# Discovering Environmental Causes of Disease: from Exposure Biology to the Exposome

**S.M. Rappaport**

**University of California, Berkeley**

CEB Center For Exposure Biology

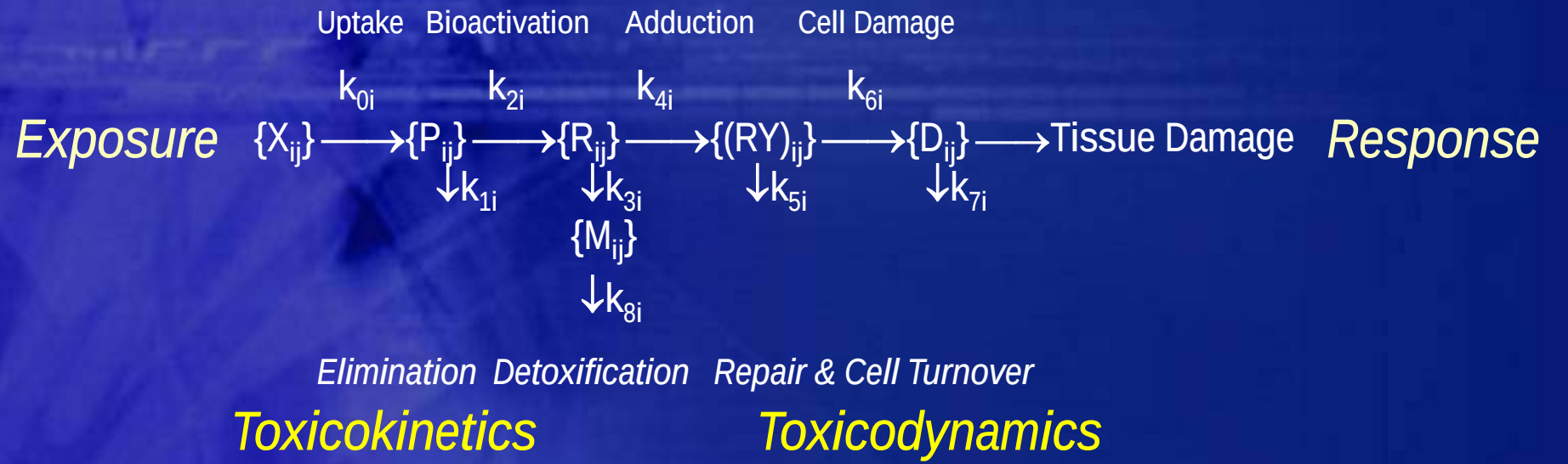
# Exposure biology

Epidemiologists wait for people to die or get sick before they can study them. Exposure biology connects exposures to hazardous chemicals with *early* effects of these exposures inside the body.

# *Exposure biology requires two types of information -*

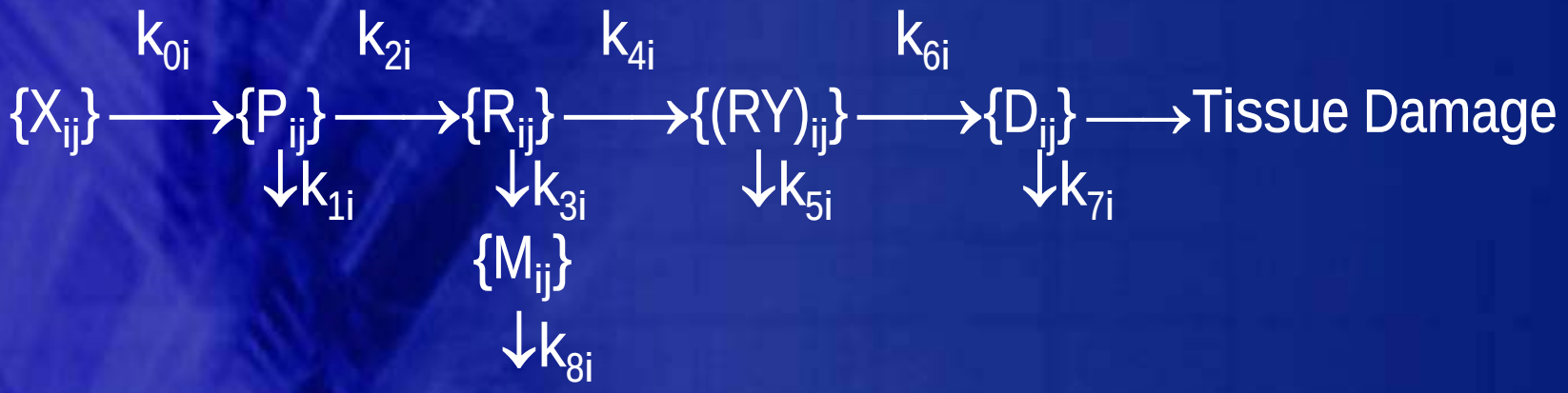
1. *Exposure levels*
2. *Biological response (biomarkers)*

*With such data we can do interesting things…*




Lin, Kupper, and Rappaport *Occ Environ Med*, 62: 750-760 (2005)

# Biomarkers




*Exposure* → *Biomarkers* → *Death or disease*

*Exposure biology* *Molecular epidemiology*

# The exposure biology of *benzene*

- A blood (hematopoietic) toxin
  - First linked with bone marrow toxicity in 1896 (Santesson, *C. Arch Hyg Berl* 31: 337) and with leukemia in 1928 (Delore, P. and Borgomano, C. *J Med Lyon* 9: 227)
- Mechanism unknown but related to metabolism
  - Uncertain risks, particularly at low exposures
- Dose-related metabolism poorly characterized in humans

# Benzene exposure and hematopoietic damage in U.S. lacquer workers (ca. 1925)

THE JOURNAL OF INDUSTRIAL HYGIENE

OLUME X JUNE, 1928 NUMBER 6

SPRAY PAINTING HAZARDS AS DETERMINED BY THE PENNSYLVANIA AND THE NATIONAL SAFETY COUNCIL SURVEYS*

HENRY FIELD SMYTH, M.D., DR.P.H.
AND
HENRY F. SMYTH, JR., B.S. IN CH.E.

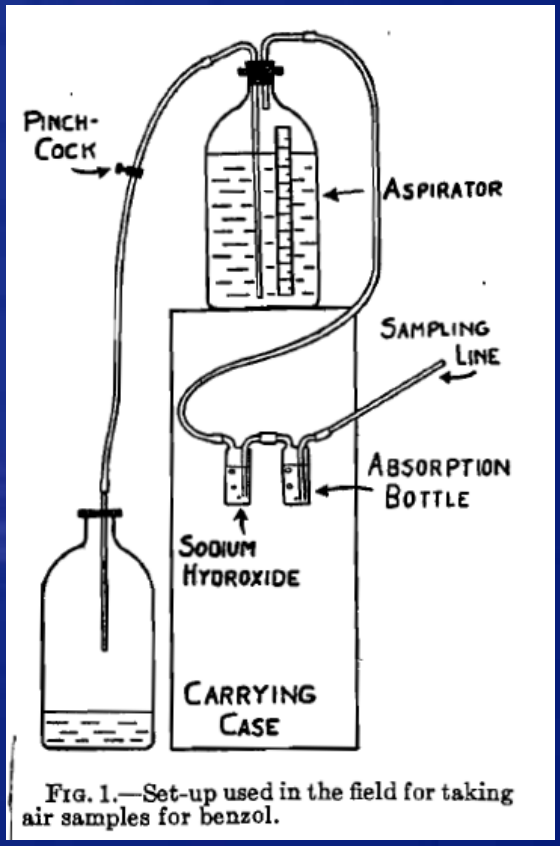



FIG. 1.—Set-up used in the field for taking air samples for benzol.

# Benzene exposure and hematopoietic damage in U.S. lacquer workers (ca. 1925)

TABLE 2.—RANGES OF BENZOL CONCENTRATIONS FOUND IN FIELD SAMPLES, PENNSYLVANIA SURVEY

| GROUP NO. | CONCENTRATION SHOWN ON TITRATION | PROBABLE TRUE CONCENTRATION[1] | NO. OF FIELD SAMPLES |
|---|---|---|---|
| | *p.p.m.* | *p.p.m.* | |
| 0 | 0 | ..... | 28 |
| 1 | 0– 30 | 0– 50 | 28 |
| 2 | 31– 50 | 51– 100 | 8 |
| 3 | 51– 75 | 101– 200 | 12 |
| 4 | 76–125 | 201– 500 | 8 |
| 5 | 126–175 | 501–1,000 | 2 |
| 6 | 176–230 | 1,001–2,000 | 2 |
| 7 | over 230 | over 2,000 | 1 |
| Total | ..... | ..... | 89 |

[1] Based on curve of accuracy of the sampling method, as discussed in the text.

