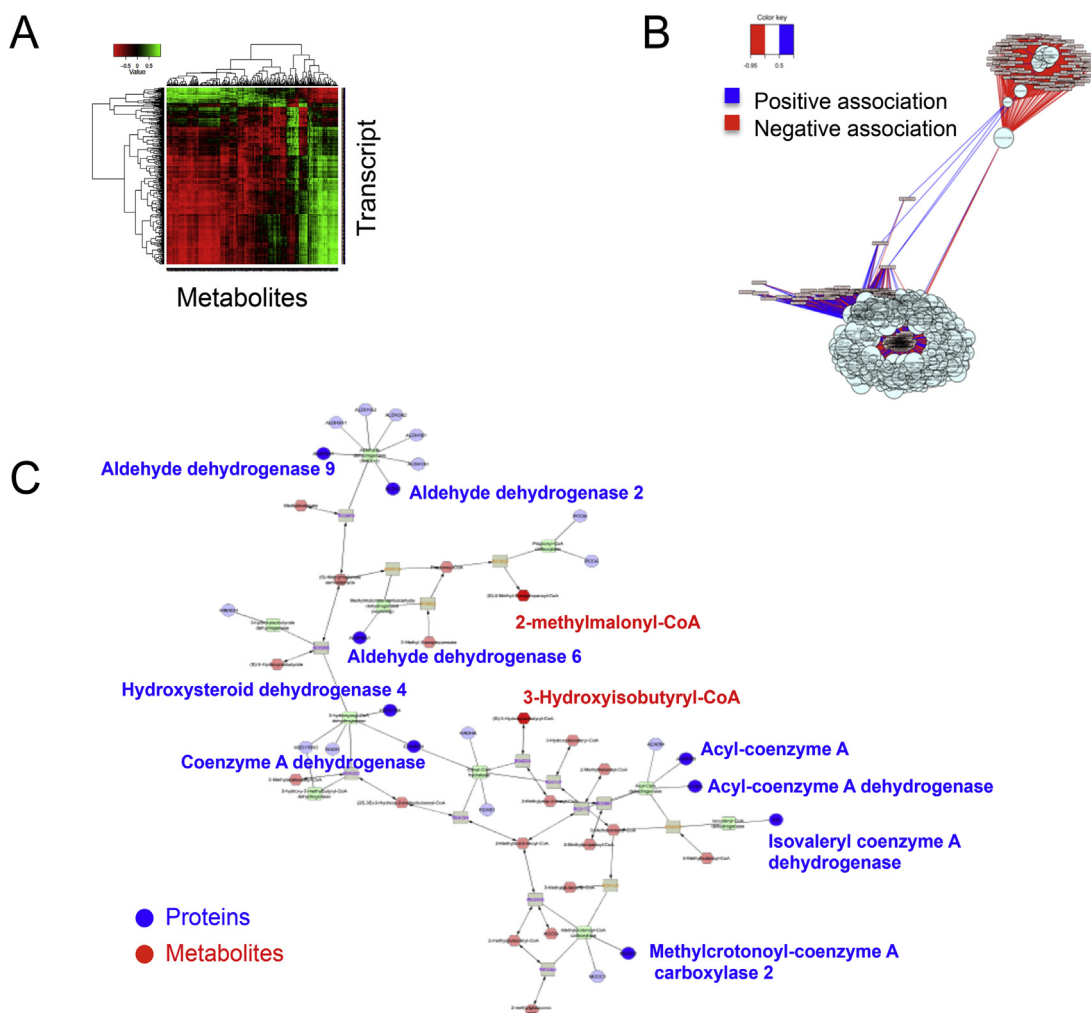


Fig. 7. Integration of high-resolution metabolomics with gene expression and redox proteomics analyses. A. A heatmap of associations of top correlations between transcriptome and metabolome analyses reveals strong associations with exposure to the fungicide, maneb, and the herbicide, paraquat. B. Network associations showed four strongly associated hubs of transcripts and metabolites. C. Metabolic changes occurred in pathways in which mitochondrial proteins were oxidized in response to cadmium. Data for A and B are from [10], and data for C are from [9].

lites responding most to Cd were acylcarnitines and branched chain amino acid metabolites (Fig 7C). An additional example is provided for integration of the microbiome and metabolome [5]. Although there are currently only a limited number of studies in which metabolomics is combined with other omics approaches in environmental research, the results available show that this is a powerful approach which can be used in cell culture and in vivo to study environmental mechanisms of disease.

