

Analytical precision, biological variation, and mathematical normalization in high data density metabolomics

Yevgeniya I. Shurubor,^a Ugo Paolucci,^{a,c} Boris F. Krasnikov,^{a,b} Wayne R. Matson,^c and Bruce S. Kristal,^{a,b,d}

^aDementia Research Service, Burke Medical Research Institute, 785 Mamaronck Avenue, White Plains, NY 10605, USA

^bDepartment of Neuroscience, Weill Medical College of Cornell University, 1300 York Avenue, NY 10021, USA

^cESA, Inc., 22 Alpha Road, Chelmsford, MA 01824, USA

^dDepartment of Biochemistry, Weill Medical College of Cornell University, 1300 York Avenue, NY 10021, USA

^eDepartment of Neurology, Westchester Medical Center, 95 Grasslands Road, Private Mail Box # 599 Valhalla, NY 10595, USA

Received 25 August 2004; accepted 1 September 2004

Metabolic serotypes sensitive to caloric intake may enable sera metabolomic profiles to validate epidemiological parameters and predict disease risk in humans. This long-range goal is complicated by the lack of known state markers and the requirement for simultaneous monitoring of multiple small changes. Therefore, analytical precision for appropriate high data density studies using HPLC separations coupled with coulometric array detectors was evaluated over a two month period in pooled rat sera samples (previously collected and stored at -80°C), and in authentic biochemical standards. In sera, mean coefficients of variation (CV) of retention time and ratio accuracy within the established metabolic serotype varied within $\pm 1\%$ and $\pm 3\%$, respectively. In sets of purified standards, the same parameters fluctuated, correspondently, in ranges of $\pm 0.1\%$ and $\pm 1\%$. Median CV of the metabolite concentrations were $\sim 13\%$ in standards and $\sim 11\text{--}19\%$ in sera, and varied non-monotonically with the analytical system status and experimental design. These parameters were shown to be sufficiently controlled so as not to dominate intra-group biological variability in serum metabolomics studies. Continuation of experimental runs across an analytical breakpoint (column replacement) was associated with disproportionate changes in metabolite concentrations, independent of maintained analytical precision. These changes were sufficient to shift overall profile localization in megavariable projection analyses. We developed a mathematical approach to normalize this break and use partial least squares projection to latent structure discriminant analysis to confirm validity of this normalization approach. This generally applicable mathematical correction helps enable longer term high data density studies by removing a critical source of systemic variation.

KEY WORDS: metabolomics; HPLC; electrochemical detection; analytical precision; serum biomarkers; metabolic serotype.

1. Introduction

Historically, each advance in analytical capacity has improved our insight into disease. Examples of this trend range from applications of “simple” advances in clinical chemistry such as the ability to measure glucose to monitor diabetes, to the use of paper and thin layer chromatography to analyze disorders of amino acid metabolism and the lipidopathies, to the application of the modern methods of molecular biology to identify genetic abnormalities that predispose individuals to diseases such as breast cancer (such as mutations in *BRCA 1* and *2* (Futreal *et al.*, 1994; Miki *et al.*, 1994; Wooster *et al.*, 1994)). Evidence to date suggests that “omics” technologies will continue this trend.

Our previous research into the modulation of metabolic serotypes by calorie intake has (i) identified analytically valid metabolites (which differed in serum concentration between groups of rats undergoing dietary restriction (DR (Weindruch and Walford, 1988; Yu, 1994) and rats fed *ad libitum*) (Vigneau-Callahan

et al., 2001), (ii) demonstrated proof of principle (in that the metabolic profiles could be used to classify rats accurately within the cohorts) (Shi *et al.*, 2002c), and (iii) validated the use of these markers in independent cohorts of rats (Shi *et al.*, 2002b). Secondary studies identified and partially eliminated inter-cohort differences in these markers and demonstrated that they could comprise useful expert systems; that is, they could be used to develop algorithms capable of objective prediction (Paolucci *et al.*, 2004a, 2004b; Shi *et al.*, 2004).

Analytical precision functions as a major limitation in both high data density and high throughput studies. This problem is, of course, exacerbated when the groups of interest lack distinguishing state markers (i.e., single or relatively few markers that unambiguously define groups, such as a metabolic compound that is present only in individuals who have taken a certain drug) and can only be distinguished by small differences in the serum concentrations of multiple metabolites, for example. Decades of study on the analytical issues involved in high precision measurement of single analytes have revealed that multiple small changes can occur in analytical systems during investigations that take place

*To whom correspondence should be addressed.
E-mail: bkristal@burke.org

over long time periods. The system that we use, in which metabolites are separated by HPLC and detected by coulometric array detectors, has a potentially wide application for the analysis of small, redox active molecules in biological samples, including sera and other fluids, tissues, cells, and mitochondria (Alburges *et al.*, 1993; Huettle and Gerhard, 1997; Kristal *et al.*, 1998; Nurmi and Adlercreutz, 1999). This technique allows highly reproducible simultaneous quantitation of more than 1000 chemically active compounds having pg mL^{-1} and ng mL^{-1} concentrations in biological samples of interest (Matson *et al.*, 1984; Acworth and Gamache, 1996; Milbury, 1997). But the analytical precision of these measurements is affected by variations in pumping flow rate, mobile phase preparation, and the precision of the auto-sampler work, as well as by the deterioration or chemical contamination of the sensors. Analytical precision is also affected by contamination of the HPLC column and by the act of repacking the material within the column. Even when studying single analytes, or analytes that move in parallel through the column, resolving these analytical complications can be difficult. The multiplicity of measurements in metabolomics studies, where the differences in metabolites between biological groups of interest are not necessarily synchronous or unimodal, suggests a need to re-examine the nature of analytical chemistry in the -omics world. This problem has multiple facets, one facet being an understanding of these issues as they relate to individual variables (e.g., Fiehn *et al.*, 2000), and another as they relate to the effects of this variability on the overall profiles. The studies presented here attempt to fill this latter gap in the area of metabolomics investigation.

2. Materials and methods

2.1. Animal husbandry

Data presented here were collected as part of a large, ongoing study in nutritional modulation of disease and aging processes (Vigneau-Callahan *et al.*, 2001; Shi *et al.*, 2002b, 2002c, 2004; Paolucci *et al.*, 2004a, 2004b). Blood was collected from ~550 rats, both male and female, aged 6–30 months that had been maintained on different dietary regimens. Male and female Fisher 344 \times Brown Norway F₁ rats were obtained monthly from the National Institute on Aging colony at Harlan (Indianapolis, IN). All animals were individually housed. DR feeding regimens were implemented at 6 weeks of age. Sera were collected following sacrifice by decapitation to avoid known differential effects of anesthesia on parameters of interest. All animal experiments were performed under institutionally approved protocols and complied with the Guide for the Care and Use of Laboratory Animals.

