Validated Biomarkers of Caloric Restriction in Rats: Markers of Disease Risk in Humans?

Bruce S. Kristal, PhD

Dept of Neurosurgery, Brigham and Women’s Hospital
Dept of Surgery, Harvard Medical School

Industry Affiliations (Disclosure):
ESA (collaborative, inactive); Mead Johnson (Consultant)
Metabolon: Consultant/Equity, (IP agreements via/Cornell Med)
“Caloric Intake” and Disease

Humans:
Increased BMI associated with increased risk of neoplasia, type II diabetes, cardio- and cerebro-vascular disease ...

Laboratory Rodents:
Low calorie diets increase longevity and delay morbidity
Caloric Restriction (CR)

CR is an experimental paradigm in which the dietary/caloric intake of a group of animals is reduced relative to that eaten by *ad libitum* fed controls.
Caloric restriction is the most potent, most robust, and most reproducible known means of reducing morbidity and mortality in mammals.
How do we study complex biological/clinical problems?

How do we address such questions in humans, where our ability to manipulate and analyze the system is limited?
High Throughput and/or Data Density Studies

- Genomics/SNPs
- mRNA expression arrays
- Proteomics
- Small metabolites
Metabolomics:
The -omics face of biochemistry

Measurement of changes in populations of low molecular weight metabolites under a given set of conditions

Fiehn
What we measure -- biochemically

Metabolites – small molecules

Pathways (eg, purine catabolites)

Interactive pathways (eg, amino acid metabolism)

Compound classes (eg, lipids)

Conceptually linked systems
  eg antioxidants, redox damage products
What we measure -- conceptually

Biochemical constituents
Excretion products
Precursor – product
Balances (eg, redox systems)
“collection depots”
Flux
Snapshot view of biochemistry
Integrated signal from genome and environment
Short and long term status
Temporal image
Sub-threshold changes (eg (toxicology, nutrition))
Metabolomics – Some Advantages

Sensitivity

“silent phenotypes”/sub-threshold effects

Discovery

Knowledge base (ie, metabolic pathways)

Limited repertoire – simplifies possibilities

(2500 non-lipid endogenous metabolites??)

Metabolome integrates signal

Nature and Nurture -- genome and environment

Measurement of system status/defects

Metabolome has the fastest response time
Metabolomics – Some Disadvantages

Too Sensitive?

cohort effects, site effects, time effects
sample handling
individual metabolites responsive to multiple factors
genes, environment, health status, location
experiment design must account for all factors
controlled or fuzzy, multiple sources

Practical

Set-up costs
Possible need for multiple platforms (NMR, MS, HPLC)
eyearly industry dominance – lots of propriety data
incompatible data standards
Data Validation, Data Normalization, Missing Data Decisions, Inclusion/Exclusion Criteria

Subgroups, Class-specific models

Outlier removal ↔ scaling ↔ transformations

Unsupervised: Clustering SOMs PCA
Supervised: kNN SIMCA PLS PLS-DA Random Forest
Machine learning: Neural Nets GAs GPs

Overfit tests, Internal validation, optimization, External validation, optimization, 2º validation
Survival Data, 1987 Cohort, Casein Diet

Survival Rate vs. Days

- AL
- DR
Hypothesis:

Long-term, low-calorie diets induce changes in metabolism that persist throughout the lifespan.
Predictions

- CR alters the sera “metabolome”
- There exists a “CR Serotype”
- …Part of “CR serotype” reflects beneficial physiological status --- ie, serotype defines health without reference to disease…
Goals

1) Insights into the mechanism of CR

2) Recognize CR in other organisms
   (e.g., non-human primates)

3) Biochemically determine the effective, long-term caloric intake of an individual
   (e.g., for epidemiological studies)

4) Identify predictive markers of disease
   (e.g., to intervene/prevent/focus resources;
    focus on diseases where intervention is possible)
Experimental Design

AL vs DR

Analytical Issues

Biological Issues
Experimental Design

Model: F344 x BN F₁ Rat

Overall Design:
- AL/CR, male/female, 5 different ages
- Different extents and duration of diets
- Total experiment ~36 groups, 82 cohorts.

