Environment-wide Association Studies (EWAS) for a more complete view of disease risk

7/10/2014

Chirag J Patel
chirag@hms.harvard.edu
@chiragjp
environmental exposure informatics

http://cbmi.harvard.edu

evironment + genome = health
Complex disease is a function of genes and environment...

...and we’re exposed to many environmental factors...

and lack methods and data to comprehensively and systematically connect the environment with disease.
... and the case is different with genetics (e.g., genomics)!

over 1,400

*Genome-wide Association Studies (GWAS)*

NHGRI GWAS Catalog

https://www.genome.gov/
...but, much disease risk may be due to differences in environment\textsuperscript{1,2,3}.

3. Lindorff et al., PNAS. (2009)
... but we lack a search engine to discover new exposures in disease.
(see: GWAS!)
**Big Environmental Exposure: Human Exposome**

Human Exposome Project to “E”WAS?

**what to measure?**

- Radiation
- Stress
- Life-style
- Infections
- Drugs
- Diet
- Pollution

**how to measure?**

- Exposome
  - Reactive electrophiles
  - Metals
  - Endocrine disrupters
  - Immune modulators
  - Receptor-binding proteins

**how to analyze in relation to health?**

“A more comprehensive view of environmental exposure is needed ... to discover major causes of diseases...”

... but there is no “microarray” for environmental exposure...
In lieu of an exposome chip: National Health and Nutrition Examination Survey

since the 1960s: 50 years!
now biannual: 1999 onwards
10,000 participants per survey

hundreds of exposure measures

many clinical measures & health assessments

1 http://www.cdc.gov/nchs/nhanes.htm
NHANES ascertains >300 exposures in serum and urine!

**Nutrients and Vitamins**
- *e.g.*, vitamin D, carotenes

**Pesticides and pollutants**
- *e.g.*, atrazine; cadmium; hydrocarbons

**Infectious Agents**
- *e.g.*, hepatitis, HIV, *Staph. aureus*

**Plastics and consumables**
- *e.g.*, phthalates, bisphenol A

**Physical Activity**
- *e.g.*, steps
Environmental exposures are associated with **Type 2 Diabetes?**
EWAS in Type 2 Diabetes visualized in a Manhattan Plot

Fasting Blood Glucose ≥ 126 mg/dL?
BMI, SES, ethnicity, age, sex
N=500-2000 per cohort

**EWAS** has had utility in searching for *exposures* in disease

A Nutrient-Wide Association Study on Blood Pressure

Ioanna Tzoulaki, PhD;* Chirag J. Patel, PhD;* Tomonori Okamura, MD, PhD; Queenie Chan, PhD; Ian J. Brown, PhD; Katsuyuki Miura, MD, PhD; Hirotsugu Ueshima, MD, PhD; Liancheng Zhao, MD; Linda Van Horn, PhD; Martha L. Daviglus, MD, PhD; Jeremiah Stamler, MD; Atul J. Butte, MD, PhD; John P.A. Ioannidis, MD, DSc; Paul Elliott, MB BS, PhD

*Circulation, 2012*

Systematic evaluation of environmental factors: persistent pollutants and nutrients correlated with serum lipid levels

Chirag J Patel,1,2 Mark R Cullen,3 John PA Ioannidis4,5,6 and Atul J Butte1,2*

*IJE, 2012*

Systematic evaluation of environmental and behavioural factors associated with all-cause mortality in the United States National Health and Nutrition Examination Survey

Chirag J Patel,1 David H Rehkopf,2 John T Leppert,3 Walter M Bortz,4 Mark R Cullen,2 Glenn M Chertow1 and John PA Ioannidis1*

*IJE, 2013*
**EWAS** and searching for **GxE** in disease:

**EWAS-identified** exposures + **GWAS-identified** SNPs

= *greater* risk for disease?
Screening all possible EWAS-GWAS interactions: OR for rs13266634 (SLC30A8) stratified by E +30-40% vs. no E

Adjusted for race, sex, BMI, age

VIEWPOINT

Chirag J. Patel, PhD
Center for Biomedical Informatics, Harvard Medical School, Boston, Massachusetts.

John P. A. Ioannidis, MD, DSc
Stanford Prevention Research Center, Department of Health Research and Policy, Department of Medicine, Stanford University School of Medicine, Stanford, California, Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, California, and Meta-Research Innovation Center at Stanford (METRICS), Stanford, California.

