Observing the human exposome as reflected in breath biomarkers: Heat map data interpretation for environmental and intelligence research

Joachim D. Pleil*[1], Matthew A. Stiegel[2], Jon R. Sobus[1], Qian Liu[3], and Michael C. Madden[4]

[1]Human Exposure and Atmospheric Sciences Division, NERL/ORD, U.S. Environmental Protection Agency, Research Triangle Park, NC, [2]Department of Environmental Sciences and Engineering, School of Public Health, University of North Carolina, Chapel Hill, NC, [3]Department of Biostatistics, School of Public Health, University of North Carolina, Chapel Hill, NC, [4]Environmental Public Health Division Division, NHEERL/ORD, U.S. Environmental Protection Agency, Chapel Hill, NC

* corresponding author: pleil.joachim@epa.gov

Abstract

Over the past decade, the research of human systems biology and the interactions with the external environment has permeated all phases of environmental, medical, and public health research. Similarly to the fields of genomics and proteomics research, the advent of new instrumentation for measuring breath biomarkers and their associated meta-data also provide very useful, albeit complex, data structures. The biomarker research community is beginning to invoke tools from systems biology to assess the impact of environmental exposures, as well as from internal health states, on the expression of suites of chemicals in exhaled breath. This new approach introduces the concept of the *exposome* as a complement to the genome in exploring the environment – gene interaction. In addition to answering questions regarding health status for the medical community, breath biomarker patterns are useful for assessing public health risks from environmental exposures. Furthermore, breath biomarker patterns can inform security risks from suspects via covert interrogation of blood borne chemical levels that reflect previous activities. This article discusses how different classes of exhaled breath biomarker measurements can be used to rapidly assess patterns in complex data. We present exhaled breath data sets to demonstrate the value of the graphical "heat map" approach for hypothesis development and subsequent guidance for stochastic and mixed effects data interpretation. We also show how to graphically interpret exhaled breath measurements of exogenous jet fuel components, as well as exhaled breath condensate (EBC) measurements of endogenous chemicals.

Introduction

The internal human systems biology is inextricably entwined with the external environment, and in fact, the mixture of endogenous and exogenous chemicals in biological media can be thought of as representing a continuous biological and chemical record of encounters with the surrounding environment. From the perspective of the genomic and proteomic framework, the environmental/biological interaction, or the "exposome" (Wild 2005, Rappaport and Smith 2010, Rappaport 2010), embodies the systems exposure science counterpart to systems biology research (Pleil 2009, Pleil and Sheldon 2010). The human exposome represents all exogenous and endogenous exposures from conception onwards including exposures from diet, lifestyle, and internal biology as a quantity of critical interest to disease etiology (*paraphrased from* Wild, 2005). As such, the window into the human exposome offered by exhaled breath analysis is that subset of exogenous and endogenous compounds measureable in the exhaled breath and exhaled breath condensate.

Currently, biomarker measurements in blood, breath, and urine have a wide variety of applications including disease diagnosis, clinical monitoring, and environmental exposure assessment. In contrast to blood and urine, the exhaled breath is a particularly useful biological fluid in that sampling is non-invasive, supply is essentially unlimited, and breath does not generate infectious waste or "leftovers". As such, breath measurements have taken on an important role for population based environmental studies, and have additionally been proposed for intelligence/security applications (Pleil 2008).

Breathing has the primary role of providing oxygen to all of the body's cells and removing their waste carbon dioxide via gas exchange with blood in the lungs; thus, the breath serves as a window into each cell of the human system. Breath contains trace levels of volatile chemicals produced normally by metabolism, and compounds from the environment whether breathed in, absorbed through the skin, or ingested. It has been recognized that there are hundreds to thousands of organic chemicals that can be measured in breath (Pauling et al 1971, Phillips et al 1999). There are numerous methods and instruments for making exhaled breath measurements that can be tailored for specific compound groups (Amann and Smith 2005, Marczin and Yacoub 2002, Buszewski et al. 2007); however, the interpretation of these measurements depends on the questions that need to be answered.

In this article, we focus on the interpretation of that portion of the exposome measureable in exhaled breath using graphical techniques (heat mapping) drawn from systems biology 'omics research. We present two examples from human studies wherein groups of organic compounds were measured in exhaled breath and their patterns explored within the context of observed meta-data (e.g. gender, exposure status, activity). Specifically, we explore data from a study of exposure to jet fuel wherein we measured exogenous compounds (alkanes and single-ring aromatics) in gas-phase breath, and from a second

study wherein we measured normally occurring endogenous compounds (volatile aldehydes and alcohols) in exhaled breath condensate (EBC).

Methods

Breath sampling and analysis

Exhaled breath sampling and analysis can be implemented in a number of different ways. In general, methods are split between gas-phase breath samples and EBC (liquid) samples. Briefly, gas-phase breath can be collected directly from the subject into a container (bag, canister, glass bulb, etc.) or transformed into a liquid via cold-surface condensation. Subsequent laboratory analysis is generally performed using gas chromatography – mass-spectrometry (GC-MS) techniques tailored specifically to the targeted analyte groups. A number of clinical applications have implemented real-time breath analysis wherein the subject breathes directly into the instrument and the analysis is performed "on the fly" using instruments such as proton transfer reaction (PTR) MS, selected ion flow tube (SIFT) MS, or a variety of optical methods based on infrared tunable diode lasers. Most breath applications for environmental and health assessments rely on GC-MS to gather a wide spectrum of compounds; the real-time MS and optical methods generally target one, or a few, compounds for specific diagnostic or tracking purposes. In addition to the broad coverage presented in the two books mentioned earlier (Amann and Smith 2005, Marczin and Yacoub 2002), the most recent advances and conceptual overviews of the broad spectrum of breath measurements are found in the literature (Pleil 2010, Pleil et al 2008, Pleil 2008, and references therein).

Breath data applications:

As outlined in the introduction, breath analysis has some basic applications primarily revolving around environmental exposure assessment and medical/clinical health assessment. A third application being explored in the intelligence gathering community is a hybrid of environmental and health research for the purposes of threat assessment. Although the methodologies are driven primarily by which breath-borne compounds are probative, they can certainly overlap. What differentiates the applications is often more based on which questions need to be answered.

