Title

Application of toxicogenomic profiling to evaluate effects of benzene and formaldehyde: from yeast to human

Journal

Annals of the New York Academy of Sciences, 1310(1)

ISSN

0077-8923

Authors

McHale, Cliona M
Smith, Martyn T
Zhang, Luoping

Publication Date

2014-03-01

DOI

10.1111/nyas.12382

Peer reviewed

Published in final edited form as:
Ann N Y Acad Sci. 2014 March ; 1310(1): 74–83. doi:10.1111/nyas.12382.

Application of toxicogenomic profiling to evaluate effects of benzene and formaldehyde: from yeast to human

Cliona M. McHale, **Martyn T. Smith**, and **Luoping Zhang**
Genes and Environment Laboratory, Division of Environmental Health and Sciences, School of Public Health, University of California, Berkeley, Berkeley, California

Abstract

Genetic variation underlies a significant proportion of the individual variation in human susceptibility to toxicants. The primary current approaches to identify gene–environment (GxE) associations, genome-wide association studies (GWAS) and candidate gene association studies, require large exposed and control populations and an understanding of toxicity genes and pathways, respectively. This limits their application in the study of GxE associations for the leukemogens benzene and formaldehyde, whose toxicity has long been a focus of our research. As an alternative approach, we applied innovative *in vitro* functional genomics testing systems, including unbiased functional screening assays in yeast and a near-haploid human bone marrow cell line (KBM7). Through comparative genomic and computational analyses of the resulting data, we have identified human genes and pathways that may modulate susceptibility to benzene and formaldehyde. We have validated the roles of several genes in mammalian cell models. In populations occupationally exposed to low levels of benzene, we applied peripheral blood mononuclear cell transcriptomics and chromosome-wide aneuploidy studies (CWAS) in lymphocytes. In this review of the literature, we describe our comprehensive toxicogenomic approach and the potential mechanisms of toxicity and susceptibility genes identified for benzene and formaldehyde, as well as related studies conducted by other researchers.

Keywords

benzene; formaldehyde; yeast; human; functional toxicogenomics

Introduction

For many environmental exposures that are toxic to humans, limited information is available about genetic determinants of susceptibility and mechanisms of action. For many years, we have studied the toxicity of benzene, a long-established cause of acute myeloid leukemia (AML) and myelodysplastic syndromes (MDS); and formaldehyde, a recently classified human leukemogen.[1,2] Both chemicals are hematotoxic at low levels of occupational exposure and target hematopoietic stem cells circulating in peripheral blood.[3,4] However, the biological mechanisms, genes, and pathways involved in benzene[5] and formaldehyde[6] toxicity and susceptibility are poorly understood. The main approaches to identify gene–environment (GxE) associations, genome-wide association studies (GWAS) and candidate gene association studies, require large exposed and control populations or an understanding of toxicity genes and pathways. Thus, their application in the study of GxE associations for benzene and formaldehyde is limited.

Address for correspondence: Luoping Zhang, School of Public Health, University of California, 50 University Hall, MC 7360, Berkeley, CA 94720-7360. luoping@berkeley.edu.

Here, we describe an alternative comprehensive toxicogenomic approach we, along with Chris Vulpe's group in the Department of Nutritional Science and Toxicology at the University of California, Berkeley, have taken in yeast and humans to discover genes and molecular pathways involved in toxicity induced by benzene and formaldehyde. We developed and applied innovative *in vitro* functional genomics testing systems, including unbiased functional screening assays in yeast[7,8] and, more recently, in a near-haploid human bone marrow cell line established from a patient with chronic myeloid leukemia (KBM7), to identify genes that modulate susceptibility to benzene and its metabolites and to formaldehyde. We have validated the roles of several genes in mammalian cell models.[9,10,11,12] In populations with well-characterized occupational exposure to benzene, we have profiled the human peripheral blood transcriptomes using microarray[13,14] and next-generation RNA sequencing (RNA-Seq) technology[15] and we have examined chromosome-wide aberrations using an innovative technique called chromosome-wide aneuploidy studies (CWAS).[16,17] Table 1 summarizes the details of our functional genomics and omics approaches, listing the characteristics, advantages and limitations, and chemicals tested for each. These complementary approaches have increased our understanding of the mechanisms of leukemogenesis and potential determinants of susceptibility associated with benzene and formaldehyde.

