A Strategy for Analyzing Population-Based Data to Inform Public Health Decisions on the Individual Level

18 April 2012

Nestlé Institute of Health Sciences
Jim Kaput PhD
Head, Clinical Unit
James.Kaput@RD.Nestle.com
Facts and Challenges

20th Century Design
Diets
Physiology
Genetics

Strategies & Analyses
Model Systems
Oomics
21st Century Design
Middle-Out, a new strategy

Summary
## Facts & Challenges 20th Century Science

A, B, C = variants of one gene or A, B, C = variants of many genes

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Pheno</th>
<th>Pheno</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>8</td>
<td>3</td>
</tr>
</tbody>
</table>

\[ \sum \text{Phenotype} / 16 = \text{Average Phenotype} \]

\[ \sum \text{Phenotype} / 14 = \text{Average Phenotype} \]
Facts & Challenges  20th Century Science

Why a distribution of health or needs within a population?

Why a distribution within cases or requirements?

Is risk/intake as calculated for population useful for the individual?

What path to knowledge?

http://science.cancerresearchuk.org/cri/research/population_studies/?version=2
Facts & Challenges  Complexity, Heterogeneity, Diversity

**Nutritional**
- Composition of agri-foods varies
- Culture & food preparation
- Food processing

**Health & Disease**
- Variable pathways to each

**Genetic**
- Humans are the same but different
- History & culture alter populations

Nestlé Institute of Health Sciences
## Facts & Challenges

**Variable pathways to disease**

<table>
<thead>
<tr>
<th>T2DM Treatments</th>
<th>Drug</th>
<th>Target</th>
<th>Effective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pharmacologic intervention required if glycemic control not achieved with diet and exercise within 3 months</td>
<td>Sulfonylurea</td>
<td>T2DM &lt; 5yr</td>
<td>~50% ²</td>
</tr>
<tr>
<td></td>
<td>Meglitinides</td>
<td>T2DM &lt; 5yr &amp; Elevated PPG</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>Biguanide</td>
<td>Obese insulin resistant</td>
<td>~75%</td>
</tr>
<tr>
<td></td>
<td>α-Glucosidase</td>
<td>Elevated postprandial glucose (PPG)</td>
<td>2nd line</td>
</tr>
<tr>
<td></td>
<td>Thiazolidinediones</td>
<td>Obese insulin resistant</td>
<td>2nd line</td>
</tr>
</tbody>
</table>

- ~85% require interventions
- Pancreas, Liver, Intestine, Adipose
- ~8% T2DM in U.S

1. [http://www.aafp.org/PreBuilt/monograph_diabetestreatment.pdf](http://www.aafp.org/PreBuilt/monograph_diabetestreatment.pdf)
2. ~10% failure/yr

Facts & Challenges Genetic Diversity

1000 Genomes: 300 – 400 variants affecting 250 – 300 genes resulting in loss of function (LOF) per person


Lu et al EJHG 17, 967 (2009)
Facts & Challenges  20th Century Logic

- **Insanity**
  - Doing the same things over and over and expecting different results

**Albert Einstein**

**Distribution of risk**

**Proportion of population at specific risk**

**Human genetic, nutritional, physiological variations**
Facts and Challenges

20th Century Design
- Diets
- Physiology
- Genetics

Strategies & Analyses
- Model Systems
- Omics

21st Century Design
- Middle-Out, a new strategy

Summary
Facts & Challenges  Research Models

Assessment of research models for testing gene-environment interactions

Amy L. Inselman a,*, Deborah K. Hansen a, Hyung-yul Lee a, Noriko Nakamura a, Bai Tang Ning a, Jacqueline Pontes Monteiro b, Vijayalakshmi Varma a, Jim Kaput a

Q1

ARTICLE INFO

Abstract

Throughout the last century, possible effects of exposure to toxicants, nutrients or drugs were examined primarily by studies of groups or populations. Individual variation in responses was acknowledged but could not be analyzed due to lack of information or tools to analyze individual genetic make-up and lifestyle factors such as diet and activity. The Human Genome, Epidemiology, 1000 Genomes and Human Variome Projects are identifying and cataloging the variation found within humans. Advances in DNA sequencing technologies will soon permit the characterization of individual genomes in clinical and basic research studies, thus allowing associations to be made between an individual genotype and the response to a particular exposure. Such knowledge and tools have generated a significant challenge for scientists to design and conduct research studies that account for individual genetic variation. However, before these studies are done in 30 humans, they will be performed in various in vivo and in vitro models. The advantages and disadvantages of all of the model systems that are being used or developed in relation to individual genetic make-up and responses to xenobiotics are discussed.
Concept  Understanding Genotype x Environment

A different effect of a **genotype** on disease in persons with different **environmental** exposures