From an environmental health perspective, exhaled breath analysis may help us understand for any subject:

- Have you been exposed to a particular chemical?
- Where did it come from?
- What is your body burden (dose)?
- How rapidly are you eliminating chemicals from your body?
- What are the short and long term risks from your exposures?

From a medical, clinical, or diagnostic standpoint, breath biomarkers may answer:

- Do you have asthma, cancer, or heart disease?
- How well are your liver and kidneys functioning?
- Is the medical treatment making you better?
- How is your anesthesia progressing?
- Are you rejecting a transplant?

And, from an intelligence gathering or threat assessment perspective, breath analysis may tell us:

- Where have you been lately?
- Have you been in contact with dangerous chemicals/explosives?
- Are you under stress or lying?
- Are you infected?
- Are you suffering from radiation poisoning?

In general, the organic content of the breath provides a direct window into the human exposome. Furthermore, all biomarkers (including breath biomarkers), provide an objective and empirical record of the current exposome condition, and, especially with respect to threat analysis, provide little opportunity for the subjects to alter or obscure their chemical status. By considering the applications and questions posed as listed above, one can quickly design the appropriate sampling and analysis scheme.

Limitations of breath data statistics

Breath measurements are generally made simultaneously for a suite of compounds along with a variety of meta-data. Often there are 20 or more specific compounds measured for each subject in addition to observations such as weight, height, gender, age, ethnicity, health state, location, timing, and exposure status. However, human studies statistics are often limited by the number of available subjects. Thus, one of the most difficult tasks in assessing the value of sparse grouping data is getting an overall view of patterns, outliers, probative groups, and parameter associations that may not be readily found using a-priori statistical approaches. A number of different mathematical and graphical statistical techniques have been applied to complex data (including genomic and proteomic data) that employ comparisons of summary statistics, variable clustering, principal components analysis, and forms of multivariable regression modeling (Harrell 2002, Domany 2003, Rayne et al. 2005, Pleil and Lorber 2007, Sobus et al. 2010, Wehniger et al. 2007). We note that this article focuses only on observational techniques; detailed statistical analyses are beyond the scope of this work.

The initial approach with any complex data structure is to determine the number of successful entries achieved within groups. Despite proper planning, real-world studies often suffer from missing values due to laboratory mishaps, left-censored data due to samples below detection limits, loss to follow-up of subjects, or simply from random fluctuations. The desire of the researcher is to collect more samples as a solution, but there are always practical limitations from financial, time, and strategic constraints. Furthermore, it is difficult to develop an appropriate data strategy in the first place if little is known about variability, underlying distributions, and range of the measured quantities. As such, brute force summary statistical tables and complex mathematical techniques may not give the best initial view of a complex environmental or biomarker data structure.

Graphical data interpretation via heat maps

Numerous graphical approaches are used to assess and explain complex data; most employ x-y scatter plots, repression analysis (linear or non-linear behavior plots), or consolidations of data groups into means and variability parameters (box and whiskers plots, column comparisons, etc.). Often these types of representations are difficult to interpret especially when there are dozens of potentially interacting variables to be graphed. Such complex data visualization has been successfully established for genomics and proteomics research with the development of the heat map approach (Aumann 2007). A three-dimensional graphics approach for human breath was used in an investigation of GC x GC – ToF MS (2-dimensional gas chromatography, time of flight mass spectrometry) for exploring effects of cardiac surgery (Mieth et al. 2010). We have now adapted these heat map tools for exploring the human exposome as expressed as relationships among environmental measurements, breath biomarkers, and host factor data including time series (Pleil et al. 2011).

Heat maps are visual representations of quantitative data on two axes; the x-axis generally reflects individual samples and the y-axis consists of groups of measured parameters. The field between the axes is comprised of an array of contiguous boxes color coded to reflect quantitation. As such, heat maps are a flexible visualization tool for grouping data and exploring patterns. They can be created using various software products; in our case, we employ MatLab (MathWorks, Natick, MA, USA). The basic data structure underlying a heat map is comprised of an independent variable (human subjects, environmental samples, days or dates, etc.) and an array of dependent variables (environmental chemicals, biomarker chemicals, biological parameters, etc.). The dependent parameters are usually continuous variables that are further coded by color to reflect their quantitative value. Both axes are generally further sub-grouped. For example, the samples on the x-axis can be ordered by time collected, by distance from a source, or sub-divided by human host factors such as gender or age; the dependent variables can

be grouped by parameters such as pre- and post-treatment, geographical site, season, compound class, or chemical structure.

The construction of any particular heat map is based purely on choices made by the investigator to explore a relationship of interest between sample groups and measurements. The advantage is that all data points can be plotted without regard to underlying distributions or other parametric restrictions to get a qualitative feel for the information. Potential outliers can be efficiently recognized, and re-arrangements of data groups or sample order can be quickly implemented to investigate specific relationships.

Breath data example: exogenous compounds from Jet Fuel study

The US Air Force (USAF) organized and conducted a series of studies in conjunction with US Environmental Protection Agency (EPA) and US National Institute of Occupational Safety and Health (NIOSH) of over 20 active Air Force bases in the United States regarding environmental and occupational exposure to the military jet fuel designated JP-8 (USAF 1996, Zeiger & Smith 1998, Egeghy et al. 2003, Witten et al. 2010, Serdar et al. 2003, Pleil et al 2000). In this paper, we focus on exhaled breath samples analyses of 12 major JP-8 constituents (C9 to C12 n-alkanes and 8 single-ring aromatic compounds) from 130 human subjects (99% male) representing a subset of seven bases (coded Dye, Lan, Sey, Mon, LiR, Hur, and Pop). Subjects were preselected into exposed (n=96) or control (n=34) groups based upon whether they worked directly with fuel systems maintenance or not. Single alveolar breath samples were collected pre- and post- shift work (4-hr duration) using 1-liter volume evacuated stainless steel canisters and analyzed by gas chromatography–mass spectrometry (Pleil & Lindstrom 1995a, 1995b). Summary statistics and an evaluation of data variance components have been published (Pleil 2009). All samples were collected and analyzed under Institutional Review Protocols (IRB) approved by USAF with informed consent of the subjects.