Functional genomics screening methods to study genetic susceptibility to environmental toxicants

Humans vary in their susceptibility to toxicants and much of this variation likely occurs through genetic variation.[18] Identification of susceptible individuals is important for early detection and prevention of exposure-related disease and for chemical exposure risk assessment. Currently, GWAS and candidate gene–association studies are the main genome-wide methodologies used to identify genes that contribute to human susceptibility to disease. In GxE studies, these methods test for an association of all genes or a subset of genes or pathways, respectively, with a toxicant-related phenotypic outcome on the disease causal pathway, proximal to exposure. Because GWAS test a huge number of variants by genotyping or genomic sequencing and require multiple test correction, large disease and control populations are necessary to identify significant associations. GWAS of benzene-and formaldehyde-associated AML, which has a low incidence and long latency, are not feasible or are prohibitively expensive. As fewer genes hypothesized to play a role in toxicant response are tested in the candidate gene-association study approach, smaller study populations are required and the studies are less expensive.

We have applied candidate gene-association studies to study susceptibility to benzene and formaldehyde toxicity. We identified associations between genetic polymorphisms in genes encoding specific metabolic enzymes,[19] cytokines,[20] DNA repair and genomic maintenance proteins,[21, 22] and cellular adhesion proteins with known or potential roles in niche function,[23] innate immunity[24] and increased susceptibility to benzene-induced hematotoxicity. Information on individual susceptibility to formaldehyde is currently limited. Polymorphisms in glutathione S-transferase genes and the DNA repair gene *XRCC1* were reported to modulate the genotoxic effects (DNA and chromosome damage) of formaldehyde in an occupationally exposed population.[25,26] In a separate study, human formaldehyde metabolism was shown to be related to aldehyde dehydrogenase 2 genotypes in exposed workers.[27] However, the targeted candidate gene approach is limited by our incomplete understanding of the genes and pathways likely to be important.

In a functional genomics approach, phenotype (e.g., cell growth) is directly measured, enabling the investigation of links between specific genes and their products in the cellular response to a toxic substance.[28] We have applied innovative *in vitro* functional genomics

screening approaches to identify human susceptibility genes involved in benzene and formaldehyde toxicity, including a system assessing DNA damage and repair pathways in chicken DT40 B lymphocytes[29,30] and unbiased functional screening assays in yeast[7] and near-haploid human cells (KBM7).[31] In these approaches, large-scale gene disruption or insertion techniques are employed to generate null allele mutants in specific pathways, e.g., DNA repair–deficient clones in DT40 cells, or in all non-essential genes, e.g. parallel deletion analysis (PDA) in yeast and insertional mutagenesis in human haploid KBM7 cells. Comparison of cellular proliferation in wild-type and mutant cells exposed to toxic chemicals enables the ready identification of genes that are essential for cell survival. The yeast and haploid cell systems are high-throughput cell survival bioassays, in which the responses of multiple mutant variants are analyzed in parallel. Comparative genomic and computational analyses are applied to identify corresponding human susceptibility genes. We have applied a functional genomics approach that combines functional screening (yeast and human haploid cells) with targeted validation of candidate genes in mammalian cells to identify human susceptibility genes involved in benzene and formaldehyde toxicity. In the following sections, we describe our approach and similar approaches by other groups.

DNA repair pathways in chicken DT40 B lymphocytes

The DT40 cell line was established from the transformation of chicken B lymphocytes with an avian leucosis virus.[32] Through the use of different methods to render cells conditionally null for essential genes, a unique higher-eukaryotic system for comprehensive reverse genetic analysis of gene function was established.[29] The DT40 cell line exhibits highly efficient gene targeting and possesses a remarkably stable phenotype and karyotype. In addition, its rapid growth rate facilitates phenotypic analysis and its lack of functional p53 enables the analysis of genome instability. Isogenic mutant clones of all known DNA damage–response pathways are available.[33] Together, these features make the DT40 system useful for testing DNA-damaging agents.