Studying the Elusive Environment in Large Scale

It is possible that more than 50% of complex disease risk is attributed to differences in an individual’s environment. Air pollution, smoking, and diet are documented environmental factors affecting health, yet these factors are but a fraction of the ‘exposome,’ the totality of the exposure load occurring throughout a person’s lifetime. Investigating one or a handful of exposures at a time has led to highly fragmented literature of epidemiologic associations. Much of that literature is not reproducible, and selective reporting may be a major reason for the lack of reproducibility. A new model is required to discover environmental exposures associated with disease while mitigating possibilities of selective reporting.

To remedy the lack of reproducibility and concerns of validity, multiple personal exposures can be assessed simultaneously in terms of their association with a condition or disease of interest; the strongest associations can then be tentatively validated in independent data sets (eg, as done in references 2 and 3). The main advantages of this process include the ability to search the list of exposures and adjust for multiplicity systematically and report all the probed associations instead of only the most significant. The EWAS vantage point, intervening on β-carotene (Figure, D) seems a futile exercise given its complex relationship with the EWAS. However, eventually for most environmental correlates, there may be unsurpassable difficulty establishing potential causal inferences based on observational data.

- evaluate new ‘omics technologies
- new informatics methods

**longitudinal** and **linkable** health data

JAMA, 2014

i2b2?
ENVIRONMENT-WIDE ASSOCIATION STUDY (EWAS) FOR TYPE 2 DIABETES IN THE MARSHFIELD PERSONALIZED MEDICINE RESEARCH PROJECT BIOBANK


314 self-report E (PhenX toolkit) ~2000 cases and controls T2D via eMERGE algorithm

https://www.phenxtoolkit.org/
Searching for maternal exposures associated with preterm birth
Mothers in NHANES and attending Stanford Clinics

<table>
<thead>
<tr>
<th>Factor</th>
<th>N(cases)</th>
<th>OR [95% CI]</th>
<th>pvalue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urinary Bisphenol A</td>
<td>109(10)</td>
<td>1.9[1.4,2.6]</td>
<td>0.002</td>
</tr>
<tr>
<td>Serum Iron</td>
<td>761(62)</td>
<td>1.6[1.2,2.1]</td>
<td>0.005</td>
</tr>
<tr>
<td>Urinary Cesium</td>
<td>245(20)</td>
<td>1.9[1.2,3]</td>
<td>0.009</td>
</tr>
<tr>
<td>Urinary 1-hydroxypyrene</td>
<td>179(11)</td>
<td>1.8[1.1,2.8]</td>
<td>0.02</td>
</tr>
<tr>
<td>Serum Beta-cryptoxanthin</td>
<td>586(51)</td>
<td>1.7[1.1,2.5]</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Screened 253 exposures in >100 mothers (NHANES)

Bruce Ling
Andy Hu
Ting Yang
Atul Butte

Searching for maternal exposures associated with preterm birth: Mothers in NHANES and attending Stanford Clinics

16 moms with preterms
21 moms with normal births

collected urine during gestation

identified via EHR

Location, location, location:
Exposure assessment via geolocation of patients

(latitude, longitude)

Income/SES
Air pollution (e.g., PM 2.5)
“Green space”
Climate

John Brownstein
**EWAS** can enable a *comprehensive view* of exposures in disease by studying the environment...

...and feasible to capture this view in the **EHR**

Understanding the role of the **exposome** will lead to a more precise medicine.
Acknowledgments

Harvard
Isaac Kohane
Nathan Palmer
Sek Won Kong
Siu Shan Wang
Sunny Alvear
Erik Corona
Adam Brown
Denise McGinnis
Jackson Scholl
Arjun Manrai
Isamar Trancoso

Stanford
John Ioannidis
Atul Butte
David Rehkopf
John Leppert
Glenn Chertow
Mark Cullen
Keiichi Kodama
Tim Assimes

CDC/NCHS
Ajay Yesupriya

Imperial (UK)
Ioanna Tzoulaki
Paul Elliott

Berkeley
Steve Rappaport

Georgia Tech
Greg Gibson

March of Dimes Prematurity Research Center (Stanford)
PhRMA Foundation Informatics Fellowship
NIH/NIEHS K99/R00 Pathway to Independence Award

chirag@hms.harvard.edu
@chiragjp