

Applying New Biotechnologies to the Study of Occupational Cancer—A Workshop Summary

Mark Toraason,^{1,*} Richard Albertini,² Steven Bayard,³ William Bigbee,⁴ Aaron Blair,⁵ Paolo Boffetta,⁶ Stefano Bonassi,⁷ Steven Chanock,⁵ David Christiani,⁸ David Eastmond,⁹ Samuel Hanash,¹⁰ Carol Henry,¹¹ Fred Kadlubar,¹² Frank Mirer,¹³ Daniel Nebert,¹⁴ Stephen Rapport,¹⁵ Kathleen Rest,¹⁶ Nathaniel Rothman,⁵ Avima Ruder,¹ Russell Savage,^{1,*} Paul Schulte,^{1,*} Jack Siemiatycki,¹⁷ Peter Shields,¹⁸ Martyn Smith,¹⁹ Paige Tolbert,²⁰ Roel Vermeulen,⁵ Paolo Vineis,²¹ Sholom Wacholder,⁵ Elizabeth Ward,^{22,*} Michael Waters,²³ and Ainsley Weston²⁴

¹National Institute for Occupational Safety and Health, Cincinnati, Ohio, USA; ²University of Vermont, Burlington, Vermont, USA;

³Occupational Safety and Health Administration, Washington, DC, USA; ⁴University of Pittsburgh Cancer Institute, Pittsburgh,

Pennsylvania, USA; ⁵National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda,

Maryland, USA; ⁶International Agency for Research on Cancer, Lyon, France; ⁷National Institute for Research on Cancer, Genoa, Italy;

⁸Harvard School of Public Health, Boston, Massachusetts, USA; ⁹University of California, Riverside, California, USA; ¹⁰University of

Michigan, Ann Arbor, Michigan, USA; ¹¹American Chemistry Council, Arlington, Virginia, USA; ¹²National Center for Toxicological

Research, Jefferson, Arkansas, USA; ¹³Health and Safety Department, International Union, United Auto Workers, Detroit, Michigan,

USA; ¹⁴University of Cincinnati, Cincinnati, Ohio, USA; ¹⁵University of North Carolina, Chapel Hill, North Carolina, USA; ¹⁶National

Institute for Occupational Safety and Health, Washington, DC, USA; ¹⁷University of Montreal, Montreal, Canada; ¹⁸Georgetown

University, Washington, DC, USA; ¹⁹University of California, Berkeley, California, USA; ²⁰Emory University, Atlanta, Georgia, USA;

²¹University of Torino, Torino, Italy; ²²American Cancer Society, Atlanta, Georgia, USA; ²³National Institute of Environmental Health

Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, North Carolina, USA;

²⁴National Institute for Occupational Safety and Health, Morgantown, West Virginia, USA

As high-throughput technologies in genomics, transcriptomics, and proteomics evolve, questions arise about their use in the assessment of occupational cancers. To address these questions, the National Institute for Occupational Safety and Health, the National Cancer Institute, the National Institute of Environmental Health Sciences, and the American Chemistry Council sponsored a workshop 8–9 May 2002 in Washington, DC. The workshop brought together 80 international specialists whose objective was to identify the means for best exploiting new technologies to enhance methods for laboratory investigation, epidemiologic evaluation, risk assessment, and prevention of occupational cancer. The workshop focused on identifying and interpreting markers for early biologic effect and inherited modifiers of risk. **Key words:** biomarkers, chemical exposure, epidemiology, gene–environment interactions, genomics, occupational cancer, polymorphisms, proteomics, risk assessment, toxicogenomics. *Environ Health Perspect* 112:413–416 (2004). doi:10.1289/txg.6343 available via <http://dx.doi.org/> [Online 14 January 2004]

