OMICS A Journal of Integrative Biology
Volume 8, Number 3, 2004

Development of Biomarkers Based on Diet-Dependent Metabolic Serotypes: Practical Issues in Development of Expert System-Based Classification Models in Metabolomic Studies

HONGLIAN SHI,[1] UGO PAOLUCCI,[1] KAREN E. VIGNEAU-CALLAHAN,[2] PAUL E. MILBURY,[3] WAYNE R. MATSON,[2] and BRUCE S. KRISTAL[1,4]

ABSTRACT

Dietary restriction (DR)–induced changes in the serum metabolome may be biomarkers for physiological status (e.g., relative risk of developing age-related diseases such as cancer). Megavariate analysis (unsupervised hierarchical cluster analysis [HCA]; principal components analysis [PCA]) of serum metabolites reproducibly distinguish DR from ad libitum fed rats. Component-based approaches (i.e., PCA) consistently perform as well as or better than distance-based metrics (i.e., HCA). We therefore tested the following: (A) Do identified subsets of serum metabolites contain sufficient information to construct mathematical models of class membership (i.e., expert systems)? (B) Do component-based metrics out-perform distance-based metrics? Testing was conducted using KNN (k-nearest neighbors, supervised HCA) and SIMCA (soft independent modeling of class analogy, supervised PCA). Models were built with single cohorts, combined cohorts or mixed samples from previously studied cohorts as training sets. Both algorithms over-fit models based on single cohort training sets. KNN models had >85% accuracy within training/test sets, but were unstable (i.e., values of k could not be accurately set in advance). SIMCA models had 100% accuracy within all training sets, 89% accuracy in test sets, did not appear to over-fit mixed cohort training sets, and did not require post-hoc modeling adjustments. These data indicate that (i) previously defined metabolites are robust enough to construct classification models (expert systems) with SIMCA that can predict unknowns by dietary category; (ii) component-based analyses outperformed distance-based metrics; (iii) use of over-fitting controls is essential; and (iv) subtle inter-cohort variability may be a critical issue for high data density biomarker studies that lack state markers.

[1]Dementia Research Service, Burke Medical Research Institute, White Plains, New York.
[2]ESA Inc., Chelmsford, Massachusetts.
[3]Antioxidants Research Laboratory, Jean Mayer USDA Human Nutrition Res. Center on Aging at Tufts University, Boston, Massachusetts.
[4]Departments of Biochemistry and Neuroscience, Cornell University Medical College, New York, New York.

INTRODUCTION

THE BENEFICIAL EFFECTS of dietary or caloric restriction in laboratory rodents (Kristal and Yu, 1994; Weindruch and Walford, 1988; McCay et al., 1935) and the detrimental effects of obesity on human health (Willett et al., 1999) reveal the influence of long term caloric intake and balance on morbidity and mortality. Indeed, over-nutrition may be second only to smoking as a preventable cause of cancer. We are therefore identifying serum profiles that can distinguish between *ad libitum* fed (AL) and dietary restricted (DR) rats. Our work is based on the belief that these profiles will be of use in understanding DR, in clarifying epidemiological relationships between caloric intake/balance and morbidity/mortality, and in the ability to predict relative risk of certain human diseases. Serotypes are being identified using HPLC coupled with coulometric electrochemical array detectors (Matson et al., 1984; Milbury, 1997; Vigneau-Callahan et al., 2001). Previous research has identified analytically valid metabolites (Vigneau-Callahan et al., 2001), demonstrated proof of principle (classification accuracy in the cohorts in which the markers were developed) (Shi et al., 2002b), and validation of the basic profiles in independent cohorts in independent cohorts (Shi et al., 2002a). These studies were conducted using unsupervised approaches based on hierarchical cluster analysis (HCA) and principal components analysis (PCA). (Unsupervised approaches are used to mathematically describe a data set independently of prior information such as class membership. Supervised approaches utilize such prior information to help inform the analysis, often for the purposes of building predictive models for subsequent classification of unknowns.) The next stage in the identification of metabolic serotypes is to build expert systems (i.e., mathematical classification models that can be used to predict the class or category of an unknown). In our direct example, the category would refer to a rat diet; in the planned long-term extension of our work, this will refer to the relative risk of developing a certain disease based on the serum profile of an individual.

