

Exposome-wide association studies (EWAS) for discovering environmental causes of disease

S.M. Rappaport

University of California, Berkeley

*Research support from NIEHS and the
American Chemistry Council*

CEB



Some background

- About three fourths of all people die from chronic diseases, mainly CVD and cancer
- These diseases result from a combination of genetic (G) and environmental (E) factors
 - Known risk factors account for less than half of chronic-disease risks
- Elaboration of the G matrix with modern GWAS has been stunningly comprehensive
 - but has explained relatively little CVD and cancer risks
- Elaboration of the E matrix relies on questionnaires, geographic information and targeted measurements
 - much as it did a century ago

The exposome – promoting discovery of environmental causes of disease

Christopher Wild defined the 'exposome', representing *all* environmental exposures (including diet, lifestyle, and infections) from conception onwards, as a complement to the genome in studies of disease etiology

Wild, C.P., *Cancer Epidemiol Biomarkers Prev* 14 (8), 1847-1850 (2005).

Editorial

Complementing the Genome with an "Exposome": The Outstanding Challenge of Environmental Exposure Measurement in Molecular Epidemiology

Christopher Paul Wild

Molecular Epidemiology Unit, Centre for Epidemiology and Biostatistics, Leeds Institute of Genetics, Health and Therapeutics, Faculty of Medicine and Health, University of Leeds, Leeds, United Kingdom

EPIDEMIOLOGY

Environment and Disease Risks

Stephen M. Rappaport and Martyn T. Smith

Although the risks of developing chronic diseases are attributed to both genetic and environmental factors, 70 to 90% of disease risks are probably due to differences in environments (1-3). Yet, epidemiologists increasingly use genome-wide association studies (GWAS) to investigate diseases, while relying on questionnaires to characterize "environmental exposures." This is because GWAS represent the only approach for exploring the totality of any risk factor (genes, in this case) associated with disease prevalence. Moreover, the value of costly genetic information is diminished when inaccurate and imprecise environmental data lead to biased inferences regarding gene-environment interactions (4). A more comprehensive and quantitative view of environmental expo-

sure is needed if epidemiologists are to discover the major causes of chronic diseases.

An obstacle to identifying the most important environmental exposures is the fragmentation of epidemiological research along lines defined by different factors. When epidemiologists investigate environmental risks, they tend to concentrate on a particular category of exposures involving air and water pollution, occupation, diet and obesity, stress and behavior, or types of infection. This slicing of the disease pie along parochial lines leads to scientific separation and confuses the definition of "environmental exposures." In fact, all of these exposure categories contribute to chronic diseases and collectively rather than individually affect disease risk.

To develop a more comprehensive and quantitative view of environmental exposures, we need to identify the major causes of chronic diseases.

A new paradigm is needed to assess how a lifetime of exposure to environmental factors affects the risk of developing chronic diseases.

chemicals that alter critical molecules, cells, and physiological processes inside the body. Thus, it would be reasonable to consider the "environment" as the body's internal chemical environment and "exposures" as the amounts of biologically active chemicals in this internal environment. Under this view, exposures are not restricted to chemicals (toxicants) entering the body from air, water, or food, for example, but also include chemicals produced by inflammation, oxidative stress, lipid peroxidation, infections, gut flora, and other natural processes (5, 6) (see the figure). This internal chemical environment continually fluctuates during life, due

School of Public Health, University of California, Berkeley, CA 94720-7356, USA. E-mail: srappaport@berkeley.edu

460

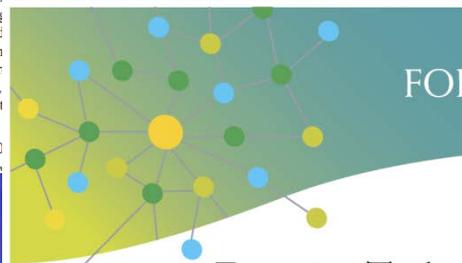
22 OCTOBER 2010 VOL 330
Published Online



**EMERGING SCIENCE
FOR ENVIRONMENTAL
HEALTH DECISIONS
WORKSHOP**

The Exposome: A Powerful Approach for Evaluating Environmental Exposures and Their Influences on Human Disease