2.2. Serum preparation

Blood was collected by terminal exsanguination after decapitation in accordance with standard animal usage guidelines. After collection, blood samples were placed on ice for 30 min and then centrifuged for 10 min at 2000g. Samples were collected as described below for individual and pool analyses, placed in vacuum tubes, frozen in liquid nitrogen, and stored at $-80\text{ }^{\circ}\text{C}$ until analysis.

For pool samples Sera: samples (100 μL each) drawn from multiple animals, were mixed on ice, vortexed, and aliquoted into new tubes (125 μL per tube). Each sample (125 μL) was precipitated with 500 μL acetonitrile (An)/0.4% glacial acetic acid (HAc) at $-20\text{ }^{\circ}\text{C}$, vortexed 20 s, and centrifuged 15 min at 12,000g at $-2\text{ }^{\circ}\text{C}$. A 500 μL aliquot of the supernatant was evaporated to dryness under vacuum in a Centrивap Concentrator (Labconco). Samples were re-dissolved in 100 μL of mobile phase A and placed in an auto-sampler vial. Past work suggests that this protocol helps to preserve some reactive species from decay (Milbury, 1997).

During HPLC-ECD analysis all samples were maintained in a closed auto-sampler platform with a circulating coolant at $4\text{ }^{\circ}\text{C}$. Injections proceeded in two steps: (i) 30 μL was injected to wash the needles, syringe, and tubes and (ii) 50 μL was injected for direct analysis.

Stock solutions of known standard analytes (primarily tyrosine, tryptophan, and purine metabolites, along with some neurotransmitters, antioxidants, and oxidative damage products) were prepared by dilution of 100 mg in 100 mL of normal saline containing 1% phosphoric acid (pH of 3) and stored at $-80\text{ }^{\circ}\text{C}$ until analysis.

2.3. HPLC and data processing methodology

Chromatographic analysis was completely controlled by CoulArray software installed on a Pentium computer. Total computerization of all analytical procedures (pumping, gradient method, injections, detections, data storage) allows samples to be run with minimal operator activity over long periods of time.

HPLC separations and electrochemical detection were conducted according to methodology developed in ESA, Inc essentially as precisely described (Matson *et al.*, 1984; Matson *et al.*, 1987, 1990; Beal *et al.*, 1990, 1992; LeWitt *et al.*, 1992; Ogawa *et al.*, 1992; Acworth and Gamache, 1996; Milbury, 1997; Kristal *et al.*, 1998; 1999; Shi *et al.*, 2002a). Analyte separation was performed on two C18 columns (META250, 5 μm ODS, 250 \times 4.6 mm I.D., ESA Inc.) in series preceded by a changeable guard column. Detection of metabolites was provided by a 16-channel coulometric electrode array with potentials incremented by 60 mV steps (from 0 to 900 mV) to detect most reducible/oxidizable compounds. Both electrodes and columns were contained in a temperature-controlled chamber at $35\text{ }^{\circ}\text{C}$.

To conserve the original status of the reactive analytes, a novel mobile phase pair, consisting of mobile phase A [11 g/L of pentane sulfonic acid (PSA) brought to pH 3.00 with acetic acid] and mobile phase B [0.1 M lithium acetate brought to pH 3.00 with HAc in 80/10/10 methanol/An/isopropanol (MeOH/AN/IPA)], was used. The chromatographic method involved the use of a mobile phase gradient over the course of a 123 min run, during which the flow rate changed from 0.7 to 1.30 mL/min (table 1). All metabolites discussed in this manuscript eluted within 5–100 min after beginning the gradient run. The mobile phase pair and the gradient method both help to improve solubility and remove protein and peptide residuals from the analytical columns and the detector (7).

The parameters for initial processing of signal (electrical current) data into peak information, subsequent data modification, executing pre-stretching and stretching procedures as well the peak identification parameters were chosen according to ESA guidelines as detailed in “Release Notes for CEAS 5.12.” To execute the stretching procedure the “model” files with highest peaks height and best resolution in sera and standard sets were chosen. Concentration of each peak in selected files was set at 100.

After the completion of automatic recognition program work, each peak from the serum metabolic profiles was manually inspected. After analysis of the chromatograms was complete, all data files were transferred into Microsoft Excel for data management and statistical consideration.

2.4. Megavariate analysis

Principal components analysis (PCA) and partial least squares projection to latent structures discriminant analysis (PLS-DA) were performed using SIMCA-P10.5

Table 1
Gradient profile of the chromatographic run

List	Time (min)	Activity	Value
1	0.00	Flow %B	0% B 1.00 mL/min
2	0.10	Auto zero	On
3	0.56	Auto-sampler injection	1.0 s
4	1.00	File	Start
5	30.00	Flow %B	12% B 1.00 mL/min
6	35.00	Flow %B	20% B 1.00 mL/min
7	55.00	Flow %B	48% B 0.70 mL/min
8	90.00	Flow %B	100% B 0.99 mL/min
9	95.00	Flow %B	100% B 1.20 mL/min
10	101.00	Flow %B	100% B 1.20 mL/min
11	101.10	Flow %B	0% B 1.20 mL/min
12	104.00	Flow %B	0% B 1.20 mL/min
13	107.00	Flow %B	0% B 1.00 mL/min
14	112.00	File	Stop
15	112.50	Clean cell	On
16	114.00	Flow 5B	0% B 1.00 mL/min
17	115.50	Clean cell	Off
18	123.00	Flow %B	0% B 1.00 mL/min

(Umetrics, Kinnelon, NJ). Data were unit-variance (z-score) scaled. Components were determined using the autofit procedure. We note that we follow the convention of using megavariate rather than multivariate to refer to analysis of datasets involving variables that are highly correlated within the X block – the analysis of which thus emphasizes reducing large data sets into a few latent variables (Erickson *et al.*, 2001).