The National Occupational Research Agenda (NORA) was established by the National Institute for Occupational Safety and Health (NIOSH) in 1996 with input from more than 500 organizations and individuals. Since its inception, NORA has become a prototype for advancing research in the area of worker safety and health. It is the largest single source of support for 21 occupational research priority areas. The goal of the cancer research methods priority area is to identify, evaluate, and recommend new technologies designed to better control and help investigators understand occupational carcinogenesis. This ongoing effort has coincided with a revolution in biology. The human genome map has just been completed, and it has arrived with a wealth of new biologic methods that drove or were driven by the goal to complete the mapping. Despite this increase in new technologies and methods, they have yet to be applied fully to occupational cancer research (Ward et al. 2003). To address this gap, a workshop, “Applying New Biotechnology to the Study of Occupational Cancer,” was sponsored by NIOSH in conjunction with the National Cancer Institute (NCI), the

National Institute of Environmental Health Sciences (NIEHS), and the American Chemistry Council (ACC). The workshop brought together researchers studying worker populations and those developing and validating new biotechnologies. The workshop focused on four topics: *a*) the challenge of applying new biotechnologies to the study of occupational cancer, *b*) markers of early biologic effect, *c*) inherited modifiers of risk, and *d*) applying genetic biomarkers to human studies.

Challenge of Applying New Biotechnologies to the Study of Occupational Cancer

Epidemiology in the 21st century. A primary challenge in applying new biotechnologies to occupational cancer research is including them within the framework of classic studies of occupational exposure and effect. Historically, occupational studies have helped identify many of the recognized environmental carcinogens, resulting in reduced exposure for both workers and the general population. Studies of carcinogens in occupational populations were facilitated by the availability of records of employees

and the high and prolonged exposures they experienced, which often resulted in high relative risks for specific cancers. Challenges to occupational cancer epidemiology in the 21st century relate to the changing nature of the workplace and the complexity of the exposures. As a result of regulations and industry efforts, exposure levels are much lower than in the past. Many exposures are mixtures, and many industries involve exposures to an ever-changing and diverse array of substances. These changes create the need for more sensitive measures to detect cancer risks. To move the field of occupational cancer research forward, it will be necessary to *a*) conduct more studies of occupational cancer among women and minorities, as these populations have been ignored in the past; *b*) perform quantitative exposure assessments, as qualitative exposure assessments that rely on general classification of occupation are not good enough; *c*) examine interactions between occupational exposures and nonoccupational exposures, as cancer is a multi-faceted disease; *d*) focus on biologic tissues and mechanisms of action and incorporate gene–environmental assessments into traditional exposure disease paradigms used

Address corresponding to M. Toraason, NIOSH C23, 4676 Columbia Parkway, Cincinnati, OH, 45226 USA. Telephone: (513) 533–8207. Fax: (513) 533–8138. E-mail: mtoraason@cdc.gov

*These authors were responsible for developing the summary based on speaker presentations at the workshop.

The workshop was cosponsored by the National Institute for Occupational Safety and Health, the National Cancer Institute, the National Institute of Environmental Health Sciences, and the American Chemistry Council.

The authors declare they have no competing financial interests.

Received 18 March 2003; accepted 14 January 2004.

in epidemiology; and *e*) integrate epidemiology, toxicology, genetics, and quantitative exposure assessment.

The promise of new biotechnologies. The progression from exposure to disease is typically expressed as a continuum of environmental exposure → internal dose → biologically effective dose → early biologic effect → altered structure and function → and finally clinical disease. Each step is affected by a person's susceptibility, and the continuum provides multiple opportunities for application of biomarkers for early prediction of disease. Although applicable to biomarkers of exposure, the new technologies apply primarily to biomarkers of early effect and susceptibility. Biomarkers that measure early effect and susceptibility can be used in selecting study cohorts, assessing participant compliance, or determining intervention effectiveness. The effective use of biomarkers include optimizing reliability, precision, accuracy, and validity. Not all biomarkers are suitable for all purposes and are likely to be imperfect in any single setting. The greatest potential for new biomarkers of early effect in occupational hazard assessment lies in toxicogenomics, which can be defined as a field of study that examines how the entire genome responds to toxicants or other environmental hazards. Toxicogenomics applies genomics, gene and protein expression profiling, metabolite profiling or metabonomics, and bioinformatics to understand gene-environment interactions and disease. The many genomic-related technologies, often referred to simply as "omics," allow exploration of multiple interactions between genetic and environmental factors. This exploration will improve the understanding of mechanisms of action, clarify the use and limitations of surrogate models, enhance predictive toxicology and screening, and better characterize susceptible populations.