HCA and PCA are exploratory data analysis methods that enable investigators to appreciate the major sample or variable correlations within megavariate data sets, but neither PCA nor HCA directly enables prediction of class membership of a given sample. In contrast, classification algorithms such as K-Nearest Neighbor (KNN; Cover and Hart, 1967) and Soft Independent Modeling of Class Analogy (SIMCA; Wold, 1976) construct models based on pre-defined (i.e., pre-assigned, supervised) samples, for example, an AL rat. These samples are thus used to teach or "train" the algorithm to recognize specific classes, and thus form what is called the "training set." These trained algorithms are then used to classify one or more unknowns, termed the "test set," for example, a set of AL and DR animals from an independent cohort. KNN, which is based on the same mathematical theory as HCA (KNN at $k = 1$ is HCA, in which k represents the number of mathematically closest observations ["neighbors"] polled), constructs models where the k nearest neighbors "vote" for membership in their own class. In other words, classification is accomplished by assigning observations (e.g., assigning a rat to either the AL or DR group) in such a way as to minimize differences within each cluster. Likewise, SIMCA is based on the mathematical principles that underlie PCA. While PCA calculates principal components on a whole data set, SIMCA generates principal component models for each training set class. Whereas PCA can only describe a dataset and thus provide visual information as to the relationship between a new sample and members of the training sets, SIMCA predicts class membership of a new sample, or indicates that a new sample is not a member of the training class(es). Again, in our current work and direct example, "class" would refer to a rat diet, in the planned long term extension of our work, this will refer to the relative risk of developing a certain disease based on the serum profile of an individual.

KNN and SIMCA are complementary approaches, although both are examples of what are termed similarity techniques. Similarity techniques presuppose that objects closer together in mathematical space (e.g., smallest Euclidian distance) are more likely to be from the same group. Alternative approaches, known as separability and probability techniques, overfit (i.e., the ability and tendency of expert system algorithms to derive models that are too specific for the training set data, and thus of limited future applicability) sample poor-variable rich data sets, and thus are not useful for our studies. KNN is a distance-based metric (separations are made based on absolute [ie, scalar] differences between two observations in a n-dimensional space, while SIMCA is component-based (separations are made based on both absolute and directional [i.e., vectorial] differences between two observations in a n-dimensional space, where n = the num-

ber of variables in the study (metabolites in our study). KNN functionally weights all variables equally, whereas SIMCA functionally weights variables according to their distributions within and between groups. KNN can be particularly well-suited to a certain sample poor environments/distributions, and it can function even with only one training set sample per category. The main advantage of SIMCA is that it can recognize and utilize relationships that are unique to each class. Therefore, SIMCA may be particularly effective where the primary discriminators between classes are weighted relatively distinctly on different components than those that distinguish intra-class individuals. Based on the nature of the techniques and their advantages, KNN and SIMCA were chosen as the classification methods to be used in this study.

MATERIALS AND METHODS

Animal husbandry

The husbandry, including diet, of the male and female Fischer 344 × Brown Norway F_1 rats obtained monthly from the National Institute on Aging colony at Harlan (Indianapolis, IN) and used in these studies has been described previously (Vigneau-Callahan et al., 2001; Shi et al., 2002b; Shi et al., 2002a).

Animal cohorts

Animals were grouped into cohorts based on entry into our animal colony, and were comprised as follows:

Male cohort A: 5 AL and 8 DR rats
Male cohort B: 6 AL and 6 DR rats
Female cohort A: 6 AL and 5 DR rats
Female cohort B: 8 AL and 8 DR rats
Female cohort C: 8 AL and 8 DR rats

