Kidney injury biomarkers and urinary creatinine variability in nominally healthy adults

M.A. Stiegel[1,2], J.D. Pleil[3]*, J.R. Sobus[3], M.M. Angrish[2], and M.K. Morgan[3]

[1]Department of Environmental Sciences and Engineering, Gillings School of Public Health, University of North Carolina, Chapel Hill, NC, 27599, USA
[2]ORISE Research Fellow, US EPA, Research Triangle Park, NC
[3]National Exposure Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, USA

Abstract

Environmental exposure diagnostics use creatinine concentrations in urine aliquots as the internal standard for dilution normalization of all other excreted metabolites when urinary excretion rate data are not available. This is a reasonable approach for healthy adults as creatinine is a human metabolite that is continually produced in skeletal muscles and presumably excreted in the urine at a stable rate. However, creatinine also serves as a biomarker for glomerular filtration rate (efficiency) of the kidneys so undiagnosed kidney function impairment could affect this commonly applied dilution calculation. The United States Environmental Protection Agency (US EPA) has recently conducted a study that collected approximately 2600 urine samples from 50 healthy adults, ages 19-50 years old, in North Carolina in 2009-2011. Urinary ancillary data (creatinine concentration, total void volume, elapsed time between voids), and participant demographic data (race, gender, height and body weight) were collected. A representative subset of 280 urine samples from 29 participants was assayed using a new kidney injury panel (KIP). In this article, we investigated the relationships of KIP biomarkers within and between subjects, and also calculate their interactions with measured creatinine levels. The aims of this work were to document the analytical methods (procedures, sensitivity, stability, etc.), provide summary statistics for the KIP biomarkers in "healthy" adults without diagnosed disease (distribution, fold range, central tendency, variance), and to develop an understanding as to how urinary creatinine level varies with respect to the individual KIP proteins. Results show that new instrumentation and data reduction methods have sufficient sensitivity to measure KIP levels in nominally healthy urine samples, that linear regression between creatinine concentration and urinary excretion explains only about 68% of variability, that KIP markers are poorly correlated with creatinine ($r^2 \sim 0.34$), and that statistical outliers of KIP markers are not random, but are clustered within certain subjects. In addition, we interpret these new adverse outcome pathway (AOP) based *in vivo* biomarkers for their potential use as intermediary chemicals that may be diagnostic of kidney adverse outcomes to environmental exposure.

* corresponding author: pleil.joachim@epa.gov

Introduction

The discipline of environmental exposure science establishes the continuum from environmental sources of chemicals to human uptake, internal dose, and eventual to adverse health outcome (Lioy 1990, Sobus et al. 2011, Tan et al. 2012, a 2012a, Bean et al. 2014). Two of the main concerns are to deduce the potential impact of exogenous chemicals on human health and to establish more accurate risk assessments (Cohen-Hubal, et al. 2010, Kavlok and Dix 2010). This is becoming an overwhelmingly important issue as current scientific research suggests: *"…although chronic diseases are primarily environmental (i.e., not genetic) in origin, the particular environmental causes of these diseases are poorly understood"* (Rappaport 2012). In fact, it is now generally accepted that 70-90% of all chronic disease risk is attributable to environmental factors and that new diagnostic strategies such as environment-wide association studies (EWAS) are required to understand the gene-environment risk paradigm (Rappaport 2010, Patel et al. 2010, Lind et al 2013). The goal is to interpret not only the internal dose from the environment, but also the subtle metabolic effects on the human exposome, systems biology, and health outcomes (Pleil and Sheldon, 2011, Edwards and Preston 2008). Ultimately, the hope is to identify the impact of exogenous chemicals on the biochemical adverse outcome pathways (AOP) at the molecular level and to diagnose their initiation in human *in vivo* from *in vitro* studies (Vinken 2013, Ankley 2010, Pleil et al. 2012b, Villeneuve et al. 2014a,b).

A mainstay of achieving the "environment-to-effect linkage" is using biomonitoring of endogenous and exogenous chemicals within the human system. This is a complex task in that it is not generally known, *a priori*, which of the thousands of potential biochemicals are useful in preclinical diagnosis of the AOP. Therefore, biomarker measurement strategies fall into two categories, "discovery" and "targeted" analyses (Pleil and Stiegel, 2013). In discovery analyses, one analyzes and documents as many compounds as possible in biological media without preconception or prior knowledge to begin to understand what the concentration distribution of any particular compound looks like in the general population. This defines the homeostatic or sustainable range (Pleil 2012). The targeted approach subsequently focuses on specific compounds that have probative value to diagnose subtle changes in metabolism or biochemistry that may indicate an activation of an AOP. These approaches rely on measurements from biological media such as blood, breath, and urine (Pleil 2012, Paustenbach and Galbraith 2006, Angerer et al. 2007, Au 2007, Pleil 2008). In fact, the U.S. Centers for Disease Control and Prevention (CDC) have been operating the National Health and Nutrition Examination Survey (NHANES) since 1971 (http://www.cdc.gov/nchs/nhanes/history.htm). Current surveys include measurements of over a 1000 exposure and health related chemicals in human biological media) which are now being used to address public health concerns from environmental exposures through various interpretive, graphical, and statistical analyses (Wambaugh et al. 2013, Sobus et al. 2014, Pleil and Sobus 2013).

Human exposure assessment studies traditionally use creatinine concentration to correct for urinary dilution from differences in hydration state based on the assumption that creatinine excretion is constant and thus proportional to the excretion of other urinary metabolites (Barr et al. 2005). In medical diagnostics, however, blood-borne creatinine also serves as one of the biomarkers for glomerular filtration rate (efficiency) of the kidneys (Becker and Friedman 2013). Therefore, the practice of using creatinine to normalize metabolite concentrations in urine

without considering kidney function impairment may obscure the true dilution factor. This could lead to biased environmental exposure assessments based on excreted chemical metabolites such as those from pesticides, polycyclic aromatic hydrocarbons, chemical fire retardants, the microbiome and others (Hill et al. 1995, Lu et al. 2014, Meeker et al. 2013, Pleil et al. 2014a).

We note that this research also provides value beyond environmental health assessments. In the pharmaceutical industry, for example, one of the major obstacles to marketing a promising drug is renal toxicity of the excreted compound or its metabolites (Fuchs and Hewitt 2011, Choudhury and Ahmed 2006). Understanding and monitoring preclinical bioindicators of effect is an important feature of drug development, however, this is traditionally done with homogeneous rodent (mouse, rat, etc.) models (Vaidya et al. 2006, Bonventre et al. 2010, Slocum et al. 2012). As such, the cross-sectional human research presented here bridges across species and identifies the "healthy" or "normal" ranges and variability of kidney injury markers in the heterogeneous human population. This is also important for interpreting health state at the cellular level of organization; new *in vitro* assays are now incorporating human primary cell lines for toxicity testing and so developing a knowledge base of molecular level indicators of damage is crucial for interpreting results, and then ultimately linking these results back to systemic *in vivo* health outcomes (Angrish et al. 2015). In essence urinary excretion biomarkers may link key events with adverse outcomes across complex biological space (Ankley et al. 2010; Villeneuve et al. 2014a,b).

In environmental assessments, the selected subject pool from the general public is nominally healthy (free of diagnosed disease) and consequently falls into the "control" category in the vernacular of medical assessments (Hudman et al. 1997, Pleil 2012). We have found, however, that there is unexplained variability in creatinine concentrations (presumably beyond hydration/dilution) in some subjects and are now exploring undiagnosed sub-clinical kidney function impairment as a potential explanatory parameter, or at least as a way to flag outlier samples. We are implementing novel immunochemistry panels developed by Meso Scale Discovery (MSD, Gaithersburg, MD) to gain an understanding of subtle perturbations in kidney function. Despite the fact that these diagnostics were originally designed to identify clinical levels of kidney injury, the MSD instrumentation and kidney injury panel (KIP) reagents have sufficient documented sensitivity to detect target biomarkers in the nominally healthy population. Another aspect of this work is to observe the *in vivo* results for eventual application to high throughput *in vitro* studies of chemical toxicology. Herein the approach would be to adapt the specific patterns and levels of the KIP markers and use these as *in vitro* diagnostics in toxicity testing (e.g. Houck et al. 2009).

In this work, we perform four related tasks:

- Document analytical methods (procedures, sensitivity, stability, etc.) for analyzing seven KIP proteins.

- Provide summary statistics (distribution, fold range, central tendency, variance, co-variance, etc.) for the KIP biomarkers in "healthy" adult subjects (those without diagnosed disease).

- Develop an understanding as to how urinary creatinine levels vary with respect to the individual KIP proteins and within the context of urinary excretion rate.

- Interpret the KIP results for future work extending to spot-measure dilution corrections and for potential applications to linking cross-sectional *in vivo* studies to future *in vitro* studies of adverse outcome pathways (AOP).

Methods

"Ex-R" study samples

All urine samples were derived from "The Pilot Study to Estimate Human Exposures to Pyrethroids Using An Exposure Reconstruction Approach (Ex-R study)". This study investigated the dietary and residential exposures of adults to pyrethroid insecticides using environmental sampling (dust, food, water, and surface wipes) and urinary biomonitoring. The study was conducted at the United States Environmental Protection Agency's (EPA) Human Studies Facility in Chapel Hill, North Carolina and at participants' homes within a 40-mile radius of this facility between November 2009 and May 2011. Fifty adults (ages of 19-50 years old) were selected from a pool of volunteers screened for pre-existing conditions that could affect urine output (i.e. heart disease, diabetes, kidney disease) and recruited into the ExR study.

Ex-R urine samples were collected using a repeat-sampling strategy wherein subjects self-collected individual (whole) voids for 24-hour periods distributed across a number of weeks. Subjects also recorded the time and date for each void allowing us to subsequently calculate urinary excretion rate. We have available data including gender, ethnicity, age and weight, and height. For this work, we explore only physiological excretion parameters and endogenous compounds from a subset of the available urine samples.

Human subject protocols

All study samples and meta-data were collected under approved Institutional Review Board (IRB) protocols. The Ex-R study protocol and procedures to obtain informed consent from the adult participants followed the guidelines set forth by the Scientific and Ethical Approaches for Observational Exposure Studies report (US EPA 2008) and were reviewed and approved by the US EPA's Human Subjects Approving Official and the University of North Carolina's Institutional Review Board (IRB study number 09-0741). Adult participants signed informed consent documents prior to participating in the Ex-R study.

KIP analyses sample selection

We selected a subset of 280 urine samples representing 29 Ex-R study subjects. The samples were chosen using a stratified, random design that included both genders and all ethnicities (except Native American) represented in our study as well as longitudinal distribution for within and between day, week, and month variability. The sample set is comprised of a range of 5 to 17

samples/subject (average 9.8/subject). Table 1 shows the breakdown of subject ethnicities and numbers of samples analyzed.

Table 1. Number of subjects and urine sample by ethnicity from random stratified procedure

	Black	White	Hispanic	Asian	no response		totals
Male Subj. #	2	5	2	1	1		11
Samples #	17	52	16	10	9		104
Female Subj. #	7	4	3	1	3		18
Samples #	67	31	27	16	35		176

The ethnicities were self-reported; "black" and "white" were defined as "non-hispanic". Some subjects declined to provide ethnicity information. The subject subsets reflect the general breakdown of the total volunteer population.

KIP analyses

The analytical platform was a SECTOR® Imager 2400 MULTI-SPOT assay (Meso Scale Discovery, MSD, Gaithersburg, MD). We used four MSD "7-plex" KIP panels with 96-well plates to investigate seven proposed urinary biomarkers of kidney injury: alpha glutathione s-transferase (αGST), calbindin, clusterin, kidney injury molecule-1 (KIM-1), osteoactivin, trefoil factor 3 (TFF3), and vascular endothelial growth factor (VEGF). Of the 96 wells on each plate, 16 were dedicated to calibration, on average 70 for real world samples, and 10 for duplicate analyses. Table 2 shows the health parameters associated with these biomarkers, "healthy" concentration ranges, and median values according to the manufacturer, MSD. Most markers rise with increasing damage; however, TFF3 and VEGF are inversely correlated with kidney injury. The literature is not consistent in the interpretation of specific effect location and function of the different kidney function biomarkers, so we present this table as a general overview and further note that some of the information is derived from animal studies. We note that we expect all biomarkers to be present at some level in even the healthiest of people; it is the change from a norm that is of probative value and so the tiniest detectable amounts are of importance.

Table 2. Description of KIP markers and their expected occurrence from the literature

KIP Marker	Location(s) of Expression	activity/relevance	healthy range*	healthy median*
			pg/ml	pg/ml
aGST	kidney: distal and proximal tubules (Fuchs) primarily in proximal tubular cells (Branten)	kidney injury marker for acute hepatoxicity, detoxification of xenobiotics (Giffen)	0 - 2100	320
Calbindin	kidney: collecting duct and distal tubules (Fuchs)	calcium binding, acute kidney injury (Bolt)	0 - 13000	4500
Clusterin	kidney: distal and proximal tubules (Fuchs)	ubiquitous tissue distribution, membrane lipid recycling apoptosis (Jones), marker for histopathological damage in rat (Lock)	0 - 200000	24000
KIM-1	kidney: proximal tubules (Fuchs), apical membrane of dilated tubules (Lock)	found in blood and urine, predictive value for acute renal injury (Lock) marker for renal ischemia-reperfusion injury (Vaidya)	0 - 2200	310
Osteoactivin	kidney: tubular epithelium (Nakamura) bone synthesis: osteoblast regulation (Sondag)	elevated levels are markers of acute injury, lesion chronicity, polycystic kidney disease (Patel-Chamberlin) protects injured muscle from fibrosis, late osteoclast differentiation (Sheng)	0 - 600	240
TFF3	kidney: renal tubules (Yu)	mucosal protection, related to breast cancer (Ahmed), decreases in response to renal tubular injury (Yu)	0 - 530	<LOQ
VEGF	kidney: distal and proximal tubules (Fuchs, glomerular podocytes and proximal tubular epithelium (Doi)	signaling protein involved in vascular growth, wound healing, repair of ischemic and inflammatory disease (Hoeben) decreased expression indicates damge to glomerular and peritubular capillaries in kidney (Doi)	0 -1400	450

Note: annotated references - Fuchs and Hewitt 2011, Branten et al. 2000, Giffen et al. 2002, Bolt et al. 2005, Lock 2010, Vaidya et al. 2006, Sondag et al. 2014, Patel-Chamberlin et al. 2011, Sheng et al. 2008, Yu et al. 2010, Ahmed et al. 2012, Doi et al. 2010, Hoeben et al 2004.

*parameters provided by manufacturer's literature: Kidney Injury Panel 3 (human) Kit, www.mesoscale.com.

Data analysis – instrument performance

The creatinine concentration measurements were performed according to protocols developed and validated for the Ex-R study. Specifically, we used a moderate throughput modified kinetic Jaffe method that was tested against commercial laboratory analyses and confirmed using liquid chromatography–time of flight mass spectrometry (LC-ToF-MS) (Andersen et al. 2014). We performed standard summary statistical calculations to evaluate sensitivity, calibration stability, and analytical precision.

The KIP analyses were similarly evaluated. We had previously established detailed evaluations of MSD multiplex well plates and instrument sensitivity for a 10-plex cytokine panel that showed that between-plate variance is unremarkable compared to total variance (Stiegel et al. 2014). We repeated a set of analyses to demonstrate similar results for the KIP analyses and calculated the analogous performance statistics.

Data analysis – KIP and creatinine summary statistics

The first step was to determine a suite of summary statistics including mean, standard deviation, median, coefficient of variation, range, 95th percentile, geometric mean (GM) and geometric standard deviation (GSD) to explore the general distribution of the creatinine and KIP measurements. We also calculated "fold range" parameters to describe how many times greater the upper 97.5% was than the lower 2.5% value. We further explored the underlying distributions using graphical tools and comparison to theoretical functions according to procedures developed previously (Pleil et al., 2014b). This information is critical for assessing creatinine concentrations and KIP biomarkers in future studies as the summary statistics can be used as a baseline for comparison.

