Scientific Challenges to the Exposome Approach

John D. Groopman$^1$ and Gerald N. Wogan$^2$

$^1$Department of Environmental Health Sciences, Johns Hopkins Bloomberg School of Public Health

$^2$Department of Biological Engineering, Massachusetts Institute of Technology
Exposome: Vision

• To identify, characterize and quantify the exogenous and endogenous exposures and modifiable risk factors that predispose and predict health effects across an individual’s life span.
Exposome: Mission

• To foster the integration of the diverse scientific disciplines and technologies in a quest to measure all environmental exposures from conception onwards; including exposures from diet, lifestyle and endogenous sources, as a quantity of critical interest to disease etiology. These metrics will contribute to the implementation of individualized prevention strategies.

Modified from Wild, CP. CEBP 14: 1847-1850, 2005.
Exposome: Goals

A. To improve the fundamental chemistry and analytical infrastructure to enhance exposure assessment and quantitation of toxicologically relevant biomarkers related to human health.

B. To advance knowledge defining the relation between the exposome, genetic and epigenetic changes that drive the changing phenotype across the lifespan.

C. To apply the foundational knowledge of the exposome to address public health policy and practice at the national and international setting in environmental health.
Exposome: Genetics

Advancing technologies to individuals
Human Genotyping: Major Technology Advances

<table>
<thead>
<tr>
<th>Year</th>
<th>SNPs per assay</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997</td>
<td>1</td>
</tr>
<tr>
<td>2001</td>
<td>10</td>
</tr>
<tr>
<td>2002</td>
<td>1,000</td>
</tr>
<tr>
<td>2004</td>
<td>50,000</td>
</tr>
<tr>
<td>2006</td>
<td>500,000</td>
</tr>
<tr>
<td>2007</td>
<td>1,000,000</td>
</tr>
<tr>
<td>2010</td>
<td>&gt;&gt;1,000,000</td>
</tr>
</tbody>
</table>

Goal: Association studies with 2,000-20,000 samples
(2 billion - 20 billion genotypes)

Courtesy of E. Lander, MIT/Broad
Confirmed genetic contributors to common human diseases (Dec 2007)

<table>
<thead>
<tr>
<th>Condition</th>
<th>Gene(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cholesterol</td>
<td>Age Related Macular Degeneration</td>
</tr>
<tr>
<td>Obesity</td>
<td>Crohns Disease</td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>Type 1 Diabetes</td>
</tr>
<tr>
<td>QT interval</td>
<td>Asthma</td>
</tr>
<tr>
<td>Atrial Fibrillation</td>
<td>Restless leg syndrome</td>
</tr>
<tr>
<td>Type 2 Diabetes</td>
<td>Gallstone disease</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>Multiple sclerosis</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>Rheumatoid arthritis</td>
</tr>
<tr>
<td>Colon cancer</td>
<td>Glaucoma</td>
</tr>
<tr>
<td>Height</td>
<td>Celiac Disease</td>
</tr>
<tr>
<td>Uric Acid</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>Gene(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>PPARγ, IBD5, NOD2, CTLA4, KCNJ11, PTPN22</td>
</tr>
<tr>
<td>2001</td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td></td>
</tr>
</tbody>
</table>

Courtesy of E. Lander, MIT/Broad
DNA Sequencing Output (currently 1-2 Gbases per machine per day)

Human haploid genome is 3 Gbases (3 billion bases)

Cancer is a genetic disease

30 to 40 years
Sequence analysis of cancer genomes

<table>
<thead>
<tr>
<th></th>
<th>Breast</th>
<th>Colon</th>
</tr>
</thead>
<tbody>
<tr>
<td># of Genes</td>
<td>18,191</td>
<td>18,191</td>
</tr>
<tr>
<td># of Transcripts</td>
<td>20,857</td>
<td>20,857</td>
</tr>
<tr>
<td># of Primer pairs</td>
<td>198,088</td>
<td>198,088</td>
</tr>
<tr>
<td># of Tumors</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Total Sequence</td>
<td>341 Mb</td>
<td>341 Mb</td>
</tr>
<tr>
<td>Somatic Mutations</td>
<td>1243</td>
<td>942</td>
</tr>
<tr>
<td>Affected genes</td>
<td>1026</td>
<td>769</td>
</tr>
</tbody>
</table>

Courtesy of Victor Velculescu, Johns Hopkins
Mutation spectra of breast and colon cancers

Breast cancers
(n=1157 base substitutions)

Colon cancers
(n= 893 base substitutions)

C:G → T:A
(54% - 64%)

T:A → A:T
(7%)

T:A → G:C
(5%)

T:A → C:G
(9%)

C:G → T:A
(37%)

C:G → A:T
(15%)

C:G → G:C
(28%)

T:A → C:G
(8% - 10%)

C:G → A:T
(12% - 17%)

C:G → G:C
(7% - 10%)

T:A → A:T
(5%)

C:G → T:A
(54% - 64%)

Courtesy of Victor Velculescu, Johns Hopkins
Cancer is a genetic disease

Not all of these mutations are required for cancer

30 to 40 years
Exposome: Initial Thoughts

• Complex diseases, e.g. cancers, are all individualized at the molecular level.

• All cancers have individual kinetic trajectories for development. Opportunities for prevention and therapy.