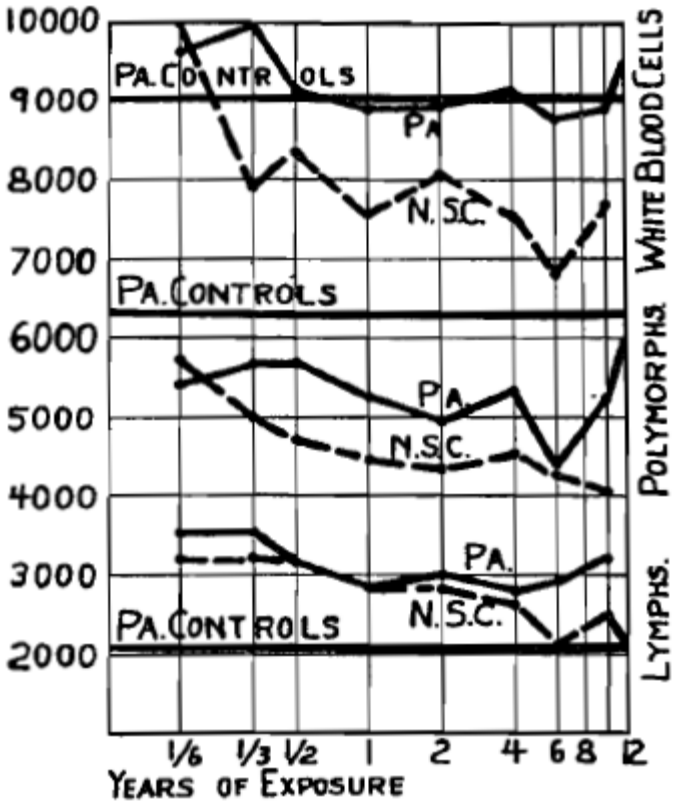


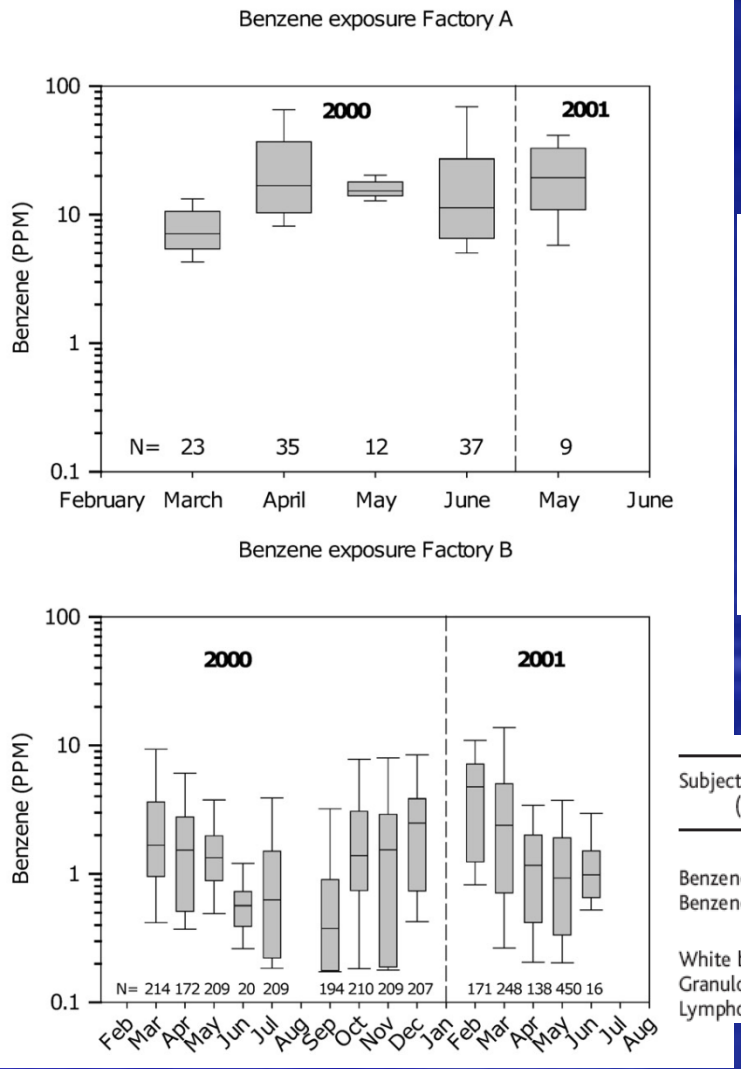

FIG. 4.—Influence on white cell counts of continuing exposure to lacquer fumes.

# Benzene exposure and hematopoietic damage in Chinese workers

Rothman *et al*., *PNAS, 1995*
Rothman *et al*., *Am J Ind Med, 1996*
Rothman *et al*., *Env Health Perspect 1996*
Rothman *et al*., *Cancer Res, 1997*
Rothman *et al*., *Occ Env Med, 1998*
Smith *et al*., *Cancer Res*, 1998
Smith *et al*., *PNAS*, 2000
Yeowell-O'Connell *et al*., *Carcinogenesis*, 1998
Waidyanatha *et al*., *Chem Biol Interact*, 1998
Yeowell-O'Connell *et al*., *Cancer Epidemiol Biomarkers Prev*, 2001
Waidyanatha *et al*., *Carcinogenesis*, 2001
Waidyanatha *et al*., *Analyt Biochem*, 2004
Rappaport *et al*., *J Chromatog B*, 2002

*Vermeulen et al., Ann Occup Hyg,* 2004
*Lan et al., Science*, 2004
*Lan et al., Zhang et al., Chem-Biol Interact*, 2005
*Lan et al., Cancer Res*, 2005
*Vermeulen et al., PNAS, 2005*
Shen *et al., Carcinogenesis*, 2006
Lan *et al., Carcinogenesis*, 2009
McHale *et . al, Genomics*, 2009
Rappaport *et al., Cancer Res,* 2002
Kim *et al., Carcinogenesis*, 2006
Kim *et al., Cancer Epidemiol Biomarkers Prev,* 2006
Kim *et al*., *Pharmacokinetics Genomics*, 2007
Lin *et al*., *Env Health Perspect,* 2007
Rappaport *et al*., *Env Health Perspect* 2009
Rappaport *et al*., *Chem Biol I nteract,* 2010

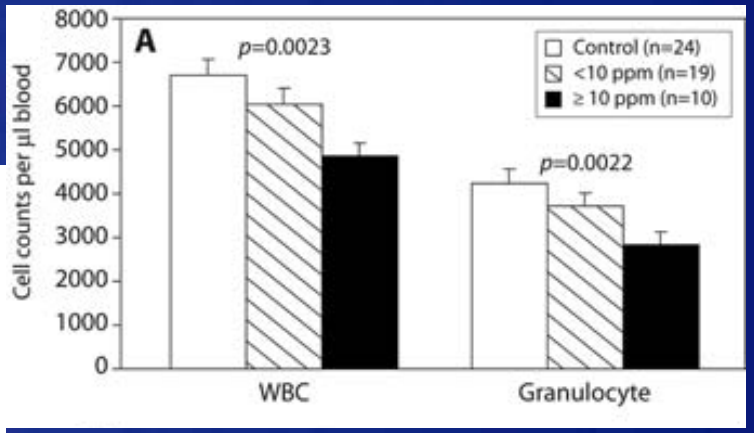
# Benzene exposure and hematopoietic damage in Chinese workers




*Vermeulen et al., Ann Occup Hyg, 48:105 (2004)*

3000 personal measurements

3M Organic Vapor Monitor




| Subject category (*n*)* | Controls (140) | <1 ppm (109) | 1 to <10 ppm (110) | ≥10 ppm (31) | *P* for <1 ppm vs. controls† |
|---|---|---|---|---|---|
| | | *Benzene exposure* | | | |
| Benzene air level (ppm)‖ | <0.04 | 0.57 (0.24) | 2.85 (2.11) | 28.73 (20.74) | |
| Benzene urine (µg/liter)¶ | 0.382 (1.24) | 13.4 (18.3) | 86.0(130) | 847(1250) | |
| | | *Peripheral blood cell counts#* | | | |
| White blood cells (WBC)** | 6480 (1710) | 5540 (1220) | 5660 (1500) | 4770 (892) | <0.0001 |
| Granulocytes | 4110 (1410) | 3360 (948) | 3480 (1170) | 2790 (750) | <0.0001 |
| Lymphocytes†† | 2130 (577) | 1960 (541) | 1960 (533) | 1800 (392) | 0.018 |