6. Universal exposure surveillance: an interim solution for exposome research

Chemical space is infinite, and the sequence and diversity of possible human exposures is infinite. Yet within these realms of infinity, the earth is finite, human lifespan is finite and a human exposome is finite. In the face of infinite possibilities, the challenge is not to partition infinity but rather to develop realistic and achievable goals for human exposures within the time frame of human lifespan.

Humans can live over 100 years so sequencing an individual's lifelong exposures is a long-term project. Multiple options are available to move this forward, such as analysis of archival samples that have been collected over decades for some cohorts and longevity studies in relatively short-lived species, like domestic dogs or non-

human primates. These approaches are useful to gain information on internal doses of environmental chemicals but do not effectively capture external environmental chemical burden. Thus, a better scientific foundation may be obtained through development of a universal exposure surveillance system. In this context, "universal" does not mean "comprehensive", but rather means a standardized way to routinely capture and assemble information on the range and prevalence of chemicals across geographies and biologic systems.

A practical initiative would be to establish a universal exposure surveillance program with two specific goals: (1) to provide a periodic census of detectable chemicals within specific geographies and (2) to provide a periodic census of detectable chemicals within humans. Such a program need not be daunting. A broad spectrum of analytical tools is available, geographic regions and populations are extensively mapped, and effective organizational structures are in place to establish priorities according to available resources.

Routine air and water surveillance methods are already in place so that minimal additional costs would be needed for acquisition for samples of these environmental media. Sampling of soil and built environments will require more strategic planning, but also can build upon existing surveillance methods. A key difference in analytical strategies, however, is that instead of targeted analyses of a small number of chemicals of concern, effective analytical tools

are needed to broadly survey chemical space. Development of a geographic grid system for regular interval measurements of chemicals would allow generation of maps that would reveal changes in the chemical environment, analogous to the maps of temperature, rainfall, particulate levels, and other commonly available environmental metrics.

In this regard, the healthcare system provides an already established system that could be used with HRM to obtain regular census of the human body burden of environmental chemicals. As described above, HRM is well developed for use as a general surveillance tool. High-resolution mass spectrometers are widely available and have been applied to study a spectrum of disease states and physiologic conditions. These include studies of cardiovascular disease, obesity, diabetes, neurodegenerative disease, lung disease, renal disease, liver disease, eye disease, infections and immunity and aging. The important conclusion is that high-resolution FTMS and Q-TOF methodologies are robust, affordable and widely available.

An immediate interim solution to sequencing the exposome would be to use this technology to support a periodic census of detectable chemicals within humans. Within the healthcare system, this could be done with minimal collection costs. Discarded blood from many collection purposes could be used. For instance, de-identified samples collected from mandatory drug screenings or HIV testing could be used; prenatal or newborn disease screening samples could be used; samples for toxicology testing in emergency rooms could be used. With appropriate incentives to replace current blood chemistry analyses with high-resolution technology, environmental biomonitoring could be obtained as a byproduct of many clinical analyses with minimal increase in cost.

Such measures would necessarily be limited by current state-of-the-art in exposure science and biomonitoring research. But the key practical issue is that before a cumulative measure of environmental influences and biological responses throughout the lifespan can be achieved, a rigorously standardized system is needed to measure environmental influences and biologic responses. The HRM platform provides such a tool and therefore could be useful for initiation of a universal exposure surveillance system.

7. A framework to sequence the human exposome

A universal exposure surveillance system using HRM as a core element could provide short-term benefits as a resource for predictive modeling of human exposures and health trends. The long-term benefits are greater, however, in that the structure provides a core upon which a program to sequence the human exposome becomes practical. Central arguments can be summarized as follows:

7.1. Use plasma as a common reference

Blood plasma is commonly available during routine health examination and provides a logical place to start with creation of an exposome-sequencing project. Plasma is a fluid that is central in physiology and supports nutrient supply, elimination of wastes and hormonal communication in homeostatic regulation. Urine provides a suitable alternative, especially for children, but has a disadvantage in not providing plasma metabolic data to monitor biologic effects of exposures.