Breath data example: endogenous compounds in EBC

US EPA and partners have conducted numerous exposure studies under controlled conditions to ascertain kinetic and statistical associations between environmental exposures and biological effects in metabolism and excretion or other perturbations (e.g. Prah et al. 2004, Sobus et al 2008, Pleil et al. 2008, Krishna et al. 1998, Kongerud et al. 2006). As part of these studies, we have had the opportunity to collect numerous pre- and post-exposure EBC samples from a series of human subjects for exploratory analysis of a variety of organic constituents including proteins, cytokines, and endogenous volatile compounds (Sawyer et al 2009, Swanson et al. 2009, Pleil et al. 2008, Hubbard et al. 2009). For this paper, we use a subset of EBC measurements comprised of 9 endogenous alcohols and 6 aldehydes measured in each of 44 samples from 10 human

subjects. These were collected in groups representing pre-, post-, and 24 hr post- exposures for 2-hours to either diesel exhaust or purified air. All subjects performed scripted exercise activity consisting of ~75 watts on a stationary bicycle alternating in 15 minute intervals. EBC samples of 1.5 to 2 ml volume) were collected using methodology developed at EPA (Sawyer et al. 2009, Hubbard et al. 2009); initial scoping experiments were also performed with commercially available RTube™ EBC collector (Respiratory Research, Inc., Austin, TX, USA) (Pleil et al. 2008). GC-MS analytical methods and a detailed mathematical and statistical description of these data have been published (Hubbard et al. 2009). All samples were collected and analyzed under Institutional Review Protocols (IRB) approved by University of North Carolina, Chapel Hill, NC, USA and US EPA with informed consent of the subjects.

Results and Discussion

The most important factor for utilizing the breath exposome for environmental and intelligence applications is a thorough understanding of the distinction between "What is normal?" and "What is an outlier?" Those samples that do not fit into the unremarkable, expected exposome pattern or have an obvious individual compound outlier could be indicators of an environmental upset or an internal biological disturbance. For example, the appearance of a group of exogenous chemicals could point to a particular source or recent location (e.g. jet fuel from an air base, solvents from a drug laboratory), a specific marker compound could indicate recent handling of explosives, or a shift of the normal pattern of metabolites could indicate mental stress, radiation poisoning, or infection.

The timing between environmental contact and data collection is also important, especially for reconstructing recent activities. Very volatile compounds such as small alkanes and aromatic hydrocarbons are absorbed quickly, but are also released quickly, whereas metabolites experience biological damping or delays that can make them linger in the system (Sohn et al. 2004, Sahmel et al. 2010, Barton et al. 2006, Pleil et al. 2007). Dermal and ingestion exposures may have longer excretion times than the same compounds from inhalation exposure (Prah et al. 2004, Kim et al. 2007). The absorption, metabolism, distribution, and excretion (ADME) parameters can vary from person to person as can the relative amount of exposure of individuals nominally within the same environment (Grossman 2009, Jamei et al. 2009, Pleil 2009). The current state of the science relies primarily on controlled exposure studies for determining the ADME kinetics of environmental compounds, and on cross sectional or observational studies that determine the normal range and patterns of metabolites in unremarkably stressed human subjects.

All of these effects can potentially be observed in the complex breath exposome given the appropriate sampling and analytical methodologies and the observational tools. We present in

the following sections two examples where the heat map approach can be implemented to very quickly answer some specific and relevant questions assessing exposures and potential threats.

Jet fuel data heat map

Before developing specific heat maps, we need to decide how to initially arrange the data to begin to tell the exposure story. The initial questions for exploring jet fuel exposure based on exhaled breath are:

- Which JP-8 constituents indicate JP-8 exposure as found in breath?
- In what order should samples be arranged?
- In what order should compounds be arranged?
- What range/distribution and color scheme is best for the quantitative display?

Based on previous work (Pleil et al. 2000), we had recognized that the sum of the n-alkanes (nonane, decane, undecane, and dodecane) found in breath can serve as an ordinal marker (or fingerprint value) for cumulative exposure, and that the aromatics (benzene, toluene, ethylbenzene, o-xylene, m,p-xylene, 1,3,5-trimethylbenzene, 1,2,4-trimethylbenzene, and 4-ethyl toluene) are all lesser constituents of jet fuel found in breath. Based on this preliminary information, we ordered the samples (x-axis) within the exposed and control groups, respectively, from lowest to highest post exposure based on the n-alkane fingerprint. We then ordered the compounds (y-axis) by mean value of post exposure level keeping the alkanes separate from the aromatics. We had found that the overall range of concentrations was from about 0.01 ppbv to about 1,000 ppbv, and that, within compounds, the concentration values were log-normally distributed.

The first heat map (Fig. 1) was constructed to partition the axes based on the grouping of controls vs. exposed (x-axis) and pre- vs. post-shift work on the y-axis, and to reflect the within group orders established for sample rank (x-axis) and compound rank (y-axis). The quantitation color scheme was chosen on a log-scale as ranging from dark blue = lowest (0.01 ppbv), to yellow = midrange (10 ppbv), through dark red = highest (1000 ppbv). The color scheme and the group annotations are included as part of the illustration.

From Figure 1, we can make a number of quick observations about JP-8 exposure:

- The proposed 12 compounds all reflect JP-8 exposure to some extent (compare post-shift to pre-shift samples of the exposed subject cohort).
- Overall, the alkanes are more prominent markers in breath than the aromatics.
- Benzene and toluene are both higher in the pre-shift samples than all other JP-8 constituents.

- Benzene and toluene vary the least from left to right with respect to JP-8 exposure ranking.
- Some post-shift controls have slightly higher levels than their pre-shift counterparts.
- A subset of the control subjects exhibits slightly higher JP-8 constituent levels than the lowest levels found in the exposed group.

The first two observations support that exhaled breath reflects the initial analyses of JP-8 constituents. The observations regarding benzene and toluene indicate that there are likely other sources of these compounds at the Air Force bases. However, the last two observation regarding the comparisons of certain post and pre-shift samples, as well as control and exposed cohort samples are puzzling and cannot be interpreted using this heat map configuration.