*Lan et al., Science, 306:1774 (2004)*

# Benzene metabolism

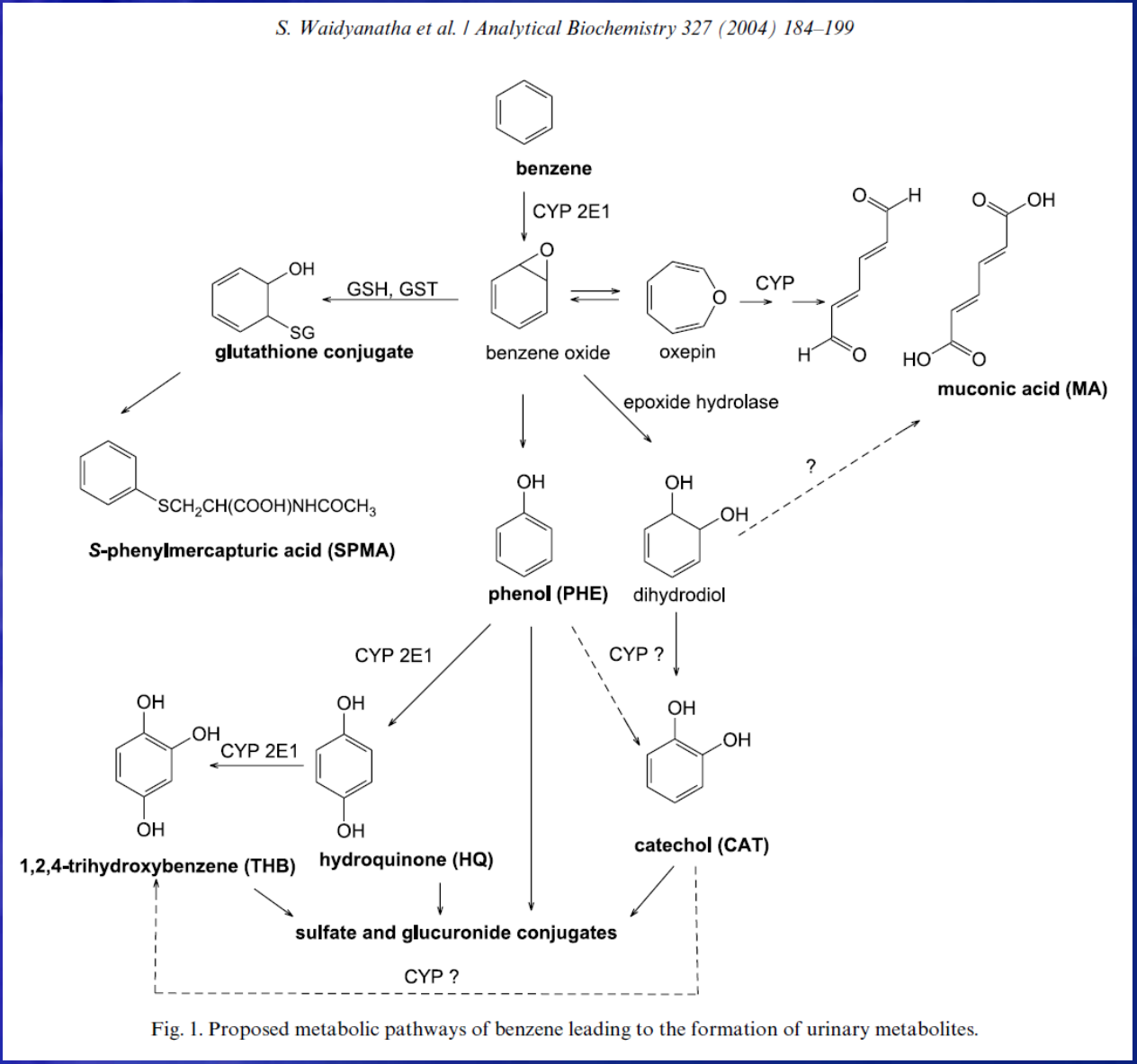



Fig. 1. Proposed metabolic pathways of benzene leading to the formation of urinary metabolites.

# Benzene metabolites in shoe workers

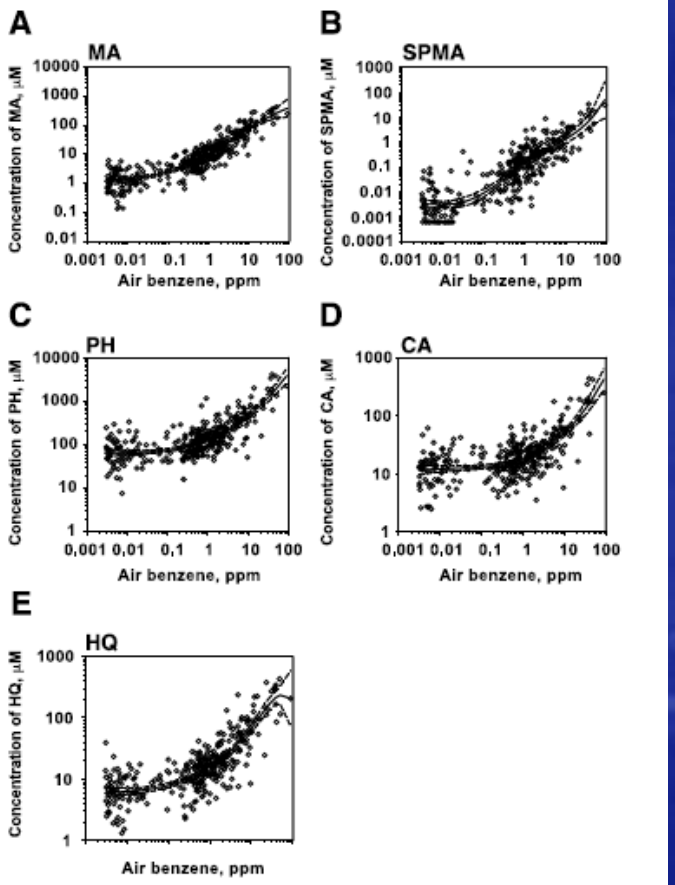
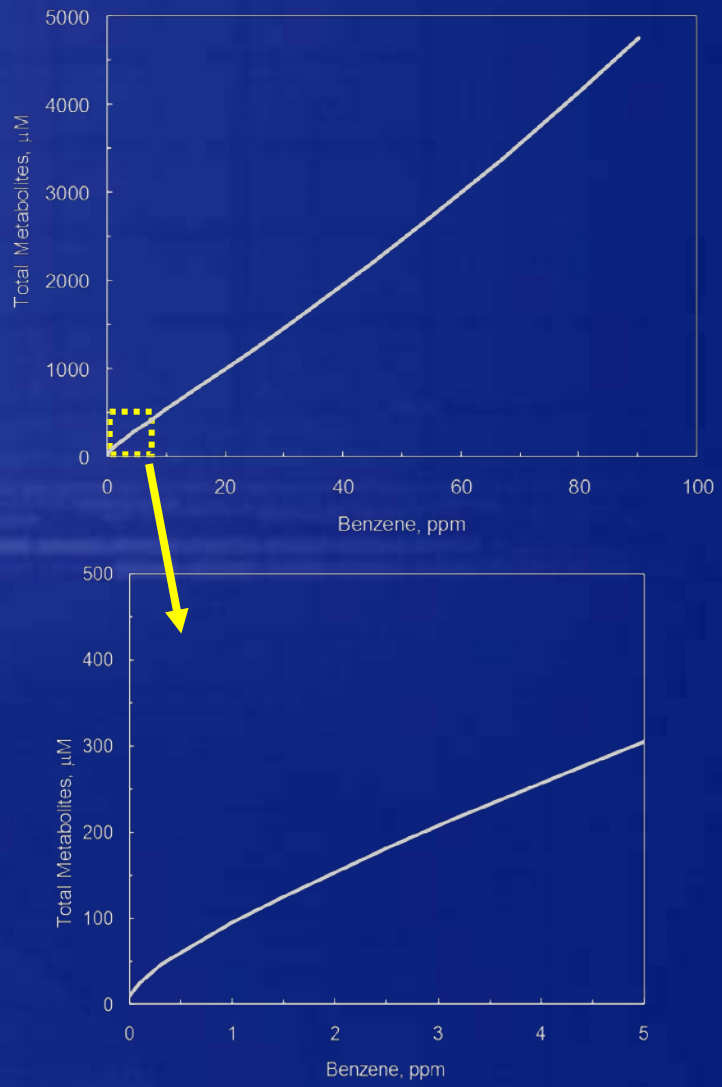


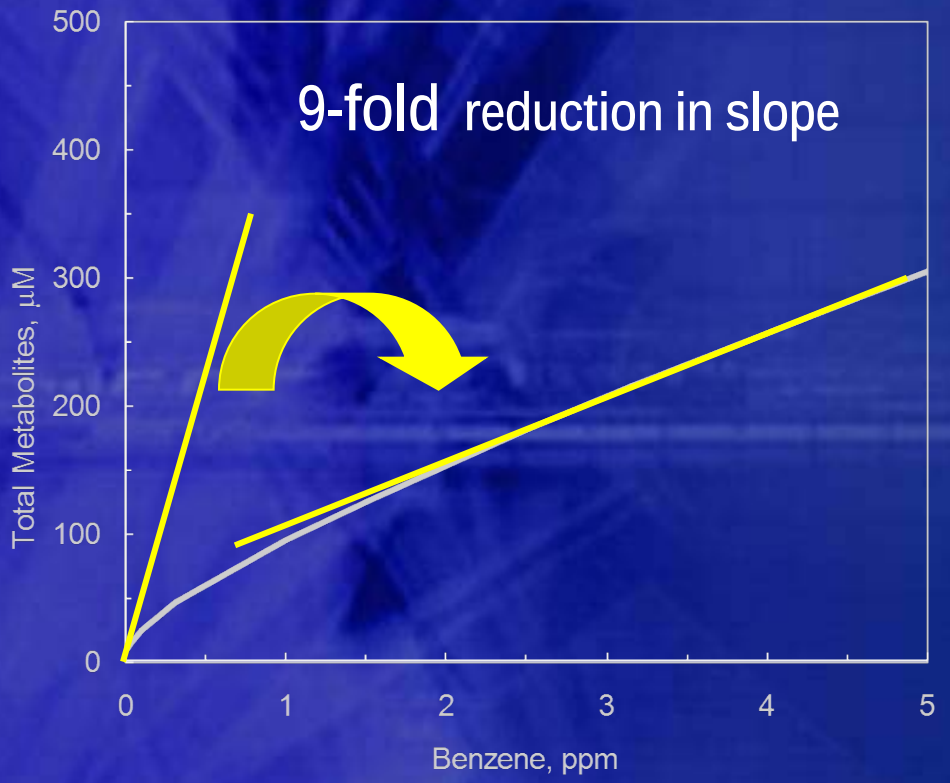

**Figure 2.** Scatter plots of levels of benzene metabolites versus benzene exposure for 326 workers with the greatest air exposures to benzene (*circles*). Expected mean trends (*solid curves*) and their 95% confidence intervals (*dashed curves*) from natural spline models. MA, *E,E*-muconic acid; PH, phenol; CA, catechol; HQ, hydroquinone.