7.2. Use HRM upon plasma collection to obtain advanced blood chemistry analysis

An advantage of HRM is that a very broad spectrum of chemicals is detected, including many environmental chemicals and representative chemicals from most metabolic pathways. Chemicals are

detected that are derived from environmental and other exposures, including diet, microbiome, supplements and drugs. Metabolites in plasma provide an integrated measure of the biologic response of an individual to exposures, including behavioral factors such as exercise and emotional stress. Comparison of data to reference plasma for normal populations shows how an individual differs from a population. The potential utility of such comparisons is enhanced by the recent demonstration that HRM can be used to quantify chemicals over more than seven orders of magnitude absolute abundance [11]. Furthermore, this blood chemistry profile provides metabolite measurements of an individual within the context of ongoing exposures and disease processes so that longitudinal analyses can be used to evaluate lifelong changes in exposure and metabolism. This translates the concept of time-point measurements into a cumulative lifelong measurement as implied by the term “sequencing the exposome”.

7.3. Use advanced data extraction algorithms to maximize information capture

Advanced analytical methods are available to systematically measure thousands of chemicals in plasma. As outlined above, these can be queried for associations with disease and, for unidentified chemicals, can be retrospectively studied for identification. While uncertainties exist concerning the best platform for cost, coverage and quantitative reproducibility, arguments are sufficiently sound to choose a simple protein removal step and a combination of ionization and chromatography methods, with at least triplicate analyses.

7.4. Deliver useful clinical measures as part of health records

HRM provides more information than provided by contemporary blood chemistry measurements. The data can be provided in a quantitative form that is meaningful for evaluation of nutrition, metabolism and disease. This information can be useful for health management and could be developed for this purpose and provided back to the individual for personal study and reference. In this regard, it is important to note that chemicals of abuse will also be measured. Consequently, procedures will be needed to ensure privacy and protect individual liberties. Most importantly, as indicated above, collection of uniform exposure surveillance data for an individual over time provides a lifecycle framework for study of the exposome.

7.5. Preserve raw data and extracted data in accessible, cumulative repositories

Databases of chemical information are less expensive to maintain and more stable against decay than stored blood samples. Thus, ongoing analysis of samples with a cumulative data structure should be established. An accessible central repository of data would enable generalized integration of multiple exposure metrics by providing a central reference. For instance, the data include measures of commonly used blood chemistry measures, such as glucose, bilirubin, creatinine, urea, amino acids, vitamins and coenzymes, steroids, free fatty acids, acylcarnitines, phosphatidylcholines, etc. Thus, interactions of any measured external exposure could be evaluated against a common reference. This could be useful to monitor changes in response to new food products, health policies or disease trends.

7.6. Maintain data reconciliation center independent of repositories

Data reconciliation efforts will be needed to determine comparability of data between facilities and over time. The contribution of geography to the plasma metabolome is currently unknown, so efforts will be needed to distinguish analytical differences and exposure differences. Although it may seem necessary to have this in place prior to initiation of a universal surveillance system, improvements based upon quality control and on-going evaluation will be needed regardless. Thus, concern about quality provides no valid excuse for delay, only a reason to invest in quality control and improvement as part of an exposome strategy. The methods are good enough to initiate a program; certainly, experience will provide improvements.

8. A call to action

Sequencing the first human genome was a lengthy process and expensive, yet today analysis is routine and affordable for many purposes. Sequencing the human exposome will also be a lengthy process and expensive. However, this can be initiated with a practical framework using existing technologies, providing future generations the information necessary to use this to improve individual health and also protect against insidious accumulation of low abundance chemicals that can adversely affect humans and ecologic systems.

Christopher Wild recognized the inherent weaknesses in epidemiologic approaches and genomics to deliver real understanding to the causes of most chronic disease [44]. The exposome is a necessary complement to the genome to understand causes and improve prevention of human disease. The genome has been sequenced. The opportunity lies before us to sequence the exposome. The time has arrived to seize that opportunity.

Transparency document

The [Transparency document](#) associated with this article can be found in the online version.

Acknowledgements

Development of high-resolution metabolomics was a consequence of the visionary leadership of R. Wayne Alexander, Michael M.E. Johns and Kenneth L Brigham at Emory University, the supportive collaborations of Young-Mi Go, Arshed A. Quyyumi, Thomas R. Ziegler, Gary W. Miller and many other colleagues at Emory, Georgia Tech and elsewhere, the mass spectrometry genius of Frederick H. Strobel and Jan Pohl, the bioinformatics wizardry of Tianwei Yu, Karan Uppal, and Shuzhao Li, and the many talented research team members, especially Douglas I Walker, ViLinh Tran, Jennifer M Johnson, Youngja Park, Yongliang Liang, and Quynh Soltow. Research support was provided by NIH grants ES023485, HL113451, AG038746, ES019776, ES016731, NIAID Contract HHSN272201200031C, Department of Defense, Emory University, Georgia Research Alliance, and the Woodruff Foundation. These funding sources had no involvement in the collection, analysis or interpretation of data for this article; these sources had no involvement in writing the manuscript and no role in the decision to submit this article for publication.