Technical and policy issues. The fields of toxicology and epidemiology are crucial for the assessment and management of the impact of chemicals on the safety, health, and welfare of workers. To realize this, a unified research agenda is needed for developing new technologies that will be used within a framework of toxicologic and epidemiologic principles. To accomplish this, the involvement of stakeholder communities is needed to address social, legal, and ethical issues (Henry et al. 2002). Technical and policy issues that need to be addressed include *a*) opportunities for shared learning in the public domain, *b*) accessibility to publicly held gene expression databases, *c*) understanding of the predictive capabilities of the technologies before widespread

application, *d*) availability of prevalence data, *e*) privacy and confidentiality concerns, *f*) security and discrimination issues, *g*) counseling for coping with genetic information, *h*) use and premature use of "omics" data, and *i*) defining the regulatory positions on "omics" data. Researchers need to focus on methods to assess gene expression in large populations, address statistical and bioinformatics issues, and use a multidisciplinary approach. These actions will lead to better integration of toxicology and epidemiology.

Markers of Early Biologic Effects

Rationale for assessing intermediate biomarkers. Epidemiologists have begun employing early markers of effect because of the challenges in using cancer as an outcome measure in occupational epidemiologic studies. There is minimal ambiguity when clinical disease is an end point, but there are limitations when studying cancer. The foremost problem is latency—the 10, 20, or even 30 years between exposure and disease. This latency moved researchers to develop the field of molecular epidemiology 20 years ago. The growth in molecular epidemiology was due to the promise that a new generation of biologic markers, with particular application to occupational cancer, would allow one to identify excess risk early in the natural history of a disease and provide an opportunity for preventive action. Other potential benefits of early markers of disease include the ability to enhance exposure assessment, especially low-dose exposures and low-risk populations, identify risks from single agents within complex exposures, estimate the total exposure from multiple sources, and provide data today that predict tomorrow's effects. While these benefits are important, in reality, many individual biomarkers may never provide a definitive answer linking exposure to disease. These markers may have the greatest impact in providing additional information to the weight of evidence that suggests a particular exposure is a potential risk.

A good example is the *p53* mutations in angiosarcomas associated with vinyl chloride-exposed factory workers. These lesions are specific to the tumors in persons with vinyl chloride exposure and are not evident in liver angiosarcomas of persons without vinyl chloride exposure. Therefore, these *p53* lesions serve as a molecular fingerprint of exposure. Other examples are not so clear. Attempts to use the glycophorin A locus somatic cell mutation as an end point of a specific locus mutation arising from exposures to benzo[*a*]pyrene or styrene was

confounded by the high background of this mutation in cigarette smokers. Therefore, two things are needed to reach exposure-specific inferences. The first is a prevalent and specific genetic lesion that can be identified in an occupationally exposed group, and the second is a low background of the lesion in the general population.

Validation and linking intermediate biomarkers to cancer. The difficult steps to validation of early biomarkers begin with animal studies and includes studies that ensure biomarker reliability before moving to human subjects with case-control and cohort approaches. Validating biomarkers as predictors requires large study populations in order to investigate events that are generally uncommon. The premier examples are recent cohort studies on DNA adducts and chromosomal aberrations (Bonassi et al. 2000). Biomarkers validated through longitudinal human studies can be used efficiently to estimate the risk of cancer in populations in which epidemiologic studies cannot be performed.