*Dose-specific benzene metabolism (mean trend of total metabolites adjusted for background levels)*




*S. Kim et al., Cancer Epidemiol Biomarkers Prevent , 2006, 15(11): 2246-2252.*

# Dose-specific benzene metabolism



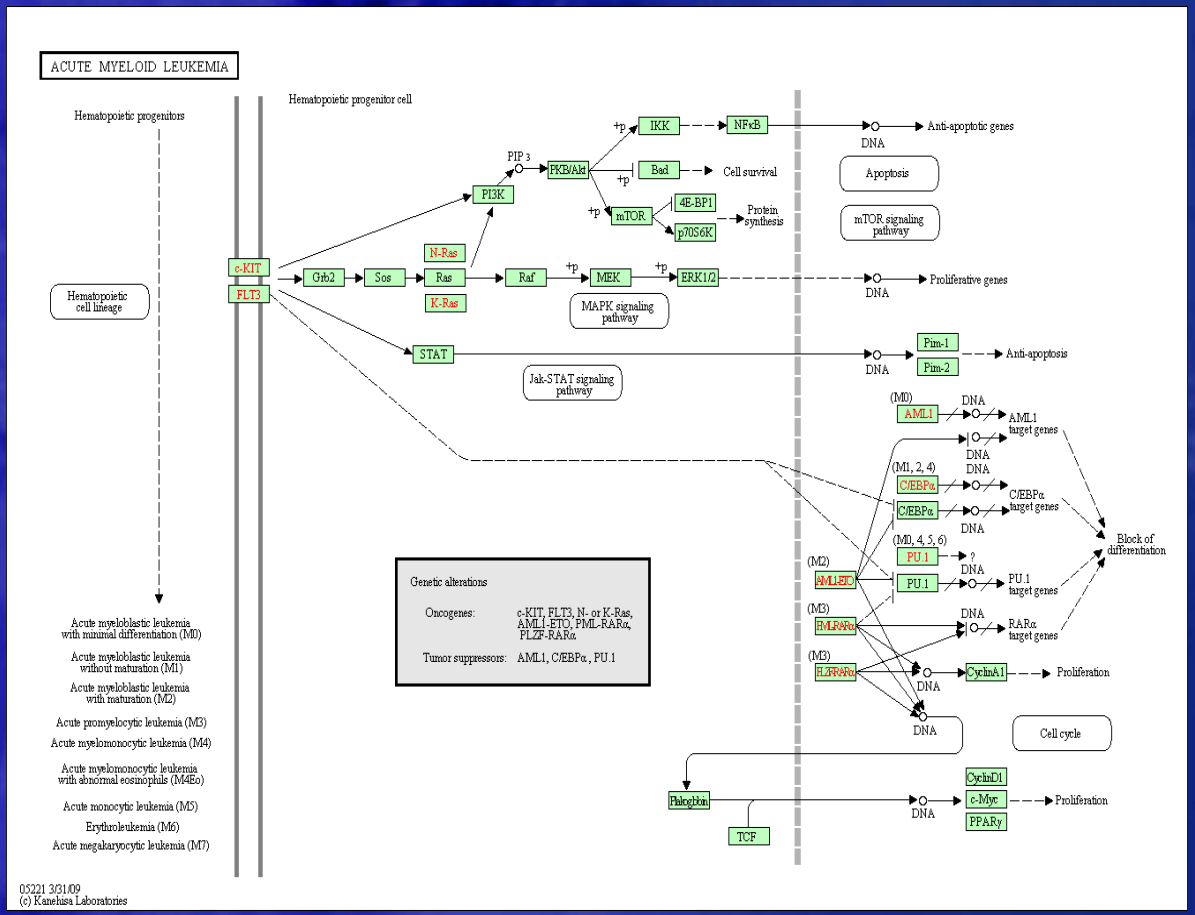

*Evidence for an unknown high-affinity pathway responsible for more than 70% of metabolism below 1 ppm*

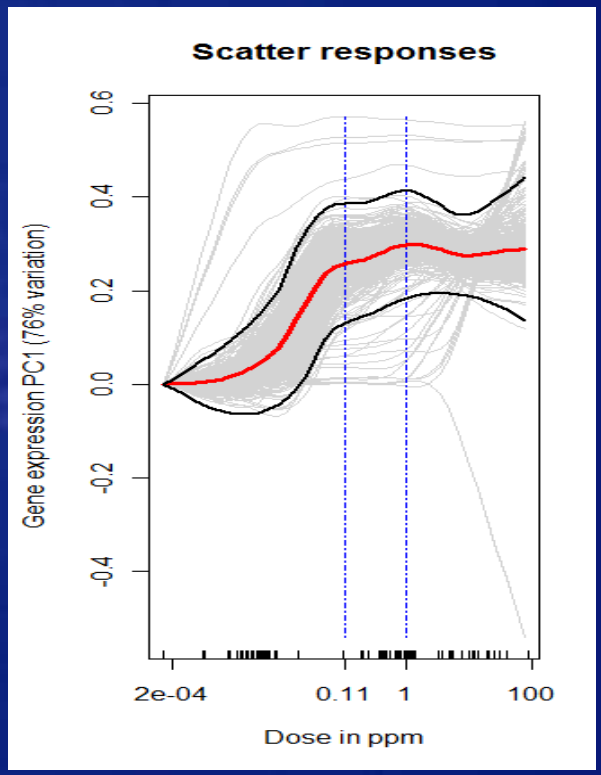
**Rappaport *et al. Env Health Perspect* 117: 946-952 (2009)**
**Rappaport *et al. Chem Biol Interact* 184: 189-95 (2010)**

*Mean trend based upon data from: S. Kim et al., Cancer Epidemiol Biomarkers Prevent, 2006*

# Acute myeloid leukemia (AML) pathway

1st principal comp. for AML pathway response in 125 Chinese shoe workers and controls



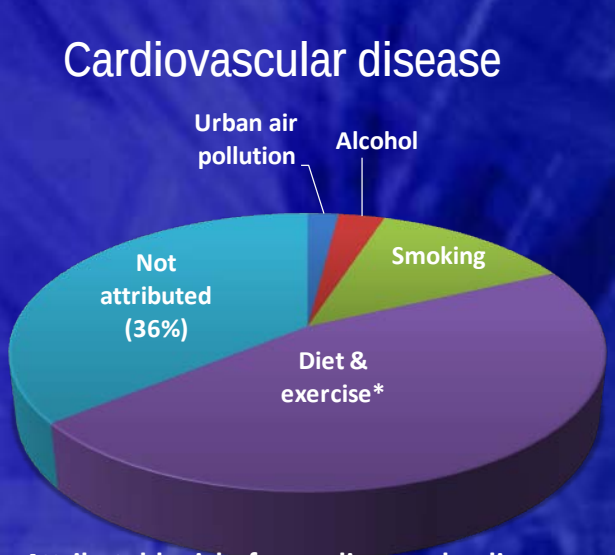

R. Thomas, L. Zhang, M. Smith *et al.*, unpublished

SM Rappaport

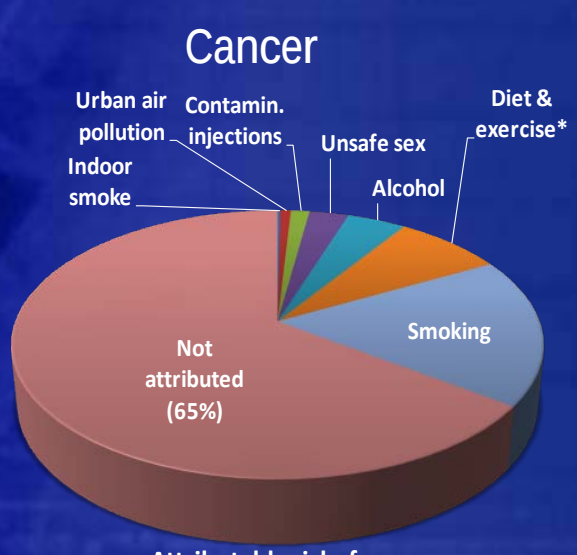
# Strengths and limitations of knowledge-driven research

- *Provides information about human health effects and kinetics of known causes of disease, such as exposure to benzene or particular genes*
- Requires testable hypotheses related to causative factors, kinetics, dose-response, mechanisms, etc.
- *Cannot provide information about unknown genetic or environmental factors (disease etiology)*

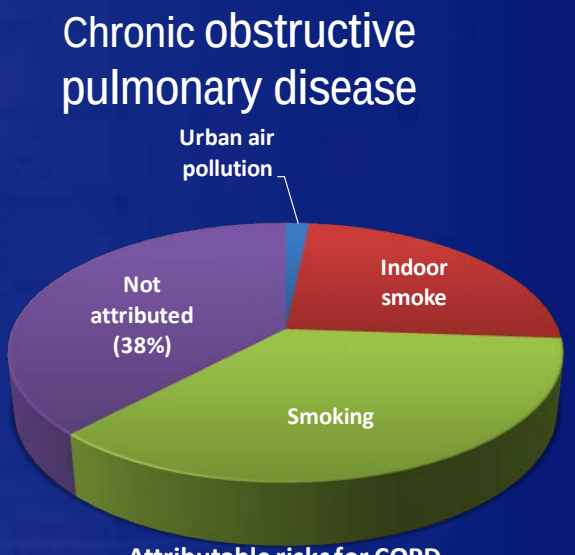
# Attributable risks of environmental factors




Attributable risks for cardiovascular disease (worldwide, joint PAF=64%)




Attributable risks for cancer (worldwide, all tumor types, joint PAF=35%)




Attributable risks for COPD (worldwide, joint PAF=62%)

*70% of chronic-disease risk due to unknown exposures*



SM Rappaport

Data from Ezzati *et al.*, "Comparative Quantification of Mortality and Burden of Disease Attributable to Selected Risk Factors," *Global Burden of Disease and Risk Factors*, Chapter 4, WHO, 2006.