3. Results and discussion

Our previous work identified biomarker profiles that distinguish dietary group in rats (Shi *et al.*, 2002b, 2002c, 2004; Paolucci *et al.*, 2004b). In contrast to most previous biomarker studies, which generally followed single markers (or a small number of markers), our studies are reliant on simultaneous quantitative validity and reproducibility of measurement across multiple small molecule markers. In this manuscript, we document this validity in our own system in the presence of certain experimental problems that inherently could associate with this class of metabolomic study. In particular, we were forced to replace analytical columns in the middle of an experimental series; the results shed light on the impact of this type of analytical complication on multivariate analysis of the metabolome. We then document a solution to this discrete analytical breakpoint.

3.1. Design of experiment

The parameters of interest in our study were (i) retention time (RT) of each metabolite on the column during a given chromatographic run, (ii) dominant channel (CH, the electrochemical channel which has the greatest absolute reactivity), (iii) ratio accuracy (RA, the comparative reactivity on the dominant and sub-dominant channels), and (iv) concentration (C) of a metabolite in a particular serum sample. These represent the major analytical parameters related to our sera metabolomic profile. The variability in these parameters for a given type of sample was assessed within single runs, in the context of adjacent runs (termed the “two-day” data set), and in the context of the larger, intact 2 months study (termed the “two-month” data set). In total, we report data from 75 runs of rat sera and standards. Of these 75 runs, the “two-day” data set was comprised of 20 pooled serum samples run consecutively during a 2 day period. The “two-month” data set was comprised of 35 pooled serum samples and 20 standard samples that were interspersed between ~550 individual sera samples (the data from which are not addressed here). All samples in this data set were run within a single 2 month period in an automated, nonstop regime, specifically consisting of 1 standard, 1 pool, and 10 individual sera samples.

Non-monotonic changes in these parameters associated with the simultaneous replacement of both analytical columns were also observed and studied. We note that changes in the RT, RA, and CH parameters have implications for the qualitative recognition of peaks of interest, whereas shifts in C, and the distribution of these shifts, is critical for future mathematical utilization of these datasets.

Model pools for these mathematical normalizations were chosen for both “two-day” and “two-month” data sets. Within the runs, RTs of each peak were normalized using a previously described stretching algorithm (Vigneau-Callahan *et al.*, 2001). This procedure facilitated the process of matching peaks, especially in the context of the aforementioned column replacement. Because the column replacement was observed to affect the parameters under study, the total “two-month” data set was subdivided to “1st” and “2nd” sets of data that correspond, respectively, to runs completed before or after replacement of the analytical columns. Both sets were initially analyzed with respect to the model pool in the 1st data set (figure 1). Even with stretching, the number of peaks correctly identified automatically by the CEAS software was greater in the 1st set as compared to the 2nd set (decreasing from 38 ± 4 to 23 ± 3 , $p < 0.0001$, excluding the outlier on run 2). Because applying the stretching algorithm to mathematically normalize RT helped to increase automated peak identification, but was unable to restore automated peak identification

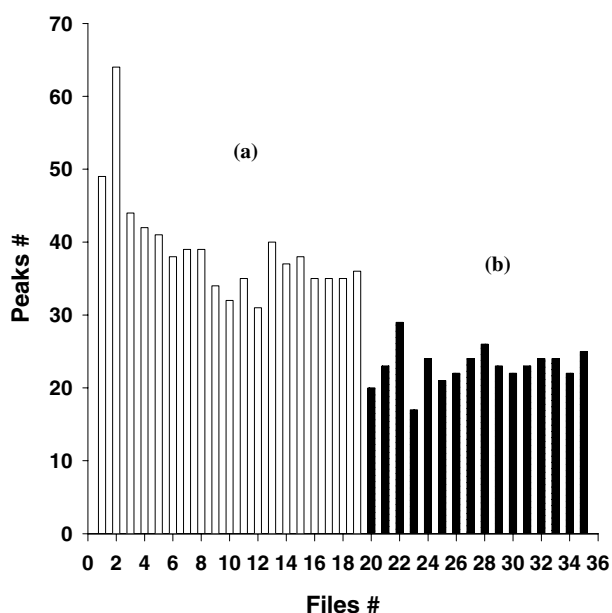


Figure 1. Automated peak matching is decreased by analytical column replacement. Data shown reflects the number of peaks (of 80 total in target run) automatically assigned by the CEAS software after stretching the chromatograms to compensate for shifts in RT. Run # indicates sequential run order. Data reported as clear bars reflect runs completed on the 1st set of columns ($n=19$); data reported as black bars were runs completed on the 2nd set of columns ($n=16$). The difference in recognized peaks was significant at $p < 0.001$ by Mann-Whitney U test.

completely, we further explored the consequences of the analytical difficulties encountered (caused by column replacement) on RT.

3.2. Retention time

Chromatographic RT is one of the primary criteria for establishing the identity of the metabolite represented by a given peak. Before mathematical compensation for chromatographic drift the average RT variability (across 80 metabolites and 35 sera runs) for the total “two-month” data set was approximately $\pm 3\%$ with minimal and maximal absolute variability of individual biomarkers in the range of ± 0.5 – 10% .

Closer examination of within-run variability revealed that RT variability in the 1st set of columns was essentially limited to the first 25 min of the gradient run [mean coefficient of variation (CV), $\pm 5\%$, figure 2a]. Analytes that eluted later (25–100 min after the beginning of the run) exhibited significantly smaller RT variability (mean CV, $\pm 0.5\%$). This result appears to have been due to column contamination, and led