FEBRUARY 25-26, 2010 . WASHINGTON, DC
THURSDAY, 8:30-5:00, FRIDAY, 8:30-NOON . NAS BUILDING, 2100 C STREET, NW, AUDITORIUM



**EMERGING SCIENCE
FOR ENVIRONMENTAL
HEALTH DECISIONS
AGENDA**

Emerging Technologies for Measuring Individual Exposomes

DECEMBER 8-9, 2011 . THURSDAY, 8:30-5:00, FRIDAY, 8:30-NOON*
HOUSE OF SWEDEN EVENT CENTER, 2900 K STREET, NW, WASHINGTON, DC

THIS WORKSHOP WILL BE WEBCAST.

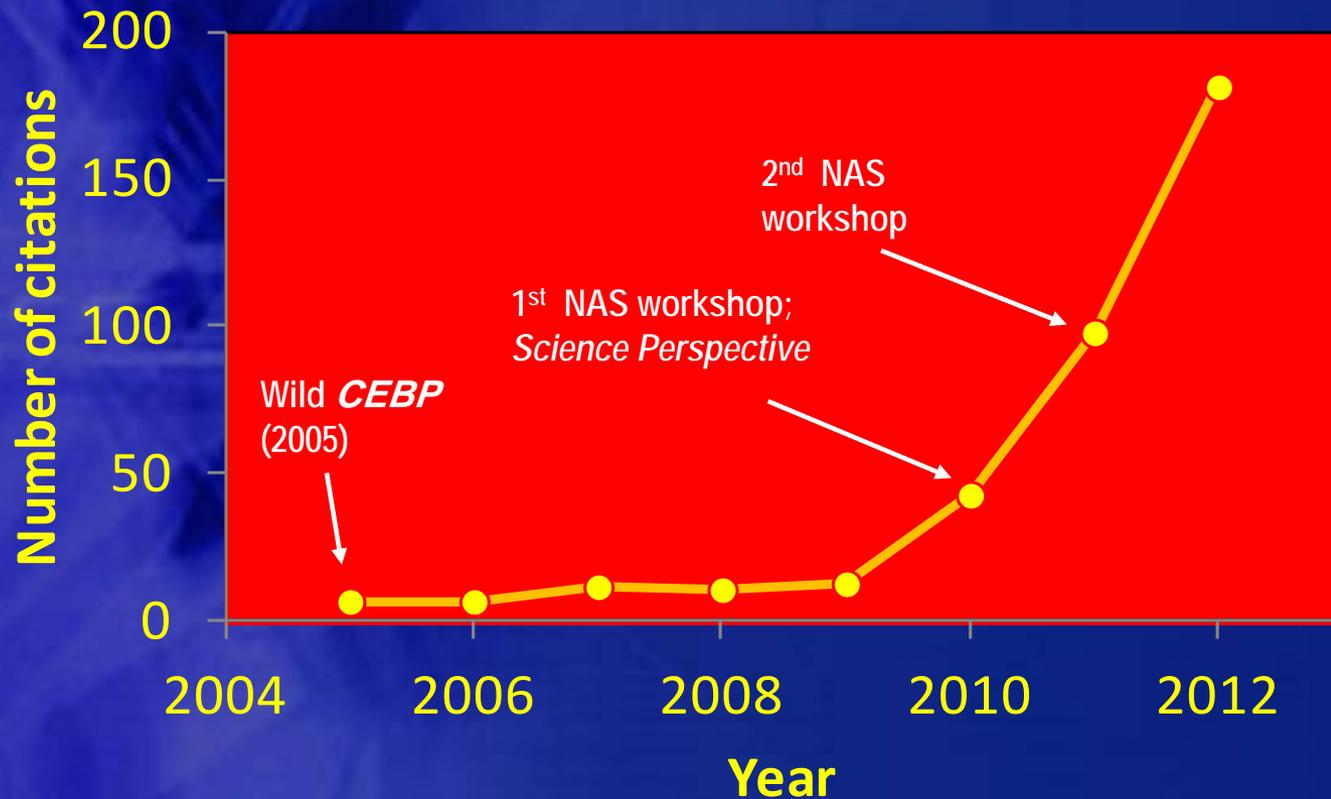
Christopher P. Wild defined the "exposome," representing all environmental exposures from conception onwards (including exposures from diet, lifestyle, and exogenous sources) as a quantity of critical interest to disease etiology (Wild, 2005). He argued that if we succeed in identifying the combined effects of genetic and environmental factors on chronic diseases, we will develop 21st-century tools to characterize exposure in human populations.