# *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.

*The exposome concept promotes data-driven research to investigate environmental causes of disease.*

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

Cancer Epidemiol Biomarkers Prev 2005;14(8). August 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**

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

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 exposure 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 can contribute to chronic diseases and should be investigated collectively rather than separately.

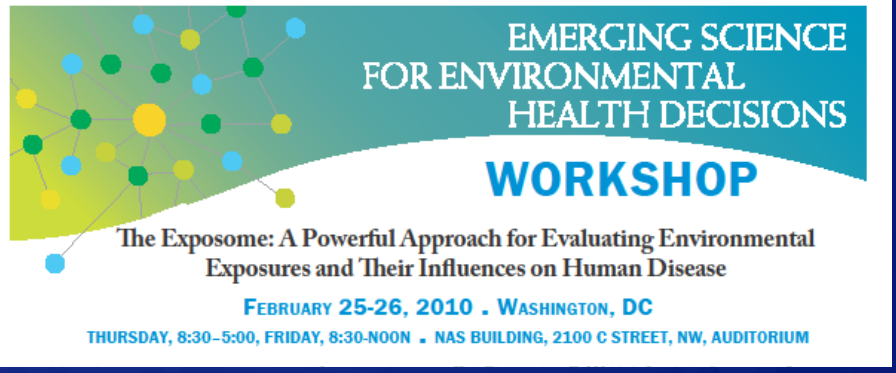
To develop a more cohesive view of environmental exposure, it is important to recognize that toxic effects are mediated through chemicals that alter critical molecules, cells, and physiological processes inside the body. Thus, it would be reasonable to consider the "environment" as the body's internal 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 to changes in external and internal sources, aging, infections, life-style, stress, psychosocial fa

Th
ity of
tion o

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



Published by AAAS




*Journal of Exposure Science and Environmental Epidemiology* (2011) 21, 5–9


www.nature.com/jes

REVIEW

## Implications of the exposome for exposure science

STEPHEN M. RAPPAPORT

*School of Public Health, University of California, Berkeley, California, USA*

*Emerging Technologies for Characterizing Individual Exposomes* (NAS-Dec. 8-9 2011)

# Capturing exogenous and endogenous exposures

EPIDEMIOLOGY

## Environment and Disease Risks

**Stephen M. Rappaport and Martyn T. Smith**

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

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 exposure 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 can contribute to chronic diseases and should be investigated collectively rather than separately.

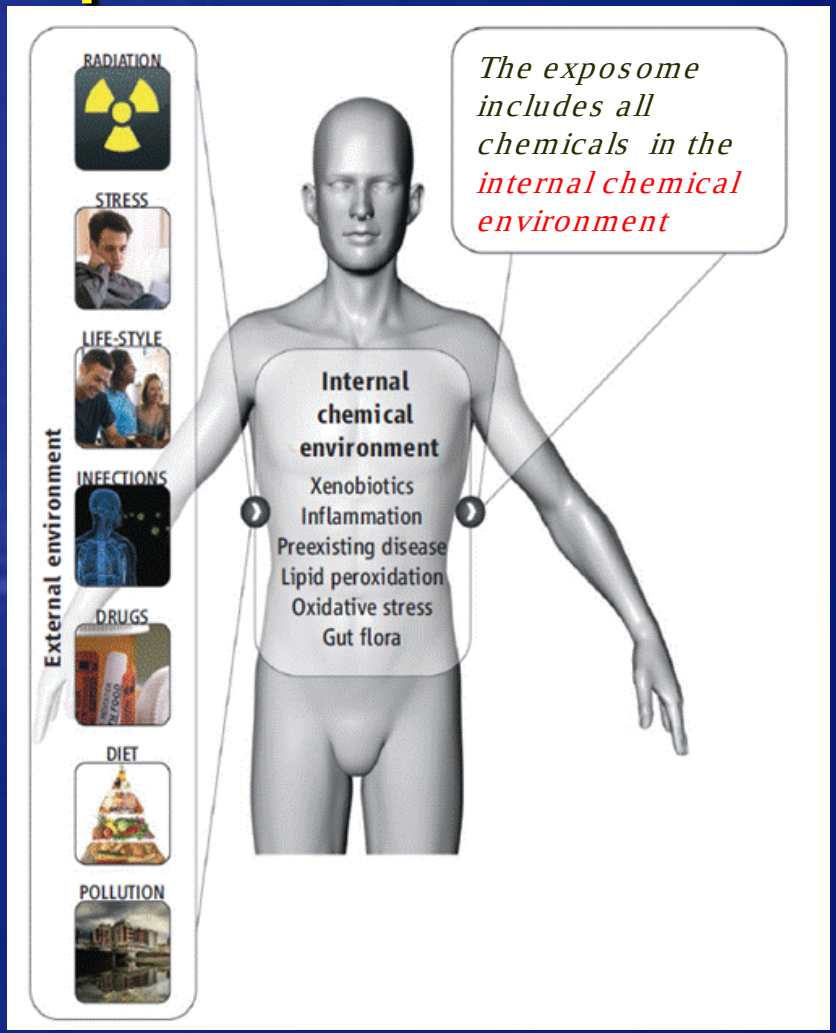
To develop a more cohesive view of environmental exposure, it is important to recognize that toxic effects are mediated through chemicals that alter critical molecules, cells, and physiological processes inside the body. Thus, it would be reasonable to consider the "environment" as the body's internal 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 to changes in external and internal sources, aging, infections, life-style, stress, psychosocial factors, and preexisting diseases.

The term "exposome" refers to the totality of environmental exposures from conception onwards, and has been proposed to be a

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



Published by AAAS

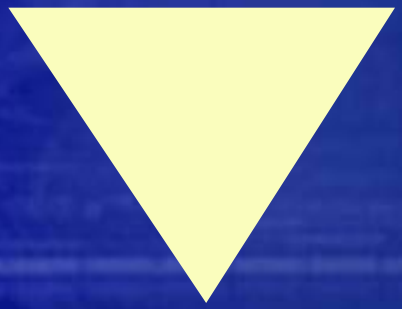



*S.M. Rappaport and M.T. Smith, Science, 2010: 330:460-461*

# Polar views of 'environmental exposure'

Exposome: '*everything except the genes*'