In multivariate analysis, covariance structure is an important factor for modeling existing data, and also for designing future experiments. We explored the co-variance among the 7-KIP biomarkers using a variable clustering analysis technique (proc VARCLUS, SAS, Cary, NC) that had been applied previously for pattern recognition in complex environmental and breath biomarkers data (Pleil et al. 2007, Pleil et al. 2011a). This information is critical for choosing individual KIP markers for streamlining future investigations and to find potential common mechanisms in kidney function.

Data analysis – visualization and regression

We used heat-mapping as a data visualization tool to detect broad (qualitative) patterns and factors among various data and meta-data. Heat maps provide a unique perspective of measurement data in that they show each individual measurement as part of a pattern. Although initially developed for displaying genetics data, heat maps are now adapted for understanding multivariate and repeat measurement patterns in environmental and biomarker data structures (Pleil et al. 2011b,c). Herein, we ordered the samples by urinary excretion rate (high to low) concentration and then stratified by gender. The ethnicity host factor was only investigated qualitatively as there were insufficient numbers of subjects per group for a more rigorous analysis.

We assessed the variability of KIP biomarkers with respect to creatinine and urinary output. Although we expected some within subject auto-correlation, we chose to treat the KIP markers as independent measures with respect to the creatinine level. We used raw data uncorrected for volume and/or excretion rate as the dependent variables to observe distinct relationships with hydration based dilution. We subsequently investigated the more obvious relationships using simple scatterplots and regressions and explored variable clustering within data strata. Heatmaps were created using MatLab software (MathWorks, Natick, MA, USA); summary statistics were calculated using Excel spreadsheet software (Microsoft Corporation, Redmond, WA, USA); regression statistics and graphs were created with GraphPad Prism software (Graphpad Software Inc., San Diego, CA, USA); and variable clustering, uni-variate distributions, and multi-variate regression calculations were performed using SAS software (SAS, Cary, NC).

Data analysis – individual level

Analytical results were inspected at the individual subject level via the repeat (longitudinal) data; recall that each subject had, on average, about 10 samples collected. From these data, we calculated within subject variance components, intra-class correlation coefficients (ICC), and developed a series of "box and whiskers" graphs for example compounds to illustrate variability. These analyses were further parsed by gender, but other meta-data (i.e., ethnicity) were too sparse for detailed within subject calculations. Linear regressions were performed in GraphPad Prism, and multivariate regressions were performed using SAS.

Data modeling – investigating urinary output

The final step in data interpretation was to model creatinine measurements with and without KIP marker data to assess the efficacy of predicting total urine flow in the absence of actual measurements. This was intended only as a demonstration, not as a full model for externally derived sampling results. Herein, various linear and multi-linear models were calibrated and tested to compare expected and actual mg/min urine collections per sample. All multivariate regressions were performed using SAS Proc Mixed.

Results and Discussion

Analytical performance

The MSD analytical platform was tested for KIP sensitivity (level of quantitation (LoQ)), precision using calibration (spiked) synthetic and real-world samples, and calibration curve range, linearity, and coefficient of determination (r^2). Summary statistics for KIP performance are given in Table 3 along with a comparison of creatinine analyses. The most important feature is that the calibration curve stability, as reflected by r^2, is better than 0.99 for all analytes. LoQ and precision will be interpreted later in the discussion of real-world data; the values in Table 2 for LoQ were calculated from calibration data as 3xσ above mean blank value and precision is calculated as the % standard error of the mean (SEM) based on all available replicate analyses across all concentration levels. The KIP compounds are grouped by their covariance clusters (discussed later); we note that creatinine performance data are given in the conventional units of mg/dl.

Table 3. Summary statistics for KIP and creatinine instrumentation

Compound	LoQ	Precision	Calib high	Calib
	(pg/ml)	(%) SEM	(pg/ml)	r^2
Clusterin	72.13	1.15	18400	0.9993
KIM-1	0.592	0.39	18400	0.9949
Osteoactivin	14.85	3.30	37000	0.9969
Calbindin	16.94	1.73	22900	0.9995
VEGF	1.564	4.98	2380	0.9973
αGST	2.196	7.92	1720	0.9984
TFF3	4.176	3.18	1890	0.9901
Creatinine	2.29*	1.15	400*	0.9999

* mg/dl

Data visualization

Figures 1 a,b,c show heat maps of the speciated KIP and creatinine results based on different representations of spot measures. The horizontal axis represents the 280 samples, regardless of subject, based on urinary output (ordered left to right from highest to lowest urinary excretion rate). The vertical axis is ordered by compound name in alphabetical order, except for creatinine, which is placed on the bottom for contrast. These three maps are arranged to explore different patterns in the data.

In Fig 1a, the color scheme of the heat map fields represents the concentration in pg/ml for KIP measures and is indicated in log format on the extreme right axis. Creatinine measurements were added in the lowest row as a comparison with units of mg/dl. These were arbitrary initial choices designed to explore the overall character of the data. From this data visualization (Fig 1a), distinctive quantitative differences are expressed among compounds (blue is low, red is high), and a trend of concentration from left to right indicates that Calbindin, and Osteoactivin are likely correlated with Clusterin, and that VEGF, Kim-1, TFF3, and α-GST appear less dependent on this order. Overall, all (but α-GST) look to have about 3-orders of magnitude dynamic range.

In Figure 1b, the concentrations (pg/ml for KIP and mg/dl for creatinine) have been normalized to the 95% fold range within analyte and expressed as the percentile within that range. The color scheme (left to right) is now proportional to this percentile and allows pattern comparisons among compounds independent of absolute values of concentration. In each row, the lowest measurements are assigned to dark blue and the highest to dark red. In this data visualization (Fig 1b), the patterns indicate that α-GST and TFF3 concentrations are essentially independent of urinary excretion rate, that creatinine is highly (inversely) correlated with urinary excretion rate as expected, and that the remaining KIP markers anti-correlate to some lesser extent with hydration.

In Figure 1c, the excretion rate data (pg/min for KIP and mg/min for creatinine) have been normalized to the 95% fold range within analyte and expressed as the percentile within that range. The color scheme is the same as for Figure 1b. In this visualization, the expected outcome is a randomized pattern within rows, which would suggest that mass/time excretion of the biomarkers is independent of the amount of urine volume. However, this hypothesis is not fully realized, in fact only creatinine, KIM-1 and Clusterin appear independent of urinary output, whereas α-GST and TFF3 exhibit a strong correlation with urine output, and the remaining KIP markers Calbindin, VEGF and Osteoactivin show a lesser, but noticeable correlation. This suggests that some of the biomarkers are expressed in urine via diffusion in addition to expected glomerular filtration and active tubular secretion.

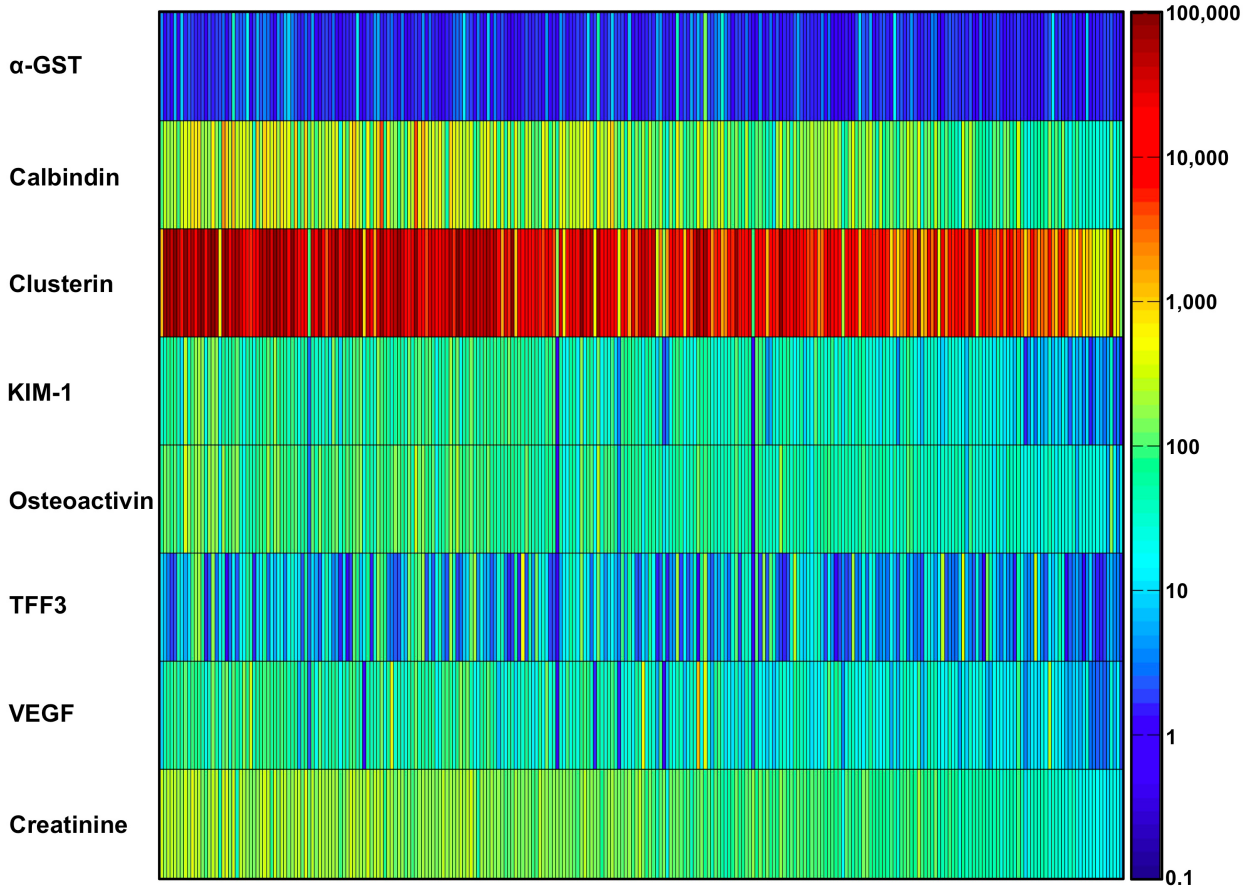


Figure 1a. Heat map of concentration data, KIP biomarkers are color coded in units of pg/ml and creatinine is color coded in units of mg/dl. All samples are ordered left to right by urinary excretion, from highest (most dilute) to lowest (most concentrated).

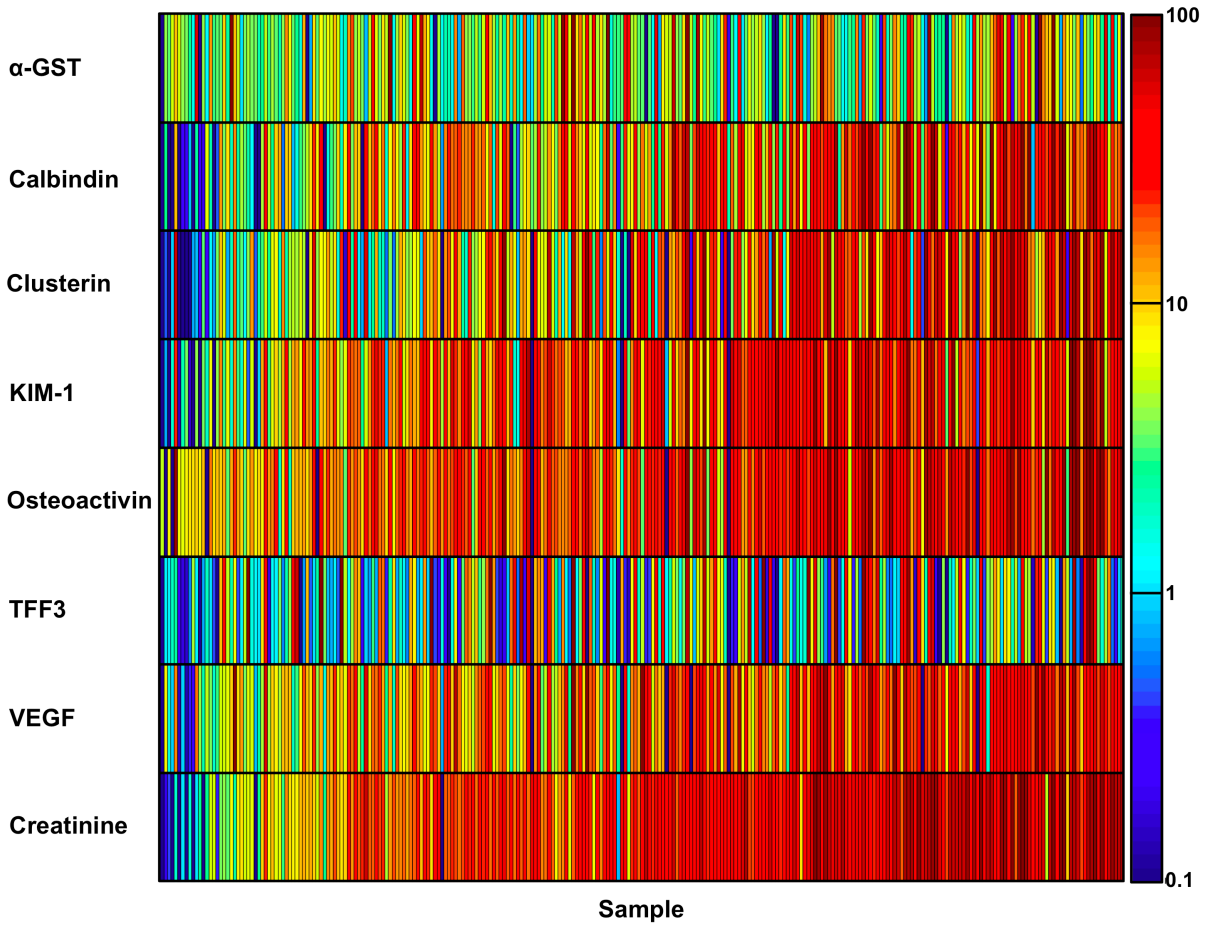


Figure 1b. Heat map of concentration space data normalized within each compound's range. All samples are ordered left to right by urinary excretion, from highest (most dilute) to lowest (most concentrated).