To build KNN and SIMCA models, samples were broken into training sets and test sets as follows: (i) For single cohort studies in male rats, algorithms were developed on a training set consisting of all samples in a single male cohort (A or B), and algorithms were tested on all samples in the opposite male cohort (i.e., B or A, respectively), and for single cohort studies in female rats, algorithms were developed on a training set consisting of all samples in a single female cohort (A, B, or C), and algorithms were tested on all samples in the other female cohorts (B or C, A or C, A or B, respectively); (ii) for studies of the overall population, algorithms were built and tested on a the intact set of either all male rats in cohorts A and B, or all female rats in cohorts A, B, and C; (iii) to examine equivalent separation issues without complications from cohort specificity, we created artificially mixed cohorts in both male and female datasets. For mixed cohort studies in male rats, two mixed cohort datasets were created. Male Mix 1 consisted of data from 3 DR/3 A cohort A and 3 DR/3 AL from cohort B. Male Mix 2 consisted of data from 5DR/2AL from cohort A and 3DR/3AL from cohort B. Algorithms were trained on one dataset, and when indicated, tested on the opposite. Two mixed cohort datasets were also created for mixed cohort studies in female rats. Female Mix 1 consisted of data from 3 DR/3 AL from cohort A, 4 DR/4AL from cohort B, and 4 DR/4 AL from cohort C. Female Mix 2 consisted of data from the rest of the three female cohorts (3 AL/2 DR in cohort A, 4 AL/4 DR in cohort B, and 4 AL/4 DR in cohort C. Algorithms were trained on one dataset, when indicated, tested on the opposite.

HPLC methodology

HPLC separations and coulometric array detection was conducted essentially as described previously using an ESA CoulArray system (ESA, Inc., Chelmsford, MA) (Matson et al., 1984; Milbury, 1997; Vigneau-Callahan et al., 2001; Milbury, 1997). Briefly, a gradient HPLC separation coupled with a 16-channel coulometric electrode array was used to determine levels of analytes relative to a standard.

Statistical analysis

Data were analyzed using the programs CEAS 504 (ESA, Inc., Chelmsford, MA), Pirouette 2.7/3.0 (Infometrix, Inc., Woodinville, WA), and SIMCA-P (Umetrics, Kinnelon, NJ).

RESULTS AND DISCUSSION

KNN and SIMCA were used to construct classification models based on previously defined metabolites identified in male (29 metabolites) and female (61 metabolites) rats.

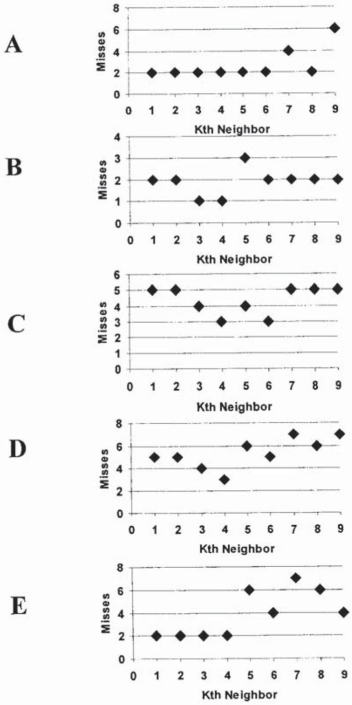

FIG. 1. Accuracy testing in KNN training sets (males). Optimal value of K is unstable. The number of misses indicates the number of samples whose predicated category did not match their *a priori* category. As in A, the miss is two at k = 1, which indicates there were two misclassified samples. Models were built with training sample sets: cohort A [13 samples] (**A**); cohort B [12 samples] (**B**); combined samples of cohorts A and B [25 samples] (**C**); Mix 1 [12 samples] (**D**) and Mix 2 [13 samples] (**E**).

KNN analysis

In this study, there are two categories, DR and AL. Different training sample sets were composed from samples in the two male cohorts and three female cohorts as described in Materials and Methods. KNN categorizes each test set member based on its proximity in mathematical space to previously classified samples. The predicted class of an unknown depends on the class of its k nearest neighbors. Each of the k closest training set samples votes once for its class, and the unknown is then assigned to the class with the most votes. A KNN prediction assigns each unknown to one and only one of the categories defined in the training set. The ability to select a single value for k that gives an assignment representative of related k values is essential for KNN to have utility in a given study. KNN models were constructed with single, combined, and mixed cohort training sets from both males and females (Figs. 1 and 2, Tables 1–4). Optimal values for k (number of neighbors to be considered) could not be consistently determined *a priori* (Figs. 1 and 2). When the optimal value for k was chosen *after* the data were examined, the predictive accuracy within the training set was 94 ± 6% (mean ± SD) within single cohort training sets, and 87 ± 8% in mixed/combined cohort training sets. To test the robustness of the models, these models were used to predict the category to which independent unknown samples belonged. Again, we were unable to determine an optimal value for k *a priori*. When the optimal value for k was chosen *after* the data were examined, accuracy in predictions based on a single cohort training set was 59 ± 8%. These data indicate that there exist inter-cohort differences sufficiently large that the algorithms used can find separations between the cohorts that are com-