References

- [1] G.M. Buck Louis, R. Sundaram, Exposome: time for transformative research, *Stat. Med.* 31 (22) (2012) 2569–2575.

- [2] L.G. Burgess, K. Uppal, D.I. Walker, R.M. Roberson, V. Tran, M.B. Parks, E.A. Wade, A.T. May, A.C. Umfress, K.L. Jarrell, B.O. Stanley, J. Kuchtey, R.W. Kuchtey, D.P. Jones, M.A. Brantley Jr, Metabolome-wide association study of primary open angle glaucoma, *Invest. Ophthalmol. Vis. Sci.* 56 (8) (2015) 5020–5028.
- [3] M.B. Comisar, A.G. Marshall, Fourier-transform ion-cyclotron resonance spectroscopy, *Chem. Phys. Lett.* 25 (2) (1974) 282–283.
- [4] S. Costello, M. Cockburn, J. Bronstein, X. Zhang, B. Ritz, Parkinson's disease and residential exposure to maneb and paraquat for agricultural applications in the central valley of California, *Am. J. Epidemiol.* 169 (8) (2009) 919–926.
- [5] S. Cribbs, K. Uppal, S. Li, D.P. Jones, L. Huang, L. Tipton, A. Fitch, R. Greenblatt, L. Kingsley, D. Guidot, E. Ghedin, A. Morris, 2015. Correlation of the Lung Microbiota with Metabolic Profiles in Bronchoalveolar Lavage Fluid in HIV Infection Microbiome, In press.
- [6] S.K. Cribbs, Y. Park, D.M. Guidot, G.S. Martin, L.A. Brown, J. Lennox, D.P. Jones, Metabolomics of bronchoalveolar lavage differentiate healthy HIV-1-infected subjects from controls, *AIDS Res. Hum. Retroviruses* 30 (6) (2014) 579–585.
- [7] Q. Cui, I.A. Lewis, A.D. Hegeman, M.E. Anderson, J. Li, C.F. Schulte, W.M. Westler, H.R. Eghbalnia, M.R. Sussman, J.L. Markley, Metabolite identification via the madison metabolomics consortium database, *Nat. Biotechnol.* 26 (2) (2008) 162–164.
- [8] J.K. Frediani, D.P. Jones, N. Tukvadze, K. Uppal, E. Sanikidze, M. Kipiani, V.T. Tran, G. Hebbbar, D.I. Walker, R.R. Kempker, S.S. Kurani, R.A. Colas, J. Dall, V. Tangpricha, C.N. Serhan, H.M. Blumberg, T.R. Ziegler, Plasma metabolomics in human pulmonary tuberculosis disease: a pilot study, *PLoS One* 9 (10) (2014) e108854.
- [9] Y.M. Go, J.R. Roede, M. Orr, Y. Liang, D.P. Jones, Integrated redox proteomics and metabolomics of mitochondria to identify mechanisms of Cd toxicity, *Toxicol. Sci.* 139 (1) (2014) 59–73.
- [10] Y.M. Go, K. Uppal, D.I. Walker, V. Tran, L. Dury, F.H. Strobel, H. Baubichon-Cortay, K.D. Pennell, J.R. Roede, D.P. Jones, Mitochondrial metabolomics using high-resolution fourier-transform mass spectrometry, *Methods Mol. Biol.* 4 (2014) 3–73.
- [11] Y.M. Go, D.I. Walker, Y. Liang, K. Uppal, Q.A. Soltow, V. Tran, F. Strobel, A.A. Quyyumi, T.R. Ziegler, K.D. Pennell, G.W. Miller, D.P. Jones, Reference standardization for mass spectrometry and high-resolution metabolomics applications to exposome research, *Toxicol. Sci.* (2015) (in press).
- [12] Y.M. Go, D.I. Walker, Q.A. Soltow, K. Uppal, L.M. Wachtman, F.H. Strobel, K. Pennell, D.E. Promislow, D.P. Jones, Metabolome-wide association study of phenylalanine in plasma of common marmosets, *Amino Acids* 47 (3) (2015) 589–601.