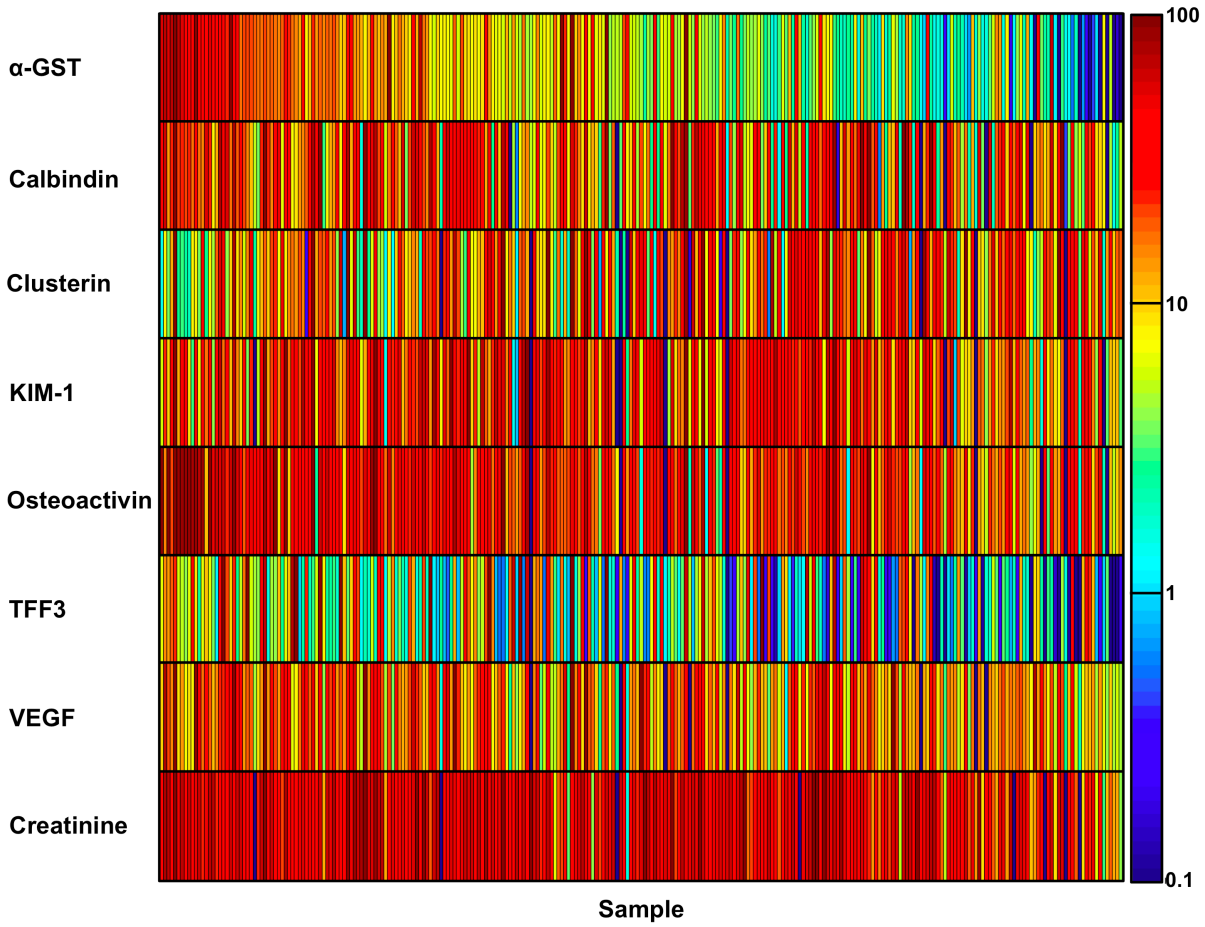
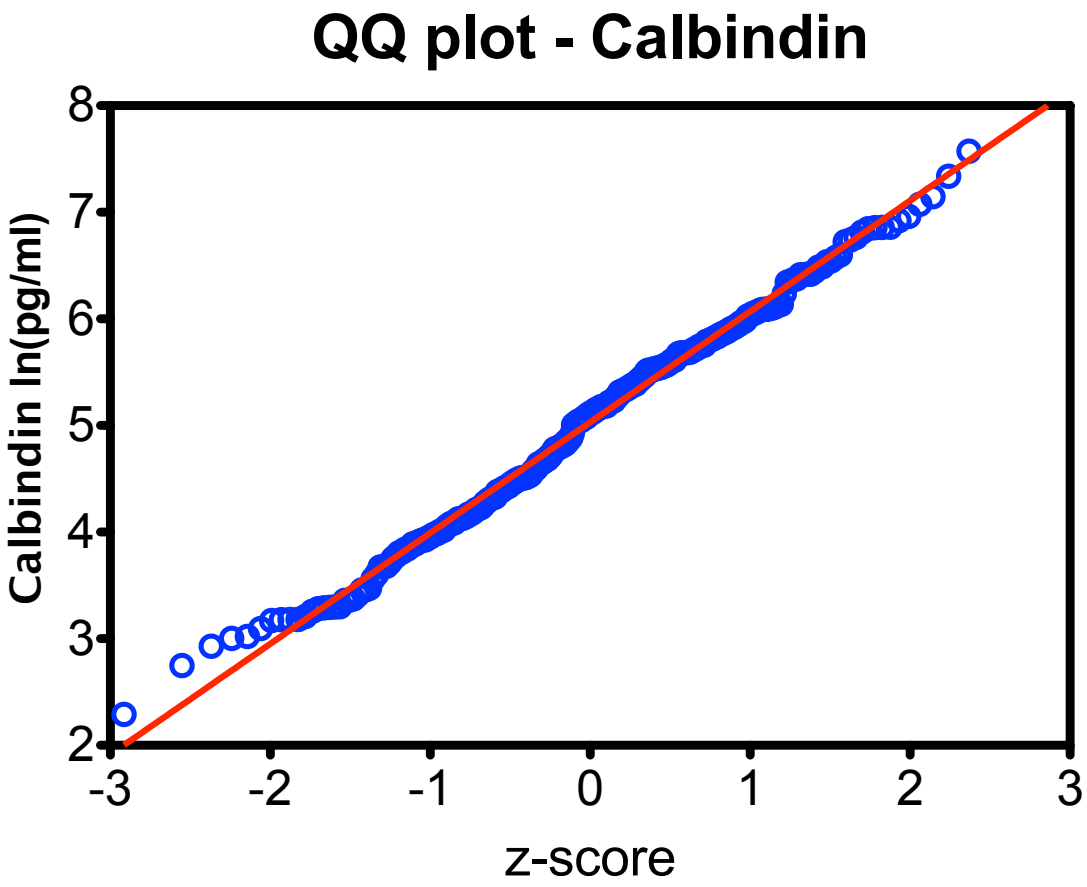


Figure 1c. Heat map of excretion space data normalized within each compound's range. All samples are ordered left to right by urinary excretion, from highest (most dilute) to lowest (most concentrated). Data is expected to be independent of dilution factor.

`

KIP variables and underlying distributions

For most measurements in the environmental and biological sciences (including biomarker measurements) the underlying distributions are expected to be lognormal (Limpert et al. 2001, Halloy and Whigham 2005). We expect our own measurements to fall into this category because it is possible to have occasional high numbers, but no values below zero. To confirm this supposition, we analyzed the log-transformed data using the SAS "proc univariate" algorithm and visually inspected the resulting bar graph distributions and quantile-quantile (Q-Q) plots (Henderson 2006); for this exercise, we treated each measurement as independent although they were grouped by subject. If the points lie along a straight line for normality, then the (log-transformed) data are considered normal. We note that the much more stringent Shapiro-Wilk (SW) tests were ambiguous due presumably to measurement variability and some outlier points, but not due to the basic nature of the data. Figures 2a,b illustrate this effect with graphs for Calbindin which passes SW and for Osteoactivin which does not. The basic observable difference between the Q-Q plots is reflected at the low end, which is likely an artifact of "left-censoring" or variability at the ultra-trace level. The remaining points are almost exactly on the normality line and so reflect the general normal trend.



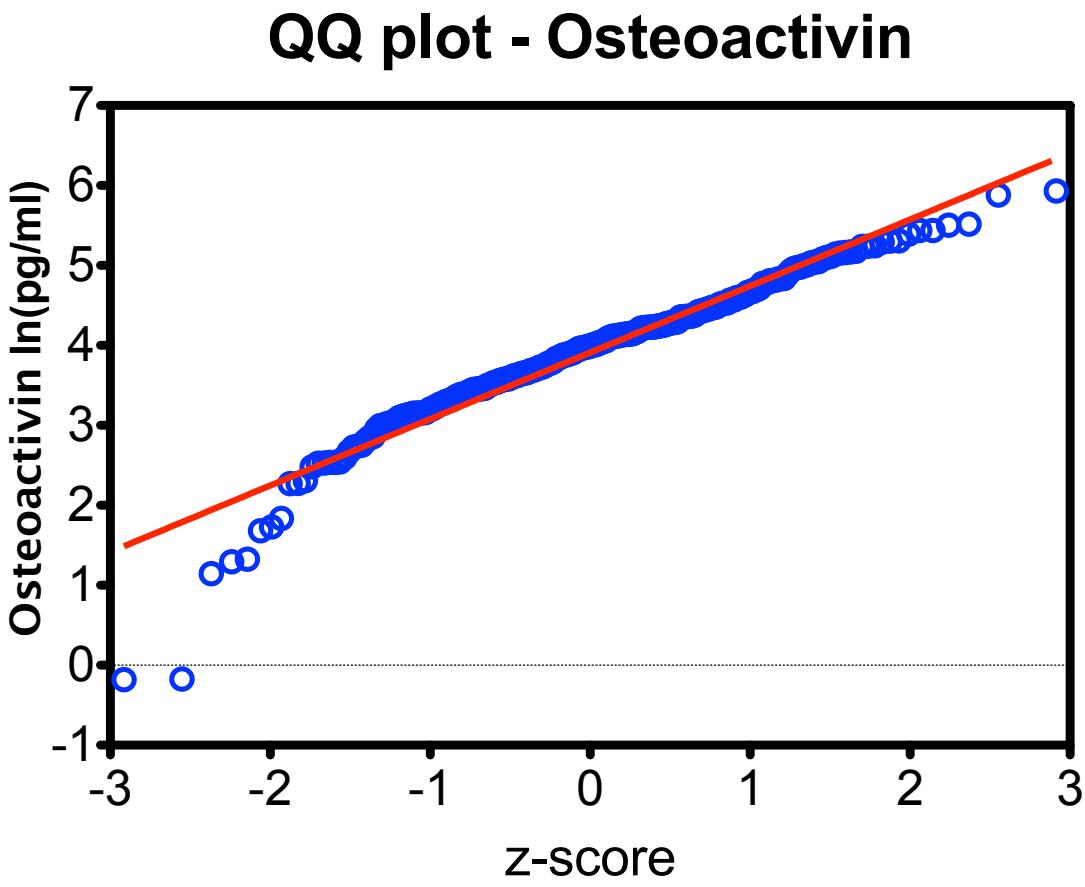


Fig. 2 a,b. Comparison of QQ plots for a) Calbindin and b) Osteoactivin confirming (observationally) the hypothesis that they are lognormally distributed. We note that the main difference is a series of eight outlier points (low end tail) in the Osteoactivin plot. Logged Calbindin data pass the SW test for normality at $\alpha = 0.05$ whereas Osteoactivin data do not pass the test.

Overall, the KIP and creatinine real-world data exhibit similar lognormal behavior. Table 4 shows a summary of the QQ plot statistics for context. Here we present the SW test statistic, whether or not the data passed the normality test, the standard deviation of the residuals (sy.x), the sy.x in pg/ml space, and the r^2 of the QQ plot regression. The results are essentially indistinguishable from each other (with the exception of αGST, which exhibits wider variance), yet only Calbindin passes the strict SW test. Furthermore, back in real-world space, the deviations from normality for all compounds as reflected by sy.x expressed in pg/ml or mg/dl for KIP and creatinine respectively, are small compared to the data range. Based on these observations, we opted to use log-transformed KIP data to perform all subsequent statistical analyses.

Table 4. Normality tests and QQ plot parameters of log-transformed data

Compound	SW test	SW test	QQ reg.	QQ reg.	QQ reg.
	Stat	pass	sy.x	sy.x (pg/ml)	r^2
Clusterin	0.9652	no	0.2978	1.347	0.9660
KIM-1	0.9463	no	0.2495	1.283	0.9470
Osteoactivin	0.9429	no	0.2075	1.231	0.9415
Calbindin	0.9924	yes	0.0917	1.096	0.9923
VEGF	0.9788	no	0.1608	1.174	0.9783
αGST	0.7585	no	0.3481	1.416	0.7570
TFF3	0.9494	no	0.3337	1.396	0.9509
Creatinine	0.9625	no	0.1492	1.161*	0.9634

* mg/dl

KIP variables - analytical results

In Table 5, we show the summary statistics for all KIP, creatinine, and excretion rate measurements including detection limit based completeness. We found that between-plate and within plate instrumentation variability for KIP measurements is negligible in contrast to the overall dynamic range from human samples (recall that precision SEM is in the 0.4% – 8% range). By extending the observations from the heat map and the numerical results in Table 5, we find that 2 of the 7 markers (Clusterin and TFF3) demonstrate a dynamic range of about 3-orders of magnitude and that the remaining biomarkers (KIM1, Osteoactivin, Calbindin, α-GST, and TFF3) have dynamic range of about 2-orders of magnitude in nominally healthy people. For confirmation, we list the calculated 95% confidence fold range (FR_{95}) defined as the 97.5 percentile divided by the 2.5 percentile (Pleil 2009). We note that the statistics for completeness of data (% > LoQ) are quite low for α-GST and TFF3 at 28% and 62%, respectively. This is a bit misleading because we could actually calculate defensible values below the nominal LoQ using 5-parameter logistic (5pl) models to interpret low-end signal values and achieve complete data structures for statistical analysis (Richards 1959, Stiegel et al. 2014).

The GM and GSD are calculated directly from measurements and serve to parameterize the data distribution (Pleil et al. 2014b). As an example, Figure 3 shows the distribution bar graph and the corresponding probability density function (pdf) for Calbindin; this is a direct analog for the QQ plot shown in Figure 2a. These figures can be made relatively easily to investigate the lognormality of data sets once the GM and GSD are available. We further point out that the GM for excretion rate herein is 0.964 ml/min, which falls between the nominal default values of 0.903 (women) and 1.001 (men) ml/min found in the literature (Boeniger et al. 1993).

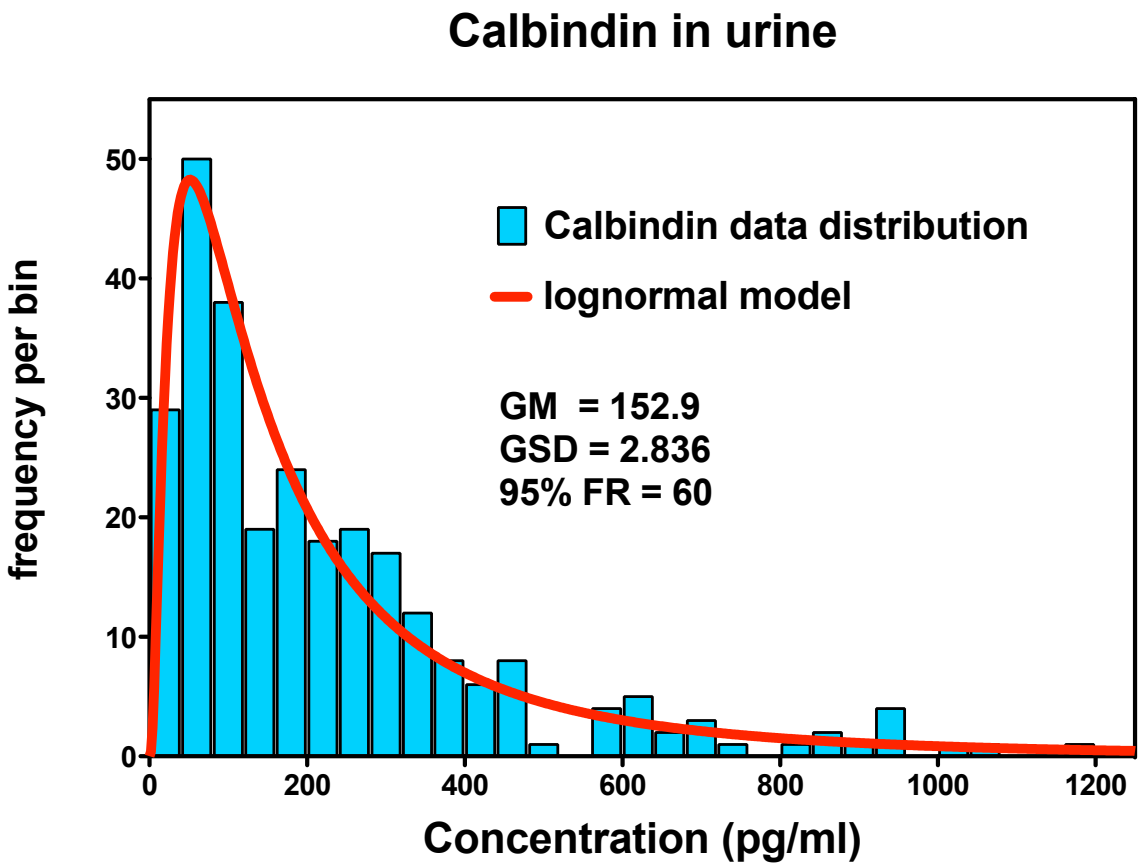


Figure 3. Calbindin example: frequency distribution and reconstructed probability density function (pdf) based on the calculated GM and GSD.