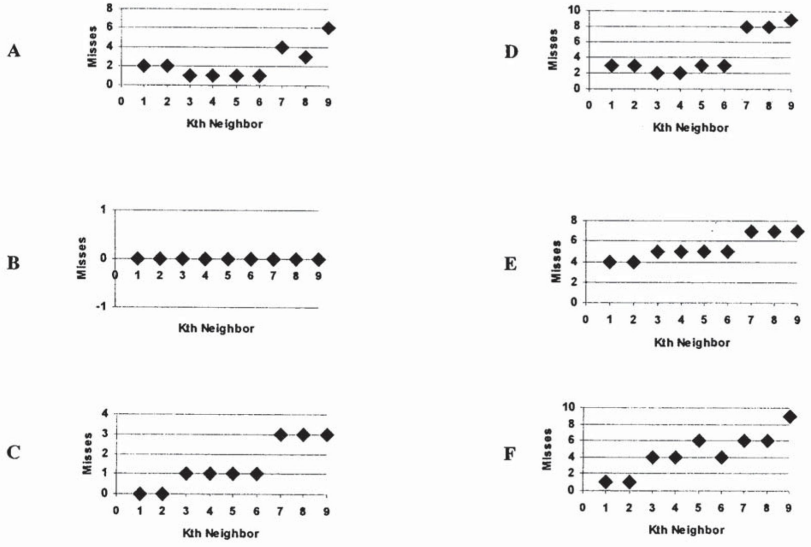

FIG. 2. Accuracy testing in KNN training sets (females). Optimal value of K is unstable. Models were built with training sample sets: cohort A, [11 samples] (**A**); cohort B, [16 samples] (**B**); cohort C [16 samples] (**C**); combined samples of cohorts A, B and C [43 samples] (**D**); Mix 1 [22 samples] (**E**) and Mix 2 [21 samples] (**F**). See description in Figure 1.

TABLE 1. KNN CLASSIFICATION OF MALE SAMPLES: TRAINING SET STUDIES

Models		*pDR*	*pAL*	*Accuracy*
KM_A	aDR	8	0	
	aAL	2	3	85%
KM_B	aDR	5	1	
	aAL	0	6	92%
KM_{AB}	aDR	12	2	
	aAL	1	10	92%
KM_{M1}	aDR	5	2	
	aAL	1	4	75%
KM_{M2}	aDR	6	1	
	aAL	1	5	85%

Algorithms were trained and tested on the same datasets. Models KM_A and KM_B were built with single male cohorts A and B, respectively. KM_{AB} was built with combined samples of male cohorts A and B. KM_{M1} and KM_{M2} were built with mixed cohort male sample sets, Mix 1 and Mix 2, respectively. pDR and pAL were predicted DR and AL classes, respectively. aDR and aAL were actual DR and AL classes, respectively. Accuracy = 100* (total correct/total samples).

parable in magnitude to those found between AL and DR. This is a weakness resulting in part from the distance-based (scalar) nature of KNN and in part from "over-fitting." Accuracy based on mixed cohorts was 89 ± 5%, indicating both that real AL/DR differences exist and that over-fitting can be addressed directly and readily using broader training sets (as expected).

SIMCA analysis

Unlike KNN, which is based on the distances between pairs of samples, SIMCA develops principal component models for each training set category. The independent variable set (*x*-block, the serum metabolite concentrations in our study) of each member of the test set is then projected into the principal component space of each class, and the unknown is assigned to the class it best fits. Unknowns insufficiently close to the principal component space of a class are considered non-members. While KNN assigns every unknown

TABLE 2. KNN CLASSIFICATION OF FEMALE SAMPLES: TRAINING SET STUDIES

Models		*pAL*	*pDR*	*Accuracy*
KF_A	aAL	6	0	
	aDR	1	4	91%
KF_B	aDR	8	0	
	aAL	0	8	100%
KF_C	aDR	8	0	
	aAL	0	8	100%
KF_{ABC}	aDR	22	0	
	aAL	2	19	95%
KF_{M1}	aDR	9	2	
	aAL	2	9	82%
KF_{M2}	aDR	11	0	
	aAL	1	9	95%