70% of chronic-disease risk



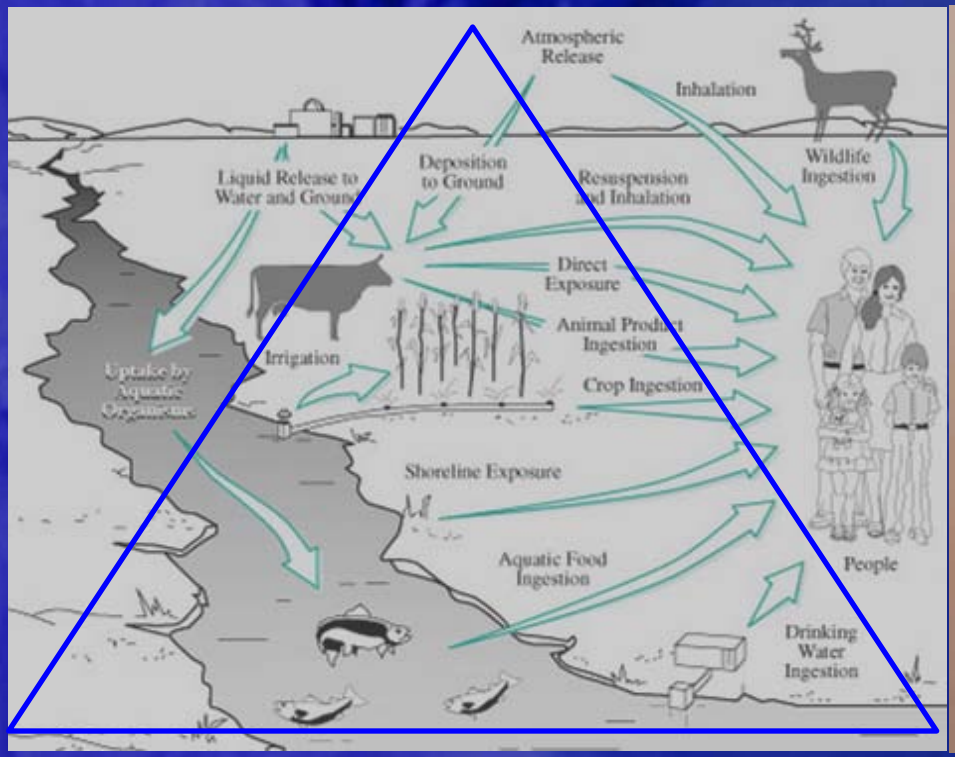
5% of chronic-disease risk

Environmental science: *'pollutants in air and water'*

# Characterizing the exposome: bottom up or top-down?

## *Bottom-up strategy*

Identify important exogenous exposures




Measure all chemicals in air, water & food

Background graphic: United States Department of Energy, Hanford site

## *Top-down strategy*

Measure all chemicals in blood

Identify all important exposures

Background graphic: http://www.flickr.com/photos/paulieparker/246707763/

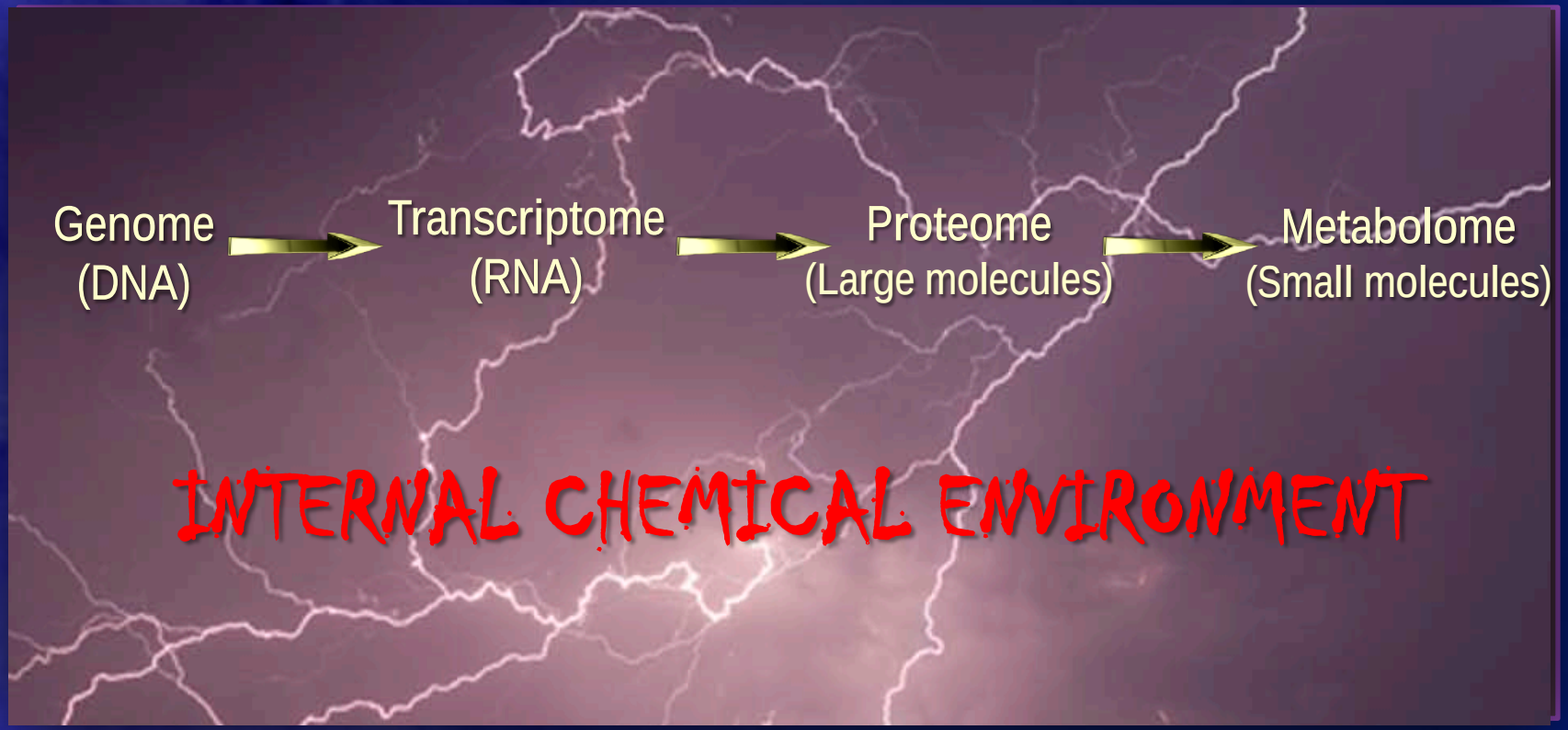
S.M. Rappaport, *JESEE, 2011, 21: 5-9.*

# Exposome-wide association studies (EWAS)

By applying EWAS to biospecimens from healthy and diseased subjects, it should be possible to discover causal environmental (i.e. non-genetic) exposures.

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

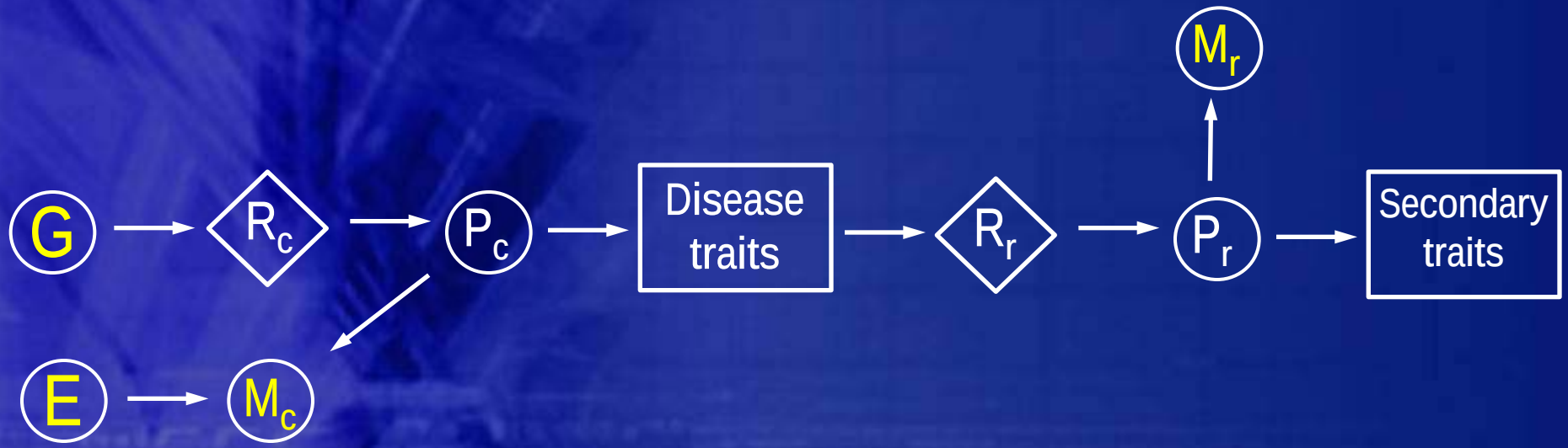
# The molecular basis of life (and disease)



# Omic connections

Causal pathway (c)

Reactive pathway (r)

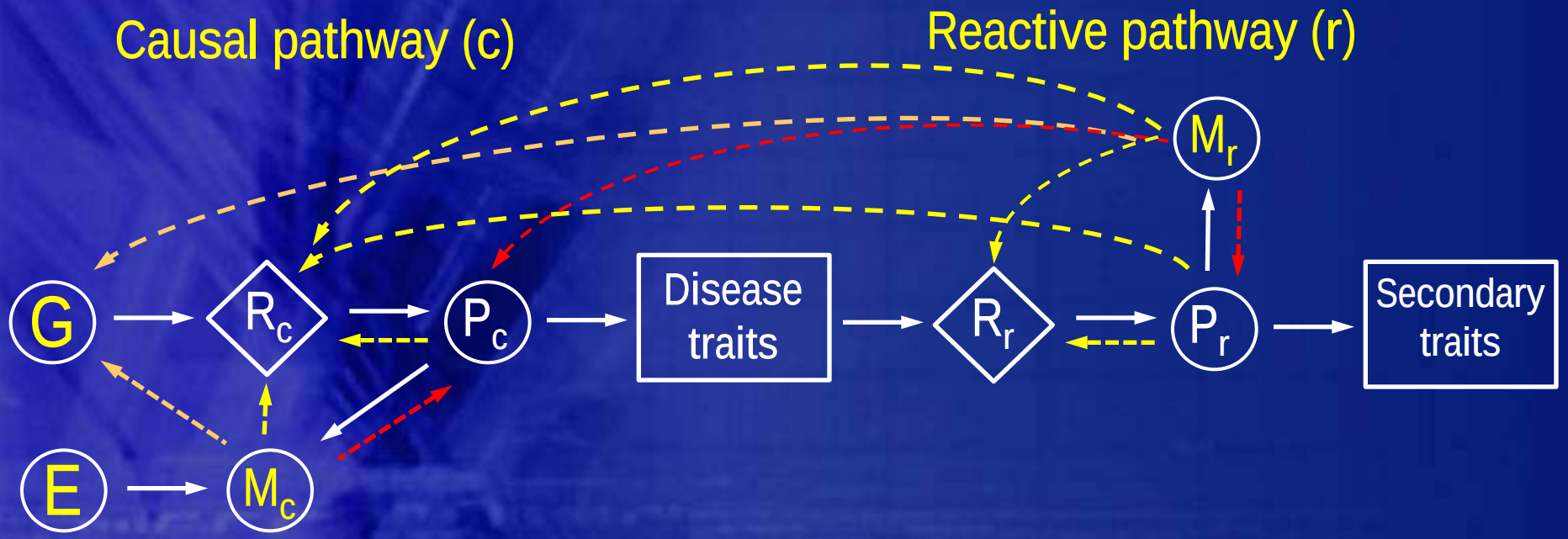



G = genome
E = environment
R = transcriptome (gene expression)
P = proteome (protein expression)
M = metabolome (all small molecules and metals)



Motivated by: E. Shadt *et al., Nat Gen,* 2005, 37: 710-717

# Omic connections

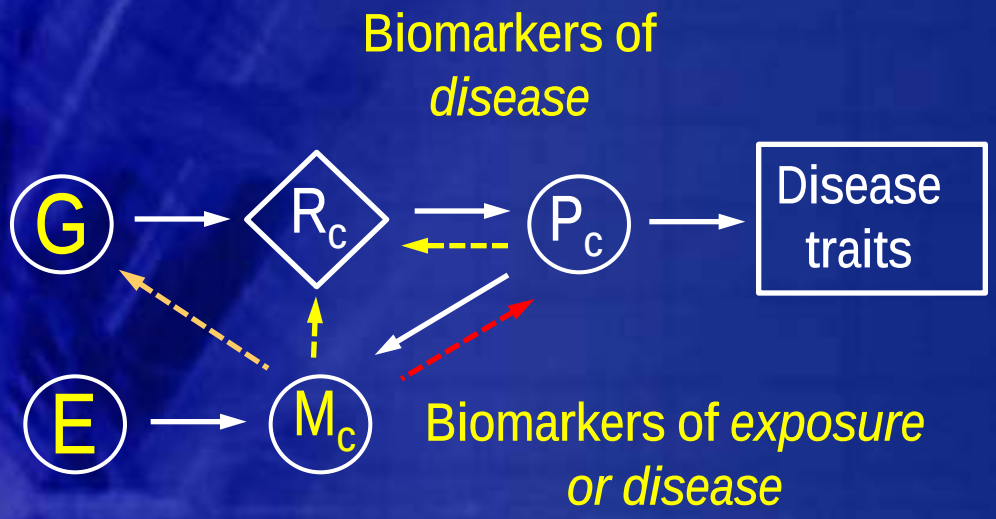



G = genome
E = environment
R = transcriptome (gene expression)
P = proteome (protein expression)
M = metabolome (all small molecules and metals)