Using the DT40 screening model, Yamazoe *et al*.[29] found a requirement for the homologous recombination (HR) pathway but not the non-homologous end joining (NHEJ) pathway in the repair of DNA damage induced by formaldehyde. DNA–protein crosslinks (DPC) are a type of DNA damage strongly induced by formaldehyde. Fanconi anemia patients are sensitive to DNA intrastrand crosslinking agents[34] owing to deficiencies in the FANC–BRCA DNA damage–response pathway. Using the DT40 reverse genetic model, Ridpath *et al*. reported that the FANC–BRCA pathway is essential to counteract DPC caused by formaldehyde and other aldehydes, as evident from the hypersensitivity to plasma formaldehyde levels of DT40 mutants deficient in pathway components.[35] FANCD2-deficient DT40 cells were the most sensitive to formaldehyde.

Validation of FANCD2 in formaldehyde toxicity in human lymphoblast cells

Ridpath *et al*. also showed that the FANC–BRCA pathway plays a role in formaldehyde toxicity in human cells, indicating the relevance of DT40 findings to human cells. Specifically, isogenic clones from colorectal cancer cell line RKO, harboring disrupted *FANCC* or *FANCG* genes, were found to be hypersensitive to formaldehyde compared to parental cells.[35] Recently, using human lymphoblast cell models of FANCD2 deficiency (PD20 cells) and sufficiency (PD20-D2 cells), we showed that the FANCD2 protein and the Fanconi anemia pathway are essential to protect against formaldehyde toxicity.[9] Treatment of the cells with formaldehyde (0–150 μM) for 24 h resulted in dose-dependent increases in DPC, cytotoxicity, micronuclei, chromosome aberrations, and apoptosis in both cell lines, with greater increases in FANCD2-deficient PD20 cells. Induction of BRCA2 in PD20 cells was compromised, potentially reducing the capacity of these cells to repair DPCs. Multiple

additional FANC–BRCA and related proteins are involved in cellular resistance to DPC induced by formaldehyde[36] and could thus influence inter-individual susceptibility.

Thus, the DT40 system has proven potential to identify genes involved in formaldehyde toxicity. However, despite multiple assay improvements, including the development of a quantitative and high-throughput version of the DT40 screening assay,[30,37] the DT40 system is a genetic, not a genomic, screen.

Yeast PDA approach

Following the complete sequencing of the *Saccharomyces cerevisiae* genome, the systematic deletion of all identified genes for functional genomics studies became possible. Using a PCR-based approach and unique 20 bp "barcodes," open reading frames were disrupted to create a set of knockout yeast strains lacking both non-essential and essential genes.[38] The incorporation of barcodes enabled multiple deletion strains to be pooled and assayed for growth in a single assay. The parallel deletion analysis (PDA) method quantitatively analyses the fitness of every deletion strain tested simultaneously.[39,40] In this approach, the molecular barcodes are amplified from all pooled culture strains present in a single PCR reaction and hybridized to microarrays containing oligonucleotide probes complementary to each barcode. The signals generated from each probe are proportional to the number of cells of the corresponding strain present in the sampled culture. Comparison of hybridization signals from exposed and unexposed control cells yields a fitness score that can be used to identify deletion strains with significantly altered growth[41,42] and thus genes that are involved in toxicant response. *S. cerevisiae* is a good model for human and higher eukaryote disease and toxicity testing, as yeast has functional orthologs of many human disease genes.[43–45] Extensive annotation of the yeast genome and the availability of bioinformatics tools such as software for gene ontology (GO) and network analysis and ortholog identification, together with the quantitative fitness data from PDA, make this a powerful approach for understanding mechanisms of toxicity.[7]

Yeast PDA of phenolic benzene metabolite toxicity

Using yeast PDA, we sought to identify yeast genes and pathways that modulate the cellular toxicity of three phenolic metabolites of benzene, hydroquinone (HQ), catechol, and 1,2,4-benzenetriol.[8] We exposed pools of 4607 yeast deletion mutants to equitoxic concentrations of the metabolites that induced 20% growth inhibition of the wild-type strain (IC_{20}) and 50% and 25% of the IC_{20}, for 5 and 15 generations of growth (5 g and 15 g). Using differential strain sensitivity analysis (DSSA), we identified yeast genes required for tolerance to these compounds. The numbers of genes identified was correlated with dose and generations of growth. We identified multiple genes that have human orthologs and therefore could be novel targets or modulators of benzene toxicity in humans. We found that the primary mechanism of toxicity of these benzene metabolites is induction of oxidative stress, in agreement with findings in previous studies in other organisms.[46–48] Other processes identified as important for benzene metabolite tolerance such as the vacuole, iron homeostasis, and lipid peroxidation may result from oxidative stress. *IRA2*, a modulator of ras signaling, was required for tolerance to HQ in yeast. Several homologous recombination (HR)–deficient mutant strains were sensitive to treatment with HQ. One yeast HR gene thus identified, *SGS1*, is an ortholog of the human genomic maintenance gene *WRN*. Polymorphic variants of *WRN* were previously associated with increased susceptibility to benzene-induced hematotoxicity in an exposed population.[22,49]