Algorithms were trained and tested on the same datasets. KF_A, KF_B, and KF_C were built with single female cohorts A, B. and C, respectively. KF_{ABC} was built with combined samples of female cohorts A, B, and C. KF_{M1} and KF_{M2} were built with mixed cohort female sample sets, Mix 1 and Mix 2, respectively. pDR and pAL were predicted DR and AL classes, respectively. aDR and aAL were actual DR and AL classes, respectively. Accuracy = 100* (total correct/total samples).

TABLE 3. KNN CLASSIFICATION OF MALE SAMPLES: TEST SET STUDIES

Model → unknown	*Accuracy*
$KM_B \rightarrow$ cohort A	62%
$KM_A \rightarrow$ cohort B	58%
$KM_{M2} \rightarrow$ Mix 1	83%
$KM_{M1} \rightarrow$ Mix 2	92%

Algorithms were trained on one dataset and tested on data from an independent dataset. KM_A and KM_B were trained on data from single male cohorts A and B, respectively. KM_{M1} and KM_{M2} were trained on mixed cohort male sample sets, Mix 1 and Mix 2, respectively. Accuracy = 100* (total correct in the test set/total samples in test set).

sample to exactly one pre-defined category, SIMCA assignments have three possible outcomes: (1) the sample fits only one pre-defined category; (2) the sample does not fit any pre-defined categories; or (3) the sample fits into more than one pre-defined categories.

As with KNN, SIMCA models were initially constructed with single cohort sample sets and combined cohort sample sets. SIMCA correctly assigned all members of each training set cohort with 100% accuracy (not shown). Analysis of the residuals between the samples and the models of each class confirmed that DR and AL class models of both female and male samples were well-fitted to themselves and well-separated from each other.

SIMCA had considerably lower accuracy predicting across cohorts, but displayed satisfactory performance in mixed cohort prediction (Tables 5 and 6). This again indicates that models built with single cohorts are cohort specific; that is, they predict well with intra-cohort samples, but are not usable to predict

TABLE 4. KNN CLASSIFICATION OF FEMALE SAMPLES: TEST SET STUDIES

Model → unknown	*Accuracy*
$KF_A \rightarrow$ cohort B	75%
$\rightarrow$ cohort C	50%
$KF_B \rightarrow$ cohort A	55%
$\rightarrow$ cohort C	56%
$KF_C \rightarrow$ cohort A	55%
$\rightarrow$ cohort B	63%
$KF_{M1} \rightarrow$ Mix 2	86%
$KF_{M2} \rightarrow$ Mix 1	95%

Algorithms were trained on one dataset and tested on data from an independent dataset. KF_A, KF_B, nd KF_C were trained on data from single female cohorts A, B, and C, respectively. KF_{M1} and KF_{M2} were trained on mixed cohort female sample sets, Mix 1 and Mix 2, respectively. Accuracy = 100* (total correct in the test set/total samples in test set).

samples in other cohorts (accuracy was 40 ± 40%), suggesting that models built with single cohorts are over-fitted. Trained with mixed cohort sample sets Mix 1, Mix 2, Mix 3 and Mix 4, SIMCA built models SM_{M1}, SM_{M2}, SF_{M1} and SF_{M2}, respectively. The four models based on mixed cohort training sets again classified training samples to actual dietary categories with 100% accuracy (not shown). Interclass residuals showed that DR and AL classes in the models were well-fitted to themselves and well-separated from each other (Tables 5 and 6). The four models predicted test sets at 89 ± 8% accuracy, without requiring adjustments in model parameters after data collection.

The requirement for mixed training sets suggests that either the algorithms are mathematically overfitting the training set data or that cohort-specific effects might predominate over those reflecting caloric intake. The latter possibility was unexpected because the DR paradigm is dominant across genetically diverse animals, and yet our cohorts show metabolomic differences despite being composed of genetically identical F1 rats. The DR paradigm is also dominant over changes in dietary constituents, and yet our cohorts show metabolomic differences despite being fed essentially identical diets (although by definition there will always be some chemical variations between multiple lots of any non-synthetic diet). Thus, our working hypothesis is that the algorithms were over-fitting the training set data. We therefore more carefully examined the inter-cohort variability in both male and female rats. PCA was used to provide a mathematical description of the metabolomes of these rats.