Approach:
- HPLC separations with coulometric array detection
- (LC/LC-MS for plasma proteomics)
- Multilayer statistical and data analysis
Experimental Design

AL vs DR

Analytical Issues

Biological Issues
Coularray

- HPLC separations coupled with coulometric array detectors
  - Sensitivity to femtomole levels of analyte
  - Resolution of co-eluting peaks
  - Qualitative characterization of peaks
    - Biochemical identity
    - Purity
HPLC-EC on Pooled Rat Sera

1075 analytically detectable peaks
sensitivity ~300 pA = ~10 fmole/125 μl sera
Analytical Stability/Biological Variability
Females—Cohort A + B – Rat sera looks like rat sera
Biologic Variability – Region of Stability

Females, channel 9, A+B Cohorts

Response (nA)

Retention time (minutes)
Biologic Variability – Region of Variability

Females, channel 4, A+B Chorts
Analytical vs Biological Variation

In Rats:
Biological variability 5 fold greater than analytical variability
Analytical variability does not influence biological variability
Primary Data Analysis

• Multivariate analyses are relatively noise-resistant
• Minimize loss of informative metabolites
  • Reduce false negatives (Type II errors)
  • Increase false positives (Type I errors)
Experimental Design

AL vs DR

Analytical Issues

Biological Issues
Does Serotype Encode Sufficient Information to Identify Diet Group?
Data Exploration and Classification Analysis

- Hierarchical Cluster Analysis (HCA)
  - Identifies natural groups in data

- Principal Component Analysis (PCA)
  - Finds linear combinations of original variables that account for maximal variation
Model Feature Selection

**Biological Model**

- **Survival Rate**
  - AL: Yellow line
  - DR: Orange line

- **Retention time (minutes)**
  - Y-axis: Response (µA)

- **T-tests, p<0.2 ?!**

- **1075 analytically detectable peaks**
- Sensitivity ~300 pA = ~10 fmole/125 µl sera

**HCA Distinguishes Female AL and DR Rats**

- **Autoscale**
  - Single: 100%
  - Complete: 100%
  - Centroid: 100%

- **Range Scale**
  - 100%

**PCA Distinguishes AL and DR Female Rats**

- **Autoscale**
  - **Range Scale**

63 variables (confirmed), Non-independent samples
**HCA Validation**

Complete

94% Accuracy

**PCA**

PCA Distinguishes AL and DR Female Rats

63 variables (confirmed), independent female Cohort #2

**HCA Simplify Model**

Removed variables can contribute to AL-DR separations

Autoscale, complete

91% Accuracy

**PCA**

PCA Distinguishes AL/DR Using Either Dataset

63 Variables

36 Variables

63/(37/36) variables (confirmed), independent female Cohort #3

37 variables (confirmed), independent female Cohort #3
Status

Proof of principle accuracy:
- HCA (100%)
- PCA (100%)

Validation Accuracy:
- HCA (94%)
- PCA (100%) - subjective rotation

Simplification –
- HCA (Fails)
- PCA (100% Accuracy)

Use larger models?
Test components vs distance
“Expert Systems/Supervised Analysis”

**KNN**
- k-nearest neighbor analysis
- Supervised HCA (HCA is KNN with K=1)
- Distance-based metric
- Strength is with small (training) datasets

**SIMCA**
- Soft Independent Modeling of Class Analogy
- Supervised PCA
- Component-based metric
- Strength is modeling flexibility (eg, group-specific interactions)
In our DR sera metabolomics data – components greatly outperform distance-based algorithms
In OUR DR SERA METABOLOMICS data – components greatly outperform distance-based algorithms
Profiles are cohort specific
male samples modeled with male/female data set

Cohort Separations

- AMAL
- AMDR
- BMAL
- BMDR
- CMAL
- CMDR

Cohort Separations
Cohort Effects

Cohort Separations

- AMAL
- AMDR
- BMAL
- BMDR
- CMA
- CMAL
- CMDR

male samples modeled with male/female data set

PLS-DA

p<0.001
Unpublished Rat Work – Proof of Concept

Fasting

Aging

Duration

Extent
Markers “Predict” Caloric Intake with High Quantitative Accuracy -- Proof of Concept --

Actual Degree of Restriction

Predicted Degree Restriction

\( r^2 = 0.877; \)
\( r^2 = 0.994 \) (means)