Yeast PDA of formaldehyde toxicity

In a pilot yeast PDA study,[50] we identified potential formaldehyde susceptibility genes and pathways. We exposed yeast homozygous deletion pools to IC_{20}, 50% IC_{20}, and 25% IC_{20} and selected sensitive strains after 15 generations. Several yeast genes involved in mRNA processing, including the alcohol dehydrogenase enzyme *SFAIP*, were required for formaldehyde tolerance. SFAIP is known to be involved in formaldehyde metabolism in yeast, and its human ortholog ADH5 has been similarly implicated in human formaldehyde metabolism.[51] Several genes involved in the repair of DNA damage were also found to be required for formaldehyde tolerance in our pilot study. Many of the identified genes also have confirmed human orthologs. *MUS81* (human ortholog: *MUS81*) encodes a crossover junction endonuclease, which is involved in replication fork stability, DNA repair, and meiotic joint molecule resolution. Other genes identified indicate a requirement for nucleotide excision repair (NER), *Rad1* (human: *ERCC4*) and homologous DNA repair, *RAD51* (human: *RAD51)*, and *RAD57* (human: *XRCC3*), suggesting that formaldehyde produces multiple forms of DNA damage. Using a less rigorous and quantitative yeast screening methodology to screen for genes involved in formaldehyde toxicity, De Graaf *et al*. identified genes involved in DNA repair and tolerance pathways that confer formaldehyde resistance under acute and chronic exposure conditions.[52] Following chronic low-dose exposure, strains containing deletions in genes mediating HR showed the greatest sensitivity. Following acute formaldehyde exposure, repair and/or tolerance of DNA–protein crosslinks was reportedly mediated by NER without the accumulation of double-strand breaks.

Validation of human orthologs of yeast PDA genes in mammalian cells

We used *in vitro* knockdown approaches to confirm and clarify the functional roles of candidate benzene susceptibility genes identified by yeast PDA, an approach which we had previously used successfully to validate arsenic susceptibility genes.[53,54] We conducted several studies of *WRN*, the human homolog of yeast *SGS1*, a HQ-resistant gene identified by PDA. First, through transient silencing using short interfering RNA (siRNA), we showed that *WRN* plays an important role in the protection of HeLa cells against toxicity of the benzene metabolite, HQ, by supporting a normal DNA-damage response to DNA double-strand breaks.[10] Second, through stable *WRN* knockdown in HL60 cells using short hairpin RNA (shRNA), we showed that loss of *WRN* increases genomic instability and enhances genotoxicity induced by HQ.[11] Third, in follow-up *in vitro* studies, we showed that additional factors modulate the effects of *WRN* deficiency on cell proliferation and genomic instability and that carcinogenesis induced by *WRN* deficiency might be influenced by p53 status.[12] As mentioned earlier, *WRN* polymorphisms were independently found to be associated with increased susceptibility to benzene-induced hematotoxicity in an exposed population.[22, 49] Together, these data provide mechanistic support for the link between *WRN* and benzene-induced hematotoxicity, and possibly benzene-induced leukemia.

As mentioned above, *IRA2*, a modulator of ras signaling, was required for tolerance to HQ in yeast. In a follow-up study, we showed that HQ toxicity was modulated by ras signaling in both yeast and human cells and that *NF1* knockout mice exhibited an increased level of DNA damage in erythroid progenitors and increased proliferation of colony forming unit–granulocyte/macrophage (CFU-GM) progenitors, which together could increase risk of myeloid disease (unpublished data). We are currently validating roles for genes identified in yeast PDA screens of formaldehyde toxicity in mammalian cells.