- [13] C.S. Ho, C.W. Lam, M.H. Chan, R.C. Cheung, L.K. Law, L.C. Lit, K.F. Ng, M.W. Suen, H.L. Tai, Electrospray ionisation mass spectrometry: principles and clinical applications, *Clin. Biochem. Rev.* 24 (1) (2003) 3–12.
- [14] Y. Hochberg, Y. Benjamini, More powerful procedures for multiple significance testing, *Stat. Med.* 9 (7) (1990) 811–818.
- [15] J.M. Johnson, F.H. Strobel, M. Reed, J. Pohl, D.P. Jones, A rapid LC-FTMS method for the analysis of cysteine, cystine and cysteine/cystine steady-state redox potential in human plasma, *Clin. Chim. Acta* 396 (1–2) (2008) 43–48.
- [16] J.M. Johnson, T. Yu, F.H. Strobel, D.P. Jones, A practical approach to detect unique metabolic patterns for personalized medicine, *Analyst* 135 (11) (2010) 2864–2870.
- [17] D.P. Jones, Redox theory of aging, *Redox Biol.* 5 (2015) 71–79.
- [18] D.P. Jones, Y. Park, T.R. Ziegler, Nutritional metabolomics: progress in addressing complexity in diet and health, *Annu. Rev. Nutr.* 32 (2012) 183–202.
- [19] M. Kanehisa, S. Goto, Y. Sato, M. Kawashima, M. Furumichi, M. Tanabe, Data, information, knowledge and principle: back to metabolism in KEGG, *Nucleic Acids Res.* 42 (2014) D199–D205 (Database issue).
- [20] T. Kind, O. Fiehn, Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1ppm, *BMC Bioinformatics* 7 (2006) 234.
- [21] P. Kirkpatrick, C. Ellis, Chemical space, *Nature* 432 (7019) (2004) 823.
- [22] S. Li, Y. Park, S. Duraisingham, F.H. Strobel, N. Khan, Q.A. Soltow, D.P. Jones, B. Pulendran, Predicting network activity from high throughput metabolomics, *PLoS Comput. Biol.* 9 (7) (2013) e1003123.
- [23] Z. Lin, J.R. Roede, C. He, D.P. Jones, N.M. Filipov, Short-term oral atrazine exposure alters the plasma metabolome of male C57BL/6 mice and disrupts alpha-linolenate, tryptophan, tyrosine and other major metabolic pathways, *Toxicology* 326 (2014) 130–141.
- [24] Y. Ma, T. Kind, D. Yang, C. Leon, O. Fiehn, MS2Analyzer: a software for small molecule substructure annotations from accurate tandem mass spectra, *Anal. Chem.* 86 (21) (2014) 10724–10731.
- [25] S. Maher, F.P. Jjunju, S. Taylor, Colloquium: 100 years of mass spectrometry: perspectives and future trends, *Rev. Mod. Phys.* 87 (2015) 113–135.
- [26] A.G. Marshall, C.L. Hendrickson, High-resolution mass spectrometers, *Annu. Rev. Anal. Chem.* (Palo Alto Calif.) 1 (2008) 579–599.
- [27] A.M. Mayampurath, N. Jaitly, S.O. Purvine, M.E. Monroe, K.J. Auberry, J.N. Adkins, R.D. Smith, DeconMSn: a software tool for accurate parent ion monoisotopic mass determination for tandem mass spectra, *Bioinformatics* 24 (7) (2008) 1021–1023.
- [28] G.W. Miller, D.P. Jones, The nature of nurture: refining the definition of the exposome, *Toxicol. Sci.* 137 (1) (2014) 1–2.
- [29] S. Moore, D.H. Spackman, W.H. Stein, Automatic recording apparatus for use in the chromatography of amino acids, *Fed. Proc.* 17 (4) (1958) 1107–1115.