\( N = 90 \)

Each group \( p < 0.05 \) vs others
O-PLS models built for better testing

female basic model. M1 (OPLS)
{Comp. 1}/t[Comp. 2]
Colored according to value in variable female basic model (Group Y new)

R²X[1] = 0.105822            R²X[2] = 0.162921

Series (Variable Group Y new)

- 1 - 1.5
- 1.5 - 2

Ellipse: Hotelling T² (0.95)
Experimental Design

AL vs DR

Analytical Issues

Biological Issues
In Rats:

- Biological variability 5 fold greater than analytical variability
- Analytical variability does not influence biological variability
Human Studies:
Analytical Controls
Duplicates (triplicates)
## Analytical Parameters

Measures and medians (Key points highlighted)

<table>
<thead>
<tr>
<th></th>
<th>Pre-polishing data</th>
<th>Post-polishing data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range, median CVs for 66 variables</td>
<td>2-54%</td>
<td>2-32%</td>
</tr>
<tr>
<td>Range, mean CVs for 66 variables</td>
<td>2-58%</td>
<td>3-38%</td>
</tr>
<tr>
<td>Overall mean CV</td>
<td>19%</td>
<td>16%</td>
</tr>
<tr>
<td>Overall median CV</td>
<td>11%</td>
<td>12%</td>
</tr>
<tr>
<td>By 13 pairs, overall mean CV</td>
<td>19%</td>
<td>16%</td>
</tr>
<tr>
<td>By 13 Pairs, overall median CV</td>
<td>9%</td>
<td>9%</td>
</tr>
</tbody>
</table>

3 variables dropped from further study
## Analytical Parameters

### Means and medians (Key points highlighted)

<table>
<thead>
<tr>
<th></th>
<th>Pre-polishing data</th>
<th>Post-polishing data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range, median CVs for 66 variables</td>
<td>2-54%</td>
<td>2-32%</td>
</tr>
<tr>
<td>Range, mean CVs for 66 variables</td>
<td>2-58%</td>
<td>3-38%</td>
</tr>
<tr>
<td>Overall mean CV</td>
<td>19%</td>
<td>16%</td>
</tr>
<tr>
<td>Overall median CV</td>
<td>11%</td>
<td>12%</td>
</tr>
<tr>
<td>By 13 pairs, overall mean CV</td>
<td>19%</td>
<td>16%</td>
</tr>
<tr>
<td>By 13 Pairs, overall median CV</td>
<td>9%</td>
<td>9%</td>
</tr>
</tbody>
</table>

### Additional study:

- Biological CV approximately 82-86%, 86%.
- Analytical CV 13-16%, (N=69).
- Signal:Noise is approximately 5:1
Variability at 24 and 48 hours
47 of 61 are stable to 48 hours
2 dropped in this study
<table>
<thead>
<tr>
<th>Analyte #</th>
<th>Percent Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>1</td>
<td>75</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>125</td>
</tr>
<tr>
<td>4</td>
<td>150</td>
</tr>
<tr>
<td>5</td>
<td>175</td>
</tr>
<tr>
<td>6</td>
<td>200</td>
</tr>
</tbody>
</table>

Processing Test

24 Hours
48 Hours
<table>
<thead>
<tr>
<th>DONOR #</th>
<th>ID processed at 0hr</th>
<th>ID processed at 24hr</th>
<th>ID processed at 48hr</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>001</td>
<td>002</td>
<td>003</td>
</tr>
<tr>
<td>21</td>
<td>017</td>
<td>022</td>
<td>023</td>
</tr>
<tr>
<td>28</td>
<td>006</td>
<td>008</td>
<td>031</td>
</tr>
<tr>
<td>35</td>
<td>012</td>
<td>013</td>
<td>014</td>
</tr>
<tr>
<td>42</td>
<td>025</td>
<td>018</td>
<td>024</td>
</tr>
<tr>
<td>49</td>
<td>007</td>
<td>033</td>
<td>004</td>
</tr>
<tr>
<td>56</td>
<td>021</td>
<td>026</td>
<td>032</td>
</tr>
<tr>
<td>63</td>
<td>030</td>
<td>005</td>
<td>015</td>
</tr>
<tr>
<td>70*</td>
<td>034</td>
<td>027</td>
<td>019</td>
</tr>
<tr>
<td>77</td>
<td>020</td>
<td>016</td>
<td>009</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NHS QC's</th>
<th>ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Postmenopausal</td>
<td>010</td>
</tr>
<tr>
<td>Postmenopausal</td>
<td>011</td>
</tr>
<tr>
<td>Premenopausal</td>
<td>028</td>
</tr>
<tr>
<td>Premenopausal</td>
<td>029</td>
</tr>
</tbody>
</table>