**Genotype X Environment Interactions**

A different effect of an **environmental** exposure on disease risk in persons with different **genotypes**


Statistical Parlance  The *main effect(s)* may be **genotype x environment interaction(s)** for chronic diseases and modifying effects
Model Systems Translational Research for Humans

Follow patients/subjects over time – evaluate

- Homeostatic assessments (clinical + omic)
- Diet and lifestyle assessments
- Genomic \(\text{(once)}\) & Epigenomic
- Changes in biomarkers due to medical or lifestyle interventions \(\text{(acute or long term)}\)

Associate changes/outcome in quantitative assessments in the context of individual genomes

Nestlé Institute of Health Sciences
Strategies  Human Study 1 – Homeostatic Challenges

Many biomarkers measure homeostasis

Homeostasis has large inter-individual variation

Perturb homeostasis, response measures “robustness”

**Oral glucose tolerance**

100g (or 75g) bolus – measure glucose in venous blood over time

Measurements during or after
Other serum metabolites
Immune function
Oxidative damage
Urine metabolites
# Model Systems: Homeostatic Challenges

**Challenge homeostatic systems**

- Functional challenge
- Nutrient challenge

**Examples**

- Oral glucose tolerance
- Lipid challenge
- Activity challenge
- Oxidative stress challenge
- OTC Drug challenge

**Dose, kinetics, and relevant physiological measures**

- Deep genotyping and deep phenotyping

**No reference population for health**

- Compare responses in differing genetic make-ups & cultures

---

**Define health and biomarkers**

Clinical Translation: The Concept

Classification Algorithms

1, 2, 3…n = Genes. A, B, C…n = Environment
Strategies Human Study 2 – CBPR Translational Research

Obesity Prevention Summer Day Camp

Community Based Participatory Research

With USDA – ARS in Little Rock

Track individuals’ nutrition, health, genetics, economics

Interventions to improve nutrition & health

McCabe-Sellers et al. Omics 12, 263 (2008)
Delta Vitamin Obesity Project

Metabolites in blood @ pre, end, post-intervention
Genomics, DNA methylation

Sequence micronutrient metabolism genes

Diet Intakes – 24 hr  Activity – Body Bugg  Skin tone – Dermometer

Correlate δ metabolite(s) to an individual’s genotype

In kids and parents

Quantitative assessment of efficacy per individual
Data Analyses Process – Traditional to Individual

**SAM/SAH Ratios**

<table>
<thead>
<tr>
<th>Metab</th>
<th>1*</th>
<th>2*</th>
<th>P value (between)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean 3</td>
<td>n = 10x</td>
<td>n = 52x</td>
<td></td>
</tr>
<tr>
<td>SAM (nm/ml)</td>
<td>1.44</td>
<td>0.77</td>
<td>&gt; 0.000</td>
</tr>
<tr>
<td>SAH (nm/ml)</td>
<td>0.88</td>
<td>0.96</td>
<td>0.022</td>
</tr>
<tr>
<td>M/H</td>
<td>1.81**</td>
<td>0.81**</td>
<td>&gt; 0.000</td>
</tr>
</tbody>
</table>

* Center cluster 1 to cluster 2 distance = 0.991
** Reference = 1.40

http://www.youtube.com/watch?v=2kYRvABhHAY
### Statistically Significant Differences in Nutrient Intakes between Subjects in SAM/SAH Clusters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>p value (Between clusters)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>(mean of 3 time pts)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Energy Intake/BW (kcal/kg)</strong></td>
<td>67.1 ± 29.5</td>
<td>46.0 ± 18.7</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>64.4 (22 – 112)</td>
<td>41.4 (16.0 – 102.9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n = 10</td>
<td>n = 51</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.50 ± 0.59</td>
<td>1.30 ± 0.40</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>1.4 (0.83 – 2.5)</td>
<td>1.3 (0.59 – 2.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n = 10</td>
<td>n = 53</td>
<td></td>
</tr>
<tr>
<td><strong>Thiamine Intake (mg/day)</strong></td>
<td>12.9 ± 3.4</td>
<td>10.9 ± 3.4</td>
<td>0.041</td>
</tr>
<tr>
<td></td>
<td>11.6 (9.5 – 19)</td>
<td>10.5 (5.4 – 23.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n = 10</td>
<td>n = 51</td>
<td></td>
</tr>
<tr>
<td><strong>β-carotene Intake (mg/day)</strong></td>
<td>1539 ± 2718</td>
<td>787 ± 1072</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td>456 (139 – 8676)</td>
<td>381 (30 – 4339)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n = 10</td>
<td>n = 51</td>
<td></td>
</tr>
<tr>
<td><strong>Iron Intake (mg/day)</strong></td>
<td>518 ± 340</td>
<td>420 ± 170</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>438(200 – 1328)</td>
<td>370 (125 - 885)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n = 10</td>
<td>n = 51</td>
<td></td>
</tr>
<tr>
<td><strong>Vit A Intake (RE) (mcg/day)</strong></td>
<td>456 (139 – 8676)</td>
<td>381 (30 – 4339)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10.9 ± 3.4</td>
<td>10.5 (5.4 – 23.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n = 51</td>
<td>n = 51</td>
<td></td>
</tr>
</tbody>
</table>