We realize that the seven bases have distinctive fuels systems procedures, different airplanes types, and varying fuels usage and so there may be between-base differences that might explain these anomalies. We subsequently created a second heat map to further partition the x-axis by Air Force base, but kept the within subgroup order the same (Fig. 2). Here we see that within any base, the control subjects almost always exhibit lower exhaled values than even their lowest exposed counterparts. However, there are some bases (i.e. Sey and LiR) that exhibit higher control subject levels than some of the exposed level subjects at Dye. With the exception of two post-shift samples at Mon, control subjects show generally consistent within-base distinction from their exposed cohort counterparts, with the exception of a few subjects at Sey, Lan, that are too similar to observe on the heat map. The question regarding control subjects with increasing post-shift levels cannot be resolved here. However, we see that most of this effect is concentrated at Mon and Lan suggesting that those two bases have more pervasive overall jet fuel concentrations that affect everyone on the base. We surmise that this could be due to weather conditions during the sampling times (perhaps stagnant air), or that control subjects work more closely with fuels systems personnel at Mon and Lan than at other bases.

Endogenous EBC data heat map

Screening for subjects with an environmentally induced metabolism disturbance, exhibiting internal stress, posing an infection risk, or with subclinical radiation damage could be accomplished by comparing a suite of normally occurring internal metabolites to pattern shifts in affected individuals. In a preliminary human study of EBC from nine subjects, 16 polar volatile organic compounds (nine alcohols, seven aldehydes) were measured using a repeat analysis sample design (Hubbard et al. 2009). Although this is as yet a very limited data set, the concept is sound; the chosen compounds are excreted as normal breath metabolites, the concentration ranges can be established for healthy human subjects, and there are specific within- and between-person variances that establish what is considered normal. Certainly case-control studies for

subjects affected by infection or radiation still need to be performed to determine what compounds are probative as there is no information currently available from controlled studies.

We ordered the x-axis of this data set by subject with sub-groupings of pre-, post- and 24 hr-post samples. The y-axis was organized by major grouping of clean air (control) and diesel exhaust (treatment) in a vertical stack, with subgrouping of endogenous biomarkers arranged in increasing relative prevalence in this order: 3-methyl-3-pentanol, 4-methyl-3-pentanol, 1-hexanol, 3-ethyl-1butanol, 1-pentanol, 3-methyl butanal, pentanal, butanal, 1-heptanol, heptanal, hexanal, 1-butanol, 2-methyl propanal, octanal, 2-methyl-1-propanol, 1-propanol (Figure 3).

From Figure 3, we make the following initial observations:

- There does not appear to be a treatment effect; that is, the clean air and diesel exhaust groupings are indistinguishable within subject (vertical direction).
- Within subject samples are essentially indistinguishable among pre-, post-, and 24-hr post-treatment samples (horizontal direction); there does not appear to be a consistent exercise effect.
- There is a large consistent between-subject variance effect as seen in intensity difference in between-subject comparisons (horizontal direction).

We interpret these initial results to mean that 100 $\mu g/m^3$ diesel exhaust exposure for 2-hrs does not affect the exhalation of the normally expressed alcohols and aldehydes, nor that the scripted exercise activity changes the internal metabolism appreciably as seen by the consistency in the vertically stacked groupings; these results were unexpected. Additionally, the large between-subject effect is puzzling; it appears to be random, and is presumably due to some fundamental difference among host factors of the subjects. We note that in the construction of the heat map that the data comparisons are balanced in the vertical direction; however, there are obviously some missing samples within subjects due to losses from unavoidable occurrences.

We chose to investigate the between-subject variance issues in a new heatmap wherein we explore subject gender as the discriminating parameter. In Figure 4, we further sub-grouped all samples by collecting pre-, post, and 24 hr post samples within gender; the y-axis was kept the same. The results of this rearrangement were striking; there is an obvious across the board gender difference in the expression of endogenous metabolites in EBC. At least in this limited sample set, male subjects appear to have appreciably higher levels of the measured polar organic metabolites in breath. Furthermore, even when pre-, post-, and 24 hr post samples were grouped together and compared as gender-based groups, we fail to see an obvious timing effect. Again, we note that the data are not completely balanced along the x-axis; this is an unavoidable consequence of using volunteer subjects who sometimes are not available for, or cannot complete, certain of the more time consuming tasks.

Conclusions and Recommendations

Interpreting the human exposome is a valuable, yet difficult, endeavor for assessing the current and past status of the environmental interaction with the human systems biology. From the standpoint of population-based studies (non-invasive samples from many subjects) or from the perspective of the intelligence community (covert sample collection), breath is an attractive biological medium. We have shown with two examples, that even complex breath biomarker data can be quickly scanned for overall patterns that may have probative value in answering specific questions. The heat map approach that we described is an excellent qualitative screening tool for quickly exploring broad hypotheses regarding relationships among exposure measurements, biomarkers, timing, meta-data, and host factors before computational efforts are expended.

Specifically, we showed that a complex environmental exposure to jet fuel affects the exposome, as seen with exhaled breath measurements. We conclude that the established pattern could be used to discern if a given individual had indeed been recently exposed. This information is valuable both from the environmental risk assessment standpoint in reconstructing exposure for assessing risk, as well as from an intelligence gathering perspective for assessing if a suspect has been in recent contact with fuel-based chemicals. We also showed that the exposome incorporates a series of endogenous compounds in EBC that presumably reflect ongoing metabolism, and that the within-person person pattern is stable in the face of modest diesel exhaust exposures and exercise activity. We further demonstrated that between-person variability in breath metabolites is likely affected by gender, but could not confirm this effect for the exogenous JP-8 compounds as the exposed subjects were all male. We conclude from this limited data that the heat map patterns can serve as a baseline guide for assessing what is unremarkable, and that we must be very cognizant regarding subject gender when interpreting the endogenous exhaled breath exposome. From an environmental perspective, the development of the unremarkable endogenous biomarker pattern is valuable for detecting if a sub-population is under some form of metabolism perturbation pressure in comparison; from the security standpoint, any major divergence from the normal exposome pattern may indicate stress or possibly a sub-clinical indication of infection or radiation poisoning.