- [30] K. Murray Kermit, K. Robert Boyd, N. Marcos Eberlin, G. John Langley, Liang Li, Yasuhide Naito, Definitions of terms relating to mass spectrometry (IUPAC recommendations 2013), *Pure Appl. Chem.* 85 (7) (2013) 1515–1609.
- [31] O. Obianyoy, Y. Liang, E.L. Burnham, A. Mehta, Y. Park, K. Uppal, F.L. Harris, D.P. Jones, L.A. Brown, Metabolic consequences of chronic alcohol abuse in non-smokers: a pilot study, *PLoS One* 10 (6) (2015) e0129570.
- [32] M.P. Osborn, Y. Park, M.B. Parks, L.G. Burgess, K. Uppal, K. Lee, D.P. Jones, M.A. Brantley Jr, Metabolome-wide association study of neovascular age-related macular degeneration, *PLoS One* 8 (8) (2013) e72737.
- [33] Y.H. Park, K. Lee, Q.A. Soltow, F.H. Strobel, K.L. Brigham, R.E. Parker, M.E. Wilson, R.L. Sutliff, K.G. Mansfield, L.M. Wachtman, T.R. Ziegler, D.P. Jones, High-performance metabolic profiling of plasma from seven mammalian species for simultaneous environmental chemical surveillance and bioeffect monitoring, *Toxicology* 295 (1–3) (2012) 47–55.
- [34] A. Peters, G. Hoek, K. Katsouyanni, Understanding the link between environmental exposures and health: does the exposome promise too much? *J. Epidemiol. Community Health* 66 (2) (2012) 103–105.
- [35] S.M. Rappaport, D.K. Barupal, D. Wishart, P. Vineis, A. Scalbert, The blood exposome and its role in discovering causes of disease, *Environ. Health Perspect.* 122 (8) (2014) 769–774.
- [36] P.L. Remington, R.C. Brownson, Control centers for disease, and prevention. Fifty years of progress in chronic disease epidemiology and control, *MMWR Surveill. Summ.* 60 (Suppl. 4) (2011) 70–77.
- [37] B. Ritz, F. Yu, Parkinson's disease mortality and pesticide exposure in California 1984–1994, *Int. J. Epidemiol.* 29 (2) (2000) 323–329.
- [38] J.R. Roede, K. Uppal, Y. Park, K. Lee, V. Tran, D. Walker, F.H. Strobel, S.L. Rhodes, B. Ritz, D.P. Jones, Serum metabolomics of slow vs. rapid motor progression Parkinson's disease: a pilot study, *PLoS One* 8 (10) (2013) e77629.
- [39] C.A. Smith, G. O'Maille, E.J. Want, C. Qin, S.A. Trauger, T.R. Brandon, D.E. Custodio, R. Abagyan, G. Siuzdak, METLIN: a metabolite mass spectral database, *Ther. Drug Monit.* 27 (6) (2005) 747–751.
- [40] Q.A. Soltow, F.H. Strobel, K.G. Mansfield, L. Wachtman, Y. Park, D.P. Jones, High-performance metabolic profiling with dual chromatography-fourier-transform mass spectrometry (DC-FTMS) for study of the exposome, *Metabolomics* 9 (1) (2013) S132–S143.
- [41] K. Uppal, Q.A. Soltow, D.E. Promislow, L.M. Wachtman, A.A. Quyyumi, D.P. Jones, MetabNet: an R package for metabolic association analysis of high-resolution metabolomics data, *Front. Bioeng. Biotechnol.* 3 (87) (2015) 1–6.
- [42] K. Uppal, Q.A. Soltow, F.H. Strobel, W.S. Pittard, K.M. Gernert, T. Yu, D.P. Jones, xMSAnalyzer: automated pipeline for improved feature detection and downstream analysis of large-scale, non-targeted metabolomics data, *BMC Bioinformatics* 14 (2013) 15.
- [43] A. Wang, S. Costello, M. Cockburn, X. Zhang, J. Bronstein, B. Ritz, Parkinson's disease risk from ambient exposure to pesticides, *Eur. J. Epidemiol.* 26 (7) (2011) 547–555.
- [44] C.P. Wild, Complementing the genome with an exposome: the outstanding challenge of environmental exposure measurement in molecular epidemiology, *Cancer Epidemiol. Biomarkers Prev.* 14 (8) (2005) 1847–1850.
- [45] The exposome: from concept to utility, *Int. J. Epidemiol.* 41 (1) (2012) 24–32.
- [46] J. Xia, R. Mandal, I.V. Sinelnikov, D. Broadhurst, D.S. Wishart, MetaboAnalyst 2.0—a comprehensive server for metabolomic data analysis, *Nucleic Acids Res.* 40 (2012) W127–W133 (Web Server issue).
- [47] M. Yamashita, J.B. Fenn, Electrospray ion-source—another variation on the free-jet theme, *J. Phys. Chem.* 88 (20) (1984) 4451–4459.
- [48] T. Yu, Y. Park, J.M. Johnson, D.P. Jones, apLC–MS—adaptive processing of high-resolution LC/MS data, *Bioinformatics* 25 (15) (2009) 1930–1936.
- [49] T. Yu, Y. Park, S. Li, D.P. Jones, Hybrid feature detection and information accumulation using high-resolution LC–MS metabolomics data, *J. Proteome Res.* 12 (3) (2013) 1419–1427.