Table 5. Summary statistics for KIP, creatinine, and excretion rate of 280 urine samples (KIP: pg/ml, creatinine: mg/dl, excretion rate: ml/min)

Compound	GM	GSD	2.5%	97.5%	95% FR	median	cv	% >LoQ*	% >LoQ**
Clusterin	9942	5.007	295	97705	332	12647	1.38	100	100
KIM-1	31.41	2.949	2.42	163	67	38.7	0.87	100	100
Osteoactivin	49.94	2.355	6.25	202	32	54.2	0.81	93	100
Calbindin	152.9	2.836	24.0	1015	42	166	1.64	99	100
VEGF	27.99	2.976	2.23	174	78	31.9	1.84	99	100
αGST	2.044	2.024	1.03	13.7	13	1.63	2.64	28	100
TFF3	9.895	4.492	1.23	220	178	7.66	1.94	62	100
Creatinine	101.4	2.178	17.6	336	19	119	0.651	100	na
Excretion rate	0.9641	2.181	0.247	6.03	24	0.896	1.083	100	na

* % > LoQ refers to the nominal instrument levels of quantitation as listed in Table 3;
** %>LOQ refers to the results from the extrapolation using the 5pl model.

KIP variables and sample correlations

One of the major problems in complex data interpretation and parameter estimation is "over-modeling". This occurs when the number of measurement variables (m) approaches the number of samples (n). For successful model building efforts, n should be greater than or equal to 10 times the number of independent variables ($n \geq 10 \times m$); on rare occasion, well-behaved data sets, wherein the underlying values are known to be perfectly normally distributed, can be modeled with $n \geq 5 \times m$ (Harrell 2002, Pleil and Lorber 2007, Pleil et al. 2011). A second important problem occurs when two or more variables actually co-vary to the extent that they provide little, if any, independent information; their inclusion in even simple multivariate regression models can result in mathematical instability (Chadeua-Hyam et al. 2013, Lee et al. 2013, Phillips et al. 2008). As the heat map indicated, some of the KIP compounds are likely correlated; the question is "by how much?"

We have implemented a technique referred to as variable clustering to assess how variables could be grouped to avoid over-correlation. As reasoned above, we used log-transformed KIP data. This approach invokes Eigen-vector projections that are calculated for all variables in n^{th} dimensional space. We used statistical software proc VARCLUS from SAS to develop a "dendrite" diagram that can be used to create clusters of variables that have certain levels of correlation (http://www2.sas.com/proceedings/sugi26/p261-26.pdf). Figure 4 shows the diagram for the 7 KIP variables wherein we see that there are two distinct covariance clusters 1) Clusterin, KIM1, and Osteoactivin, and 2) Calbindin and VEGF. α-GST and TFF3 are uncorrelated biomarkers in this analysis. This is of particular interest in that it may help design future experiments in reducing analytical effort as a single marker may represent the behavior of the cluster.

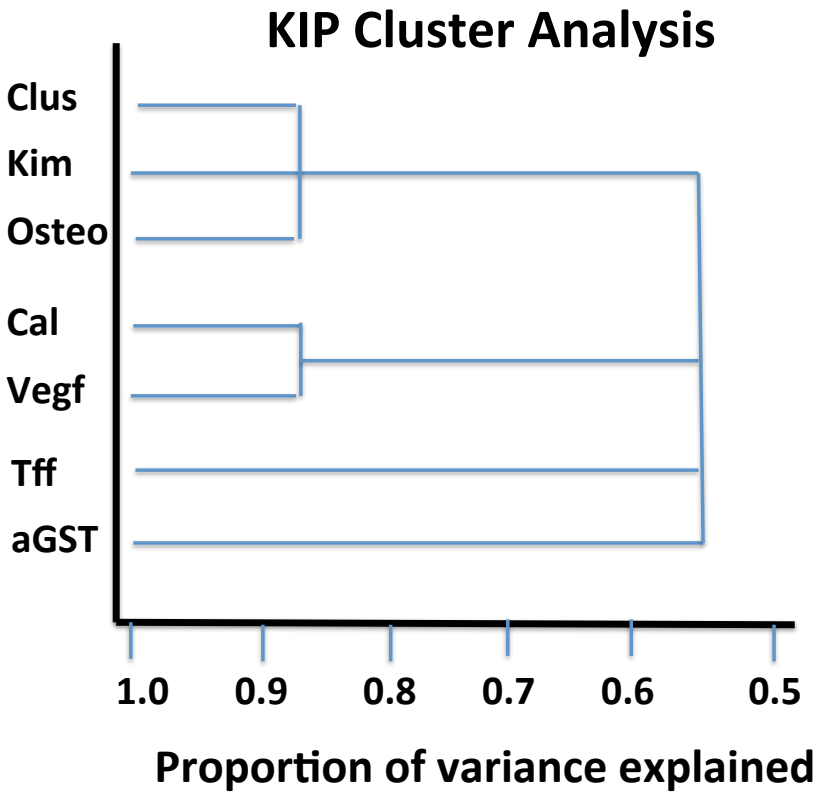


Figure 4. Dendrite diagram showing results of variable cluster analysis calculated with log-transformed data. There are two primary clusters with ~0.87 common variance, and all KIP biomarkers are correlated to some extent at ~0.55 proportion of explained variance.

Such cluster analysis is relatively new in biomarker science. For comparison, we present the more typical correlation matrix approach that implements paired correlations, rather than more global "clustered" groupings. Table 6 shows the standard correlation results:

Table 6. Standard pairwise correlation table organized by variable clusters.

	Clusterin	KIM1	Osteoactivin	Calbindin	VEGF	TFF3	aGST	Creatinine
Clusterin	1.000	0.699	0.751	0.523	0.737	0.159	0.072	0.572
KIM1		1.000	0.789	0.613	0.630	0.285	0.187	0.615
Osteoactivin			1.000	0.573	0.650	0.180	0.246	0.507
Calbindin				1.000	0.687	0.323	0.305	0.661
VEGF					1.000	0.175	0.274	0.557
TFF3						1.000	0.149	0.213
aGST							1.000	0.134
Creatinine								1.000

Cluster 1
Cluster 2
Cluster 3

Excretion rate and KIP parameter correlations

The excretion rate of individual KIP compounds is likely dependent on a number of different factors including total urine elimination volume (hydration), kidney function efficiency (health), and other host factors including gender, age, body mass index (BMI) and ethnicity. The present study structure does not have sufficient power to statistically evaluate all of the host factors, however, we could explore the effect of gender as we had 11 male and 18 female subjects represented.

Our initial "common sense" hypothesis was that kidney function/health is more likely stable within persons and more variable between persons. We tested this by calculating intra-class correlation coefficients (ICC) for each KIP constituent (and creatinine) for all raw data. In this context, if ICC >0.5, then the conjecture is likely true, as then the greater portion of the total variance is between persons. A second hypothesis was that individual urinary excretion rate could have sufficient impact on the ICC analysis of raw data, that is, some individuals may chronically excrete more or less than the mean value, and so we reran the analyses using excretion corrected data in units of pg/min. Table 7 shows the ICC results of these analyses.

Table 7. Intra-class correlation coefficients (ICC) for KIP biomarkers and creatinine for concentration and excretion rate based data.

Compound	ICC	ICC	ICC Change
	concentration data	excretion data	exc/conc
Clusterin	0.222	0.132	0.595
KIM-1	0.419	0.431	1.029
Osteoactivin	0.173	0.149	0.861
Calbindin	0.434	0.168	0.387
VEGF	0.249	0.138	0.554
αGST	0.0452	0.0690	1.527
TFF3	0.198	0.176	0.889
Creatinine	0.396	0.162	0.409

Contrary to initial conjectures, all of the calculated ICC results are less than 0.5 indicating that variability is more pronounced within individuals. Upon comparison between concentration (pg/ml) and excretion rate (pg/min) results as shown in the last column, most KIP markers (except KIM-1 and α-GST), have a further reduction in ICC when excretion rate data is used, that is, when between person variance decreases. When compared to the known effects from creatinine behavior with respect to excretion rate, this indicates that these markers are also eliminated proportionally to hydration state. Of the outliers, α-GST has extremely low ICC's (below 0.1) so regardless of its relationship with hydration, this marker does not appear to be affected by longitudinal or between person variability. The ICC's for KIM-1 are both over 0.4 indicating that this compound's behavior is much less affected by excretion rate than creatinine, and so may be considered as more independent of hydration and other overall external factors influencing variability.

Creatinine levels and KIP parameter correlations

Under standard procedures, urinary creatinine is used as a co-variate for interpreting dilution status of any spot urine sample when the true value of urine excretion rate "U_r" (ml/min) is not known (Barr et al. 2005, Muscat et al. 2011). Furthermore, it is generally assumed that urinary metabolite concentrations are inversely proportional to excretion rate in any given sample, and so creatinine concentration can be used as a surrogate for dilution. Figure 5 shows a linear regression of creatinine concentration *versus* urinary excretion rate (in log-log space) with r^2 = 0.6858; this suggests that more than 30% of the creatinine variance is attributable to sources beyond dilution such as kidney function. As such, efforts to correct for dilution using the creatinine stability hypothesis are not necessarily accurate.

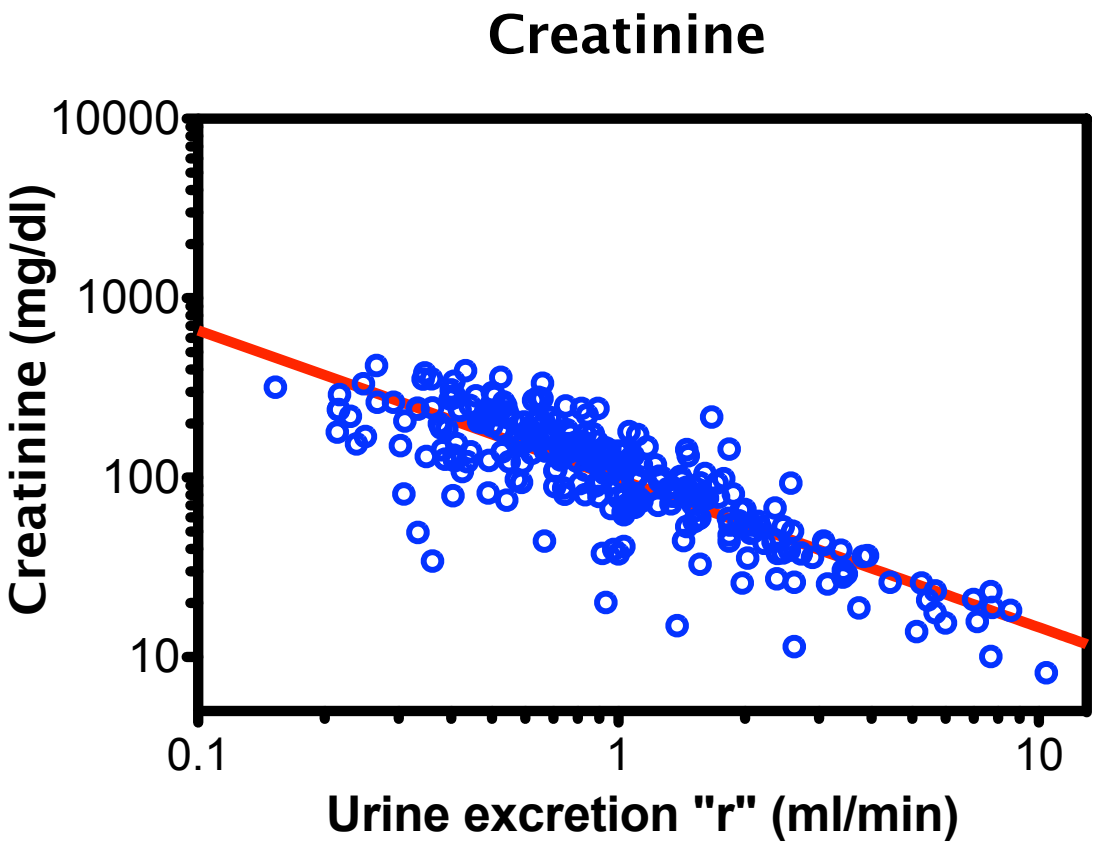


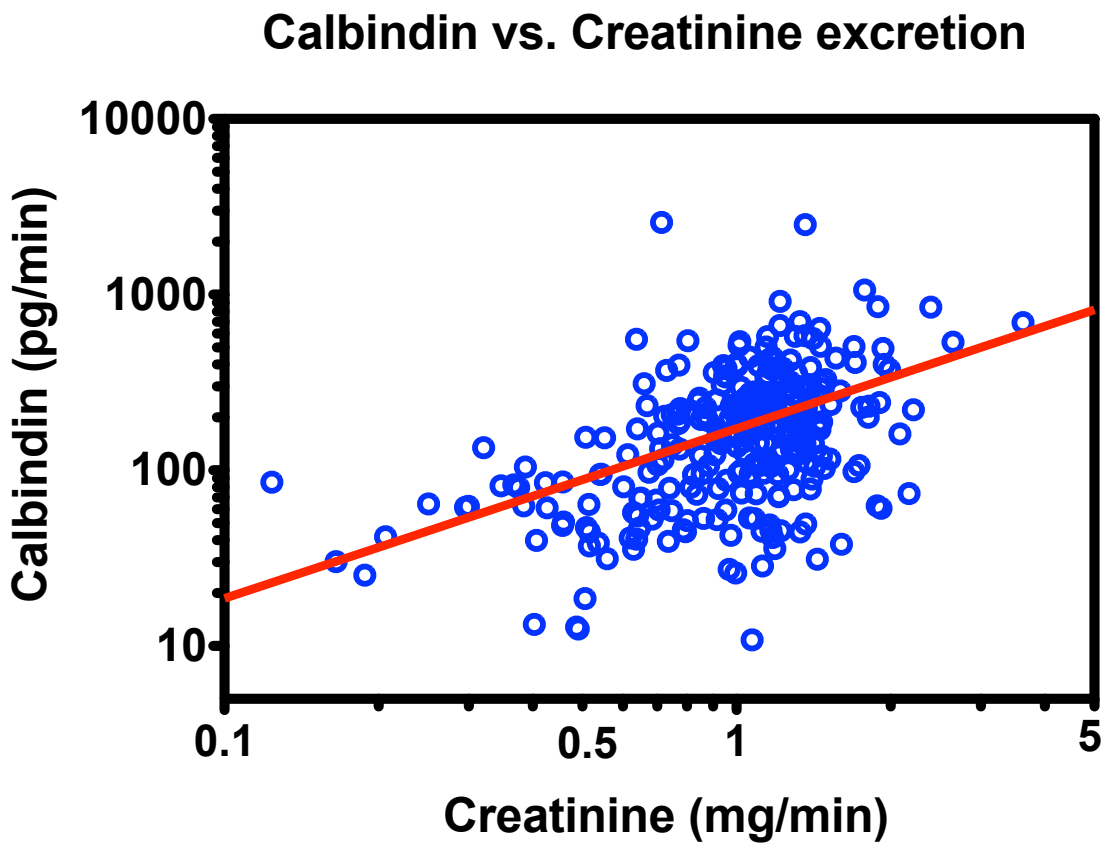
Figure 5. Creatinine level is inversely related to urinary excretion rate with $r^2 = 0.6858$ suggesting that more than 30% of creatinine variance is from sources beyond dilution/hydration.

