Using omics methods to characterize individual exposomes

S.M. Rappaport
University of California, Berkeley
EWAS and beyond

By applying EWAS* to biospecimens from healthy and diseased subjects, it should be possible to discover causal environmental exposures. But which ‘omes’ offer the most promise for EWAS and follow-up studies?

* Exposome-wide association studies
Knowledge-driven research

- Environmental health sciences use *knowledge-driven* approaches to investigate *known* environmental hazards.
  - Goals: dose-response, toxicokinetics and mechanisms
  - Approach: reductionist and deterministic, though toxicogenomics sometimes used
  - Focus: exogenous exposures (mainly air/water pollutants and food contaminants)
  - Targets: single chemicals or mixtures with known toxicity
  - Populations: assigned by exposure intensity
  - Individual risk: not generally possible
Knowledge-driven research in the environmental health sciences

- Well established for known hazards (e.g., lead, mercury, arsenic, asbestos, vinyl chloride, benzene, aflatoxin, PAHs)
- External measurements plus biological outcome (mortality or biomarkers)
- Consider benzene
  - A known cause of human hematological damage (since 1896) and leukemia (since 1928)

Exposure-disease relationship

Figure 3. Sensitivity analysis on the prediction of the ERC based on a natural spline. Graph represents nine plots of the predicted ERC based on all studies minus one. The plots are identified by the study that was excluded: 1, CAPM-NCI; 2, Dow; 3, Costantini; 5, US-Chemical (Wong et al. 1987); 6, Swaen; 7, Canada-Petal; 8, AHW; 9, UK-Petrol; and 10, Pilofilm. Plot 4 is the predicted ERC based on all available studies (blue line). Abbreviation: AHW, Australian Health Watch.

Vlaanderen et al., EHP 118: 526 (2009)
Benzene exposure and hematopoietic damage

Chromosomal damage


Benzene exposure and saturable metabolism

Benzene-related HSA adducts

Urinary benzene metabolites

Based upon data from: S. Kim et al., CEBP, 115: 2247 (2006)

Benzene exposure and toxicogenomics

Acute myeloid leukemia gene-response pathway

AML response pathway in 125 benzene workers and controls

R. Thomas, L. Zhang, M. Smith et al., unpublished
Going beyond knowledge-driven research

- The benzene example shows that environmental health sciences can effectively investigate known hazards
- But what about unknown hazards (70% of chronic-disease burden)?
Data-driven research

- Advances in analytical systems and bioinformatics (required for GWAS and transcriptomics), facilitate *data-driven* (omic) investigations of *unknown* hazards.
  - Goal: disease etiology, causal biomarkers, personalized healthcare
  - Approach: complex systems
  - Focus: exogenous and endogenous factors detectable in biospecimens
  - Targets: omic profiles (metabolome, transcriptome and proteome) leading to causal biomarkers
  - Populations: assigned by disease status
  - Individual risk: an anticipated outcome
The molecular basis of life (and disease)

INTERNAL CHEMICAL ENVIRONMENT

Which omes for finding causative exposures?
The serum exposome

**Metabolome:**
- Lipids
- Sugars
- Nucleotides
- Amino acids
- Metabolites

**Reactive electrophiles:**
- Reactive O&N species
- Aldehydes
- Epoxides
- Quinones

**Drugs**

**Cytokines and Chemokines**

**Receptor-binding agents:**
- Hormones
- Xeno-estrogens
- Endocrine disruptors

**Other small-molecule exposomes can use different biofluids (urine, breath, csf etc.)**
A protocol for EWAS and follow-up studies

DATA-DRIVEN DISCOVERY

Serum exposome
- Diseased vs. healthy (case-control studies)
- Discriminating features
- Chemical identification
- Candidate biomarker(s)

Causal biomarker(s) of exposure
- Diseased vs. healthy (prospective-cohort studies)
- Identify sources and measure exposures
- Individual risk
- Dose-response

Molecular epidemiology
- Exposure biology
- Personalized healthcare
- High vs. low exposure (Transcriptomics, proteomics & in vitro experiments)

KNOWLEDGE-DRIVEN HYPOTHESES

Systems biology and toxicogenomics

SM Rappaport
Small-molecule omics

With advances in NMR and mass spectrometry, data-driven studies of small molecules in human biofluids are becoming commonplace.

Metabolomics publications since 1999

*Figure 1.* Number of papers in PubMed covering the following search terms: metabolomic*, metabolomic*, "metabol* profiling", "metabol* fingerprinting", and/or lipidomics (left vertical axis) compared to the following set of terms in the title and in the abstract, as well as the corresponding MeSH keywords: "nuclear magnetic resonance/NMR", "mass spectrometry/MS" and "biological marker/biomarker" (right vertical axis).