THIS WORKSHOP WILL EXAMINE the concept of exposome and its importance to the etiology of chronic diseases. In doing so, we will consider the roles of epidemiologists and laboratory scientists can play in identifying resources and technologies for elaborating the exposome in human populations.

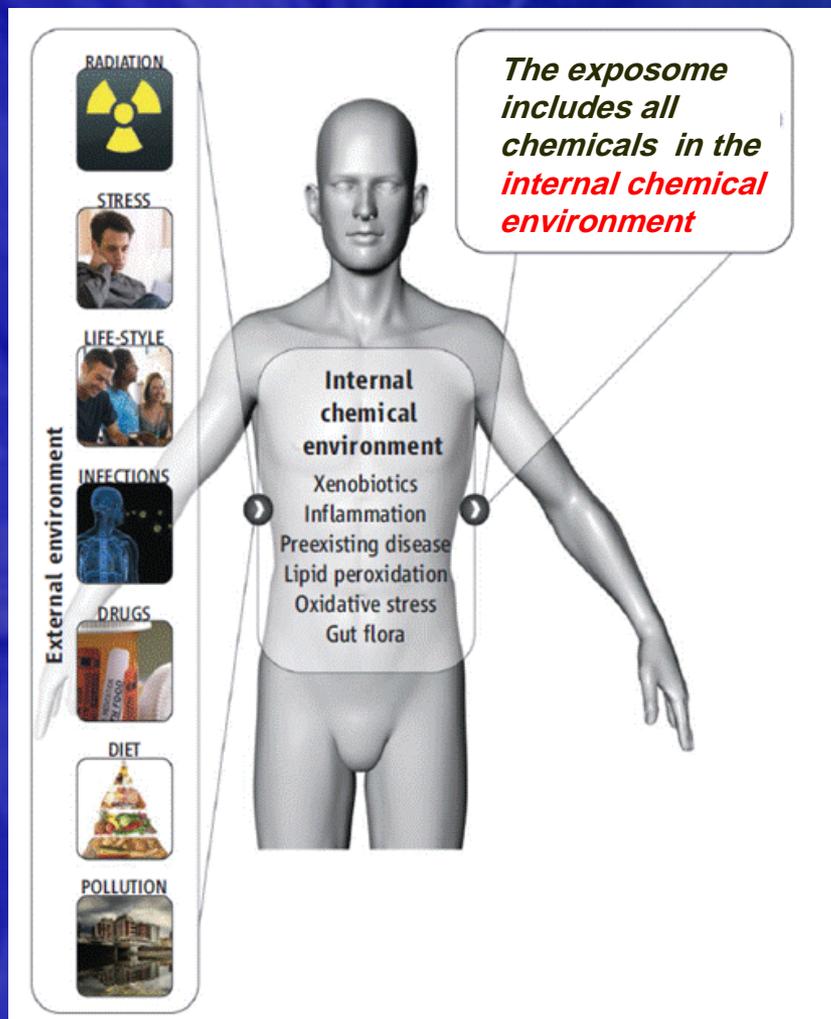
5 Exposures from the Social Environment and the Biology of Aging—*Elissa Epel, University of California, San Francisco, Department of Psychology*

6 Panel Discussion
What types of epidemiology studies and exposure

Scientific citations to 'exposome' (Google Scholar)



Functionalizing the exposome



Measurements of biomarkers can provide information about both *exogenous and endogenous* exposures

S.M. Rappaport and M.T. Smith, *Science*, 2010: 330, 460-461

Exposome-wide association studies (EWAS)

By applying EWAS to biospecimens from healthy and diseased subjects, we can discover causal environmental exposures.