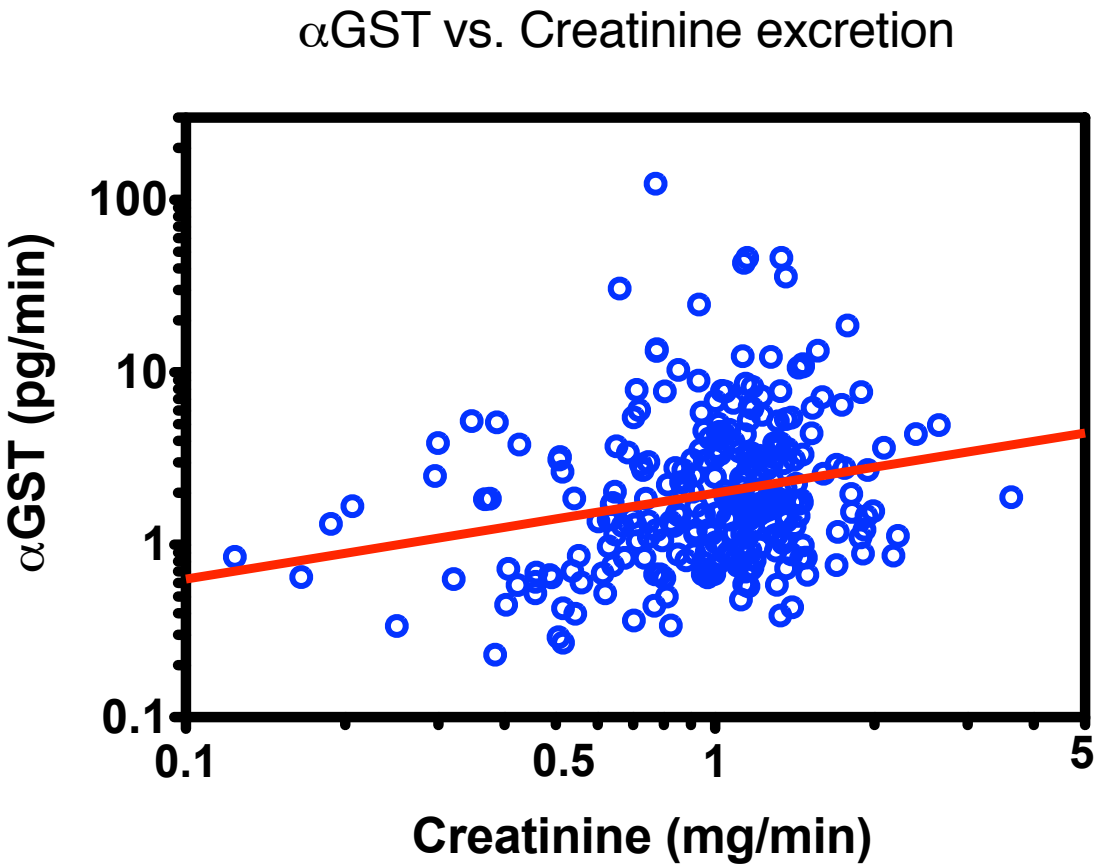
The next step is to assess how much creatinine level variance might be explained with the KIP parameters. In Table 8 we show the relationships between the different KIP biomarkers and creatinine using log-log regression in excretion rate space, that is, we compare KIP and creatinine data in units of mass/min thus removing the effect of dilution. Here, the slope parameter β represents the % change in the KIP value for every 1% change in creatinine; the r^2 parameter is an estimate of explained variance. We see that there is a certain amount of positive correlation between KIP and creatinine excretion rates independent of urinary dilution. This suggests that absolute creatinine concentration measurements could be modified to provide better dilution normalization estimates for spot-urine measures when urine volume rates are not available. Figures 6 a,b show examples for Calbindin and α-GST that represent the extremes of the relationships with creatinine excretion (r^2 values are 0.2170 and 0.0533, respectively). The general trends are positive suggesting that creatinine excretion is affected by some of the same biological mechanisms that increase KIP biomarker excretion. We cannot read too much into this observation with respect to health state, because the subjects were initially selected to be free of diagnosed disease of all kinds (including kidney). However, these results indicate that a moderate amount of variance in creatinine excretion could be explained by changes in KIP biomarkers independent of urinary output suggesting that there are common underlying biological processes involved.

Table 8. Excretion rate normalized data regressions of KIP biomarkers (pg/min) *versus* creatinine (mg/min) performed in log-log space

Compound	Correlation	Slope	S.D. of residuals	S.D. of residuals
	r^2	B_1	sy.x	sy.x (pg/ml)
Clusterin	0.0560	0.7051	1.324	3.758
KIM-1	0.1094	0.6489	0.847	2.333
Osteoactivin	0.0971	0.5675	0.792	2.208
Calbindin	0.2170	0.9067	0.788	2.199
VEGF	0.1380	0.8032	0.919	2.506
αGST	0.0533	0.4987	0.961	2.615
TFF3	0.0466	1.5360	1.536	4.646

With the present data, one can consider defining general correction equations that could help explain variability in creatinine and in excretion rates of other metabolic species. This is discussed in a later section.


Calbindin vs. Creatinine excretion
10000
1000
100
10
Calbindin (pg/min)
0.1
0.5
1
5
Creatinine (mg/min)



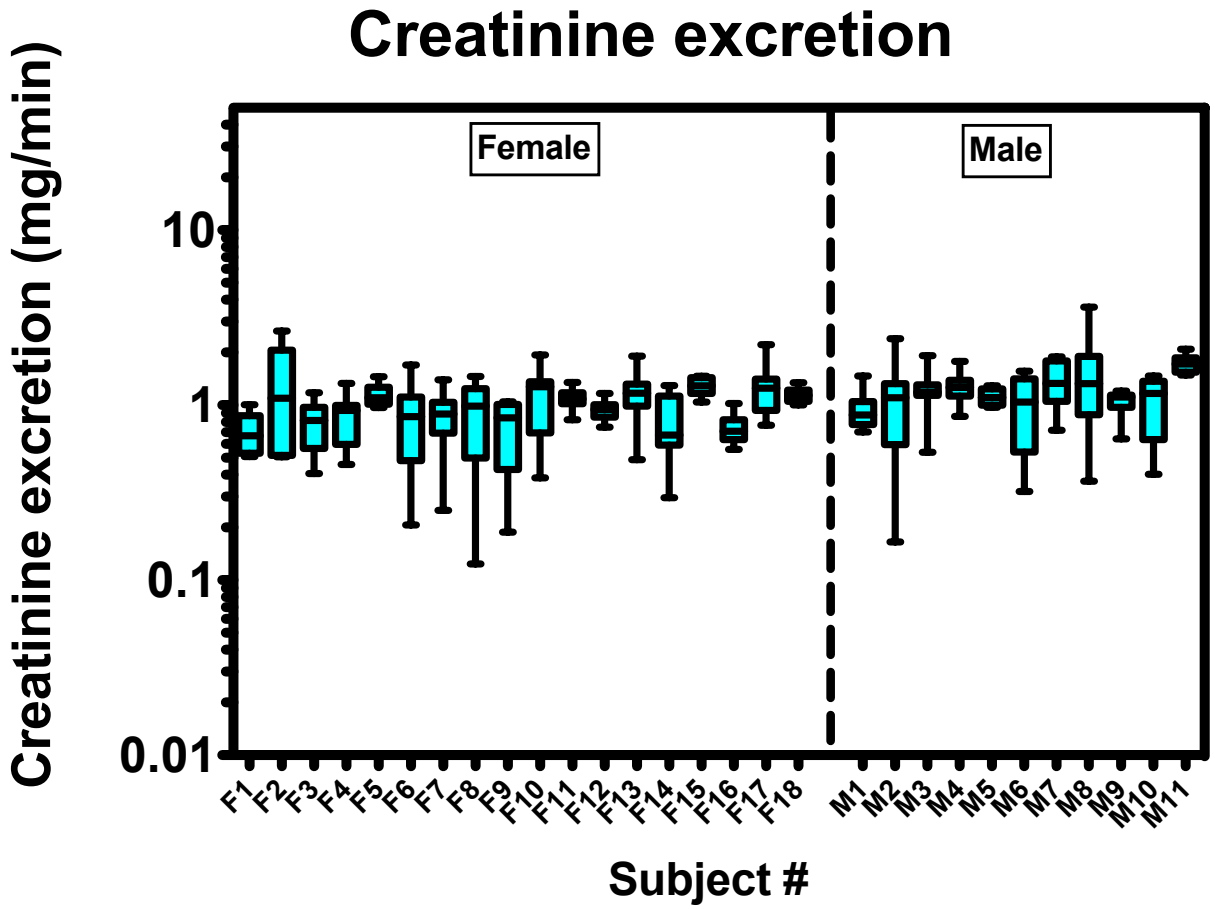
Figures 6 a,b. Relationship of excretion rates of KIP biomarkers with respect to Creatinine showing a positive correlation independent of urinary dilution. a) shows Calbindin with strongest correlation ($r^2 = 0.2170$ and slope factor $\beta = 0.9067$); b) shows αGST with weakest correlation ($r^2 = 0.0533$ and slope factor $\beta = 0.4984$).

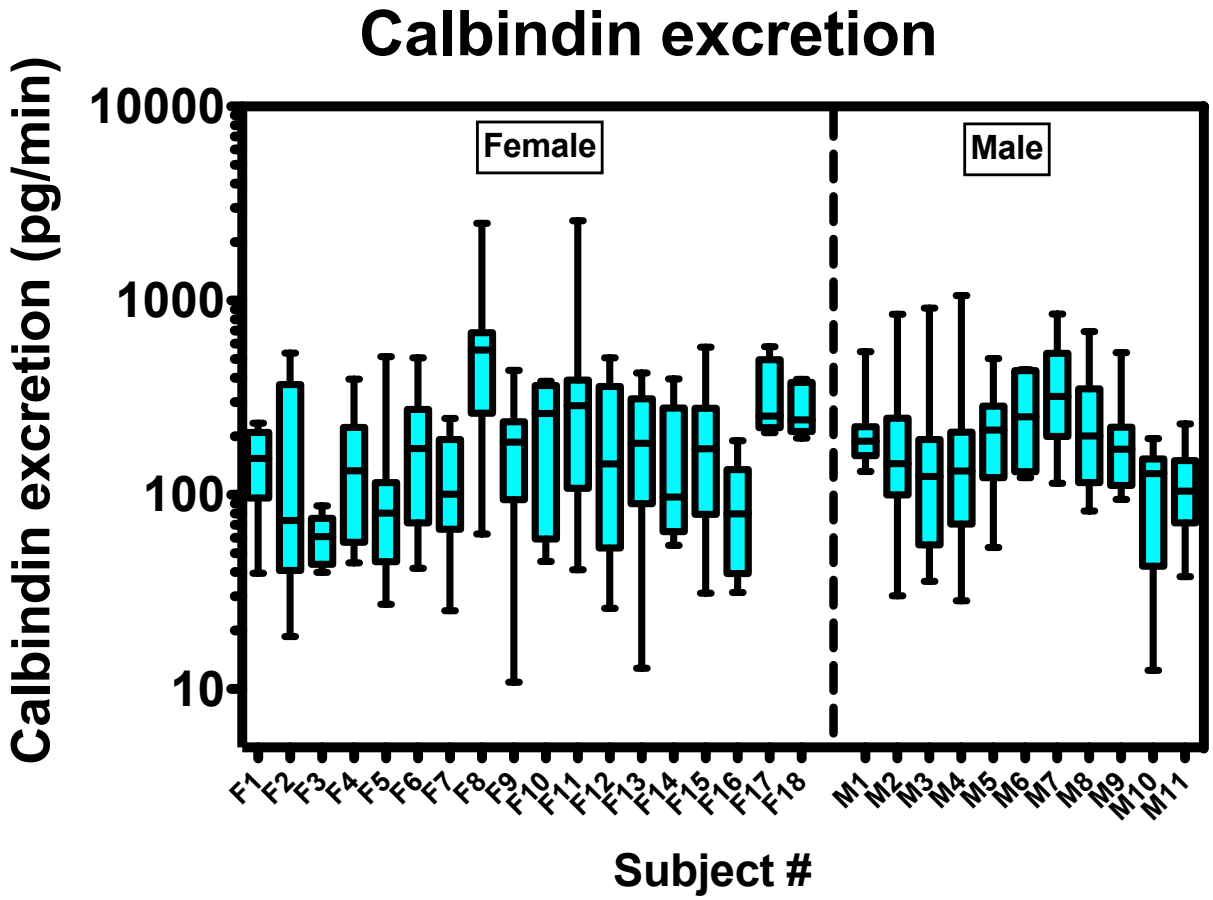
Effects at the individual level

With the exception of a brief discussion of ICC's above, to this point, all measurements have been treated as independent to assess broad behaviors. However, under the hypothesis that sub-clinical kidney injury effects may present more consistently within subjects than between subjects, we would expect to find a mean difference among subjects for KIP and creatinine excretion. We have recalculated the broad data based on within-subject statistics to investigate such phenomena by assessing the summary statistics GM, GSD, FR_{95}, etc. for each subject in excretion rate space in pg/min (KIP) and mg/min (creatinine). As examples, Figures 7 a,b,c show the box and whiskers plots for creatinine, Calbindin and KIM-1 excretion for all subjects individually, segregated by gender. There were not enough within gender "n" to assess continuous variables such as age or body mass index (BMI).

All of the ensuing discussions are based on excretion data in units of mass/time; hydration effects of the urinary dilution have been divided out. Figure 7a shows the individual subject excretion for creatinine in units of mg/min and shows that there is little if any trend difference

between genders, and that the within subject variance is relatively small. In contrast, the examples for Calbindin and KIM-1 excretion show much greater within subject variability and the gender and ICC relationships show the extremes in behavior for the KIP biomarkers. Specifically, creatinine excretion rate is relatively stable between subjects with ICC's of 0.158 (all), 0.151 (females) and 0.100 (males), and an overall fold range (FR_{95}) of 6, which is considered an unremarkable for biomarkers in general (Pleil 2009). Calbindin excretion rate exhibits similar ICCs of 0.162 (all), 0.173 (females), and 0.107 (males), but within- and between subject variability are both greater and have $FR_{95} = 33$. KIM-1 excretion rate has much more between subject variability with ICC's of 0.423 (all), 0.509 (males), and 0.306 (females), but has the similar overall FR_{95} of 34.