Nestlé Institute of Health Sciences
Statistically Significant Differences in Plasma and RBC Metabolites between Subjects in SAM/SAH Clusters

<table>
<thead>
<tr>
<th>Metabolites</th>
<th>Cluster 1 (n = 10)</th>
<th>Cluster 2 (n = 51)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAM/SAH ratio</td>
<td>1.81 ± 0.48</td>
<td>0.81 ± 0.28</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>1.61 (1.41 – 2.95)</td>
<td>0.79 (0.31 – 1.31)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(n = 10)</td>
<td>(n = 51)</td>
<td></td>
</tr>
<tr>
<td>SAM (nmol/ml)</td>
<td>1.44 ± 0.31</td>
<td>0.77 ± 0.28</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>1.34 (1.1 – 1.95)</td>
<td>0.73 (0.34 – 1.59)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(n = 10)</td>
<td>(n = 51)</td>
<td></td>
</tr>
<tr>
<td>SAH (nmol/ml)</td>
<td>0.88 ± 0.14</td>
<td>0.96 ± 0.11</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>0.87 (0.63 – 1.15)</td>
<td>0.95 (0.7 – 1.31)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(n = 10)</td>
<td>(n = 51)</td>
<td></td>
</tr>
<tr>
<td>Homocysteine (µmol/l)</td>
<td>11.1 ± 4.0</td>
<td>7.3 ± 2.8</td>
<td>0.053</td>
</tr>
<tr>
<td></td>
<td>11.9 (5.6 – 17.5)</td>
<td>6.2 (3.5 – 14)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(n = 10)</td>
<td>(n = 51)</td>
<td></td>
</tr>
<tr>
<td>Vitamin A (µg/dl)</td>
<td>98 ± 35</td>
<td>61 ± 27</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td>93 (36 – 156)</td>
<td>56 (11.5 - 134)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(n = 8)</td>
<td>(n = 42)</td>
<td></td>
</tr>
</tbody>
</table>
Data Analyses Process  Traditional to Individual

Use common statistical methods to place individual into high, medium, low groups

Group Average

Average or reference
Hierarchical Cluster Analyses of Plasma Metabolites & RBC SAM/SAH

Row = subject  
Column = metabolite  
Bar = mean of 3 time points displayed as tertile level for that metabolite
Middle Out Micronutrient Genomics System

Micronutrients

- Ascorbic
- B12 - Methionine
- Biotin
- Folate
- Niacin
- Nicotinamide
- Pyridoxal
- Pyridoxine
- Pantothenic Acid
- Riboflavin
- Thiamine
- Vit A Retinol
- Vit D2
- Vit D3
- Vitamin E
- Vitamin K2