Haploid screening systems

Human haploid cell line

Recently, an insertional mutagenesis system that screens directly for human susceptibility genes was developed from a derivative of a human chronic myeloid leukemia cell line (KBM7).[31] KBM7 has half the human diploid DNA content, apart from a disomy of chromosome 8. An insertional mutagenesis library (KBM7-Mu) covering most nonessential human genes was developed using a gene-trap retrovirus. Thus, genes whose insertional disruption allows cells to survive and proliferate in a selective environment can be identified. Using this approach, Carette *et al*. identified genes involved in susceptibility to influenza infection and diphtheria toxin and exotoxin A cytotoxicity. We have used the KBM7-Mu cell line to screen for genes conferring resistance to benzene metabolites and formaldehyde.[55] In addition, we adapted the screening method to use a semi-solid rather than a liquid medium, thus shortening the screening process by 2–3 weeks. Preliminary results have identified several genes with a potential role in susceptibility to toxicity of these exposures, and we are currently confirming our findings.

Mouse haploid embryonic stem cells

We are currently establishing a mouse embryonic stem cell (ESC)–screening system that will combine the power of a haploid genome with the pluripotency of ESCs to identify susceptibility to toxicants in relevant cell types, such as hematopoietic stem and progenitor cells in the case of leukemia, at a genomic scale as described by Elling *et al*.[56] and others.[57–61] Elling *et al*. reported the generation of haploid mouse ESC lines, which carry 20 chromosomes, express stem cell markers, and develop into all germ layers *in vitro* and *in vivo*, from parthenogenetic embryos. They also developed a reversible mutagenesis protocol that allows saturated genetic recessive screens and results in homozygous alleles.

Through the functional genomic approaches described here, we have identified and validated human genes involved in susceptibility to toxicants.

Genomic-wide approaches to understand mechanisms of toxicity

We have applied additional omics methods to further our understanding of benzene toxicity in the blood cells of occupationally exposed workers. We conducted transcriptome profiling of PBMC, by microarray and RNA-Seq, and analysis of chromosomal aneuploidy in lymphocytes across all chromosomes using fluorescent *in situ* hydrization (FISH)-based CWAS. Through these approaches, we have identified genes and chromosomes targeted by benzene exposure and gained insight into toxicity pathways and mechanisms, as described in the following sections.

Chromosome-wide association studies (CWAS)

Aneuploidy is a potential mechanism underlying benzene-induced leukemia, suggested by the detection of chromosomal aneuploidies in benzene-related leukemia, pre-leukemia patients,[62] and healthy workers exposed to benzene.[16,62–65] In these studies, classical cytogenetics and FISH were applied to analyze chromosomes hypothesized to play a potential role in benzene-induced AML. An apparently selective effect of benzene or its metabolites on monosomy and trisomy of certain chromosomes was observed in human lymphocytes *in vitro*[66,67] and in benzene-exposed workers.[63,65] In order to systematically analyze aneuploidy, we developed a novel CWAS approach, comprising a triple-color painting FISH method called OctoChrome FISH, which simultaneously detects aneuploidy in all 24 chromosomes on a single eight-square slide.[68]

CWAS studies in benzene-exposed workers

In a pilot study, we analyzed aneuploidy in the peripheral blood lymphocytes of six benzene-exposed workers (> 5 ppm) and five unexposed controls by CWAS. We found significant increases in the rates of monosomy of chromosomes 5, 6, 7, and 10 and trisomy of chromosomes 8, 9, 17, 21 and 22 in the exposed workers.[16] In a more recent study, we analyzed aneuploidy by CWAS in 47 benzene-exposed workers (< 10 ppm and ≥ 10 ppm) and 27 unexposed controls from the same study population.[17] The monosomy and trisomy rates of the 22 autosomes showed heterogeneity when plotted against continuous benzene exposure. In the benzene-exposed workers, we found statistically significant, dose-dependent increases in the aneuploidy rates of specific chromosomes, and of *a priori* defined "susceptible" sets, compared with all other chromosomes. Thus, our systematic CWAS approach confirmed that specific aneuploidies are selectively induced by benzene in a dose-dependent manner and may play roles in benzene-induced leukemogenesis. Benzene-associated aneuploidies from earlier studies were validated and new associations were revealed. The underlying mechanisms by which benzene may induce aneuploidy, such as errors in chromosome segregation, as a consequence of incorrect microtubule-kinetochore attachments or failure of the spindle checkpoint;[69] dysfunctional telomeres;[70] and chromosome architecture[71] remain to be determined. Altered expression of certain genes may also play a role in aneuploidy;[72] we have performed transcriptome profiling subjects in the same population,[73] as described in the next section.