Creatinine excretion
Creatinine excretion (mg/min)
Female
Male
10
1
0.1
0.01
F1
F2
F3
F4
F5
F6
F7
F8
F9
F10
F11
F12
F13
F14
F15
F16
F17
F18
M1
M2
M3
M4
M5
M6
M7
M8
M9
M10
M11
Subject #


Calbindin excretion
Calbindin excretion (pg/min)
10000
1000
100
10
Female
Male
F1
F2
F3
F4
F5
F6
F7
F8
F9
F10
F11
F12
F13
F14
F15
F16
F17
F18
M1
M2
M3
M4
M5
M6
M7
M8
M9
M10
M11
Subject #

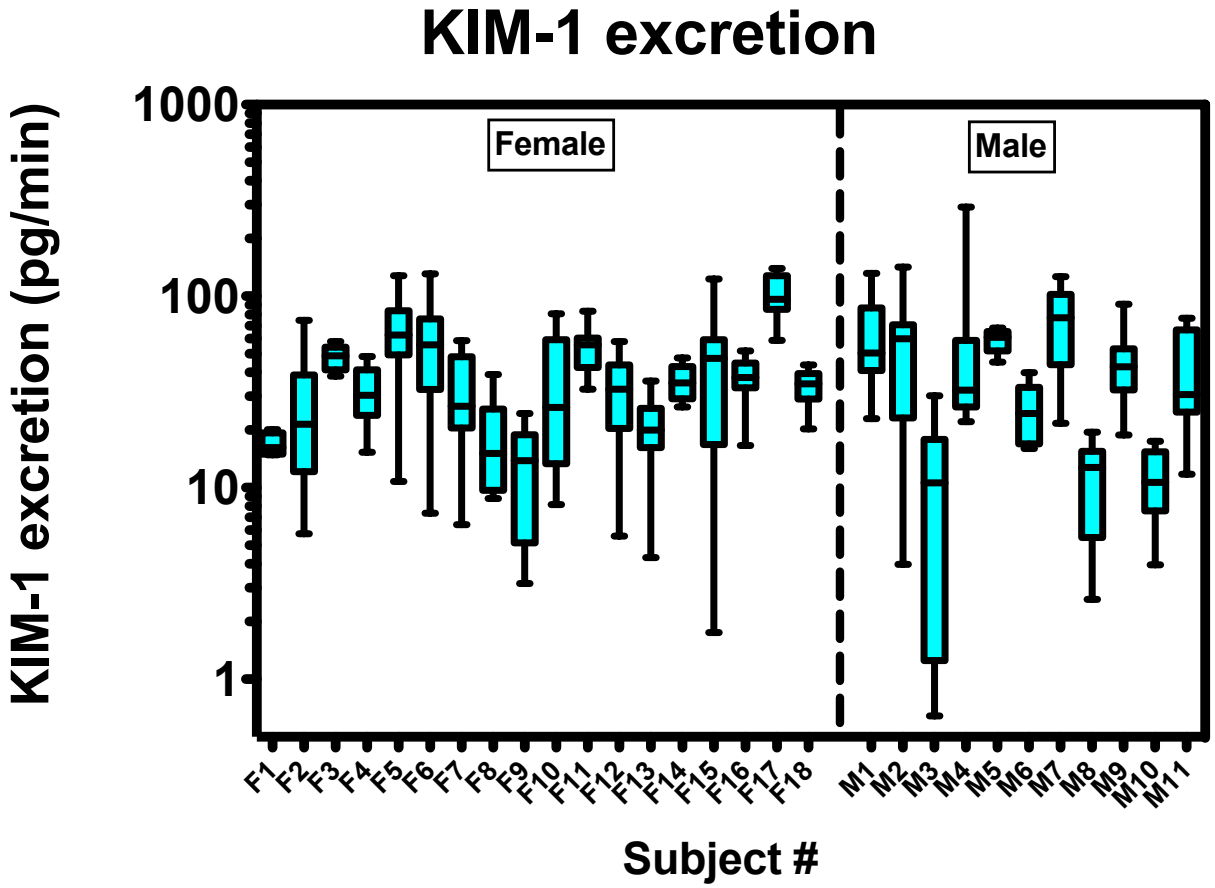


Figure 7a,b,c. Examples of (a) creatinine, (b) Calbindin, and (c) KIM-1 excretion variability among individuals with gender grouping. The y-axis of all three graphs have 4.5 orders of magnitude range (log-scale) for visual consistency. Creatinine excretion rate is very stable between subjects with ICC's of 0.158 (all), 0.151 (females) and 0.100 (males), and an overall fold range (FR_{95}) of 6. Calbindin excretion rate exhibits similar ICCs of 0.162 (all), 0.173 (females), and 0.107 (males), but within- and between subject variability are both greater and have FR_{95} = 33. KIM-1 excretion rate has much more between subject variability with ICC's of 0.423 (all), 0.509 (males), and 0.306 (females), but has the similar overall FR_{95} of 34.

Refinement of excretion rate estimates

Actual measurement of urinary excretion rate U_r (ml/min) is a burden in most studies because it requires having subjects record the elapsed time since the most recent urination and collecting the total urine volume for that sample. As such, many studies work with urine aliquots as "spot measures" and infer the metabolic excretion rate using a default parameter for urine output, typically set at 1 ml/min. As seen in the data from Table 5, this is a poor estimate for any given individual sample as there is a wide range of "normal" urinary output ranging from about 0.2 to 10 ml/min. As discussed above, creatinine concentration can be used to correct for dilution effects under the assumption that it is excreted at a stable rate. Often, external data is used to calibrate urine output (U_r) with measured creatinine data, and then this correction is applied to all

data. The simplest generic form of urine output prediction uses measurements of U_r and creatinine data to calculate a regression parameter (α_1) with equ. 1:

$$\ln(U_r(s_{i,j})) = \alpha_1 \ln(Cr(s_{i,j})) + \varepsilon_{i,j} \qquad \text{(equ.1)}$$

where $U_r(s_{i,j})$ = urinary excretion rate (ml/min) for the j^{th} sample of the i^{th} subject, $Cr(s_{i,j})$ is the measured Creatinine concentration in mg/dl, α_1 is the empirically fitted regression parameter and and the $\varepsilon_{i,j}$ term encapsulates unexplained variability. This equation can then be used to estimate U_r in the absence of empirical data. Figure 8 shows the result of such an approach where the creatinine data have been regressed against the empirical U_r data, and then presented as a scatter plot of measured vs. modeled points. The diagonal line represents perfect agreement. In this example, modeled and measured U_r correlate with $r^2 = 0.685$, which is a respectable coefficient of variation.

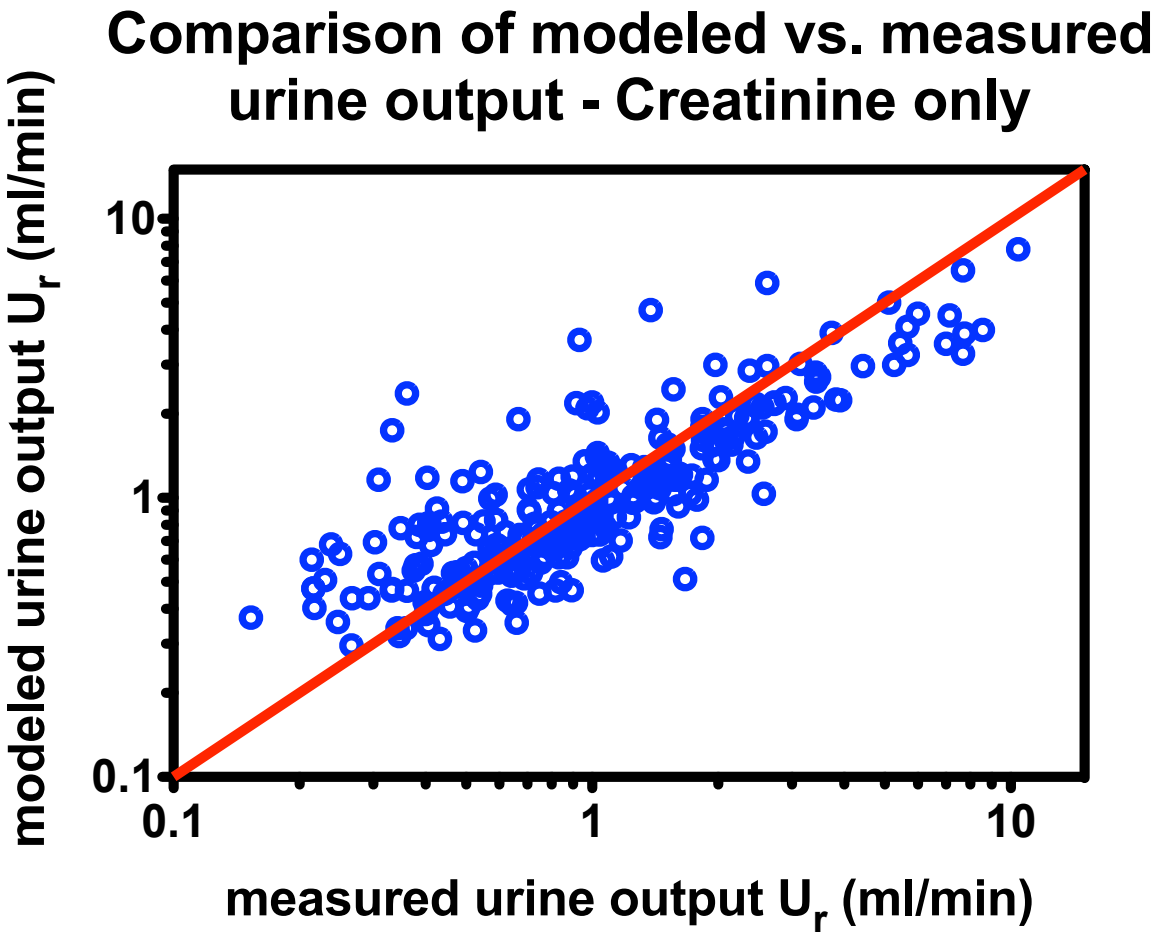


Figure 8. Scatter plot of measured U_r data vs. U_r calculated from creatinine regression model. The model explains about 67% of the total variance. Straight line is "perfect" correlation.

Given additional information about kidney function, a revised estimate can be made using KIP markers. Specifically, the new model included KIM-1 to represent the first cluster, Calbindin the

second cluster, TFF the third cluster and αGST as fourth cluster (see Fig. 4) to represent other kidney function parameters, and a gender parameter, Gen(j) for the jth subject; the model was constructed as an augmented multivariate correction factor equation for urinary excretion rate U_r of the form in equ. 2:

$$\ln(U_r(s_{i,j})) = \alpha_1 \ln(Cr(s_{i,j})) + \alpha_2 \ln(Kim(s_{i,j})) + \alpha_3 \ln(Cal(s_{i,j})) + \alpha_4 \ln(TFF(s_{i,j})) + \alpha_5 \ln(\alpha GST(s_{i,j})) + Gen(j) + b_i + \varepsilon_{i,j} \quad \text{(equ. 2)}$$

where $U_r(s_{i,j})$ = urinary excretion rate (ml/min) for the j^{th} sample of the i^{th} subject, $Cr(s_{i,j})$ is the measured Creatinine concentration in mg/dl, $Kim(s_{i,j})$, $Cal(s_{i,j})$, $TFF(s_{i,j})$, and $\alpha GST(s_{i,j})$ are the measured KIM-1, Calbindin, TFF, and αGST data in pg/ml, and Gen indicates subject gender; the α's are the empirically fitted regression parameters and the b_i and $\varepsilon_{i,j}$ terms encapsulate unexplained variability. We note that the r^2 values calculated from mixed regression models are estimated based on differences from the "null" model that represents the data mean.

Investigation of individual parameters

The mixed model results were unexpected and engendered further investigation. The full model delineated in equ. 2 was evaluated for explanatory power using individual and combinations of independent variables. Table 9 shows the r^2 results for different versions of equ 2.

Table 9. Contribution to model explanatory power using different combinations of independent variables

Model	r^2
Full: creatinine, KIP, gender	**0.7021**
creatinine only	**0.6846**
KIP only	**0.3639**
gender only	**0.0427**
creatinine and gender	**0.7018**
creatinine and KIP	**0.6848**
gender and KIP	**0.3714**

Based on this evaluation, creatinine alone is the dominant explanatory parameter and explains about 68% of the overall differences in urinary excretion volume. Of interest is that the KIP parameters alone explain only about 36% of the total variability indicating that their excretion is driven only partly by dilution (about half as much as creatinine) and so there are other factors involved. Finally, gender difference alone only explains 4% of the total variance, but when used

in addition to creatinine, improves the r^2 more (68.457% vs. 70.177%) than alternatively adding the KIP parameters. When both creatinine and KIP are in the model, there is essentially no change in total explanatory value over the creatinine only model (68.457% vs. 68.478%). This apparent disparity indicates that KIP parameters are much more closely co-varying with creatinine than is the gender parameter, or perhaps they (KIP) are more independent of true urinary output than is creatinine. In either case, adding KIP parameters to any of the models that already include creatinine provides no appreciable improvement.

The question remains: What is the mechanism for changes in urinary excretion of the KIP biomarkers beyond hydration/dilution factors? We still have not accounted for 64% of the variability. With the current information, the only remaining avenue for investigation is at the individual sample level.

Investigation of individual samples

Figure 9 shows the full model analog (equ. 2) for the creatinine only model (equ. 1) that was presented in Figure 8. Upon inspection of both of the prediction graphs for urinary output U_r (as shown in Figures 8 and 9), it appears that there is some curvature to the data. This is especially noticeable at the lower end of the graph below measured urinary output values of 0.5 ml/min. In Figure 9, the straight (red) line indicates the expected perfect correlation (slope = 1, intercept = 0); this full model reconstruction explains about 60.9% of the variance from those expected values. The curved (black) line indicates the best fit 2nd order polynomial regression, which explains about 72.9% of the variance, but serves only to illustrate the effect of the curvature. This regression shows that in general, the model conforms reasonably to the expected "perfect correlation" for most of the range of measured urinary output but becomes less accurate at the lower end.

There are some potential explanations for this behavior:

- Competing excretion mechanism (such as direct diffusion) that becomes important at low U_r (highly concentrated urine).
- An underlying pre-clinical condition (undiagnosed health state) that changes the relationship among excreted metabolites and hydration state.

As a first rudimentary attempt to deduce a mechanism, the data points that were at or below measured U_r of 0.5 ml/min ($U_r$lo, n = 56) were segregated from the remaining data ($U_r$hi, n = 224), and each set was independently evaluated for correlations with U_r and each other.

<u>*Relationships with respect to dilution*</u>: The two sets of samples demonstrate very different behaviors; individual KIP and creatinine regressions against U_r for the $U_r$lo data set show no significance in slope, whereas all (but αGST) show a significant negative slope (as expected) for the $U_r$hi dataset. The observation drawn from this result is that, in general, great care must be taken in interpreting urinary biomarkers in samples below urinary 0.5 ml/min urinary excretion rate as they no longer track with dilution. Under the assumption that more concentrated samples have longer residence time in the body, one could interpret this result to support the idea that

diffusion plays an increasingly greater relative role in metabolite excretion when there is overall less fluid to process per time.

Relationships among individual biomarkers: The two data sets were tested individually with respect to paired correlation (analogous to Table 6 results for the complete set). There is a distinct difference in character in these correlations; for example, Calbindin is much more correlated (by a factor of 2) with Clusterin, KIM1 and Osteoactivin in the $U_r$hi subset, whereas Calbindin is much less correlated (<50%) with TFF3, VEGF in $U_r$hi with respect to the $U_r$lo data set. Although they are all still positively correlated with each other, there are specific differences between the variability exhibited as a function of urine volume per time. This observation indicates that the excretion of metabolites is somehow tied to sub-clinical effects occurring within the kidney and that whatever the mechanism is, that these individual differences are important with respect to the overall gold standard of creatinine.

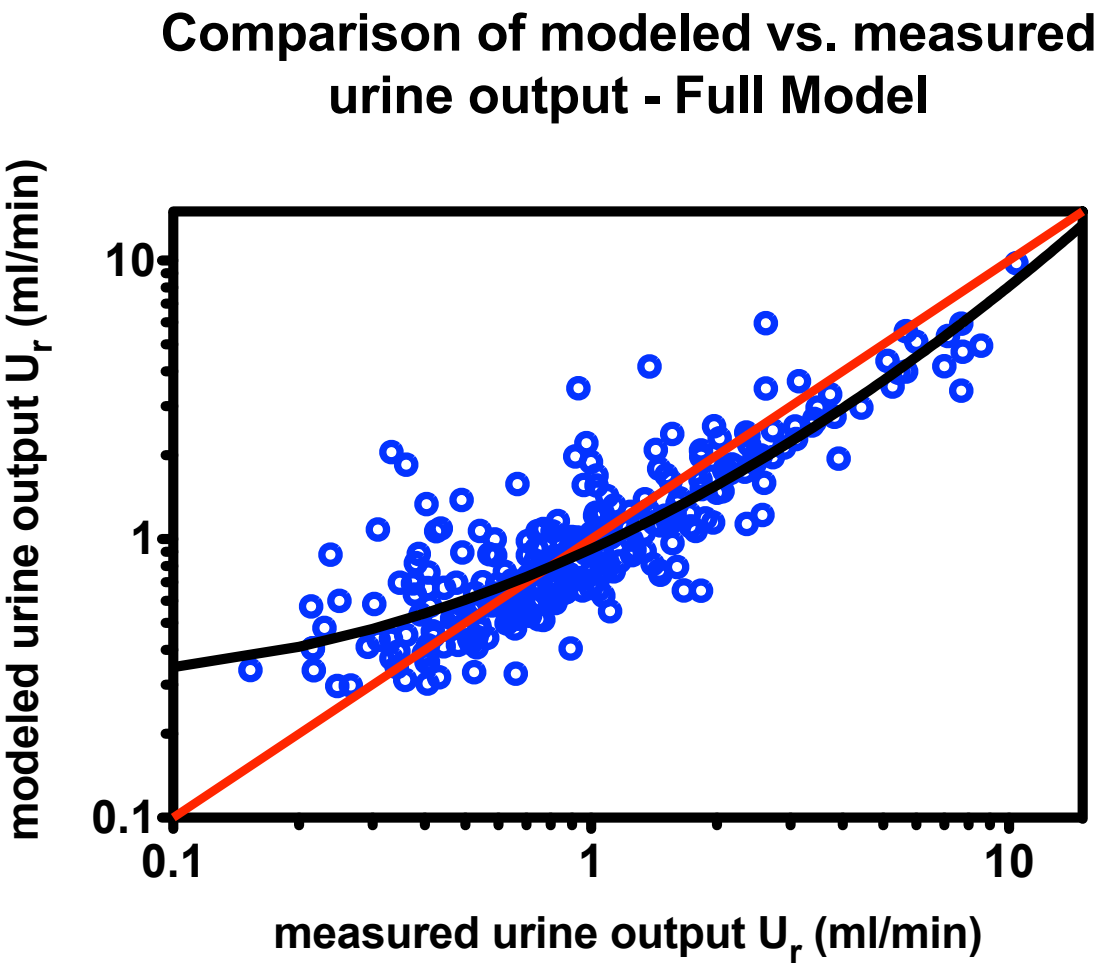


Figure 9. Scatter plot of measured U_r vs. U_r calculated from the "full model" that includes creatinine, KIP biomarkers (Cal, KIM-1, TFF3 and αGST) and gender as regression variables.