• Given the modern understanding of this biology should we still be anchoring cancer diagnoses to 19th century histopathology?
Exposome: Epigenetics
Epigenetics and Human Disease

Environmental Stress \rightarrow \text{Gene Expression} \rightarrow \text{Inherited Adverse Health Outcome}

Proof of Concept: Agouti Mice

Coat color in Agouti mice varies from black to yellow due to stochastic methylation of CpG motifs

Unmethylated agouti (expressed)

Methylated agouti (not expressed)

Folate and B12 have transgenerational effect on decreased expression of the Agouti gene

Epigenetic differences arise during the lifetime of monozygotic twins
The complexity of genetic regulation is one of the great wonders of nature, but it represents a daunting challenge to unravel. The International Human Epigenome Consortium is an appropriate response.

Human DNA methylomes at base resolution show widespread epigenomic differences

“... cost of $100,000 and ...expect to drop to $10,000 soon”
Exposome: Human biomarkers

A short aflatoxin story: advances in mass spectrometry
Aflatoxin factoids

- Discovered in UK in 1960 in moldy, toxic animal feed
- Frequent contaminant of improperly stored food crops
  - Produced by few strains of the mold *Aspergillus flavus* *(A. flavus* toxin = “*Aflatoxin*”)
  - Spores are widely distributed in soil globally
  - Mold grows on food crops after harvest, before drying

- Some relevant properties
  - Highly fluorescent, heat stable
  - Lethal to animals at high levels (“Turkey X disease”)
  - Carcinogenic to liver of animals when fed at non-toxic levels
  - Immunotoxic to animals and humans
Aflatoxin $\text{B}_1$ $\rightarrow$ Aflatoxin-8,9-epoxide $\rightarrow$ DNA $\rightarrow$ AP site $\rightarrow$ Aflatoxin – $N^7$-guanine (urine)

Aflatoxin – mercapturic acid (urine)

Aflatoxin $M_1$ (urine)

CYP1A2

GSTs

CYPs 1A2 3A4

Aflatoxin albumin adduct (serum)

other metabolites
Outline of aflatoxin isolation scheme

Sample (urine, serum, milk)
Preparative chromatography on C$_{18}$ Sep-Pak
Wash Sep-Pak with 5% Methanol
Elute aflatoxins with 80% methanol and concentrate
Apply aflatoxins to monoclonal antibody affinity column
Wash column with P$_i$/NaCl
Elute aflatoxins with 50% Me$_2$SO in P$_i$/NaCl
Regenerate column with P$_i$/NaCl

HPLC with UV detection

Rat Urine (1 mg AFB$_1$ / kg)

Human Urine (87 µg AFB$_1$ in Previous Day's Diet)

Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 7728-7731, December 1984

Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 6492-6496, October 1985
Aflatoxin DNA Adducts in People in Guangxi, P.R.C.

TOTAL AFB-N⁷-GUANINE IN URINE
GUANGXI PROVINCE, P.R.C.

Groopman et al Cancer Research 52:45-52, 1992
AFB-DNA Adducts by HPLC-MS/MS

Rat Urine

Human Urine

Exogenous and Endogenous DNA Adducts in Humans (Partial List)

- Benzopyrene and PAH DNA adducts
- Etheno adenine and cytosine (lipid peroxidation products)
- Estrogen guanine and adenine
- Aromatic and heterocyclic amine DNA adducts
- Mycotoxin DNA adducts (aflatoxin and aristolochic acid)
- ROS and RNS purine and pyrimididine adducts
- Chemotherapeutic mediated DNA damage
- Multiple 7-alkylguanines and 3-alkyladenines
Exposome: Human biomarkers

Hemoglobinomics
<table>
<thead>
<tr>
<th>COMPOUND</th>
<th>fmole/g Hemoglobin</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPB from NNK</td>
<td>29.3 ± 25.9</td>
</tr>
<tr>
<td>2-Aminonapthalene</td>
<td>40 ± 20</td>
</tr>
<tr>
<td>4-ethylaniline</td>
<td>99 ± 10</td>
</tr>
<tr>
<td>2,6-dimethylaniline</td>
<td>157 ± 50</td>
</tr>
<tr>
<td>4-Aminobiphenyl</td>
<td>166 ± 77</td>
</tr>
<tr>
<td>3,5-dimethylaniline</td>
<td>220 ± 20</td>
</tr>
<tr>
<td>o-Toluidine</td>
<td>320 ± 90</td>
</tr>
<tr>
<td>p-Toluidine</td>
<td>640 ± 370</td>
</tr>
<tr>
<td>m-Toluidine</td>
<td>6400 ± 1900</td>
</tr>
<tr>
<td>N-(2-carbamoylethyl)valine</td>
<td>19000 ± 12000</td>
</tr>
<tr>
<td>Aniline</td>
<td>41000 ± 22000</td>
</tr>
<tr>
<td>N-(2-Hydroxyethyl)valine (Acrylamide)</td>
<td>61200 ± 25000</td>
</tr>
</tbody>
</table>

Exposome: Quantitative chemical analysis

• In last 10 years sensitivity of mass spectrometry for human biomarkers (adducts) has increased 1000 fold and number of non-detectable human samples has dramatically dropped.

• Per unit cost of sample analysis is $50-100.

• Future improvements will be data driven or we have to start somewhere if we expect to finish.
Exposome: Intersecting Science and Policy/Regulation

• The science of the exposome projects have the promise of generic solutions to the issues of environmental dose and health outcomes.

• Policy and regulation is still guided by compound by compound dose-response analyses in risk assessment.