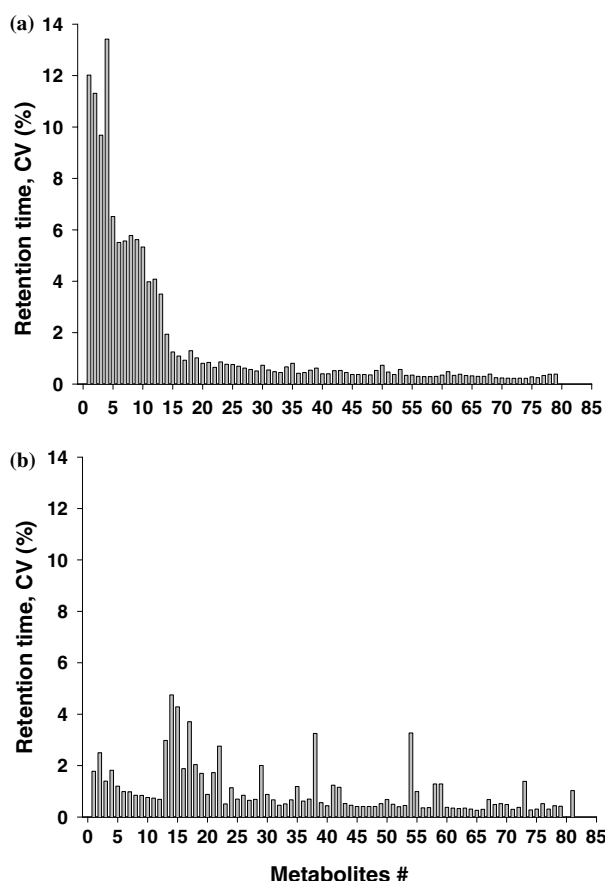


Figure 2. Chromatographic drift of RT in raw data files differed qualitatively and quantitatively between sequential analytical columns. Data presents the variability of RT within the serum “two-month” database that include 1st (a) and 2nd sets (b) of samples (prior to normalization). Metabolites # indicates the number of each from 80 measured metabolites.

directly to our replacing the columns. Analysis of RT variability within the 2nd data set was initially made in comparison to the “model” pool belonging to the 1st data set. This analysis revealed that RT variability within the 2nd data set was both smaller and more sporadic as compared with the fluctuation within the 1st data set (figure 2b). There were no apparent differences between the initial 0–25 min of the chromatographic run (mean CV, $< \pm 2\%$) and the final 25–100 min of the chromatographic run (mean CV, $< \pm 1\%$) in the 2nd set (figure 2b). A sporadic pattern of RT distribution may reflect the settling of the packing material during initial runs with this new set of columns.

Application of a mathematical form of compensation for this chromatographic drift (i.e., the stretching algorithm noted above) reduced RT variability within the total “two-month” data set (mean CV, $\pm < 1\%$, see table 2). The mean RT CV for the initial 0–25 min of gradient run in the 1st set was reduced from $\pm 5\%$ to $\pm 0.6\%$; the mean RT CV for the remaining part of chromatogram (25–100 min) was reduced from $\pm 0.5\%$ to $\pm 0.1\%$. Purified standards had mean CVs of $\pm 0.07\%$, showing the overall stability of the system as

well as the influence of the biological matrix on the RT of compounds present in sera.

RT variability within the “two-day” data set is similar to RT variability within the 2nd set of the “two-month” database. The percentage of metabolites with RT variability $\pm < 1\%$ (73%) was comparable to the scale of RT variability in 2nd data set (83%).

3.3. Dominant channels and ratio accuracy

The dominant channel is another critical parameter for qualitative metabolite identification. In general, the dominant channel response is also stable, but deterioration/contamination of the electrodes, especially of the higher potential channels, may result in shifting the dominant channel. Data from the standards suggest that the dominant channel was stable on our instrument across the run (no changes in 860 analytes measured, table 3). Data from the sera runs suggest that even in the complex biological matrix only 10–15% of compounds ever display such shifts (table 3).

RA is the final critical parameter for qualitative metabolite identification. In general, the ratio between

Table 2
Variation of RT, post-normalization

Shift in RT (as CV%)	Percent of measurements with shifted RT				
	“Two-month” data set			“Two-day” data set	
	Total*	1st set of columns*	2nd set of columns*	Standard*, 1st set of columns	2nd set of columns*
0–1	67	94	83	100	73
1–2	19	5	8	–	18
2–3	9	1	6	–	7
3–4	4	–	1	–	1
4–5	1	–	2	–	1
> 5	1	–	–	–	–
\pm CV	± 0.94	± 0.26	± 0.67	± 0.07	± 0.82

*Total: 35 runs, 80 metabolites; 1st data set: 19 runs, 80 analytes; 2nd data set: 16 runs, 78 analytes; 1st data set, standard: 20 runs, 43 analytes; 2nd data set: 20 runs, 80 analytes.

Table 3
Variation of dominant channels

Total # metabolites displayed shifted dominant channel	Percent of measurements with shifted dominant channels and statistic				
	“Two-month” data set			“Two-day” data set	
	Total*	1st set of columns*	2nd set of columns*	Standard*, 1st set of columns*	2nd set of columns*
0	81	86	88	100	90
1	5	4	7	0	4
2	2	2	0	0	3
3	2	4	1	0	1
4	0	0	1	0	0
5	2	0	1	0	1
> 5	8	6	2	0	1

Note: This table expresses the difference between predicted an actual dominant channel as it related to the percentage of the variables either never showing a change, or showing a change one or more time across all of the runs within the overall dataset.

*Total: 35 runs, 80 metabolites; 1st data set: 19 runs, 80 analytes; 2nd data set: 16 runs, 78 analytes; 1st data set, standard: 20 runs, 43 analytes; 2nd data set: 20 runs, 80 analytes.

Table 4
Variation of RA

Shift in RA (%)	Percent of measurements with shifted RA and statistic				
	“Two-month” data set				“Two-day” data set
	Total*	1st set of columns*	2nd set of columns*	Standard*, 1st set of columns	2nd set of columns*
0–1	40	56	46	79	46
1–2	32	30	25	21	19
2–3	12	6	11	–	18
3–4	0	2	2	–	3
4–5	1	2	2	–	4
>5	12	2	12	–	10
Mean \pm SD	1.0 \pm 0.03	1.0 \pm 0.02	1.0 \pm 0.03	0.97 \pm 0.01	1.0 \pm 0.03

*Total: 35 runs, 80 metabolites; 1st data set: 19 runs, 80 analytes; 2nd data set: 16 runs, 78 analytes; 1st data set, standard: 20 runs, 43 analytes; 2nd data set: 20 runs, 80 analytes.

pre-, post- and leading channels is a quite stable parameter. Again, however, the wearing of electrodes and contamination of the columns may cause a slight RA fluctuation.

Average RA variation in serum “two-month” and “two-day” data sets was 2-fold higher than the RA variation of the standards (± 2.5 as compared to $\pm 1.1\%$). Only 65–85% of identified metabolites in serum profiles had average RA CVs within $\pm 2\%$, as compared with 100% of the standards (table 4).

The observed increases in the number of metabolites that shifted dominant channels and in RA variability in the biological samples suggests again that there are effects caused by the biological matrix; these changes may reflect co-elution of metabolites.