The work presented here is only the beginning. We recommend that for this approach to become a broadly applied investigative technique, it will require substantial data sets and their related heat maps to visually demonstrate what is within the normal range, and what is somehow remarkable. Only then can individual "unknowns" be interpreted within the context of the exposome subset of interest, as measured in the breath. We further concede that qualitative interpretation is not necessarily sufficient to draw conclusions. Subsequent detailed statistical analyses will always be required to calculate the quantitative aspects (e.g. correlations, p-values, etc.), as well determining the confidence we can put into a qualitative assessment. Finally, we suggest that in addition to providing an a-priori exploration of raw data, the heat map approach can also be used in retrospect to illustrate results from statistical and multivariable analyses that

would otherwise be relegated to complex numerical tables and/or a series of two-dimensional graphs.

Acknowledgements:

The authors wish to thank Tzipporah Kormos, Sarah Bereznicki, and Myriam Medina-Vera of US EPA for their expert advice. This research has been subjected to (EPA) Agency review and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use. Interested readers are invited to contact the authors for assistance and advice in constructing specific heat maps.

References:

Aggarwal K, Lee KH. 2003. Review: Functional genomics and proteomics as a foundation for systems biology, Brief Funct Genomic Proteomic. 2(3):175-184.

Amann A, Smith D, eds. 2005. "Breath Gas Analysis for Medical Diagnostics". World Scientific Publishers, Singapore.

Aumann JT, Boorman GA, Wilson RE, Travlos GS, Paules RS. 2007. Heat map visualization of high density clinical chemistry data. Physiological Genomics 31:352-356.

Barton HA, Pastoor TP, Baetcke K, Chambers JE, Diliberto J, Doerrer NG, Driver JH, Hastings CE, Iyengar S, Krieger R, Stahl B, and Tmchalk C, 2006. The acquisition and application of absorption, distribution, metabolism, and excretion (ADME) data in agricultural chemical safety assessments, Critical Reviews in Toxicology 36(1):9-35.

Buszewski B, Kesy M, Ligor T, and Amann A, 2007. Human exhaled air analytics: biomarkers of diseases, Biomed. Chromatogr. 21:553-566.

Domany, E. 2003, Cluster analysis of gene expression data. J. Stat. Phys., 110 (3-6): 1117-1139.

Egeghy PP, Hauf-Cabalo L, Gibson R, and Rappaport SM, 2003. Benzene and naphthalene in air and breath as indicators of exposure to jet fuel. Occup Environ Med 60:969–976.

Grossman I, 2009. ADME pharmacokinetics: current practices and future outlook, Expert Opinions in Drug Metabolism and Toxicology 5(5):449-462.

Harrell, FE, 2002. Chapter 4. Regression Modeling Strategies. In *Multivariable Modeling Strategies*; Springer-Verlag, New York.

Hubbard HF, Sobus JR, Pleil JD, Madden MC, and Tabucchi S, 2009. "Application of novel method to measure endogenous VOCs in exhaled breath condensate before and after exposure to diesel exhaust", Journal of Chromatography, B 877(29):3652-3658.

Jamei M, Dickinson GL and Rostami-Hodjegan A, 2009. A framework for assessing inter-individual variability in pharmacokinetics using virtual human populations and integrating general knowledge of physical chemistry, biology, anatomy, physiology, and genetics: A tale of 'bottom-up" vs. 'top-down' recognition of variates, Drug Metabolism and Pharmacokinetics 24(1):53-75.

Kim D, Andersen ME, Pleil JD, and Nylander-French LA, Prah JD, 2007. "Refined PBPK model of Aggregate Exposure to Methyl t*ertiary*-Butyl Ether", Toxicological Letters, 169(3):222-235.

Kongerud J, Madden MC, Hazucha M, and Peden D, 2006. Nasal responses in asthmatic and nonasthmatic subjects following exposure to diesel exhaust particles, Inhalation Toxicology 18(9):589-594.

Krishna MT, Chauhan AJ, Frew AJ, and Holgate ST, 1998. Effects of 0.2 ppm ozone on biomarkers of inflammation in bronchoalveolar lavage fluid and bronchial mucosa of healthy subjects, European Respiratory Journal 11(6):1294-1300.

Marczin N and Yacoub MH, eds., 2002. Disease markers in exhaled breath: Basic mechanisms and clinical applications, NATO Science Series, 346. IOS Press, Amsterdam, Berlin.

Mieth M, Schubert JK, Groger T, Sabel B, Kischkel S. Fuchs P, Hein D, Zimmerman R, and Miekisch W, 2020. Automated needle trap heart-cut GC-MS and needle trap comprehensive two-dimensional GC/TOF-MS for breath analysis in the clinical environment, Anal. Chem. 82:2541-2551.

Pauling L, Robinson AB, Teranishi R, and Cary P, 1971. "Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography", Proceedings of the National Academy of Sciences USA, 68(10):2374-2376.

Phillips M, Herrera J, Krishnan S, Zain M, Greenberg J, and Cataneo RN, 1999. "Variation in volatile organic compounds in the breath of normal humans", Journal of Chromatography, B 877(29):3652-3658.

Pleil JD, Lindstrom AB, 1995a. Collection of a single alveolar exhaled breath for volatile organic compounds analysis. Am J Ind Med 28:109–121.

Pleil JD and Lindstrom AB, 1995b. Measurement of volatile organic compounds in exhaled breath as collected in evacuated electropolished canisters. Journal of Chromatography, B 665:271–279.

Pleil JD, Fisher JW, and Lindstrom AB, 1998. "Comparison of human blood and breath levels of trichloroethylene from controlled inhalation exposure", Environmental Health Perspectives 106(9): 573-580.

Pleil JD, Smith LB, and Zelnick SD, 2000. "Personal exposure to JP-8 jet fuel and exhaust at Air Force bases", Environmental Health Perspectives, 108(3): 183-192.

Pleil JD and Lorber MN, 2007. "Relative Congener Scaling of Polychlorinated Dibenzo-p-dioxins and Dibenzofurans to Estimate Building Fire Contributions in Air, Surface Wipes, and Dust Samples", Environmental Science and Technology, 41(21):7286-7293.