Search Networks & Pathways of each micronutrient

Genes

- 428 ALL
- 286 measured
- 26 one carbon

Find SNPs on Illumina Array

All SNPs in Micronutrient System

Analyze SNPs in each subject

- 5,295 for ALL
- 3,870 for measured
- 353 for 1-carbon

Nestlé Institute of Health Sciences

GenoGO Build 6.2.24268
(All) Micronutrient Genomics SubSystem

Freq > 0.5%  
Homozygote = 0  
Heterozygote = 1  
Homozygote = 2

184 SNPS in 81 genes
(Metabolite) Micronutrient Genomics SubSystem

Freq > 0.5%
Homozygote = 0
Heterozygote = 1
Homozygote = 2

Start list = metabolite
Top SNP/gene
One sample/child

Each row is one subject

81 SNPS in 81 genes
## Nutrient Intake by Year for 2 Subjects

<table>
<thead>
<tr>
<th>Intake Measure</th>
<th>High SAM/SAH¹</th>
<th>Child 1</th>
<th>Child 2</th>
<th>Low SAM/SAH²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2009</td>
<td>2010</td>
<td>2009</td>
<td>2010</td>
</tr>
<tr>
<td>Kcal/kg</td>
<td>67.1</td>
<td>111.6 (49)</td>
<td>52.4 (46)</td>
<td>79.6 (65)</td>
</tr>
<tr>
<td>HEI Total grain</td>
<td>4.2</td>
<td>5.0</td>
<td>4.1</td>
<td>4.6</td>
</tr>
<tr>
<td>HEI Milk</td>
<td>7.3</td>
<td>3.5</td>
<td>6.3</td>
<td>6.3</td>
</tr>
<tr>
<td>HEI Saturated Fat</td>
<td>41</td>
<td>60</td>
<td>31</td>
<td>25</td>
</tr>
<tr>
<td>HEI SOFAA</td>
<td>4.9</td>
<td>3.0</td>
<td>4.3</td>
<td>3.7</td>
</tr>
<tr>
<td>HEI Protein</td>
<td>61.6</td>
<td>100.3 (34)</td>
<td>57.3 (34)</td>
<td>49.3 (19)</td>
</tr>
<tr>
<td>HEI Carbohydrate</td>
<td>250.1</td>
<td>394.0 (130)</td>
<td>264.9 (130)</td>
<td>221 (130)</td>
</tr>
<tr>
<td>HEI Saturated</td>
<td>123.3</td>
<td>164.7</td>
<td>169.4</td>
<td>95.8</td>
</tr>
<tr>
<td>HEI Thiamine</td>
<td>27.3</td>
<td>38.7</td>
<td>28.8</td>
<td>26.6</td>
</tr>
<tr>
<td>HEI Folate DFE</td>
<td>1.5</td>
<td>2.4 (0.9)</td>
<td>1.5 (0.9)</td>
<td>1.2 (0.6)</td>
</tr>
<tr>
<td>HEI Iron</td>
<td>420.9</td>
<td>522.1 (300)</td>
<td>386.7 (300)</td>
<td>341.5 (200)</td>
</tr>
<tr>
<td>HEI Sodium</td>
<td>13</td>
<td>18.9 (8)</td>
<td>12.4 (8)</td>
<td>10.1 (10)</td>
</tr>
<tr>
<td>HEI Selenium</td>
<td>3420.9</td>
<td>5588.5 (1500)</td>
<td>3363.3 (1500)</td>
<td>3043.7 (1200)</td>
</tr>
<tr>
<td>HEI 1,2 Designate average of values in cluster 1 and 2 respectively.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEI 1,2 Designate average of values in cluster 1 and 2 respectively.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Values in ( ) are Recommended Dietary Allowances (RDA) or Adequate Intake (AI) references but not individual requirement.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plasma Mean</td>
<td>High SAM/SAH&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Child 1 2009</td>
<td>Child 1 2010</td>
<td>Child 2 2009</td>
</tr>
<tr>
<td>----------------</td>
<td>--------------------------</td>
<td>--------------</td>
<td>--------------</td>
<td>--------------</td>
</tr>
<tr>
<td></td>
<td>84.2</td>
<td>84.6</td>
<td>36</td>
<td>110.7</td>
</tr>
<tr>
<td></td>
<td>1.8</td>
<td>1.2</td>
<td>1.0</td>
<td>2.6</td>
</tr>
<tr>
<td></td>
<td>24.6</td>
<td>29.3</td>
<td>17.3</td>
<td>25.3</td>
</tr>
<tr>
<td></td>
<td>5.2</td>
<td>0.1</td>
<td>15.0</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>9.7</td>
<td>5.8</td>
<td>1.0</td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.4</td>
<td>0.3</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td>11.8</td>
<td>12.6</td>
<td>7.2</td>
<td>11.9</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>1.3</td>
<td>0.6</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>1.8</td>
<td>1.5</td>
<td>0.6</td>
<td>1.6</td>
</tr>
</tbody>
</table>

<sup>1,2</sup> Designates average of values in cluster 1 and 2 respectively

3 Retinol = 30 – 80 µg/dl
4 Vit E = 0.5 – 1.8 mg/dl
5 Vit D = 14 – 60 ng/ml
6 B6 Pyridoxine = 5 – 30 ng/ml (I do not have pyridoxal reference)
7 Riboflavin (B2) = 4 – 24 µg/dl
8 Folate = 3.0 – 20.0 ng/ml in serum
9 Homocysteine values between 15-30 µmol/l mean lack of vitamins
   Values between 31-100 µmol/l indication of heterocysteite homocysteinemia;
   Values > 100µmol/l indication of homocysteite homocysteinemia.
Network map of 81 genes associated with Metabolite Patterns
Limitations and Advantages 21st Century Science

Limitations

Genetic spectrum is narrow (admixture an advantage)

Cultural and food diversity is constrained

Physiology of individual not well defined

Advantages

Population level data

Individual level data

Outcomes raised new nutrition/health questions
Deep analyses of response to acute or long-term intervention – *in all models*

Whole genome analyses = individual

Middle – out strategies for complex systems (*interim*)

*Compare* across many populations and cells (*genetics and lifestyle*)

*Quantify* before data reduction & classification
Thank You and Input

Move more
Choose ancestors wisely
Eat better