3.4. Metabolite concentration

Two facets of overall analytical accuracy have been functionally defined in the biomarker field: (i) pre-analytical error and variability and (ii) analytical error and variability. We focus here on issues related solely to analytical variability and error, leaving our simultaneous investigations into pre-analytical variation for a subsequent report. Analytical precision is dependent on a series of analytical factors, including the reproducibility of sample preparation and mobile phase production, the precision of the auto-sampler draw and injection, and the status of the analytical system. Each of these factors has been studied broadly and deeply with respect to the analysis of single analytes, but quantitative understanding of situations involving multiple metabolites that may be differentially affected by changes in analytical parameters remains a challenging problem (Fiehn *et al.*, 2000).

If the pool concentration of each metabolite is defined to be 100, the average concentration of serum biomarkers measured in the “two-month” data set over the course of the study varied between 97 (in the 1st set)

and 107 (in the 2nd set). The mean concentration of all metabolites relative to the pool concentration in the total set was 102, with a median CV of $\pm 24\%$. Although the mean metabolite concentrations were higher in the 2nd set (i.e., 107 versus 97 in the 1st set), the median CVs were comparable in the two sets at ~ 19 – 20% . The average concentration of metabolites in the “two-day” (88; median CV, $\pm 11\%$) and standard (83; median CV, $\pm 13\%$) data sets was lower relative to their cognate pools (table 5).

The finding that the overall concentration in the 2nd set was increased relative to the average concentration in the 1st set suggests that differences in the separation of individual analytes in the two sets of analytical columns are responsible for the observed differences in concentrations. In truth, however, individual analytes demonstrate a far more complex behavior over the gradient run. The measurement error for metabolites eluted during the first (0–75) minutes of the gradient run (median CV for “two-month” data set, ~ 13 – 20% ; for “two-day” data set, ~ 8 – 17%) was small in comparison to the measurement error of the final (75–100) minutes of the gradient run (median CV, ~ 32 – 40% and $\sim 23\%$, respectively; table 6). The sharp increase of variability in the last time frame (75–100 min) of the chromatographic run corresponds to the period when the proportion of the organic modifier in the mobile phase gradient is gradually increased from 80% to 100%, a process that effectively removes residual lipids and polysaccharides from both the analytical columns and the coulometric detector (see tables 1 and 6). All metabolites in our standard eluted by 75 min, and these metabolites exhibited relatively low variability over the gradient run (median CV, ~ 9 – 15%).

3.5. Analytical precision and intra-group biological variability

It is, essentially, a trivial and self-referential statement that, for any set of biomarkers to offer high

Table 5
Variation of concentration

Range of concentration variability (%)	Percent of measurements with different concentration range and statistic				
	“Two-month” data set			“Two-day” data set	
	Total*	1st set of columns*	2nd set of columns*	Standard*, 1st set of columns*	2nd set of columns*
0–10	13	24	24	37	40
11–20	25	28	29	33	24
21–30	21	21	23	9	16
31–40	19	15	13	9	11
41–50	14	9	5	7	4
51–60	8	3	3	5	5
61–70	0	0	3	0	0
71–80	0	0	1	0	0
Mean	101.5	96.5	107.4	83.2	87.5
Median	97.7	95.3	105.8	81.7	87.10
Mean, CV	±27.7	±22.9	±22.9	±18.2	±18.5
Median, CV	±23.9	±19.9	±18.6	±13.1	±11.2
Mean, SD	±28.5	±21.5	±24.1	±14.5	±16.5
Median, SD	±23.4	±19.0	±17.7	±10.2	±10.4

*Total: 35 runs, 80 metabolites; 1st data set: 19 runs, 80 analytes; 2nd data set: 16 runs, 78 analytes; 1st data set, standard: 20 runs, 43 analytes; 2nd data set: 20 runs, 80 analytes.

Table 6
Distribution of average concentration and average variability of metabolites over the chromatographic gradient run

Subdivision of gradient run (min)	Metabolites concentration over the gradient run, mean ± median SD/± CV			
	“Two-month” data set			“Two-day” data set
	1st set of columns*	2nd set of columns*	Standard*, 1st set of columns*	2nd set of columns*
0–25	102.6 ± 12.4/±13.6	98.4 ± 15.4/±17.2	85.1 ± 11.1/±13.3	89.1 ± 9.2/±10.1
25–50	87.7 ± 15.8/±17.2	108.0 ± 18.5/±19.6	81.2 ± 12.1/±14.9	78.5 ± 10.6/±16.7
50–75	89.9 ± 19.9/±20.4	108.0 ± 16.0/±13.3	89.6 ± 7.7/±8.7	96.4 ± 6.8/±8.2
75–100	133.8 ± 35.0/±32.5	129.0 ± 37.9/±39.1		92.5 ± 24.2/±22.8

*Number of metabolites eluted during different parts of gradient run for “two-month” and “two-day” data sets: 0–25 min, 19–20; 25–50 min, 27–30; 50–75 min, 19–21; 75–100 min, 7–9. The same for standard set of data: 0–25 min, 11, 25–50 min, 26, 25–75 min, 6.

discriminatory power, it is essential that they be measured with sufficient precision. It is, however, a much more complex concept to assign specific quantitative criteria to determine if sufficient precision has been reached. Classical univariate power calculations provide one answer, by expressing N as a function of the difference in means and standard deviations for the concentration of a given biomarker in different serum samples. Cohen's d and other statistical measures of effect size likewise look at the mean and standard deviation as a way of expressing this potential difference. Both, however, deal more with defining differences between groups rather than in defining the groups *a priori* or with conducting classification analyses. Bayesian predictors can express probabilities with respect to group memberships, but appear dependent on more information about prior probability and distributions than we are comfortable making at this time, especially as we are still defining the biological meaning of the distributions that are in question.

It seems reasonable, however, to consider megavariable analogs to standard deviations and differences in means. To some extent, the standard deviation of a given group of observations' position in a projection analyses is a linear combination of the standard deviations of all of the individual metabolites that make up a profile, multiplied by their appropriate constants derived from the projection. Given this relationship, it becomes reasonable to ask whether the analytical variation – essentially the analytical precision – dominates the biological variation. This question may be addressed at two levels: inter-group (i.e., across different genders and/or diets) variability and intra-group (i.e., within a single gender and diet) variability, with the latter being the stronger and more stringent test (because the intra-group standard deviations are expected to be significantly smaller than the inter-group standard deviations).