The establishment of a correlation between chromosomal aberrations in peripheral lymphocytes and cancer has stimulated the development of new techniques to detect aberrations in a variety of exposed populations. This is because scoring of unbanded chromosomes in metaphase preparations to detect aberrations is labor intensive and prone to technical artifacts. Therefore, the micronucleus assay has become popular as it is faster, inexpensive, and can be performed on virtually any cell type. Unfortunately, an association has not been established between micronuclei in peripheral blood lymphocytes and a risk for human cancers, as is the case with chromosomal aberrations. Therefore, validation studies with micronuclei and other end points are needed. Other alternatives include fluorescent *in situ* hybridization (FISH), which is relatively fast to perform, with costs ranging from inexpensive to moderately expensive. A more comprehensive type of FISH is multicolor karyotyping, spectral karyotyping (SKY) or M-FISH, which can identify aberrations in all chromosomes. These techniques are equipment and labor intensive and remain too costly for large-scale use. To be useful for occupational cancer research in the future, cytogenetic techniques will need to incorporate automation, rapid-image analysis, and flow cytometry so that a large number of samples can be processed for modest cost.

Monitoring changes in gene expression. Carcinogens presumably disrupt gene expression, resulting in a wide interindividual variation in response. Demonstrating a

link between chemical exposure and gene expression profiles could pave the way for the use of carcinogen-induced changes in transcription as biomarkers to assess worker risk. Ideally, early biologic effect markers can be used to evaluate risk in groups of workers exposed to chemicals and other insults. Advantages of using markers of early biologic effects in cancer etiology studies are that fewer persons may be needed than in a cohort study that evaluates cancer as an outcome. Studies can be performed quickly, as they are generally cross-sectional or short-term longitudinal investigations. Also, because recent exposure often has the greatest impact on early biologic effect biomarkers, highly accurate exposure assessment can be achieved.

In identifying environmental factors in the induction of human disease, one is confronted with thousands of chemicals, dose- and time-related effects, multiple genetic substrates, and uncertainty about disease models. The simple model of a normal cell experiencing several genetic events to become cancerous is outdated. Innumerable genetic and other events must occur. To assess the effects of low-level exposure, it will be necessary to examine multiple gene-environment interactions to demonstrate that the cancer risk is related to a specific exposure.

The complexity of gene-environment interactions is exemplified by simple gene expression studies conducted in isolated cells. The response to benzo[*a*]pyrene of approximately 7,000 genes in primary epithelial cells from multiple human donors revealed altered (both increases and decreases) expression patterns (> 100% change) in more than 500 RNA species. Dose- and time-dependent changes in expression were noted in cytochrome P450 metabolism enzymes, other carcinogen metabolism genes, DNA repair genes, and cell-cycle regulation genes such as *p52* and *p21*. Analysis of the many changes in expression is enhanced by cluster analysis, which is an algorithm designed to identify patterns of expression. It groups RNA transcripts that respond to test treatments in similar fashion and organizes a map ideally suited to large data set analysis. Such analyses are necessary for rapid and effective interrogation of thousands of genes. As outcome data expand exponentially, new data analysis, storage, and mining strategies will be needed. To understand mechanisms of toxicity and predict the toxicity of new chemicals, a reference knowledge base will be needed that anchors gene expression patterns, proteomic data, and metabolite profiles to conventional toxicology and pathology determinations. Linkages need

to include information about dose, route, time, and target tissue as well as information about early, middle, and late toxicity-related changes. Databases need to be easily accessible and provide chemical-signature analyses so that unknown toxicants can be queried within the database to determine their potential toxicity.

Inherited Modifiers of Risk

Identification of relevant genes. The influence of genetic factors on susceptibility to cancer is widely recognized. Some well-known genetic risk factors, such as the *BRCA1* gene, result in a high absolute risk of cancer in persons with the gene. Susceptibility to environmental carcinogens is likely to be influenced by a multitude of genes, none of which alone has a very large effect. Moreover, the cumulative effect on susceptibility to a class of environmental toxicants may result from complex interactions of multiple genes. Historically, studies have focused on what the body does in terms of absorption, distribution, metabolism, and excretion of environmental agents (pharmacokinetics). More recently, emphasis has been on what the agent does to the body (pharmacodynamics). Environmental agents can act either as agonists or antagonists or as activators or inhibitors, thereby perturbing normal function. To be applicable to risk assessment, polymorphisms of susceptibility will have to be included in models that define a chemical's adverse effects.