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

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

# Focusing EWAS

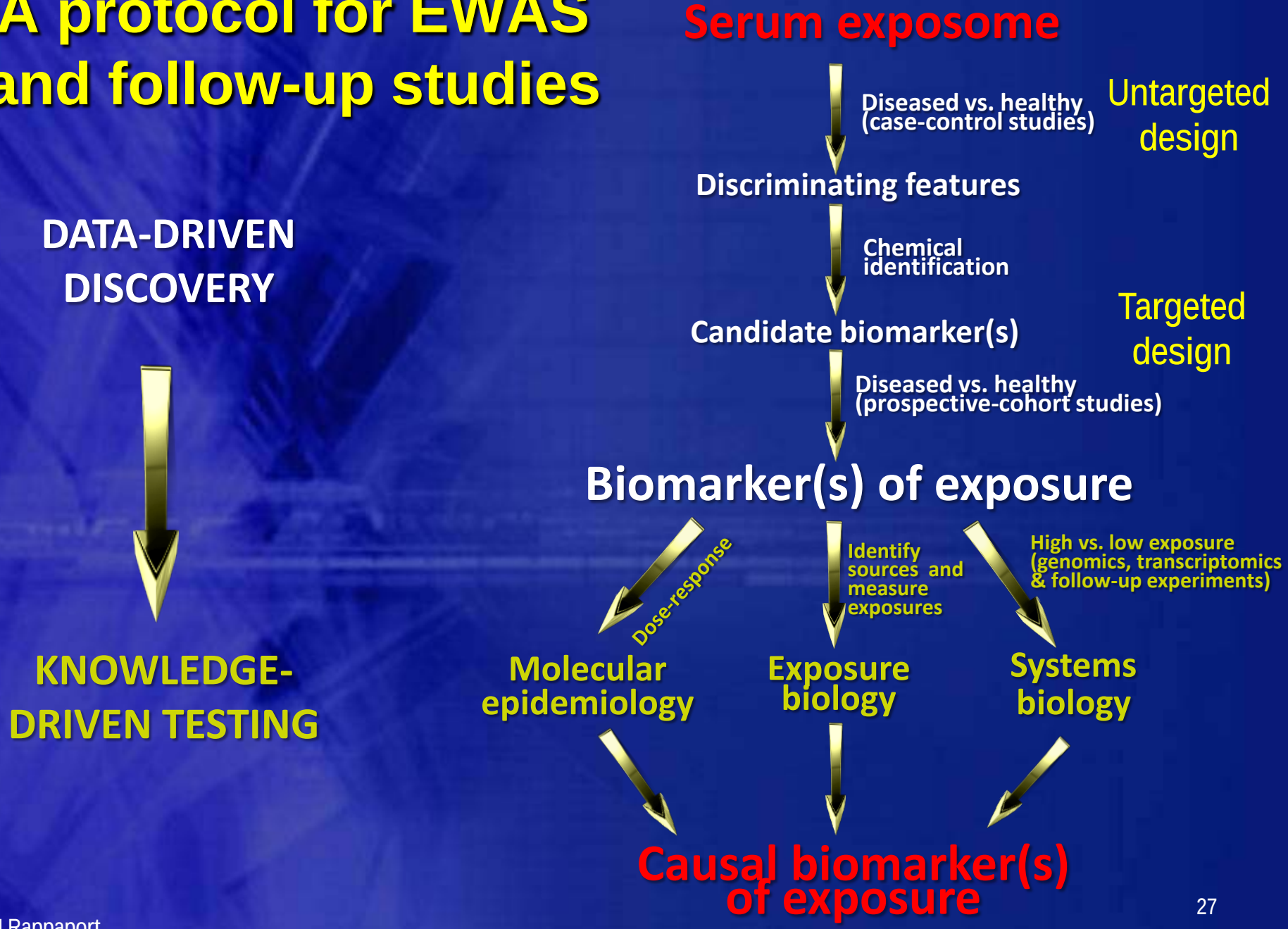



***Causal exposures operate primarily through small molecules ($M_c$) and proteins ($P_c$).***

- EWAS require metabolomics and proteomics *(e.g., serum exposome)*
- The transcriptome ($R_c$) provides no structural information about exposures and is more useful for identifying biomarkers of disease.

# The serum exposome

# A protocol for EWAS and follow-up studies



# Candidate serum biomarkers

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

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

Modest numbers of subjects

Candidate biomarkers

# An important example

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

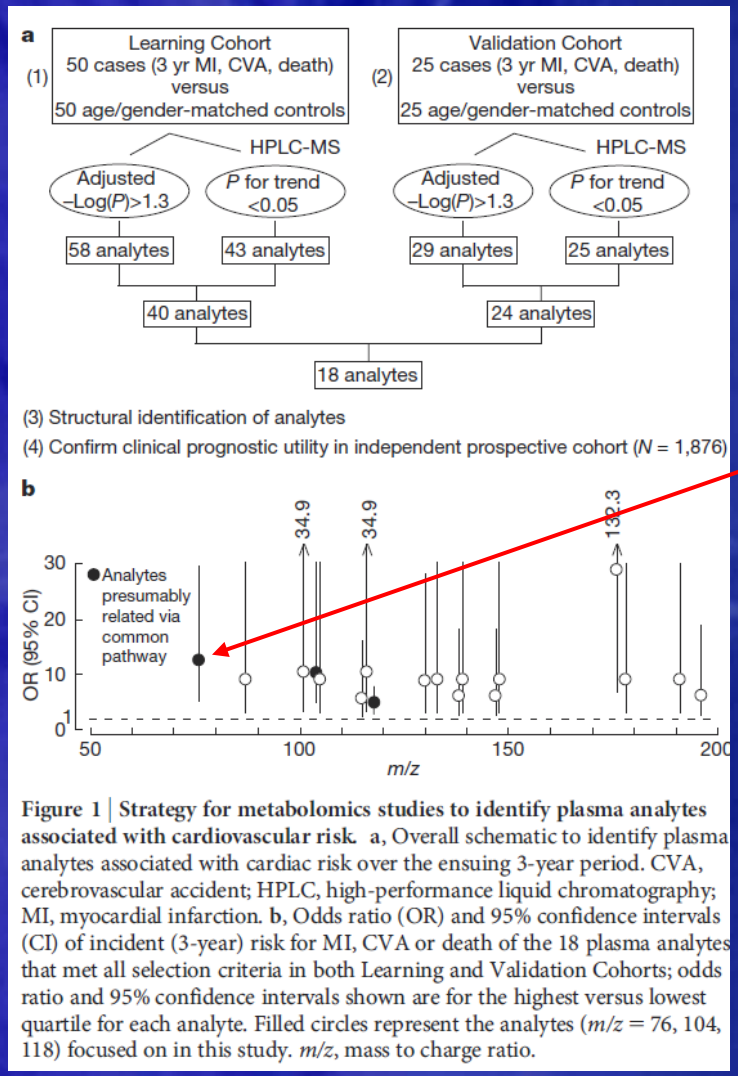
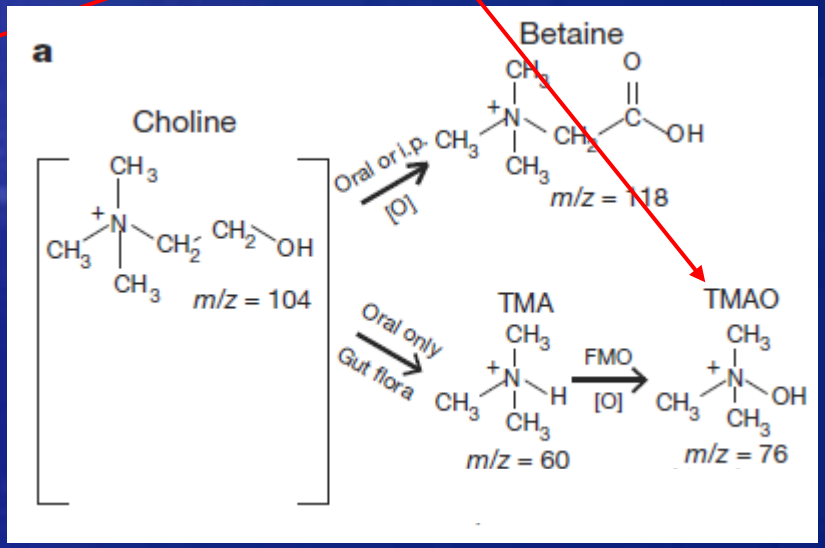



**Figure 1 | Strategy for metabolomics studies to identify plasma analytes associated with cardiovascular risk.** **a**, Overall schematic to identify plasma analytes associated with cardiac risk over the ensuing 3-year period. CVA, cerebrovascular accident; HPLC, high-performance liquid chromatography; 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.