Figure 3 shows that biological variation significantly predominates over analytical variation when one considers intra-group standard deviations of all 80 of

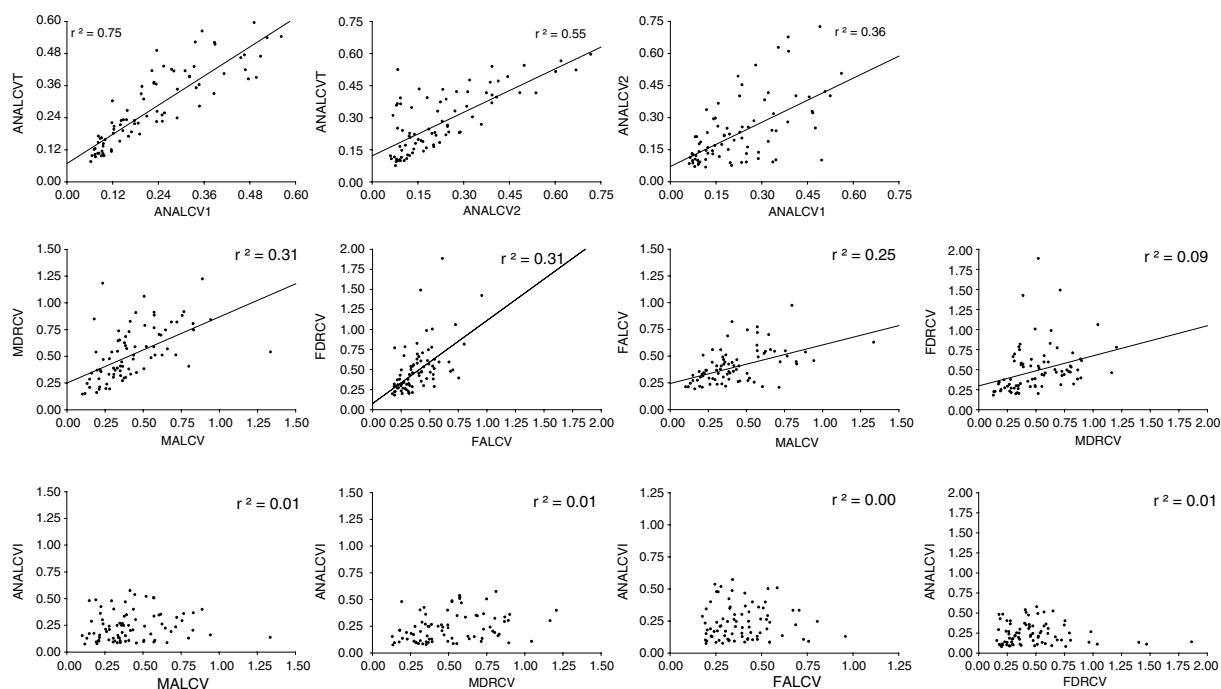


Figure 3. Biological variation dominates analytical variation. The coefficients of variation are plotted for combinations of analytical parameters (top row), biological parameters (middle row) and biological versus analytical parameters (bottom row). Regressions (see text for limitations) shown for statistically significant lines. ANALCV1, ANALCV2, ANALCVT – analytical coefficients of variation in the “two-month” set 1, “two-month” set 2, and overall “two-month” set. MAL, MDR (male *ad libitum*, dietary restricted groups); FAL, FDR (female *ad libitum*, dietary restricted groups).

the variables in the study. Standard deviations (as CVs) in the total study are related to those in the set run on either the 1st or 2nd set of columns, which are also statistically correlated (figure 3, top row, note that regressions were performed to present a visual picture, but many of the graphs shown in figure 3 violate strict regression criteria because residuals are correlated with the absolute values). Likewise, biological variation for a given analyte was well-conserved across diet and gender groups (figure 3, middle row). In contrast, no relationship was found between analytical precision and overall measured variation, suggesting that the variation was biological and not analytical in nature (figure 3, bottom row). For most analytes (>85%), analytical variation was, as expected, less than the overall measured variation (whether the exceptions were stochastic statistical artifacts or specific analytical problems has not yet been determined). Of perhaps equal importance, analytical variation does account for at least one-half of the variation in nearly one-half of the analytes, suggesting the importance of tight analytical controls in studies such as ours that lack overt state markers.

3.6. Mathematical normalizations: linear transforms in k -dimensional space

Thus, the analysis above demonstrates that the overall spread within our distributions is primarily determined biologically rather than analytically, even at

the intra-group level. As discussed, this provides a functional megavariable/projection analysis equivalent to determining the standard deviation in univariate studies. The second issue, the functional equivalent to intra-group means (with subsequent implications to inter-group differences in means) is the variation in the value of a given component on a projection plot. The analysis above suggests that the intra-group spread of this score should be primarily determined biologically, but does not speak to whether the mean of this score may be skewed relative to a hypothetical “true” value by analytical drift. The column shift described two sections ago provides a perfect test case. To address this issue, we looked at a PCA of the 35 pooled samples from the “two month” 1 and 2 studies described above. PCA readily distinguishes these groups (figure 4A), suggesting that any projection analyses using these data may be skewed by the components recognized here on PC1 and PC2. PLS-DA confirms that these distinctions are sufficient to enable total class separation with a single component (figure 4B). Linear scaling, normalization, and transformations failed to recombine the groups (not shown, but see (Paolucci *et al.*, 2004b) for an equivalent problem). In theory, one could recombine these axes by delving into the projections themselves and correcting the relevant equations. This approach, while mathematically feasible, offers potentially insurmountable complications as one moves from practical use in normalizing across generalized analytical breakpoints to