The cohort specificity of our models was examined in both male and female rats by PCA. As shown in Figure 3, both males and females cluster more tightly by cohort than by diet. This result indicates the need for developing models capable of overcoming whatever biological or analytical confound relates to cohort specificity, a requirement that was successfully addressed in the accompanying manuscripts (Paolucci et al., 2004a,b).

The last report addresses the possibility that the observed cohort to cohort variability results from changes in the concentration of compounds across the metabolome or changes in a defined subset of metabolites drawn from those that comprise the metabolome (Paolucci et al., 2004b). An alternative possibility, which appears more consistent with the data presented, is that DR induces a systematic change in the metabolome whose appearance is reflected differently in animals from different cohorts (Paolucci et al., 2004b). Such

TABLE 5. SIMCA CLASSIFICATION OF MALE SAMPLES: TEST SET STUDIES

Model → unknown		*pDR*	*pAL*	*No match*	*Accuracy*
SM_B → cohort A	Factors	4	4		
	aDR	6	0	2	
	aAL	3	2	0	62%
	Unmodeled	0	0	0	
SM_A → cohort B	Factors	3	1		
	aDR	0	0	6	
	aAL	0	1	5	8%
	Unmodeled	0	0	0	
SM_{M1} → Mix 2	Factors	1	1		
	aDR	5	2	0	
	aAL	0	5	0	83%
	Unmodeled	0	0	0	
SM_{M2} → Mix 1	Factors	2	1		
	aDR	6	0	1	
	aAL	1	5	0	85%
	Unmodeled	0	0	0	

Algorithms were trained on one dataset and tested on data from an independent dataset. "Factors" indicates the number of principal components in the models. SM_A and SM_B were built with single male cohorts A and B, respectively. SM_{M1} and SM_{M2} were built with mixed cohort sample Mix 1 and Mix 2, respectively. pDR and pAL were predicted DR and AL classes, respectively. aDR and aAL were actual DR and AL classes. Accuracy = 100* (total correct/total samples).

Table 6. SIMCA Classification of Female Samples: Test Set Studies

Model → unknown		*pAL*	*pDR*	*No match*	*Accuracy*
SF_A → cohort B → cohort C	Factors	4	3		
	aAL	7	1	0	
	aDR	0	8	0	94%
	Unmodeled	0	0	0	
	Factors	4	3		
	aAL	7	1	0	
	aDR	0	8	0	94%
	Unmodeled	0	0	0	
SF_B → cohort A → cohort C	Factors	2	2		
	aAL	5	0	1	
	aDR	5	0	0	45%
	Unmodeled	0	0	0	
	Factors	2	2		
	aAL	0	0	8	
	aDR	0	0	8	0%
	Unmodeled	0	0	0	
SF_C → cohort A → cohort B	Factors	2	2		
	aAL	0	0	6	
	aDR	0	0	5	0%
	Unmodeled	0	0	0	
	Factors	2	2		
	aAL	0	0	8	
	aDR	2	0	6	13%
	Unmodeled	0	0	0	
SF_{M1} → Mix 2	Factors	5	4		
	aAL	11	0	0	
	aDR	3	8	0	86%
	Unmodeled	0	0	0	
SF_{M2} → Mix 1	Factors	5	5		
	aAL	11	0	0	
	aDR	0	10	0	100%
	Unmodeled	0	0	0	

Algorithms were trained on one dataset and tested on data from an independent dataset. "Factors" indicates the number of principal components in the models. SF_A, SF_B and SF_C were built with single female cohorts A, B, and C, respectively. SF_{M1} and SF_{M2} were built with mixed cohort female samples Mix 1 and Mix 2, respectively. pDR and pAL were predicted DR and AL classes, respectively. aDR and aAL were actual DR and AL classes. Factors were the number of principal components in the models. Accuracy = 100* (total correct/total samples).

changes are more complex, and require initial work to first eliminate the more basic potential confounds, such as magnitude changes.