Griffiths *et al.*, Angewandte Chemie, 2010, 49, 5426-45
### Candidate serum biomarkers

(12 Examples of metabolomics applied to serum/plasma from case-control studies, reviewed by Nordstrom and Lewensohn, *J Neuroimmune Pharmacol*, 2010, 5:4-17)

<table>
<thead>
<tr>
<th>Disease</th>
<th>Type</th>
<th>No. Subjects</th>
<th>Disc. features</th>
<th>Ident. features</th>
<th>Reference.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neurological</td>
<td>Huntington's disease</td>
<td>50</td>
<td>15</td>
<td>15</td>
<td>Underwood et al. (2006)</td>
</tr>
<tr>
<td>Neurological</td>
<td>Parkinson's disease</td>
<td>88</td>
<td>17</td>
<td>3</td>
<td>Bogdanov et al. (2008)</td>
</tr>
<tr>
<td>Neurological</td>
<td>Motor neuron dis.</td>
<td>58</td>
<td>76</td>
<td>0</td>
<td>Rozen et al. (2007)</td>
</tr>
<tr>
<td>Immunological</td>
<td>Celiac disease</td>
<td>68</td>
<td>16</td>
<td>16</td>
<td>Bertini et al. (2009)</td>
</tr>
<tr>
<td>Metabolic</td>
<td>MMA/PA</td>
<td>42</td>
<td>263</td>
<td>9</td>
<td>Wikoff et al. (2007)</td>
</tr>
<tr>
<td>Cardiological</td>
<td>Ischemia</td>
<td>31</td>
<td>5</td>
<td>5</td>
<td>Barba et al. (2008)</td>
</tr>
<tr>
<td>Cardiological</td>
<td>Myocardial injury</td>
<td>72</td>
<td>13</td>
<td>13</td>
<td>Lewis et al. (2008)</td>
</tr>
<tr>
<td>Cardiological</td>
<td>Myocardial ischemia</td>
<td>36</td>
<td>23</td>
<td>6</td>
<td>Sabatine et al. (2005)</td>
</tr>
<tr>
<td>Cardiological</td>
<td>Myocardial ischemia</td>
<td>39</td>
<td>4</td>
<td>4</td>
<td>Lin et al. (2009)</td>
</tr>
<tr>
<td>Cancer</td>
<td>Kidney</td>
<td>129</td>
<td>14</td>
<td>14</td>
<td>Gao et al. (2008)</td>
</tr>
<tr>
<td>Cancer</td>
<td>Pancreas</td>
<td>190</td>
<td>3</td>
<td>3</td>
<td>Beger et al. (2006)</td>
</tr>
<tr>
<td>Cancer</td>
<td>Prostate</td>
<td>220</td>
<td>10</td>
<td>10</td>
<td>Osl et al. (2008)</td>
</tr>
</tbody>
</table>

**Modest numbers of subjects**

**Candidate biomarkers**
A recent example of biomarker discovery for colorectal cancer (U.S. incidence of CRC is 50/100,000)

Possible omic features:
900 Da × 500 features/Da ≈ 0.5M features
Candidate biomarkers of CRC

Biomarker levels are lower in CRC cases

High association with stage-I suggests biomarkers are not derived from tumors

Figure 6 Results of triple-quadrupole multiple reaction monitoring analysis of the Chiba validation sample set. (A) Scatter plots of the concentrations of hydroxylated polyunsaturated ultra long-chain fatty acids (HPO-(U)FA) 446, 448, 450 expressed as [1-13C]-chole acid equivalents in asymptomatic normal controls, and pre-treatment colorectal cancer patients. (B) receiver operating curve (ROC) analysis based upon the corresponding scatter plots in (A). Grey dotted lines indicate the 95% confidence interval. (C) Bar charts of the average concentration equivalents of HPO-(U)FA by disease stage. Error bars represent standard errors of the mean. (D) ROC analysis by disease stage.
Biomarker identification

- Structures have not been confirmed
  - Hydroxylated ultra-long-chain fatty acids ($C_{28} - C_{36}$)
  - Unique-mass spectra permit precise measurement
- Probably anti-inflammatory agents similar to resolvins, protectins and lipoxins (products of omega-3 fatty acids)

Resolvin E1
Follow up measurements of CRC-446 in post-treatment cases and controls

Figure 2 CRC-446 levels in controls and CRC patients. (A) ROC analysis based on CRC-446 concentrations across 150 Caucasian post-treatment CRC patients and 761 age-matched controls. Dotted lines represent the 95% confidence interval. Mean CRC-446 levels (± 1 S.E.M) are shown by disease stage for the 150 CRC patients (B) and by treatment combination (C). p-values based on Student’s t-test between all stages and between treatment comparisons were >0.05.

No return to normal CRC-446 levels post-treatment

Results indicate that CRC-446 is not produced by the disease and may be a causal biomarker of exposure!

SM Rappaport
Ritchie et al., BMC Gastroenterology, 2010, 10, 140
Summary of important points

- Research in environmental health sciences has been knowledge-driven
  - Good for known environmental hazards but not for unknown hazards

- Biomedical research uses data-driven (exposome) approaches to find unknown health hazards
  - Untargeted EWAS in diseased/healthy populations
  - Small molecules in serum or other biofluid
  - Seek causal biomarkers of exposure

- Causal biomarkers guide knowledge-driven research
  - Identify sources of exposure
  - Dose-response
  - Mechanisms
  - Personalized healthcare
Benzene collaborators

UC-Berkeley
Martyn Smith
Luoping Zhang
Reuban Thomas

IOM (Beijing)
Guilan Li
Songnian Yin

NYU
Qingshan Qu
Roy Shore
Beverley Cohen

UNC
Sungkyoon Kim
Suramya Waidyanatha
Karen Yeowell-O’Connell

NCI
Nat Rothman
Qing Lan
Roel Vermeulen

Work supported by NIEHS, NCI, and HEI
Best wishes from Berkeley

Major support from NIEHS through grants U54ES016115 and P42ES04705 and from the ACC Long-range Research Initiative