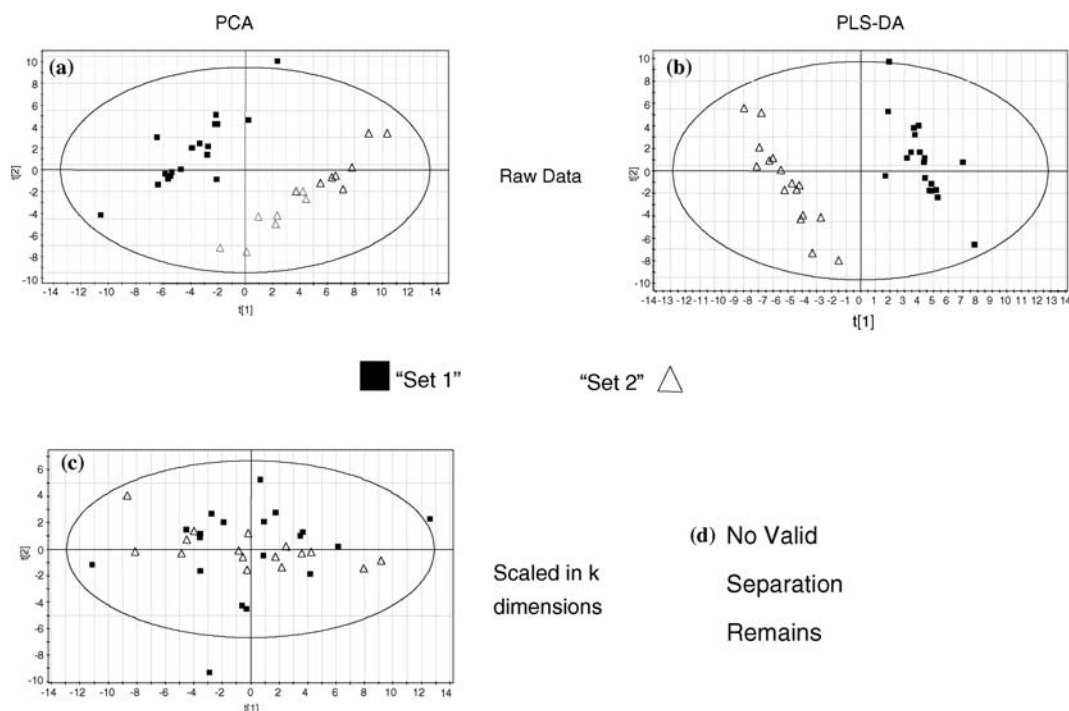


Figure 4. k -Dimensional scaling eliminates analytical shift scores plot from statistically valid PCA (panels a and c) and PLS-DA analyses. Panel d is left blank as PLS-DA finds little or no correlation between the variables and the run order (i.e., the artifactual separation has been removed). Statistical validity based on seven trials of leave-one-seventh out and $Y(\text{class})$ -variable permutation validation.

studies where the variable loadings are not determined solely by analytical drift.

In contrast, we know that the component scores are essentially derived from the projection across all k dimensions (where k is the number of variables in the study). We thus propose that simultaneously scaling linearly in every dimension should scale the projection. Therefore, the mean of each variable in “two-month” set 2 was normalized to the mean of each variable in “two-month” set 1, and the resulting k -dimensional scaling applied to each member pool in “two-month” 2. The resulting 35 pool set was analyzed by PCA. This scaling approach completely eliminated overt separations (figure 4C, first two components shown; Furthermore, none of the eleven statistically valid principal components discriminated the groups). Successful scaling of the projections was confirmed by PLS-DA, which could not generate separations even when we attempted to force components to be generated.

4. General discussion

Metabolomics is the general name for the -omics version of classical biochemical approaches to metabolism. Metabolomics preferably focuses on ‘holistic’ views of systemic biochemistry or multiple pathways rather than on one or a few metabolites of interest; also it focuses on patterns or profiles and, in consequence, often draws information from compounds that are currently undefined structurally. Major goals of the metabolomics field include broad studies in diagnostics,

drug development, disease prediction, plant growth and metabolism, etc. (Harrigan and Goodacre, 2003). Primary analytical tools include HPLC, MS and its variants, LC-MS and its variants, GC-MS, and NMR (Harrigan and Goodacre, 2003).

Each of these analytical tools has been in common use for decades, and the analytical issues related to the measurement of single analytes with these tools are also well explored. Past areas of evaluation and optimization included sample preparation, specialized instrumentation conditions (i.e., specific HPLC columns, type of detection), and run matrices (i.e., mobile phases for HPLC; methods, gradient or isocratic; flow rates; matrix for MALDI). Similarly, normalization protocols have also been optimized, and primarily utilize one of a few approaches: (i) direct measurements compared to a standard curve; (ii) linear, monotonic normalization of entire runs to internal standards; (iii) blended samples; (iv) split and/or titrated samples; (v) spiked samples; and (vi) use of stored aliquots of pools of interest over a period of years. Unfortunately, none can be directly applied to solve the complex analytical complications that arose in the current study.

It is not surprising that the collision of high data density -omics-type approaches to biochemistry and high precision is problematic. The unfortunate occurrence of several analytical complications during our study gave us the opportunity to begin to develop robust mathematical approaches designed to reduce the complications that are inherently associated with these types of approaches.

The problems occurred despite apparent stringent control of overall chromatographic conditions. Dominant channel assignments were conserved in 100% of all standards and in 85–90% of the metabolites in pool sera. The CVs for RA were within ~1.1–2.6%, and the CVs for RTs were within ~0.1–0.8%, in both standard and sera samples. The median CVs for concentration measurements were significantly higher, ~13% in the standard and ~11–20% in the “two-day” and “two-month” data sets of sera samples. Despite this higher variability, we observed non-monotonic, time-dependent changes in concentration across the chromatographic profile. These changes were further complicated by apparent co-elution and matrix effects, as well as by the necessity of replacing the analytical columns midway through the study.

5. Concluding remarks

Analytical problems arising during metabolomic analysis are further complicated by the ability of modern informatics and data mining tools to extract every nuance of signal from large datasets. Indeed, megavariable analyses showed that these changes can shift projection analyses even though they do not appear to be the driving factor for overall variation or spread of a given metabolite or intra-group profile. From experience, these mathematical tools can, for example, enable us to determine the day an animal enters a colony. Straightforward normalization techniques applied to HPLC data, such as scalar normalization to a constant peak such as tyrosine, fall short and fail to correct for analytical shifts (Paolucci *et al.*, 2004b). In contrast, the multidimensional mathematical normalizations introduced here appear to reduce overall variability and facilitate further mathematical analysis. We suggest that this approach will be generally useful for such analyses, and especially relevant when these analyses are conducted over long experimental runs.

Acknowledgments

The authors thank Dr. Tom Vogl for his many critical discussions and comments on the manuscript. This work was supported by NIH NIA R01 AG15354 and funds from Burke Medical Research Institute.