We therefore move systematically forward by directly addressing what might be termed monotonic changes in the metabolome, such as an overall magnitude shift—that is, a change in the concentration of all metabolites in one cohort as compared to another. Magnitude changes could be analytical in nature (i.e., reduced sensitivity in the electrode sensors), but this seems unlikely as many of the samples from the two cohorts were run consecutively. Magnitude changes could also result from sample treatment issues (i.e., longer storage time in the freezer for the samples from one cohort), but also this seems unlikely as we have unpublished data that demonstrate that most metabolites in the profiles are stable over years at −70°C. Magnitude changes could also be caused by environmental variation (e.g., a factor that changes hydration might alter overall solute concentration, although this seems unlikely). Changes that occur only in a defined subset of metabolites would be apparent if the following criterion were met: if one were to remove

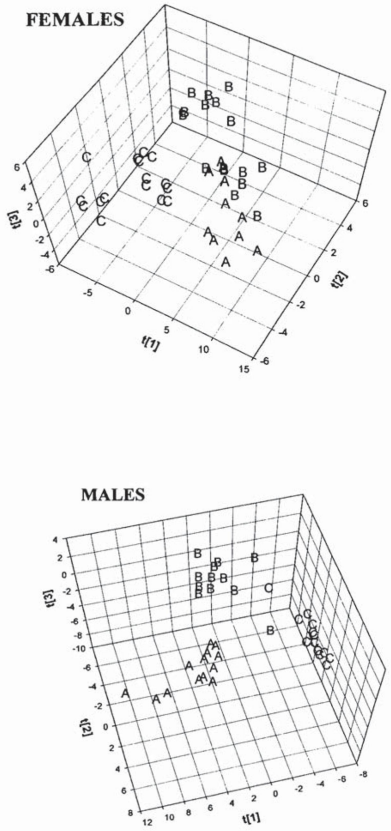

FIG. 3. Principal components analysis of male and female rats can distinguish cohort of origin. PCA plots based on 63 variables in females and 66 variables in males, rotated to show cohort distinctions. Letter labels refer to cohort of origin.

from consideration those metabolites that no longer differ between AL and DR statistically, then the remaining metabolites should largely follow the previously defined profile. These changes must be addressed by developing combined/common datasets comprised of data from multiple cohorts, a dataset that will be presented for the first time in the accompanying paper (Paolucci et al., 2004a). Generation of a single unified dataset from our current biologically and analytically distinct individual sets is primarily a problem in identification of the proper approaches to normalize our data across cohorts. Likewise, as we initially built

models in both sexes independently, we now need to determine if the metabolite profiles that separate DR and AL males also separate females into these two classes (and vice-versa). The underlying, unstated hypothesis behind this manipulation is consistent with the observation that DR extends longevity and decreases morbidity equally well in both sexes. Thus, the following report combines disparate datasets to address the general nature of the DR-mediated change in the metabolome (Paolucci et al., 2004a). The cohort-to-cohort changes identified here by SIMCA also point to the potential requirement for a modified informatics strategy.

SIMCA works, and works well, by defining the overall characteristics of a class (or, more specifically, a series of classes within a single experiment). Thus, SIMCA is at its strongest under conditions in which class descriptors are comprised of variables displaying little change across the total membership of the class in question. SIMCA is weakest when class descriptors are comprised of variables that drift with respect to multiple examples of a class. Thus, if the inter-cohort variability results from a scaling-type problem (e.g., differences in overall magnitude of variables between cohorts) or an inclusion-exclusion type problem (i.e., certain variables need to be discarded), SIMCA will, based on the evidence presented here, be an appropriate classification algorithm. In contrast, if the problem results from disproportionate changes across the metabolome, SIMCA will need to be replaced with an alternative, but similar approach. This issue will be discussed further in the accompanying reports (Paolucci et al., 2004a,b).