The straight (red) line indicates the expected perfect correlation; the full model reconstruction explains about 60.9% of the variance. The curved (black) line indicates the best 2nd order polynomial regression, which explains about 72.9% of the variance.

Summary: The measurement of individual KIP biomarkers in conjunction with urinary output and creatinine concentration indicates a need for caution for investigating the excretion rates of exogenous biomarkers. The differences found in individual responses and in correlations associated with urine excretion rate cannot be further parsed here as this is limited data set and is focused only on nominally healthy adults. Certainly, a more inclusive measurement set is required if kidney health effects on urine excretion rates were to be pursued. However, if observation or inference of urinary excretion rate indicates a sample with 0.5 ml/min or less, then dilution correction is suspect.

Using KIP biomarkers to flag outlier samples

As creatinine level is the most commonly used parameter for assessing dilution of a urine aliquot, the question arises as to how the KIP biomarkers could be used to flag individual samples where this correction may be inaccurate. Under the hypothesis that unremarkable "healthy range" levels of KIP biomarkers track with urinary dilution just like creatinine, then conceivably, when the ratio of KIP/creatinine is beyond some statistical limit, then the overall sample integrity for all other excreted compounds may be affected by undiagnosed kidney injury. Table 10 shows the summary statistics of the individual KIP biomarkers divided by their respective creatinine levels composited for all 280 samples, segregated by gender.

Table 10. Summary statistics of KIP/creatinine biomarkers ratios by gender, units: (KIP pg/ml)/(Creatinine mg/dl)

Female	aGST	Calbindin	Clusterin	KIM1	Osteoactivin	TFF3	VEGF
mean	0.0413	2.2296	205.3390	0.4338	0.6542	0.4547	0.3185
median	0.0156	1.6781	138.7525	0.3652	0.5451	0.1092	0.2449
95%	0.1001	5.0281	544.7657	0.9256	1.5137	1.9757	0.7744

Male	aGST	Calbindin	Clusterin	KIM1	Osteoactivin	TFF3	VEGF
mean	0.0385	1.8336	175.9841	0.3832	0.6510	0.1660	0.6799
median	0.0200	1.5393	88.4365	0.2967	0.5199	0.0659	0.3949
95%	0.1094	4.4169	484.9014	0.9541	1.3206	0.6451	1.4632

Each individual sample was interrogated with respect to how many of the individual ratios (7/sample) fell into the high (or low) 95% tail. The expected value is 5% by random chance for any individual KIP marker entry but the actual number of outliers per subject could be different due to clustering of sample effects and individual subject effects.

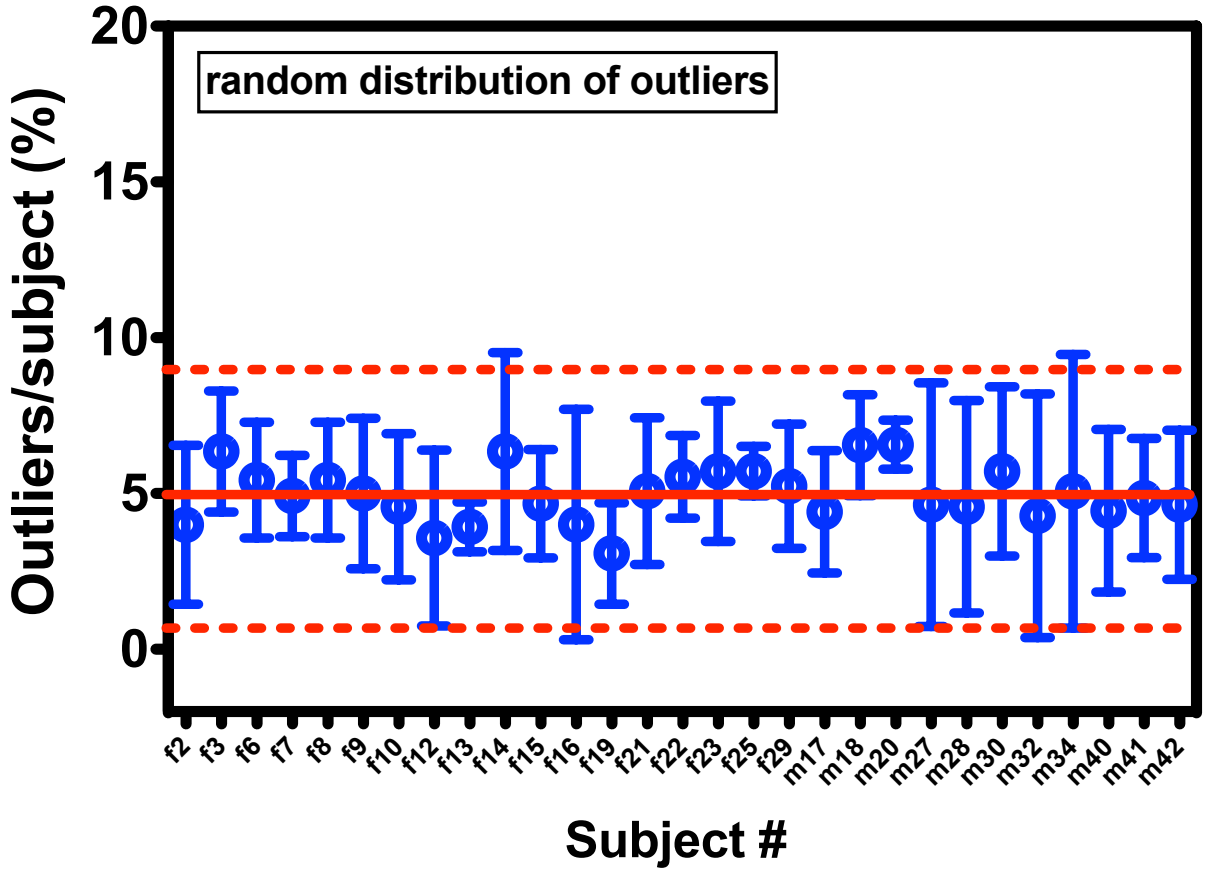
Upon reduction of data, the number of outlier occurrences tended to be clustered within subjects rather than randomly distributed across all samples. On the whole, 4 subjects (2 male, 2 female) exhibited outlier status under the definition of 2-times the expected value for the high level (>

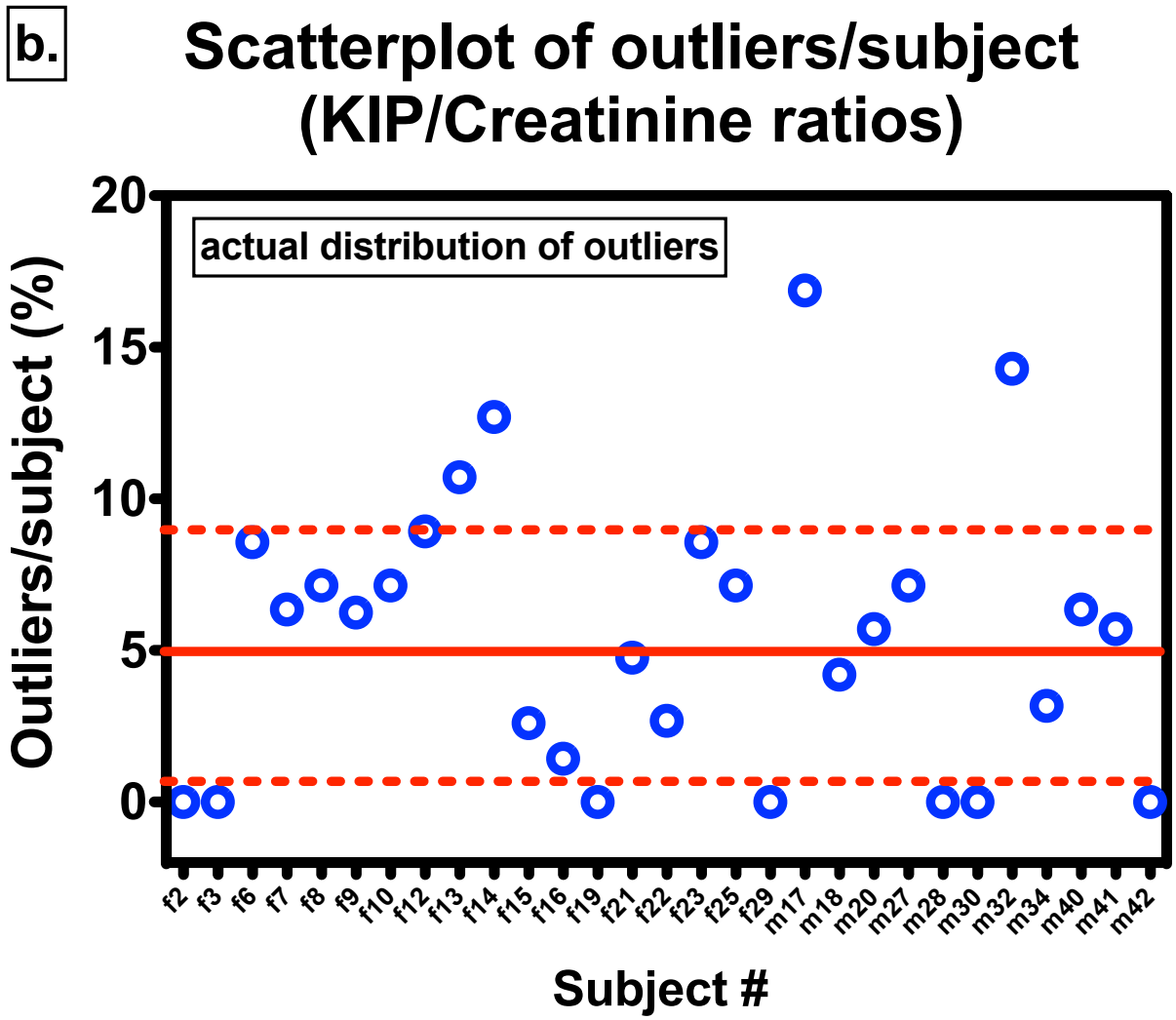
95%) and 3-subjects (1 male, 2 female) exhibited outlier status for the low level (<5%) corresponding to p <0.02. There was no overlap in classification, that is, no individual had both high and low outlier clusters. Fig. 10 a presents the results of 5 randomized trials showing the expected results if outliers were randomly distributed and Fig 10 b shows the true distribution of outliers. The solid line in each graphic shows the expected value (5% outlier/subject) from the definition; the dashed lines show the $\alpha = 0.05$ confidence levels. These results show that there are localized patterns in the KIP/creatinine ratios well-beyond random chance, and that the overall between-subject variability is broad (Fig. 10 b) in contrast to the expected random concentration effect around the mean value seen in the randomized data (Fig. 10 a).

Although there is some suggestion in the literature that decreases in TFF3 and VEGF are indicators of some specific injuries (see Table 2), it is unclear if the "low outlier" category is useful overall for flagging samples because higher KIP values generally indicate injury. Nonetheless, both outlier groups indicate that metabolites are not consistently tracking dilution status, and so the individuals exhibiting excessive outlier samples warrant further scrutiny.

a.

Scatterplot of outliers/subject (KIP/Creatinine ratios)





Figures 10 a,b. Scatterplot comparisons of the within-subject clustering of outliers defined as ratios of KIP biomarkers and the respective creatinine concentrations. Solid line indicates the defined 5% outlier value; dotted lines indicate $\alpha = 0.05$ confidence levels. Figure a) shows the outliers if they had occurred at random within subjects; these were drawn from the same distribution with 5 randomized trials. Figure b) shows the actual measurement data showing outlier clustering where 4/29 subjects are well above and 7/29 subjects are below the prediction band.

Conclusions

This investigation into human KIP biomarkers has documented a series of novel results:

- It is the first such investigation (published) on a suite of KIP biomarkers that provides a method useful for studies of nominally healthy human subjects and provides performance statistics of accuracy, dynamic range, and precision.
- It documents for the first time the actual distributions (lognormal) of the KIP biomarkers and provides the "normal range", covariance, and intra-class correlation for nominally healthy persons without diagnosed kidney disease.
- It documents for the first time the interaction of urinary (volumetric) excretion rate, creatinine concentration, and KIP biomarkers in urine samples.
- It demonstrates for the first time a rationale for caution in interpreting aliquots of urine samples using volume correction with creatinine concentration and identification of metabolic outlier samples using individual KIP concentrations.

From these observations, one can draw a number of specific conclusions that will inform other evaluations of urine samples for exogenous and endogenous compounds. This is especially important when actual urinary excretion rates are not available.

- KIP markers are not independent, but are clustered; as such, it would be possible to use a restricted analyte set to represent the whole KIP set.
- Creatinine and KIP measurements must be treated as lognormally distributed when comparing central tendencies and outlier statistics.
- Urinary dilution can be inferred to some extent ($r^2 = 0.68$) using regression analysis based on creatinine concentration, but KIP biomarkers alone explain only ~36% of the variability indicating that these endogenous markers do not track well with dilution, and that there are other, as yet unknown factors involved in nominally healthy subjects.
- Creatinine and KIP biomarkers exhibit unexpectedly low intra-class correlations (ICC) in concentration and in excretion space data indicating that within-subject variability is important, and that multiple analyses per subject are required to assess overall metabolism behavior.
- The summary statistics for creatinine, KIP biomarkers, and excretion rate presented in Table 4 can be used as a baseline for evaluating other urine samples with respect to the expected "normal" ranges in nominally healthy adults.
- The summary statistics for the ratios of KIP/creatinine provided in Table 10 could be used to flag other urine samples for potential "outlier" status, and at least indicate that the relationships among metabolites are suspect in those samples.

In summary, this work provides the first empirical exploration of pre-clinical effects of kidney injury in nominally healthy humans as a potential confounder of urinary dilution relationships to exogenous and endogenous biomarkers. By design, homogenous AOP-based HTS assays lack the intrinsic variability demonstrated here in healthy and "normal" adults. Therefore, we urge caution when translating *in vitro* cell-based and cell-free assay data to the quantitative exposome and disease in human cohorts. Although the statistics are based on a relatively small data set (280 samples, 29 subjects), the general conclusions are realistic. The next step for this work is to

monitor many more urine samples for the interaction of excretion rate, creatinine concentration and KIP biomarker concentration and include a cohort of subjects with medically diagnosed kidney injury to observe a greater dynamic range in KIP response. This would further test the hypothesis that some of the unexplained variability in creatinine and KIP excretion in nominally healthy subjects is driven by undiagnosed (pre-clinical) effects. Finally, the authors encourage that future urinary excretion studies incorporate a repeat measurement strategy that includes time since last void and total void volume for each sample so that the actual excretion rate can be calculated.