Identification of SNPs. All genes are highly polymorphic and, perhaps, every gene is capable of being an environmental susceptibility gene. DNA sequence variants include single-nucleotide polymorphisms (SNPs), insertions and deletions, or inversions and duplications of multiple bases or repetitive DNA. As many as 10 million SNPs are estimated to exist per person, many of which are population specific. Although most SNPs have no phenotypic effect, approximately 50,000 to 250,000 SNPs bear a phenotypic change. In the past, geneticists searched for significant penetrant mutations that explain rare diseases. Many of these monogenetic disorders occur against the background of SNPs that function as modifiers of outcome. Similar strategies are used to dissect the genetic contribution to complex diseases, especially those with important environmental exposures. Technical advances make possible the study of large collections of SNPs from either known genes and pathways or those distributed randomly across chromosome(s). In this regard, future studies will examine genes up and downstream to the candidate gene. The scope of studies has evolved from

single-gene design with a phenotype measurement to the promise of surveying the whole genome with "dense-SNP scans." Public SNP databases such as NCBI dbSNP database (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=snp>), which contains about 3.5 million randomly generated SNPs, will be essential to this research. The NCI has developed the Cancer Genome Anatomy Project (<http://cgap.nci.nih.gov>). Two features of this program are to identify SNPs *in silico* and validate SNPs by sequence analysis (<http://snp500cancer.nci.nih.gov>). In the future, publicly available integrated databases need to be built on environmental exposures, SNPs, important genes, and measured disease outcomes. The field will advance only with substantial collaboration and meta-analyses. Genetic databases must be crossed with those addressing exposure and disease.

Applying Genetic Biomarkers to Human Studies

A primary goal for assessing gene expression responses is the identification of candidate exposure biomarkers. Undoubtedly, perturbations in RNA expression and protein patterns will be noted in exposed persons. However, uncertainty will exist as to their use in predicting disease. It is not a given that changes in gene expression will make a difference in a risk-related outcome. Physiologically based pharmacokinetic models have demonstrated that 10-fold differences in enzyme levels may make little difference in bioactivation of a chemical, as the chemical is completely metabolized at low doses in the absence of enzyme induction. The application of data to risk assessment will be aided by the development of models of gene expression of oncogenes and tumor suppressor genes and modeling of polymorphisms of susceptibility. One potential approach will be to group chemicals with similar global gene expression profiles (GGEP) and use available cancer bioassays on these chemicals to derive relative potency parameters in dose-response models. More broadly, GGEP can be used to link chemicals that induce similar enzymes or adverse effects to derive relative potency estimates. Mutations found in oncogenes or tumor suppressor genes may be used to develop dose-response models for humans. The application of biomarkers to risk assessment will require a clear understanding of how environmental exposure indices such as air concentrations and markers of early biologic effects are linked through biomarkers of exposure.

Exposure assessment. Once intermediate markers and underlying pathways are known, the dose-response relationship and

effective agents must be identified. This requires detailed exposure assessment for which cross-sectional biomarker studies are useful. When environmental measures are not available, exposure assessments need to rely on the body burden of a compound. To understand the exposure–response relationship, it is necessary to understand the relationship between environmental concentrations of a compound and a measure of body burden or a biomarker that can be a reactive intermediate, a stable metabolite, or a macromolecular adduct. These are random variables capable of varying both within and between subjects in a population. The key to understanding biomarkers of exposure is a categorization of biomarkers by their half-lives. Short-term biomarkers have residence times or half-lives of less than 30 hr; longer-term biomarkers have half-lives greater than a thousand hours; and intermediate-term biomarkers are in between. These distinctions are for convenience in relating exposure concentrations to biomarker levels. Since intermediate and long-term biomarkers provide information about exposures over weeks to years, a small number of biomarker measurements can be sufficient to assess exposure. In some cases, new technologies will be beneficial in assessing risk of occupational exposure to complex mixtures or a variety of agents simultaneously (mixed exposures). High throughput technologies may help identify those agents within a complex mixture or mixed exposure that are responsible for observed cancer risks, the level of risk associated with the various agents, the agent driving the risk, and the mechanisms of action. These could be investigated by comparing patterns of genetic changes in tissue exposed to mixtures with known patterns for suspect agents. To achieve this capability, it will be necessary to go beyond hypothesis testing and conduct discovery-based research.