In summary, both KNN analysis and SIMCA analysis suggested that previously defined metabolites encode sufficient information to enable construction of classification models/expert systems with the potential to define biomarkers for future studies of cancer risk. Models built with single cohorts lacked power to classify samples other than samples from the same cohort, suggesting that these models are over-fitted and thus cohort specific, but models built from mixed cohort sample sets had reasonable accuracy. These data suggest that future studies will require careful attention to over-fitting concerns. The overall accuracy of SIMCA in test sets, the increased information available from SIMCA on the components of a given class, and the requirement of KNN for post-hoc optimization of k all lead us to select a component-based approach for future studies. While SIMCA was generally successful in this study, the cohort specificity issues discussed previously suggest the need to generate single, integrated databases, determine whether single profiles can be used for males and females, and consider other approaches that may optimize separation using the same basic approach (components); but with a focus on the separation rather than the groups. These issues are addressed in the accompanying manuscripts (Paolucci et al., 2004a,b).

Given the well documented power of DR to cause significant changes in physiology, the evidence presented here that cohort-specific drift occurs in the metabolome, and that this drift is sufficient to obscure cross-cohort evaluations, suggests that this cohort-specificity effect will be a common issue that must be addressed in any nutrition-metabolomics study, and very likely in other biomarker studies as well. These observations do not, however, repudiate our claim that metabolome studies yield useful information for predictive model building. On the contrary, variability resultant from factors other than dietary intervention is expected and exploratory statistical analysis serves in part to identify and account for this variability. Furthermore, the fact that no individual metabolite shows great variation in concentration between classes exemplifies the complex interdependence of metabolome components. This finding also demonstrates how metabolic serotypes as a whole better reflect physiologic modulation resulting from DR than do alterations in specific metabolites. This issue, which we believe has broad relevance for biomarker development for pre-disease status, is further discussed in the accompanying manuscripts.

ACKNOWLEDGMENTS

We thank Dr. Tom Vogl for his many critical discussions and comments on the manuscript. This work was supported by NIH NIA R01-AG15354 (B.S.K.), ESA, Inc., and the Winifred Masterson Burke Relief Foundation.

REFERENCES

COVER, T., and HART P. (1967). Nearest neighbor pattern classification. IEEE Trans Information Theory **13,** 21–27.

KRISTAL, B.S., and YU, B.P. (1994). Aging and its modulation by dietary restriction. In *Modulation of Aging Processes by Dietary Restriction*, 1st ed. B.P. Yu, eds. (Boca Raton, FL, CRC Press), pp. 1–36.

MATSON, W.R., LANGIALS, P., VOLICER, L., et al. (1984). n-Electrode three dimensional liquid chromatography with electrochemical detection for determination of neurotransmitters. Clin Chem **30,** 1477–1488.

McCAY, C.M., CROWELL, M.F., and MAYNARD, L.A. (1935). The effect of retarded growth upon the length of lifespan and upon the ultimate body size. J Nutr **10,** 63–79.

MILBURY, P.E. (1997). CEAS generation of large multiparameter databases for determining categorical process involvement of biomolecules. In *Coulometric Array Detectors for HPLC* (Utrecht, VSP International Science Publication), pp. 125–141.

PAOLUCCI, U., VIGNEAU-CALLAHAN, K.E., SHI, H., et al. (2004a). Development of biomarkers based on diet-dependent metabolic serotypes: concerns and approaches for cohort and gender issues in serum metabolome studies. OMICS **8,** 209–220.

PAOLUCCI, U., VIGNEAU-CALLAHAN, K.E., SHI, H., et al. (2004b). Development of biomarkers based on diet-dependent metabolic serotypes: characteristics of component-based models of metabolic serotypes. OMICS **8,** 221–238.

SHI, H., VIGNEAU-CALLAHAN, K.E., SHESTOPALOV, A.I., et al. (2002a). 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., et al. (2002b). 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., et al. (2001). Characterization of diet-dependent metabolic serotypes: analytical and biological variability issues in rats. J Nutr 924S–932S.

WEINDRUCH, R., & WALFORD, R. (1988). *The Retardation of Aging and Disease by Dietary Restriction* (St. Louis, Charles C. Thomas).

WILLETT, W.C., DIETZ, W.H., and COLDITZ, G.A. (1999). Guidelines for healthy weight. N Engl J Med **341,** 427–434.

WOLD, S. (1976). Pattern recognition by means of disjoint principal components models. Pattern Recognition **8,** 127–139.

Address reprint requests to:
Dr. Bruce S. Kristal
Dementia Research Service
Burke Medical Research Institute
785 Mamaroneck Ave.
White Plains, NY 10605

E-mail: Bkristal@burke.org