<http://www.flickr.com/photos/paulieparker/246707763/>

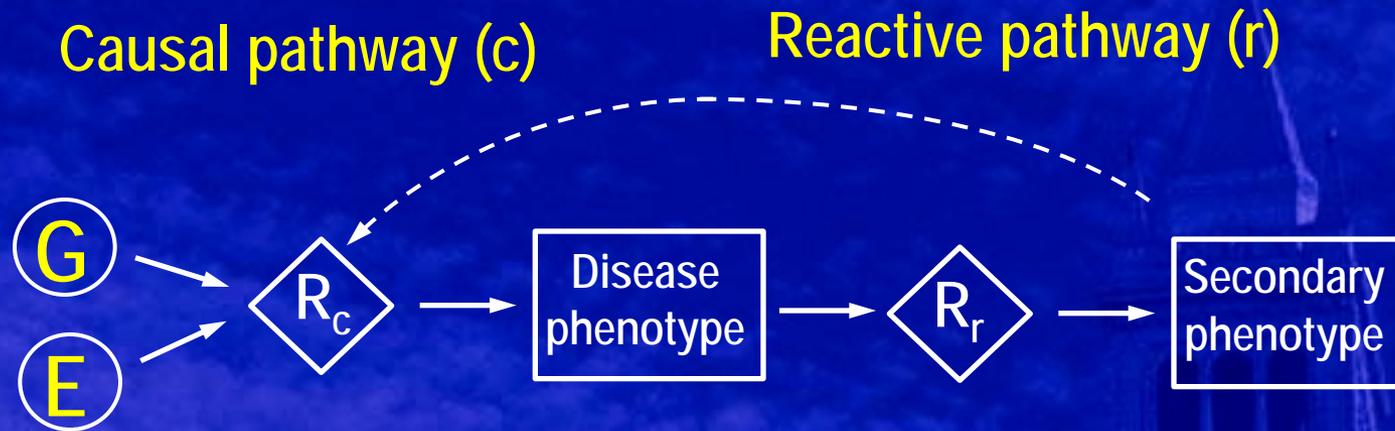
But which 'omes' offer the most promise for EWAS and follow-up studies?

The molecular basis of life (and disease)



INTERNAL CHEMICAL ENVIRONMENT

Disease pathways



G = genome

E = environment

R = transcriptome (gene expression)

S. Rappaport, *Biomarkers*, 2012, 17(6), 48: 3-9

Based on: E. Shadt *et al.*, *Nat Gen*, 2005, 37: 710-717

Adding omes

Causal pathway (c)

Reactive pathway (r)