Pleil JD, Kim D, Prah J, Ashley DL and Rappaport SM, 2007. "Exposure reconstruction for reducing uncertainty in risk assessment: Example using MTBE biomarkers and a simple pharmacokinetic model", Biomarkers 12(4):331-348.

Pleil JD, 2008. "The Role of Exhaled Breath Biomarkers in Environmental Health Science", Journal of Toxicology and Environmental Health: Part B: Critical Reviews, 11:613-626.

Pleil JD, Hubbard HF, Sobus JR, and Madden MC. 2008. "Volatile polar metabolites in exhaled breath condensate (EBC): collection and analysis", Journal of Breath Research, 2: 026001.

Pleil JD, 2009. "Influence of systems biology response and environmental exposure level on between-subject variability in breath and blood biomarkers", Biomarkers 14(8):560-571.

Pleil JD, 2010. "Breath Biomarkers Networking Sessions at PittCon 2010, Orlando, Florida", Journal of Breath Research, 4:029001 5pp.

Pleil JD and Sheldon LS, 2010. Adapting Concepts from Systems Biology to Develop Systems Exposure Event Networks for Exposure Science Research, Biomarkers (in press).

Pleil JD, Stiegel MA, Madden MC, and Sobus JR, 2011. "Heat Map Visualization of Complex Environmental and Biomarker Measurements", Chemosphere (in press).

Prah, J, Ashley D, Blount B, Case M, Leavens T, Pleil J, Cardinali F, 2004. Dermal, oral and inhalation pharmacokinetics of methyl tertiary butyl ether (MTBE) in human volunteers Toxicol Sci 77:195–205.

Rappaport SM, 2010. Implications of the exposome for exposure science, Journal of Exposure Science and Environmental Epidemiology 21(1)5:9.

Rappaport SM and Smith MT, 2010. Environment and Disease Risk, Science, 33(6003):460-461.

Rayne S, Ikonomou MG, Butt CM, Diamond ML, and Truong J, 2005. Polychlorinated dioxins and furans from the World Trade Center attacks in exterior window films from Lower Manhattan in New York City. Environ. Sci. Technol. 2005, 39, 1995-2003.

Sahmel J, Devlin K, Paustenbach D, Hollins D, and Gaffney S, 2010. The role of exposure reconstruction in occupational human health risk assessment: current methods and a recommended framework, Critical Reviews in Toxicology 40(9):799-843.

Sawyer K, Samet JD, Ghio AJ, Pleil JD, and Madden MC, 2009. "Responses measured in the exhaled breath of human volunteers acutely exposed to ozone and diesel exhaust", Journal of Breath Research, 2:037019, 9pp.

Serdar B, Egeghy PP, Waidyanatha S, Gibson R, and Rappaport SM, 2003. Urinary biomarkers of exposure to jet fuel (JP-8). Environ Health Perspect 111:1760–4.

Sobus JR, Pleil JD, Madden MC, Funk WE, Hubbard HF, and Rappaport SM, 2008. "Identification of surrogate measures of diesel exhaust exposure in a controlled chamber study", Environmental Science and Technology, 42:8822-8828.

Sobus JR, Pleil JD, McClean MD, Herrick RF, and Rappaport SM, 2010. "Biomarker variance component estimation for exposure surrogate selection and toxicokinetic inference", Toxicology Letters 199(3):247-253.

Sohn MD, McKone TE, and Blancato JN, 2004. Reconstructing population exposures from dose biomarkers: inhalation of trichloroethylene (TCE) as a case study, Journal of Esposure Analysis and Environmental Epidemiology 14(3):204-213.

Swanson KJ, Kado NY, Funk WE, Pleil JD, Madden MC, and Ghio AJ, 2009. "Release of the Pro-inflammatory Markers by BEAS-2B Cells Following in vitro Exposure to Biodiesel Extracts", The Open Toxicology Journal, 3:8-15.

USAF, 1996. United States Air Force, JP-8 composition and variability, AL/EQ-TR-1996-0006, edited by Howard T., Mayfield . Armstrong Laboratory, Air Force Material Command, Tyndall Air Force Base, FL, USA.

Wehniger A, Schmid A, Mechtcheriakov S, Ledochowski M, Grabmer C, Gastl G, and Amann A, 2007. Lung cancer detection by oroton transfer reaction mass-spectrometric analysis of human breath gas. Int. J. Mass Spectrom. 265:49-59.

Wild CP, 2005. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol Biomarkers Prev 14(8):1847-50.

Witten ML, Zeiger E, and Ritchie GD, eds., 2010. *Jet Fuel Toxicology,* Taylor and Francis, Inc., New York.

Zeiger E, and Smith L, 1998. The first international conference on the environmental health and safety of jet fuel. Environ Health Perspect 106:763–764.