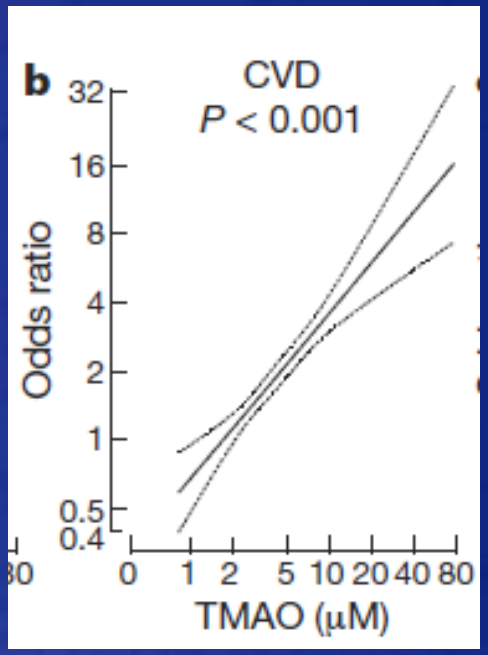
Trimethylamine oxide (TMAO)

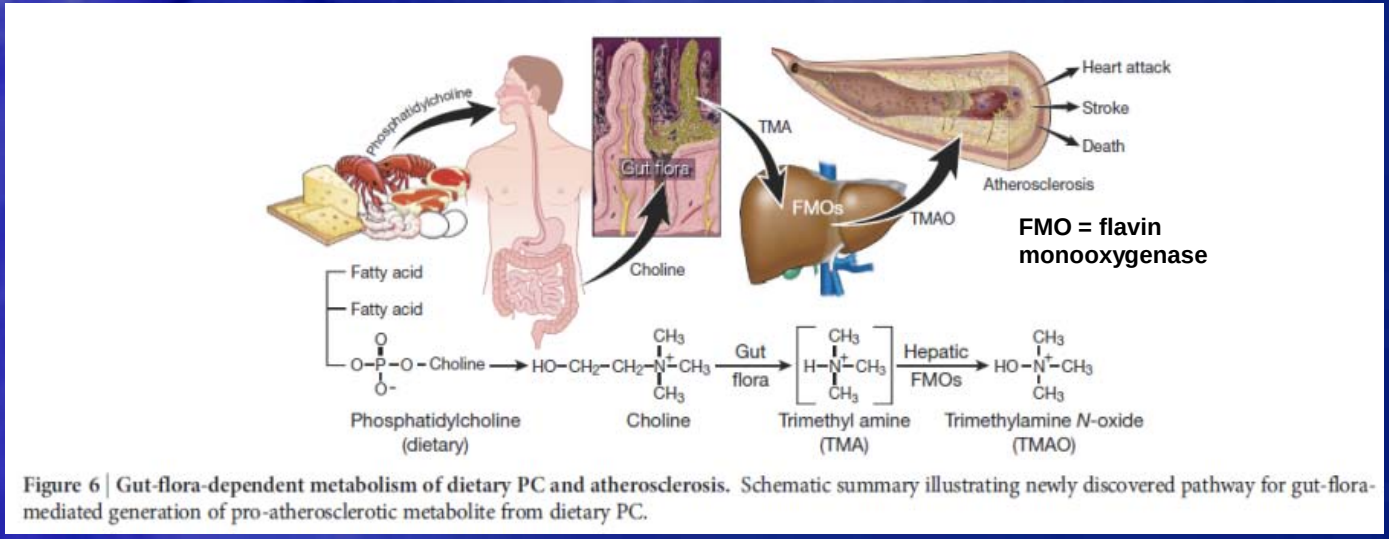



Wang *et al. Nature* (2011) 472: 57-63.

# Choline biomarkers and CVD risk

*Targeted* analyses of TMAO in serum from 1870 CVD patients and controls



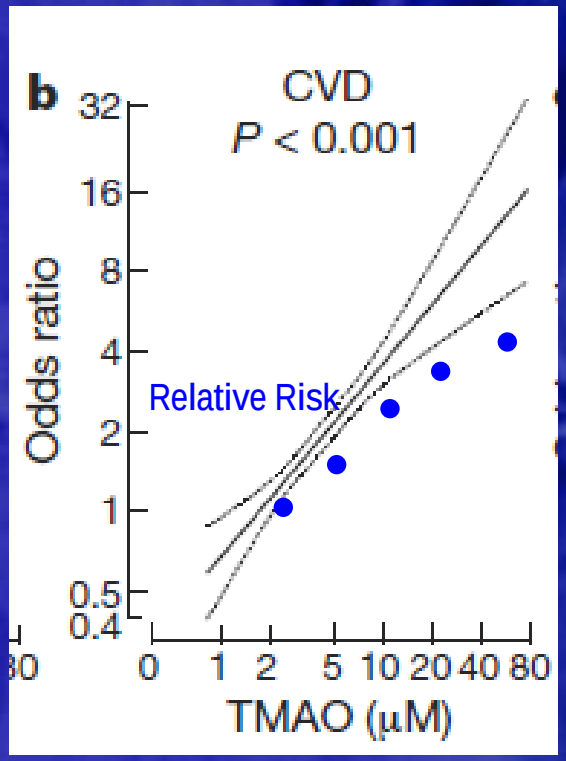

**FMO = flavin monooxygenase**

**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.

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

# Context for the findings

Cardiovascular disease risks from exposure to TMAO




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

Cardiovascular disease risks from exposure to cigarette smoke and PM (15% of global burden of CVD)

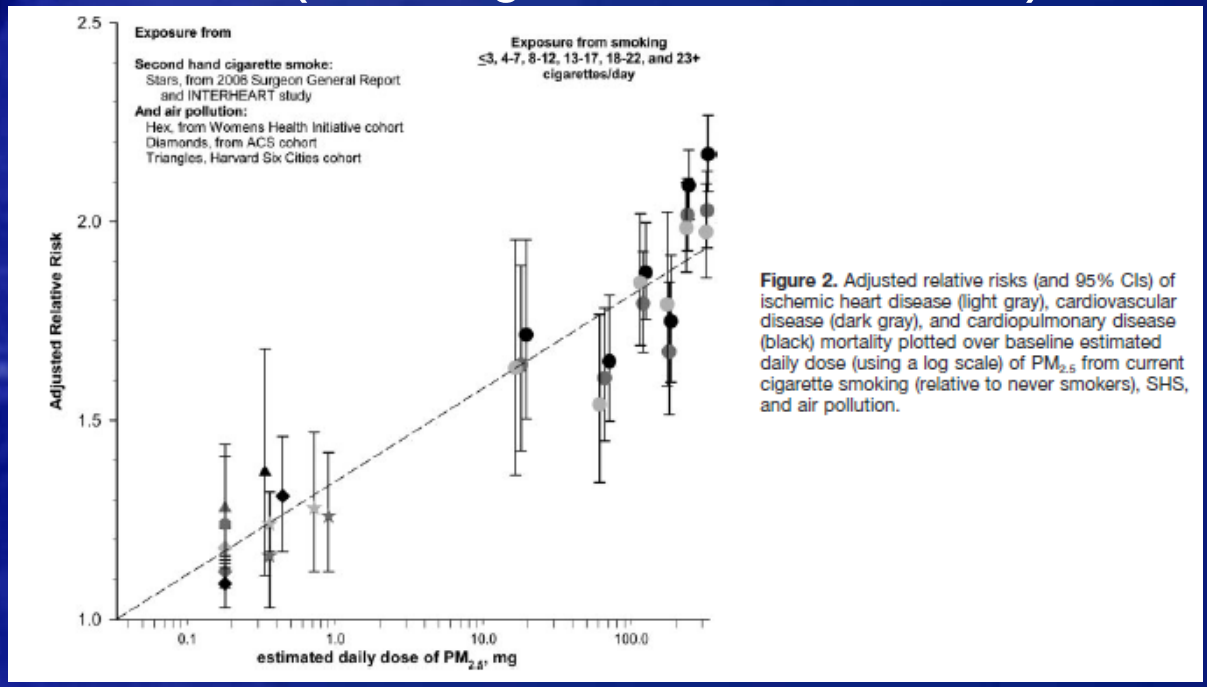



**Figure 2.** Adjusted relative risks (and 95% CIs) of ischemic heart disease (light gray), cardiovascular disease (dark gray), and cardiopulmonary disease (black) mortality plotted over baseline estimated daily dose (using a log scale) of $PM_{2.5}$ from current cigarette smoking (relative to never smokers), SHS, and air pollution.

CA Pope III *et al. Circulation* (2009) 120: 941-948.

# Two exposure agendas

- Exposure assessment
  - For causative or suspicious chemicals
  - Knowledge-driven, targeted designs
  - Provides feedback for public health
- The exposome
  - For disease etiology
  - Data-driven, untargeted designs
  - Focus on small molecules and proteins
  - Proof of concept has been established

# ‘Chemicals’ cause disease – but which ones?

- EPA priority pollutants: *129*
- CDC environmental pollutants: *300*
- Occupational exposure limits: *500*
- Human metabolome: *8,000*
- High-volume chemicals: *80,000*
- Chemicals from the human microbiome (3M genes): *300,000 ?*

# What is needed?

1. High-throughput omics (mainly metabolomics and proteomics) and targeted assays
   - State-of-the-art equipment (mostly mass spectrometry)
2. Prospective -cohort studies with bio-repositories
3. Advanced bioinformatics and statistics
4. Interdisciplinary research teams (epidemiology, medicine, *analytical chemistry*, toxicology, exposure science and statistics/bioinformatics)

# 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

SM Rappaport

Genes & Environment Laboratory