G = genome

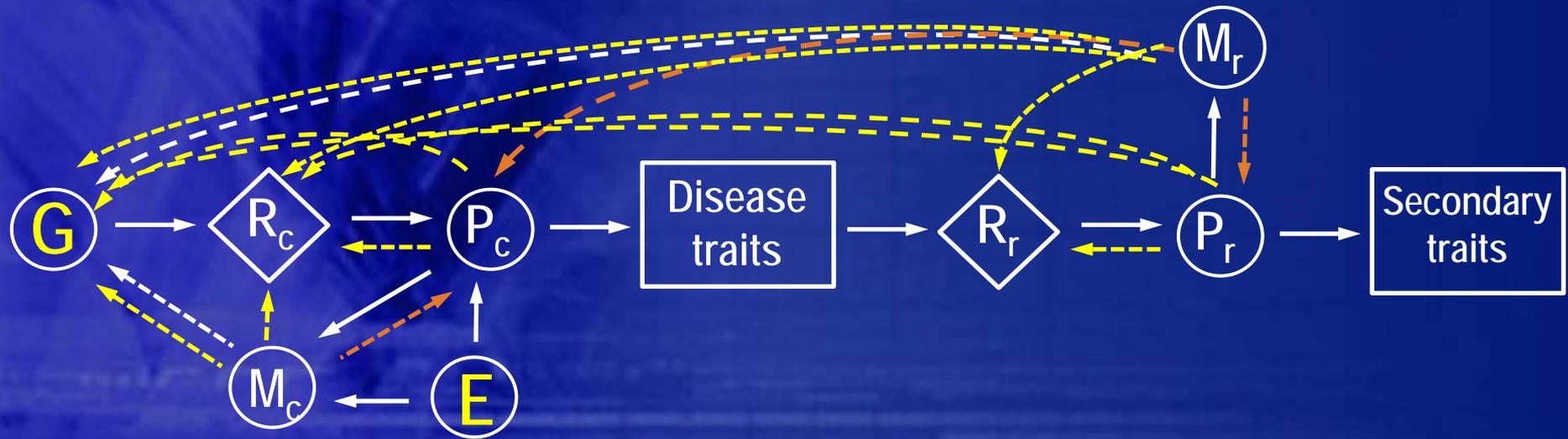
E = environment

R = transcriptome (gene expression)

P = proteome (protein expression)

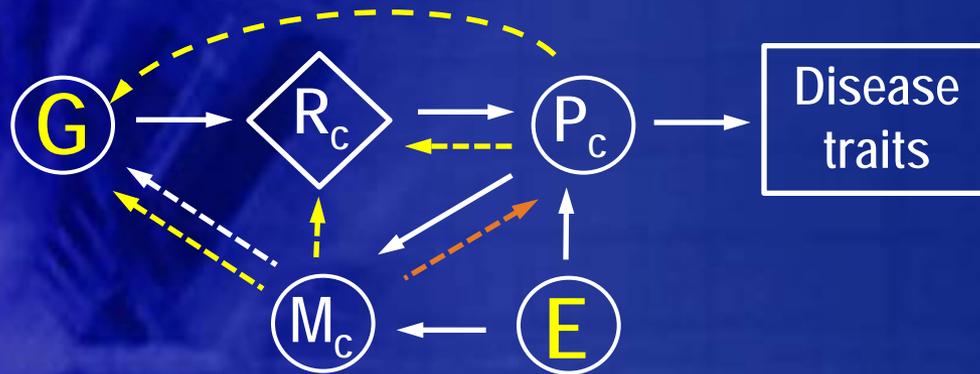
M = metabolome (all small molecules and metals)

More omic connections



- Genetic modifications (mutations)
- - - - Post-translational modifications
- . . . Epigenetic modifications

Which omes for EWAS?



If causal exposures operate primarily through small molecules (M_c) and proteins (P_c), then EWAS require metabolomics and/or proteomics.

Bioactive molecules

Reactive electrophiles:

Reactive O, N & Cl species

Aldehydes

Epoxides

Quinones

Metabolome:

Lipids

Sugars

Nucleotides

Amino acids

Metabolites

Xenobiotics

Inflammation markers:

Cytokines

Chemokines

Eicosanoids

Vasoactive amines

Growth factors

SERUM EXPOSOME

Micronutrients

Receptor-binding agents:

Hormones

Xenoestrogens

Endocrine disruptors

Microbiome
products

Metals

Drugs

Serum exposome

↓
Diseased vs. healthy
(case-control studies)
Untargeted designs

Discriminating features

↓
Chemical
identification

Candidate biomarkers

↙ ↘
Diseased vs. healthy
(prospective cohorts)
Targeted designs

Biomarkers of exposure *Biomarkers of disease*

**DATA-DRIVEN
DISCOVERY (EWAS)**

Serum exposome

Diseased vs. healthy
(case-control studies)
Untargeted designs

Discriminating features

Chemical
identification

Candidate biomarkers

Diseased vs. healthy
(prospective cohorts)
Targeted designs

Biomarkers of exposure *Biomarkers of disease*

DATA-DRIVEN
DISCOVERY (EWAS)