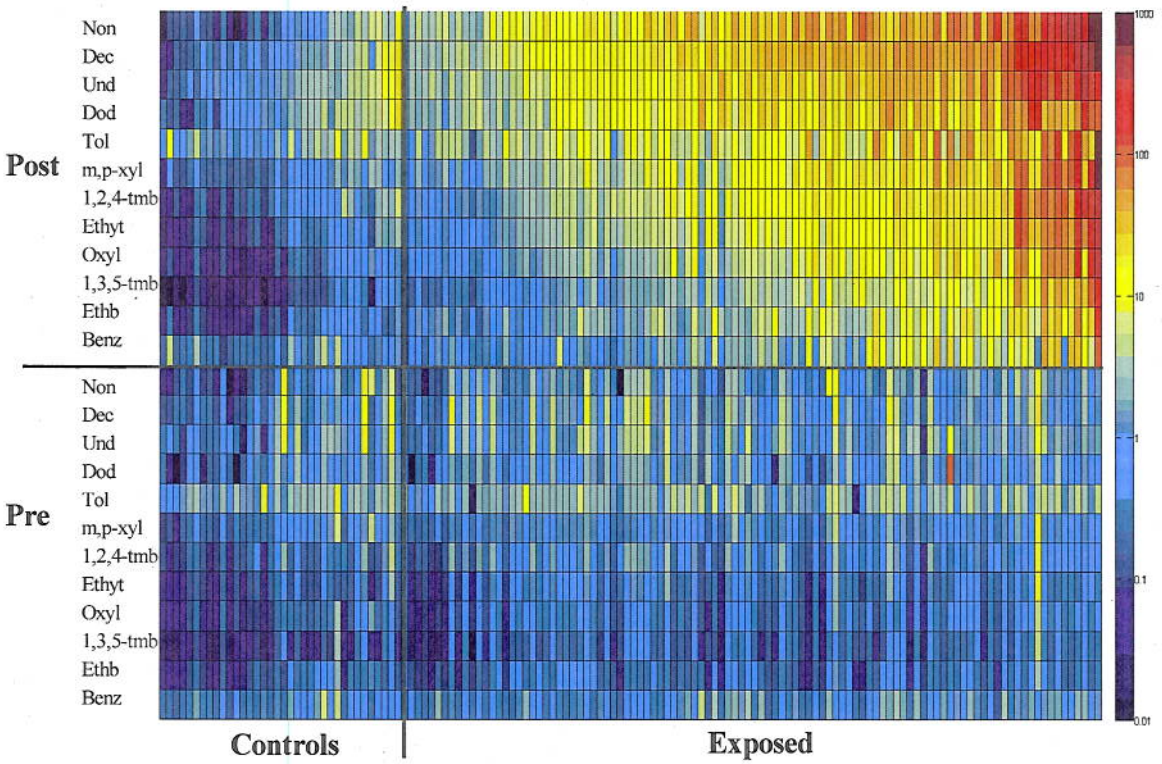

Figure 1. Heat map: jet fuel constituents in exhaled breath of 130 subjects; x-axis represents individuals partitioned into control group (n = 34) and exposed group (n=96) ordered within group from left to right based on increasing post shift n-alkanes fingerprint; y-axis represents measured compounds partitioned by pre- and post-work shift and ordered within group by increasing (upwards) relative abundance. Quantitation is reflected by color intensity ranging from 0.01 to 1000 ppbv in a log scale (dark blue = lowest, dark red = highest) as displayed in the far right column.

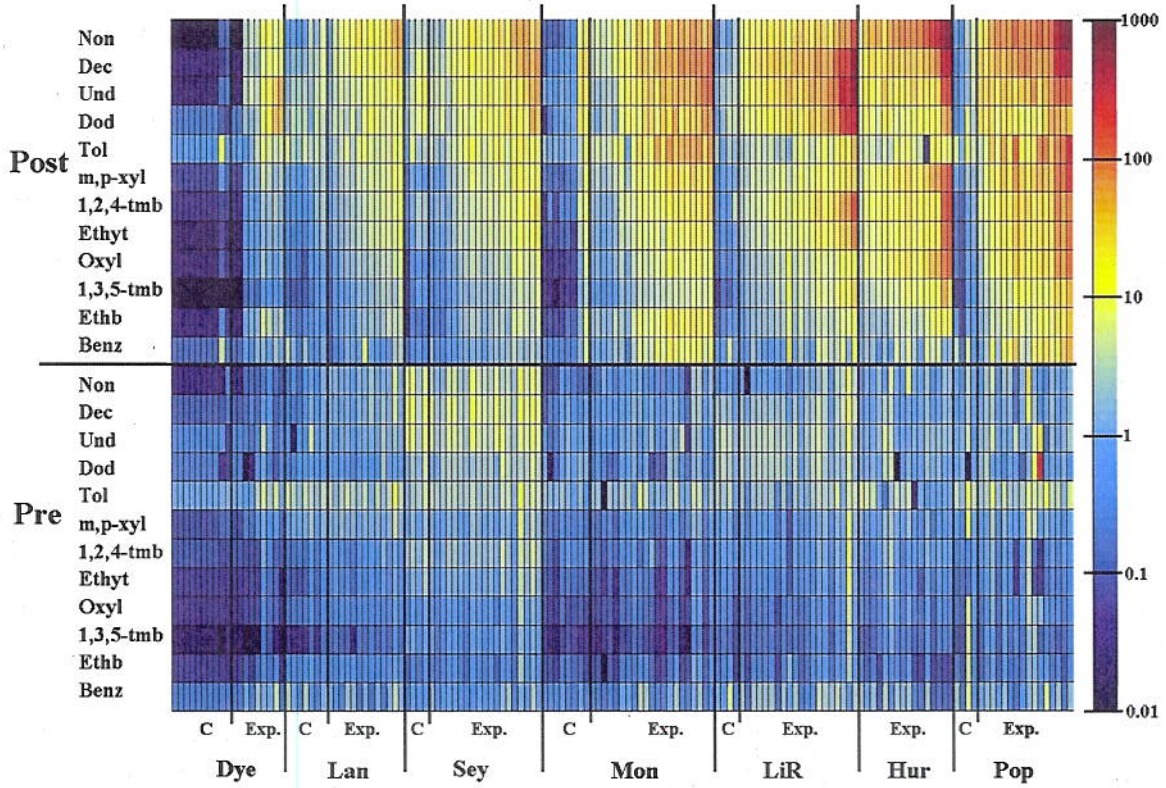

Figure 2. Heat map: jet fuel constituents in exhaled breath of 130 subjects; x-axis represents individuals partitioned into seven groups by AF base, and then sub-grouped within base by control vs. exposed status and ordered within subgroup by increasing post shift n-alkanes fingerprint. The y-axis and quantitation color scheme remain the same as in Figure 1.