**NHS QC's ID**
- Postmenopausal: 010
- Postmenopausal: 011
- Premenopausal: 028
- Premenopausal: 029

<table>
<thead>
<tr>
<th>ID</th>
<th>All Three Correct (two for Pools)</th>
<th>Two Correct</th>
<th>Not Assigned</th>
<th>Incorrectly assigned</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Other decent matches -- are there more than triplets

**34 samples Total**
- 29 correctly grouped
- 4 not grouped* two marked as possible
- 1 incorrectly assigned (r2=>0.6 to group assigned)

* Other decent matches -- are there more than triplets
Markers have high utility – even in worst case scenarios

Key point: Under worst case
all unstable metabolites included
48 hour delay
No model optimization
85% Origin ID’d correctly; only 3 % absolute error
Markers have high utility – even in worst case scenarios

Key point: Under worst case, all unstable metabolites included
48 hour delay
No model optimization
85% Origin ID’d correctly; only 3 % absolute error

Fix this, 100% accuracy
Biological Variability Affects Stability

A

24 Hour

Sample A at 48 hours (relative to control)

Sample C at 48 hours (relative to control)

Sample D at 48 hours (relative to control)

B

48 Hour

Concentration Relative to Control
Biological Variability impacts Analytical Parameters
Human Studies: Inter- vs Intra-Person Variability

- Test only best rat profile
  - 30 variables
    - Knowns (uric acid, tryptophan, tryptophol, 5-hydroxytryptophan, norepinephrine, methionine, 4-hydroxyphenyllactic acid)
  - 69 Triplets scored

- Single Profile Tested
  - (ie, true test; not re-fitted or optimized)

- ICC = 0.76
Human Studies:
Profile = ...

- Recent Food (ie, Fast)
- BMI
- Food intake
- Physiology
Human Studies: Profile = ...

- Recent Food (ie, Fast)
- BMI
- Food intake
- Physiology
What is coming (funded)

Equivalent proteomics models (NIA)

Studies in CALERIE (NIA)
(0-3-6-12-18-24 mos pre/post diet)
Some validation against double-labeled water

750 case/control pairs for breast cancer (NCI) – NHS, samples taken 2-10 years before onset

1000 case/control pairs for Type II diabetes (NIA) – NHS, samples taken 2-10 years before onset
What is coming (funded)

NIH Genes/Environment Initiative (NIEHS)
Exposure Biology Program
Biological Response indicators
- Non-fat based models for primary fat in diet
- Models for glycemic index
- Models for 24 combinations of fat/GI
- Cross of all models into breast cancer
- Cross of all models into Type II diabetes
- Ties of all models to liver mitochondrial function
Summary

Created and validated a working model of the CR serotype in both male and female rats

Can identify group of origin with high accuracy
Wide capture of metabolome
Reduced dimensionality of the problem
Identify critical markers

Profiles distinguish diet

Fasted vs AL and CR
Across lifespan
across duration and extent of diet ($r^2 = 0.88$)
Summary

Markers are passing analytical tests in human plasma

- Can identify split duplicates (100% accuracy)
- Good analytical parameters (5:1 signal/noise)
- Handles worst case shipping conditions
- Good inter/intra-individual variability
The metabolomic markers and profiles identified appear analytically and biologically suitable for studies in defined human populations such as national clinical trials and epidemiological cohorts.
Acknowledgements

• Brigham and Women’s and Burke/Cornell

• Outside Collaborators
  – Walter Willett, Sue Hankinson, Frank Hu, Paul Vouros
  – Wayne Matson, Karen Vigneau-Callahan, Paul Milbury
  – Tom Vogl

• Funding
  – NIH (NIA; NCI; NIEHS)
  – Winifred Masterson Burke Relief Foundation
  – Brigham and Women’s Hospital