KNOWLEDGE-DRIVEN
APPLICATIONS

Dose-response

Identify
sources &
measure
exposures

Genomics,
epigenomics,
transcriptomics
& experiments

Disease
stage and
response to
therapy

Molecular
epidemiology

Exposure
biology

Systems
biology

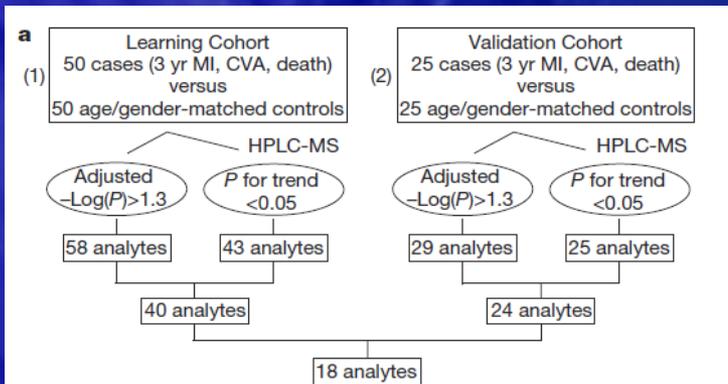
Drug
development

*Causality and
prevention*

*Diagnosis, prognosis
and treatment*

An EWAS of cardiovascular disease

From *untargeted* serum metabolomics (2,000 features), 18 *unknown features* were associated with *cardiovascular disease*. Of these, 3 were highly correlated, suggesting a common pathway.



(3) Structural identification of analytes
 (4) Confirm clinical prognostic utility in independent prospective cohort ($N = 1,876$)

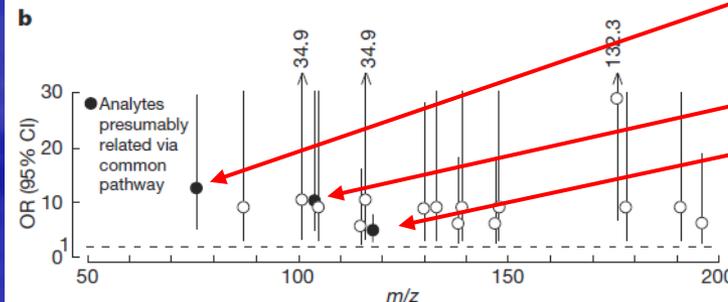
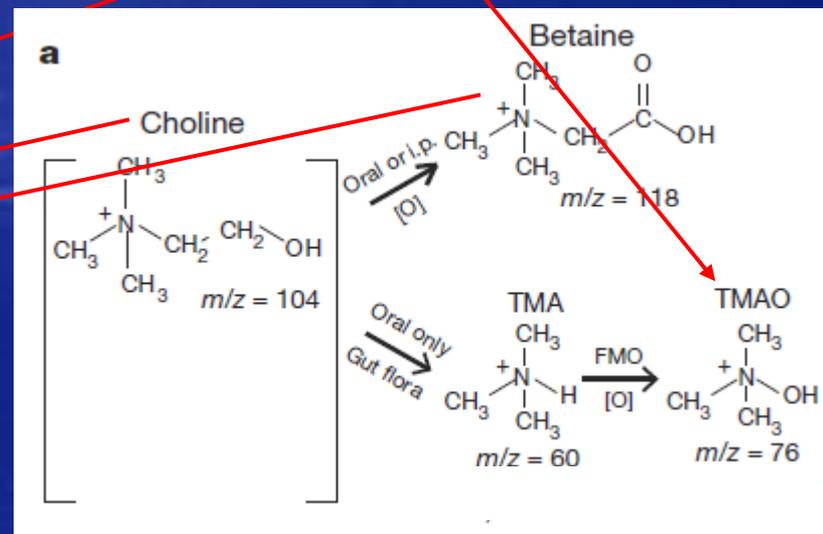