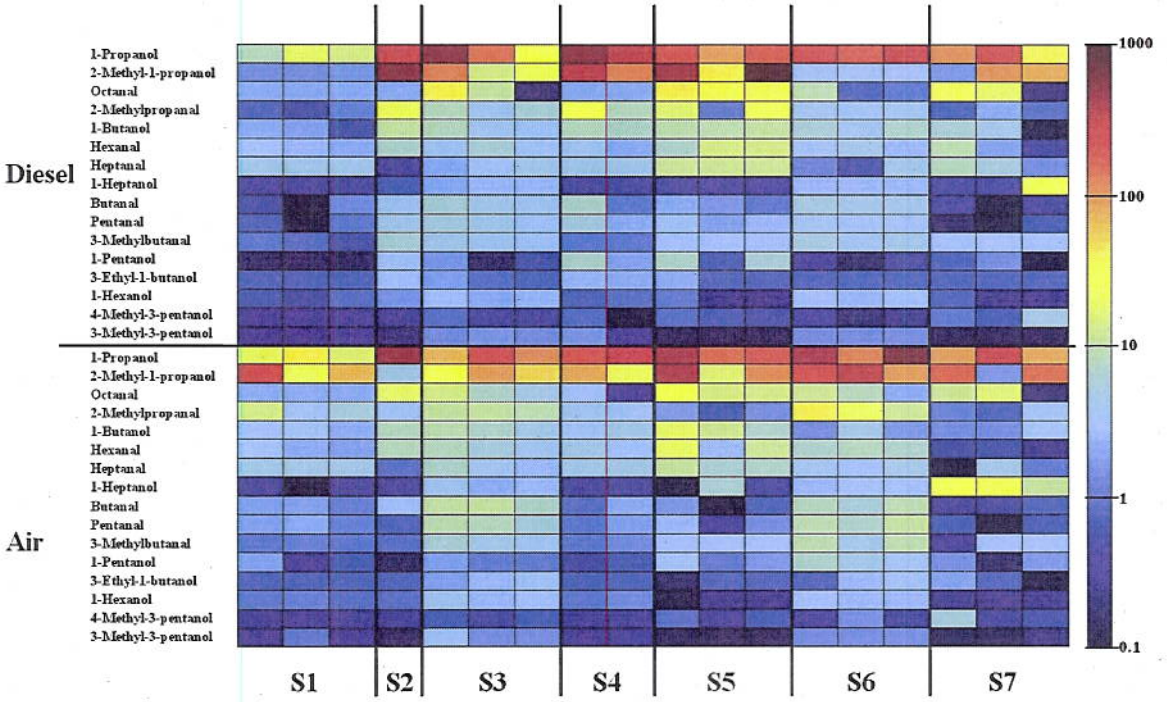

Figure 3. Heat map: endogenous biomarkers in exhaled breath condensate of seven subjects; x-axis is grouped by subject representing all vertically matched data ordered as pre, post, and 24hr post as available; y-axis in partitioned into clean air (sham) exposure and 100 $\mu g/m^3$ diesel exhaust exposure experimental groups and ordered within group by compounds in increasing (upwards) mean concentration. Quantitation is reflected by color intensity ranging from 0.1 to 1000 ppbv in a log scale (dark blue = lowest, dark red = highest) as displayed in the far right column. Only data sets with vertically matched samples could be plotted, as such, the x-axis is unbalanced with respect to samples per subject.

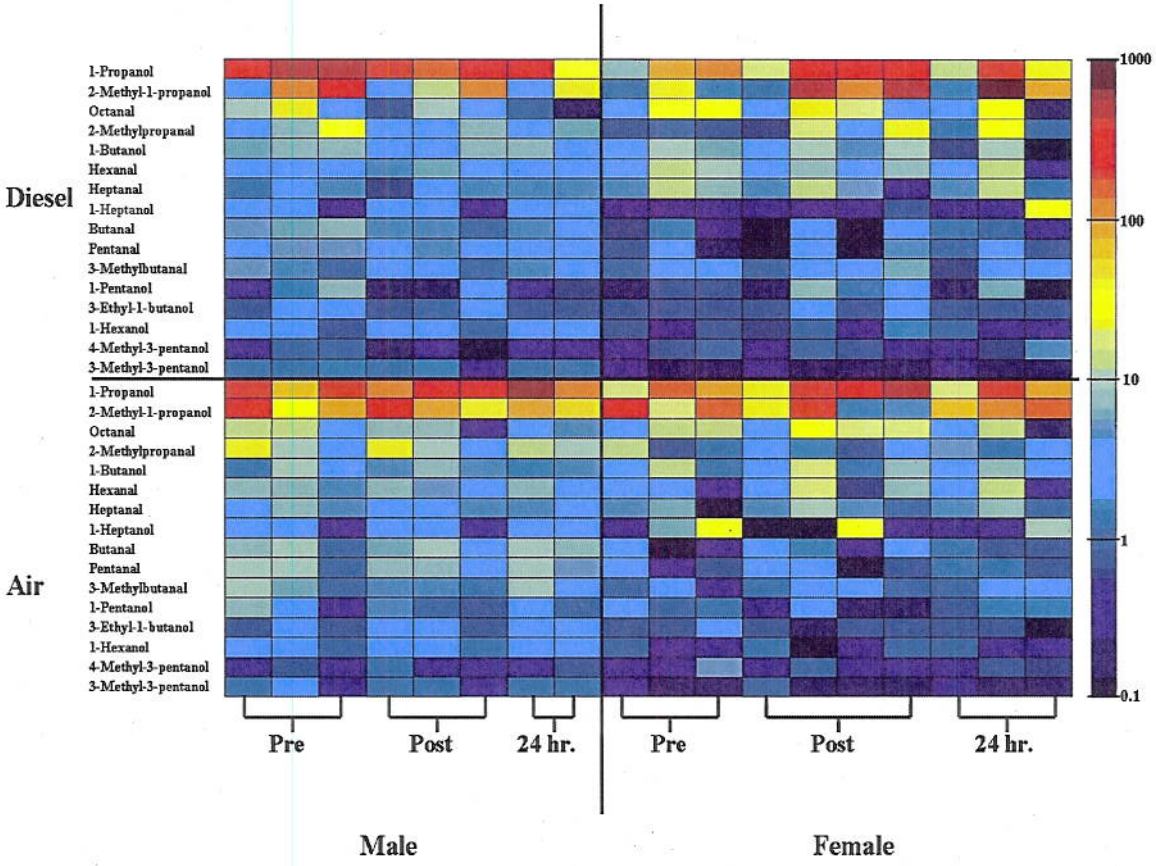

Figure 4. Heat map: endogenous biomarkers in exhaled breath condensate of seven subjects; x-axis is partitioned by subject gender and sub divided into pre, post, and 24hr post exposures representing all vertically matched data. The y-axis and quantitation color scheme remain the same as in Figure 3. Only data sets with vertically matched samples could be plotted, as such, the x-axis is unbalanced with respect to samples per time frame.