Acknowledgements

The authors are grateful to our EPA and contractor colleagues who performed the field, laboratory, and data validation work for the original ExR study; they are Erik Andersen, Rich Walker, Lillian Alston, Fu-lin Chen, and Carry Croghan. We also thank Michael Madden, Andrew Ghio, Stephen Edwards, Myriam Medina-Vera, and Timothy Buckley from US EPA, and Steve Klose and Mikhael Wallowitz from MesoScale Discovery, for valuable insights and discussions. This article was reviewed in accordance with the policy of the National Exposure Research Laboratory, U.S. Environmental Protection Agency, and approved for publication. Approval does not signify that the contents necessarily reflect the view and policies of the Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use. This work was funded in part by EPA Chemical Safety for Sustainability (CSS) Biomonitoring Program. The authors declare they have no competing financial interests.

References

Ahmed ARH, Griffiths AB, Tilby MT, Westley BR, May FEB, (2012). "TFF3 Is a Normal Breast Epithelial Protein and Is Associated with Differentiated Phenotype in Early Breast Cancer but Predisposes to Invasion and Metastasis in Advanced Disease", American Journal of Pathology 180(3):904-916.

Andersen EM, Sobus JR, Strynar MJ, Pleil JD, and Nakayama SF, 2014. "An alternative method for rapid urinary creatinine determination", Journal of Toxicology and Environmental Health A 77(18): 1114-1123.

Angerer J, Ewers U, Wilhelm M, 2007. Human biomonitoring: State of the art. Int J Hyg Environ Health 210(3-4):201-228.

Angrish MM, Madden MC and Pleil JD, 2015. Adverse outcome pathways (AOPs) in human systems biology: *in vivo* discovery for developing *in vitro* target methods, Chemical Research in Toxicology (submitted).

Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, Mount DR, Nichols JW, Russom CL, Schmieder PK, Serrrano JA, Tietge JE, Villeneuve DL, 2010. "Adverse

outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment", Environmental Toxicology and Chemistry 29(3):730-741.

Au WW, 2007. Usefulness of biomarkers in population studies: From exposure to susceptibility and to prediction of cancer. Int J Hyg Environ Health 210:239-246.

Barr DB, Wilder LC, Caudill SP, Gonzalez AJ, Needham LL, Pirkle JL, 2005. "Urinary creatinine concentrations in the U.S. population: implications for urinary biologic monitoring measurements", Environmental Health Perspectives 113(2):192-200.

Bean HD, Pleil JD, and Hill, JE, 2014. "Report: New analytical and statistical approaches for interpreting the relationships among environmental stressors and biomarkers", Biomarkers (in press).

Becker J, Friedman E, 2013. "Renal function status", American Journal of Roentgenology 200(4):827-829.

Boeniger MF, Lowry LK, and Rosenberg J, 1993. Interpretation of urine results used to assess chemical exposure wih emphasis on creatinine adjustments: A review, American Industrial Hygiene Association 54:615-627.

Bolt MJG, Cao LP, Kong J, Citrin MD, and Li YC, 2005. "Vitamin D receptor is required for dietary calcium-induced repression of calbindin-D9k expression in mice", Journal of Nutritional Biochemistry 16(5):286-290.

Bonventre JV, Vaidya VS, Schmouder R, Feig P and Dieterle F, 2010. "Next-generation biomarkers for detecting kidney toxicity", Nature Biotechnology 28(5):436-40.

Branten AJ, Mulder TB, Peters WH, Assman KJ, and Wetzels JF, 2000. "Urinary excretion of glutathione S transferases alpha and pi in patients with proteinuria: reflection of the site of tubular injury", Nephron 85:120-126.

Chadeau-Hyam M, Campanella G, Jombart, T., Bottolo, L., Portengen, L., Vineis, P., Liquet, B. and Vermeulen, R. C.H. (2013), Deciphering the complex: Methodological overview of statistical models to derive OMICS-based biomarkers. Environ. Mol. Mutagen., 54: 542–557.

Choudhury D and Ahmed Z, 2006. "Drug associated renal dysfunction and injury", Nature Clinical Practice Nephrology 2:80-91.

Cohen Hubal, E.A., Richard, A., Aylward, L., Edwards, S., Gallagher, J., Goldsmith, M.R., Isukapalli, S., Tornero-Velez, R., Weber, E., Kavlock, R., 2010. Advancing exposure characterization for chemical evaluation and risk assessment. Journal of toxicology and environmental health Part B, Critical reviews 13, 299-313.

Doi K, Noiri E, and Fujita T, 2010. "Role of vascular endothelial growth factor in kidney disease", Current Vascular Pharmacology 8(1):122-128.

Edwards SW and Preston RJ, 2008. “Systems biology and mode of action based risk assessment” Toxicological Science 106:312-318.

Fuchs TC and Hewitt P, 2011. Biomarkers for drug-induced renal damage and nephrotoxicity – an overview for applied toxicology”, American Association of Pharmaceutical Scientists Journal 13(4):615-631.

Giffen PS, Pick CR, Price MA, Williams A, and York MJ, 2002. “Alpha-Glutathione S-transferase in the assessment of hepatotoxicity—Its diagnostic utility in comparison with other recognized markers in the Wistar Han rat”, Toxicologic Pathology 30(3):365-372.

Halloy S, Whigham P, 2005. The lognormal as universal descriptor of unconstrained complex systems: a unifying theory for complexity, Complexity International 12:1-12.

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

Henderson AR, 2006. “Testing experimental data for univariate normality”, Clinic Chimica Acta 366:112-129.

Hill RH, Head SL, Baker S, Gregg M, Shealy DB, Bailey CC, Williams CC, Sampson EJ, and Needham LL, 1995. Pesticide Residues in Urine of Adults Living in the United States: Reference Range Concentrations, Environmental Research 71(2):99-108.

Hoeben A, Landuyt B, Highley MS, Wildiers H, Van Oosterom AT, and De Bruijn EA, 2004. Vascular Endothelial Growth Factor and Angiogenesis, Pharmacological Reviews 56(4):549-580.

Houck KA, Dix DJ, Judson RS, Kavlock RJ, Yang J, and Berg EL, 2009. Profiling bioactivity of the ToxCast chemical library using BioMAP primary human cell systems. J Biomol Screen 14(9):1054-66.

Hudmon KS, Honn SE, Jiang H, Chamberlain RM, Xiang W, Ferry G, Gosbee W, Hong WK, Spitz MR, 1997. “Identifying and recruiting healthy control subjects from a managed care organization: a methodology for molecular epidemiological case-control studies of cancer”, Cancer Epidemiology, Biomarkers and Prevention 6(8):556-571.

Jones SE and Jomary C, 2002. “Clusterin”, International Journal of Biochemistry and Cell Biology 3r4(5):427-431.

Kavlock R, Dix D, 2010. Computational toxicology as implemented by the U.S. EPA: providing high throughput decision support tools for screening and assessing chemical exposure, hazard and risk. Journal of toxicology and environmental health Part B, Critical reviews 13, 197-217.

Lee, Hae Woo, Carl Lawton, Young Jeong Na, et al. 2013. "Robustness of chemometrics-based feature selection methods in early cancer detection and biomarker discovery" *Statistical Applications in Genetics and Molecular Biology*, 12.2 (2013): 207-223.

Limpert E, Stahel WA, Abbt M, 2001. Lognormal distributions across the sciences: keys and clues, Bioscience 51(5):341-352.

Lind PM, Riserus U, Salihovic S, van Bavel B, Lind L, 2013. "An environmental wide association study (EWAS) approach to the metabolic syndrome", Environment International 55:1-8.

Lioy PJ, 1990. "Exposure analysis and assessment for low-risk cancer agents", Int J Epidemiol 19 Suppl 1:S53-61.

Lock EA, 2010. "Sensitive and Early Markers of Renal Injury: Where Are We and What Is the Way Forward?", Toxicological Sciences 116(1):1-4.

Lu SS, Sobus JR, Sallsten G, Albin M, Pleil JD, Gudmundsson A, Madden MC, Strandberg B, Wierzbicka A, and Rappaport SM, 2014. "Are urinary PAHs biomarkers of exposure to diesel exhaust?", Biomarkers 19(4): 332–339.

Meeker JD, Cooper EM, Stapleton HM, and Hauser R, 2013. "Urinary metabolites of organophosphate flame retardants: temporal variability and correlations with house dust concentrations, Environmental Health Perspectives 121:580–585.

Muscat JE, Liu A, Richie JP, 2011. A comparison of creatinine vs. specific gravity to correct for urinary dilution of cotinine", Biomarkers 16(3):206-211.

Nakamura A, Ishil A, Ohata C, and Kamurasaki T, (2007). "Early induction of osteoactivin expression in rat renal tubular epithelial cells after unilateral ureteral obstruction", Experimental and Toxicological Pathology 59(1):53-59.

Patel-Chamberlin M, Wang Y, Satirapoj B, Phillips LM, Nast CN, Dai T, Watkins RA, Wu X, Natarajan R, Leng A, Ulanday K, Hirschberg RR, LaPage J, Nam EJ, Haq T and Adler SG, 2011. "Hematopoietic growth factor inducible neurokinin-1 (Gpnmb/Osteoactivin) is a biomarker of progressive renal injury across species", Kidney International 79:1138-1148.

Patel CJ, Bhattacharya J, Butte AJ, 2010. "An Environment-Wide Association Study (EWAS) on Type 2 Diabetes Mellitus", PLoS ONE 5(5): e10746.

Paustenbach D and Galbraith D, 2006. "Biomonitoring and biomarkers: exposure assessment will never be the same", Environmental Health Perspectives 114(8):1143-1149.

Phillips M, Altorki N, Austin JHM, Cameron RB, Cataneo RN, Kloss R, Maxfield RA, Munawar MI, Pass HI, Rashid A, Rom WN, Schmitt P, Wai J, 2008. Detection of lung cancer using weighted digital analysis of breath biomarkers, Clinica Chimica Acta 393(2):76-84.

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, 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, 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, Stiegel MA, and Sobus JR, 2011a. "Breath Biomarkers in Environmental Health Science: Exploring Patterns in the Human Exposome", Journal of Breath Research 5:046005 (7pp).

Pleil JD, Stiegel MA, Sobus JR, Liu Q, and Madden MC, 2011b. "Observing the human exposome as reflected in breath biomarkers: heat map data interpretation for environmental and intelligence research", Journal of Breath Research 5:037104 (9pp).

Pleil JD, Stiegel MA, Madden MC, and Sobus JR, 2011bc. "Heat Map Visualization of Complex Environmental and Biomarker Measurements", Chemosphere 84:716-723.

Pleil JD and Sheldon LS, 2011. "Adapting concepts from systems biology to develop systems exposure event networks for exposure science research", Biomarkers 16(2):99–105.

Pleil JD, 2012. "Categorizing biomarkers of the human exposome and developing metrics for assessing environmental sustainability", Journal of Toxicology and Environmental Health, Part B. Critical Reviews 15:264-280.

Pleil JD, Blount BC, Waidyanatha S, and Harper M, 2012a. "Establishing Exposure Science as a Distinct Scientific Discipline", Editorial: Journal of Exposure Science and Environmental Epidemiology 22:317-319.

Pleil JD, Williams M, and Sobus JR, 2012b. "Chemical Safety for Sustainability (CSS): Biomonitoring *in vivo* data for distinguishing among adverse, adaptive, and random responses from *in vitro* toxicology", Toxicology Letters 215(3):201-207.

Pleil JD and Sobus JR, 2013. "Estimating lifetime risk from spot biomarker data and intra-class correlation coefficients (ICC)", Journal of Toxicology and Environmental Health A 76(12):747-766.

Pleil JD, Miekisch W, Stiegel MA, and Beauchamp JD, 2014a. "Extending breath analysis to the cellular level: Current thoughts on the human microbiome and the expression of organic compounds in the human exposome", Journal of Breath Research 8:029001.

Pleil JD, Sobus JR, Stiegel MA, Hu D, Oliver KD, Olenick C, Strynar MJ, Clark M, Madden MC, and Funk WE, 2014b. Estimating common parameters of log-normally distributed environmental and biomonitoring data: harmonizing disparate statistics from publications, Journal of Toxicology and Environmental Health, Part B. Critical Reviews 17:341-368.

Rappaport SM, Smith MT, 2010. "Environment and disease risk", Science 330(6003):460-461.

Rappaport SM, 2012. “Discovering environmental causes of disease”, Journal of Epidemiology and Community Health 66:99-102.

Richards FJ, 1959. A flexible growth function for empirical use. *J Exp Bot,* 10**,** 290-300.

Sheng MHC, Wergedal JE, Mohan S, and Lau KHW, (2008). “Osteoactivin is a novel osteoclastic protein and plays a key role in osteoclast differentiation and activity”, FEBS Letters 582:1451–1458.

Stiegel MA, Pleil JD, Sobus JR, and Madden MC, 2014. Multiplex analysis of inflammatory cytokines in human blood, breath condensate, and urine, ???

Sobus JR, Tan YM, Pleil JD, and Sheldon LS, 2011. “A biomonitoring framework to support exposure and risk assessments”, Science of the Total Environment 409(22):4875-4884.

Sobus JR, Dewoskin RS, Tan YM, Pleil JD, Phillips MB, George BJ, Christensen KY, Schreinemachers DM, Williams MA, Cohen-Hubal EA, and Edwards SW, 2014. Uses of NHANES biomarker data for chemical risk assessment: Trends, challenges and opportunities, Environmental Health Perspectives (submitted).

Sondag GR, Salihoglu S, Lababidi SL, Crowder DC, Moussa FM, Abdelmagid SM, and Safadi FF, (2014). “Osteoactivin Induces Transdifferentiation of C2C12 Myoblasts into Osteoblasts”, Journal of Cellular Physiology 221(7):955-966.

Tan YM, Sobus JR, Chang D, Goldsmith M, Tornero-Velez R, Pleil JD, and Dary CC, 2012. “Reconstructing Human Exposures Using Biomarkers of Exposure and Other Clues”, Journal of Toxicology and Environmental Health, B. Critical Reviews 15(1):22-38.

United States Environmental Protection Agency (USEPA), 2008, “Scientific and Ethical Approaches for Observational Exposure Studies (SEAOS)”, Office of Research and Development, Washington, DC, EPA/600/R-08/062.

Vaidya VS, Ramirez V, Ichimura T, Bobadilla NA, and Bonventra JV, 2006. “Urine kidney injury molecule-1: a sensitive quantitative biomarker for early detection of kidney tubular injury”, American Journal of Physiology 290:F517-F529.

Villeneuve, D. L., D. Crump, N. Garcia-Reyero, M. Hecker, T. H. Hutchinson, C. A. LaLone, B. Landesmann, T. Lettieri, S. Munn, M. Nepelska, M. A. Ottinger, L. Vergauwen and M. Whelan 2014a. "Adverse Outcome Pathway (AOP) Development I: Strategies and Principles." Toxicol Sci **142**(2): 312-320.

Villeneuve, D. L., D. Crump, N. Garcia-Reyero, M. Hecker, T. H. Hutchinson, C. A. LaLone, B. Landesmann, T. Lettieri, S. Munn, M. Nepelska, M. A. Ottinger, L. Vergauwen and M. Whelan 2014b. "Adverse Outcome Pathway Development II: Best Practices." Toxicol Sci **142**(2): 321-330.

Vinken, M, 2013. The adverse outcome pathway concept: A pragmatic tool in toxicology, Toxicology 312(4):158-165.

Wambaugh JF, Setzer RW, Reif DM, Gangwal S, Mitchell-Blackwood J, Arnot JA, Joliet O, Frame A, Rabinowitz J, Knudsen TB, Egeghy P, Vallero D, Cohen-Hubal EA, 2013. "High-throughput models for exposure-based chemical prioritization in the ExpoCast project", Environ Sci Technol. 47(15):8479-88.

Yu Y, Jin H, Holder D, Ozer JS, Villareal S, Shughrue P, Shi S, Figueroa DJ, Clouse H, Su M, Muniappa N, Troth SP, Bailey W, Seng J, Aslamkhan AG, Thudium D, Sistare FD, and Gerhold DL, 2010. Urinary biomarkers trefoil factor 3 and albumin enable early detection of kidney tubular injury, Nature Biotechnology 28(5):470-477.