References

- Acworth, I.N. and Gamache, P.H. (1996). *The Coulometric Electrode Array for Use in HPLC Analysis, Part 1: Theory*. Amer. Lab., May.
- Alburges, M.E., Narang, N. and Wamsley, J.K. (1993). A sensitive and rapid HPLC-ECD method for the simultaneous analysis of norepinephrine, dopamine, serotonin and their primary metabolites in brain tissue. *Biomed. Chromatogr.* **7**, 306–310.
- Beal, M.F., Matson, W.R., Storey, E., *et al.* (1992). Kynurenic acid concentrations are reduced in Huntington's disease cerebral cortex. *J. Neurol. Sci.* **108**, 80–87.
- Beal, M.F., Matson, W.R., Swartz, K.J., Gamache, P.H. and Bird, E.D. (1990). Kynurenic pathway measurements in Huntington's disease striatum; evidence for reduced formation of kynurenic acid. *J. Neurochem.* **55**, 1327–1339.
- Erickson, L., Johansson, E., Kettaneh-Wold, N. and Wold, S. (2001). *Multi- and Megavariable Data Analysis*. Umetrics AB.
- Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R.N. and Willmitzer, L. (2000). Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* **18**, 1157–1161.
- Futreal, P.A., Liu, Q., Shattuck-Eidens, D., *et al.* (1994). BRCA1 mutations in primary breast and ovarian carcinomas. *Science*. **266**, 120–122.
- Harrigan, G.G. and Goodacre, R. (2003). *Metabolic profiling: its role in biomarker discovery and gene function analysis*. Kluwer Academic Publishers.
- Huettle, P. and Gerhard, A. (1997). Multicomponent quantitations of monoamines and related compounds in biological fluids and CNS tissues by HPLC coupled with multielectrode electrochemical detector. Coulometric electrode array detectors for HPLC in Acworth, I.N., *et al.* (Eds), *Progress in HPLC-PRCE*. VSP International Science Publication, Utrecht, pp. 213–223.
- Kristal, B.S., Vigneau-Callahan, K.E. and Matson, W.R. (1998). Simultaneous analysis of the majority of low-molecular-weight, redox-active compounds from mitochondria. *Anal. Biochem.* **263**, 18–25.
- Kristal, B.S., Vigneau-Callahan, K.E. and Matson, W.R. (1999). Purine catabolism: links to mitochondrial respiration and antioxidant defenses. *Arch. Biochem. Biophys.* **370**, 22–33.
- LeWitt, P.A., Galloway, M.P., Matson, W.R., Milbury, P.M., McDermott, M. and Srivastava, D.K. (1992). Markers of dopamine metabolism in parkinson's disease: the Parkinson's Study Group. *Neurology*. **42**, 2111–2117.
- Matson, W.R., Bouckoms, A., Svendsen, C., Beal, M.F. and Bird, E.D. (1990). Generating and controlling multiparameter databases for biochemical correlates of disorders in *Basic, clinical and therapeutic aspects of Alzheimer's and Parkinson's diseases*. Plenum, New York, pp. 513–516.
- Matson, W.R., Gamache, P.H., Beal, M.F. and Bird, E.D. (1987). EC array sensor concepts and data. *Life Sci.* **41**, 905–908.
- Matson, W.R., Langials, P., Volicer, L., Gamache, P.H., Bird, E.D. and Mark, K.A. (1984). n-Electrode three dimensional liquid chromatography with electrochemical detection for determination of neurotransmitters. *Clin. Chem.* **30**, 1477–1488.
- Miki, Y., Swensen, J., Shattuck-Eidens, D., *et al.* (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science*. **266**, 66–71.
- Milbury, P.E. (1997). CEAS generation of large multiparameter databases for determining categorical process involvement of biomolecules in Anonymous, *Coulometric array detectors for HPLC*. VSP International Science Publication, pp. 125–141.
- Nurmi, T. and Adlercreutz, H. (1999). Sensitive high-performance liquid chromatographic method for profiling phytoestrogens using coulometric electrode array detection: application to plasma analysis. *Anal. Biochem.* **274**, 110–117.
- Ogawa, T., Matson, W.R., Beal, M.F., *et al.* (1992). Kynurenic pathway abnormalities in Parkinson's disease. *Neurology*. **42**, 1702–1706.
- Paolucci, U., Vigneau-Callahan, K.E., Shi, H., Matson, W.R. and Kristal, B.S. (2004a). Development of biomarkers based on diet-dependent metabolic serotypes: concerns and approaches for cohort and gender issues in serum metabolome studies. *OMICS J. Integr. Biol.* **8**, 209–220.
- Paolucci, U., Vigneau-Callahan, K.E., Shi, H., Matson, W.R. and Kristal, B.S. (2004b). Development of biomarkers based on diet-dependent metabolic serotypes: Characteristics of component-

- based models of metabolic serotype. *OMICS J. Integr. Biol.*, **8**, 221–238.
- Shi, H., Paolucci, U., Vigneau-Callahan, K.E., Milbury, P.E., Matson, W.R. and Kristal, B.S. (2004a). Development of biomarkers based on diet-dependent metabolic serotypes: practical issues in development of expert system-based classification models in metabolomic studies. *OMICS J. Integr. Biol* **8**, 197–208.
- Shi, H., Vigneau-Callahan, K.E., Matson, W.R. and Kristal, B.S. (2002). Attention to relative response across sequential electrodes improves quantitation of coulometric array. *Anal. Biochem.* **302**, 239–245.
- Shi, H., Vigneau-Callahan, K.E., Shestopalov, A.I., Milbury, P.E., Matson, W.R. and Kristal, B.S. (2002). Characterization of diet-dependent metabolic serotypes: primary validation of male and female serotypes in independent cohorts of rats. *J. Nutr.* **132**, 1039–1046.
- Shi, H., Vigneau-Callahan, K.E., Shestopalov, A.I., Milbury, P.E., Matson, W.R. and Kristal, B.S. (2002). Characterization of diet-dependent metabolic serotypes: Proof of principle in female and male rats. *J. Nutr.* **132**, 1031–1038.
- Vigneau-Callahan, K.E., Shestopalov, A.I., Milbury, P.E., Matson, W.R. and Kristal, B.S. (2001). Characterization of Diet-Dependent Metabolic Serotypes: Analytical and Biological Variability Issues in Rats. *J. Nutr.* **131**, 924S–932S.
- Weindruch, R., Walford, R. (1988). *The Retardation of Aging and Disease by Dietary Restriction*. Charles C. Thomas, St. Louis.
- Wooster, R., Neuhausen, S.L., Mangion, J., et al. (1994). Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12–13. *Science*. **265**, 2088–2090.
- Yu, B.P. (1994). *Modulation of Aging Processes by Dietary Restriction*. CRC Press, Boca Raton.