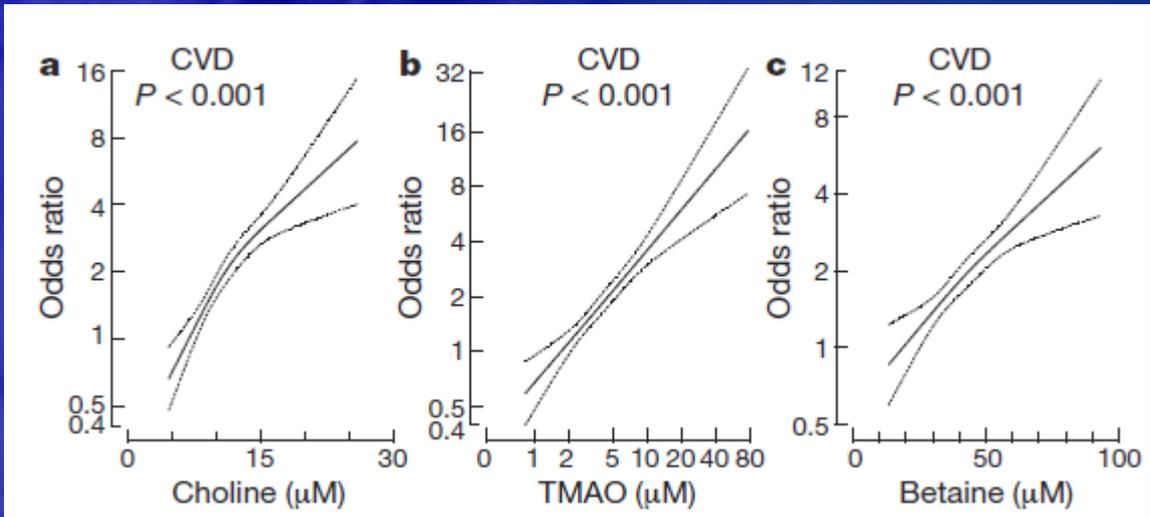
Figure 1 | Strategy for metabolomics studies to identify plasma analytes associated with cardiovascular risk. a, Overall schematic to identify plasma analytes associated with cardiac risk over the ensuing 3-year period. CVA, cerebrovascular accident; HPLC, high-performance liquid chromatography; MI, myocardial infarction. b, Odds ratio (OR) and 95% confidence intervals (CI) of incident (3-year) risk for MI, CVA or death of the 18 plasma analytes that met all selection criteria in both Learning and Validation Cohorts; odds ratio and 95% confidence intervals shown are for the highest versus lowest quartile for each analyte. Filled circles represent the analytes ($m/z = 76, 104, 118$) focused on in this study. m/z , mass to charge ratio.

Trimethylamine oxide (TMAO)



Wang et al. *Nature* (2011) 472: 57-63.

Choline metabolism: a major cause of CVD?



Targeted LC-MS/MS analyses of 1870 subjects from an independent cohort of CVD patients and controls

Dietary choline: eggs, milk, red meat, poultry, seafood and lecithin (food additive)

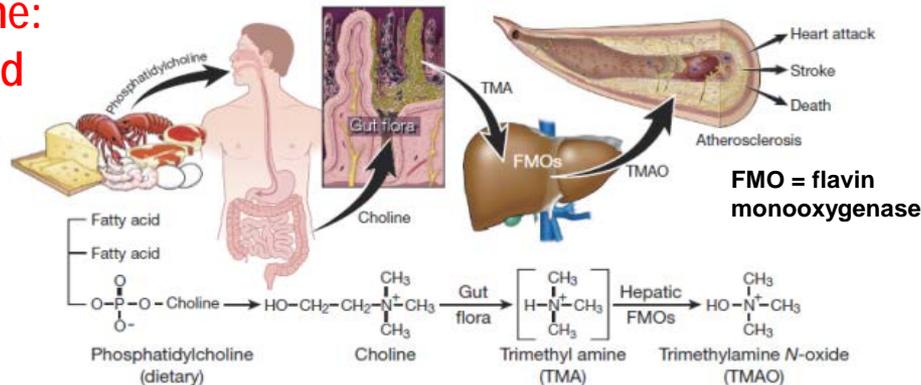


Figure 6 | Gut-flora-dependent metabolism of dietary PC and atherosclerosis. Schematic summary illustrating newly discovered pathway for gut-flora-mediated generation of pro-atherosclerotic metabolite from dietary PC.

Two biomarker-research agendas

EWAS

- For disease etiology
- Data-driven, untargeted designs
- Focus on small molecules and proteins
- To identify biomarkers
- Proof of concept has been established

Follow-up studies

- *For etiology, diagnosis and prognosis*
- *Knowledge-driven, targeted designs*
- For causative or suspicious factors
- Use biomarkers to confirm causality, etc.
- Provide feedback for public health and treatment

Best wishes from Berkeley



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

CEB

Center For
Exposure Biology



**Genes &
Environment
Laboratory**