Ethics and the use of new technologies.

As genotyping and epidemiologic studies become an integral part of occupational disease prevention and control, fear of privacy violations and discrimination in employment will increase. Issues that arise with the ability to identify markers of disease susceptibility include employment

eligibility, insurability, employer abuse, permissible exposure limits, privacy legislation, and structure of human subject review boards. According to a U.S. Congress Office of Technology Assessment report (1992), 55% of commercial insurance carriers did not consider a genetic disease trait a preexisting disease. In contrast, 75% of health maintenance organizations did. Another study reported that gene testing results were interpreted correctly only 68% of the time (Giardiello et al. 1997). These perceived disparities in perspective and potential for erroneous results heighten public concerns about genetic research. A gene for beryllium disease is an example of a genetic marker that raises legal and ethical issues. Apparently, 30% of the general population carries a gene placing them at high risk for disease, even if exposed to low concentrations of beryllium. Although pre-employment screening is possible, testing of this gene has been confined to research studies. However, no federal law prohibits employers from acquiring genetic information if a prospective employee signs a medical release. Laws need to be written to protect the public while at the same time not restricting research, which would have a negative impact on public health. Researchers should become more actively involved with the ethical and policy implications of their work. To achieve this, they should *a*) ensure correct application of research in the clinical or occupational setting; *b*) protect confidentiality and privacy; *c*) provide appropriate feedback for subjects; *d*) improve the language of informed consent forms; *e*) define guidelines for sample archiving (when to preserve or destroy links); *f*) guard against undue influence from commercial interests; *g*) reduce the stigma associated with assessing gene polymorphisms; *h*) consider the environmental and occupational regulatory implications of research findings; and *i*) contribute to the development of federal laws addressing access, disclosure, or storage of genetic information by employers.

Summary

The ability of new biotechnologies to group chemicals with similar global gene expression profiles has the potential to

provide an early warning system for suspected carcinogens before they are introduced into commerce. The challenge will be to identify the degree of similarity required to predict carcinogenicity and to distinguish pathogenic patterns from homeostatic ones. Gene expression patterns will likely be used in epidemiologic studies as surrogate end points for cancer. Attention to basic epidemiologic principles of design and analysis are still important to guard against biases and irreproducible results. To enhance risk assessments, expression patterns need to demonstrate comparability across species for extrapolation purposes, and to be robust at different doses for dose–response predictions. Before these technologies are used in humans, the ethical, legal, and social issues should be addressed along with the scientific issues. The ultimate challenge to the occupational safety and health community is how to exploit new technologies appropriately without disregarding the potential benefits of traditional “low-tech” research approaches. Meeting this challenge will require the integration of historically tested technologies with newer ones.

REFERENCES

- Bonassi S, Hagmar L, Stromberg U, Montagud AH, Tinnerberg H, Forni A, et al. 2000. Chromosomal aberrations in lymphocytes predict human cancer independently of exposure to carcinogens. European Study Group on Cytogenetic Biomarkers and Health. *Cancer Res* 60:1619–1625.
- Giardiello FM, Brensinger JD, Petersen GM, Luce MC, Hyland LM, Bacon JA, et al. 1997. The use and interpretation of commercial APC gene testing for familial adenomatous polyposis. *N Engl J Med* 336:823–827.
- Henry CJ, Phillips R, Carpanini F, Corton JC, Craig K, Igarashi K, et al. 2002. Use of genomics in toxicology and epidemiology: findings and recommendations of a workshop. *Environ Health Perspect* 110:1047–1050.
- U.S. Congress Office of Technology Assessment. 1992. Genetic Tests and Health Insurance: Results of a Survey-Background Paper. OTA-BP-BA-98. Washington, DC:U.S. Government Printing Office.
- Ward EM, Schulte PA, Bayard S, Blair A, Brandt-Rauf P, Butler MA, et al. 2003. Priorities for development of research methods in occupational cancer. *Environ Health Perspect* 111:1–12.