Applying the same approach, we are currently analyzing aneuploidy and structural chromosomal changes in formaldehyde-exposed workers.

Transcriptomics

Using gene expression microarrays, we identified genes and pathways altered by exposure to high (> 10 ppm) levels of benzene in PBMC from a small number of occupationally exposed and unexposed control workers,[13,14] matched by age, gender, and smoking status. Later, we analyzed a much larger group of subjects from this population (n = 125) who were exposed to a range of benzene levels, from < 1 ppm, the current U.S. occupational standard,[74] to > 10ppm.[73] Study design and analysis with a mixed effects model minimized potential confounding and experimental variability. Overall, benzene exposure significantly (FDR-adjusted P-values ≤ 0.05) altered expression of 3007 probes representing 2846 genes, relative to unexposed controls. Highly significant widespread perturbation of gene expression was observed at all exposure levels. Immune response was the most significant biological process identified by GO analysis. The AML pathway was among the pathways most significantly associated with benzene exposure. Other affected pathways included toll-like receptor signaling, oxidative phosphorylation, B cell receptor signaling, apoptosis, and T cell receptor signaling. In the same study, we identified a 16-gene expression signature upregulated at all levels of benzene exposure. The signature genes are involved in immune response, inflammatory response, cell adhesion, cell–matrix adhesion, and blood coagulation. Our findings suggest that chronic benzene exposure, even at levels below the current U.S. occupational standard, perturbs many genes, biological processes and pathways.

As recently reviewed by us, RNA-Seq offers some advantages over microarrays in the analysis of the human transcriptome.[75] RNA-Seq is more sensitive[76–78] and has a wider dynamic range for the quantification of gene expression[79] than microarrays. It also has the ability to detect both coding and non-coding RNAs, as well as novel transcripts and alternative splice variants. Therefore, in a pilot RNA-Seq study, we analyzed the transcriptomes of 10 workers highly exposed (> 5 ppm) to benzene and 10 unexposed control study subjects[15] previously analyzed by microarray.[73] We found that RNA-Seq was more sensitive than microarrays in the detection of transcripts expressed at lower levels. We

identified 146 statistically significant differentially expressed genes (including 29 non-coding RNAs) by RNA-Seq compared with one gene by microarray, in the 20 subjects. Differential splicing (in six transcripts, false discovery rate < 0.05), a known leukemogenesis pathway,[80] was identified as a potential mechanism of benzene toxicity. Overall, this pilot study shows that RNA-Seq can complement microarrays in the analysis of changes in transcript expression resulting from chemical exposures. We are currently expanding our benzene RNA-Seq study to analyze more samples and additional alterations in the transcriptome.

Conclusions and future directions

We have applied a comprehensive toxicogenomic approach to discover genes and molecular pathways involved in benzene and formaldehyde toxicity. Using innovative *in vitro* functional genomics testing systems in yeast, in conjunction with validation in mammalian cells by knockdown studies, we have identified and validated candidate human susceptibility genes. We are optimizing the human haploid screening system in KBM7 cells and are adopting a mouse haploid ESC method, increasing our ability to identify relevant genes.

Mechanisms of toxicity were also revealed by our functional genomic studies. Additional mechanistic information has come from transcriptomic and chromosomic studies that we have conducted in populations with well-characterized occupational exposure to benzene. Through these studies, we have identified gene expression and chromosome targets of benzene and formaldehyde, established human leukemogens. These complementary approaches have increased our understanding of the mechanisms of leukemogenesis and potential determinants of susceptibility associated with benzene and formaldehyde. In future studies, we will analyze additional omic endpoints in our benzene-exposed population (DNA methylation, miRNA expression), following up on earlier pilot studies,[81] and will measure a range of omic endpoints in populations occupationally exposed to formaldehyde. We will develop and apply bioinformatics methodologies to perform cross-omic analysis in a systems biology approach.[81]

Acknowledgments

Prof. Chris Vulpe's group in the Department of Nutritional Science and Toxicology at UC Berkeley performed all of the yeast studies described in this review. This work was supported by National Institute of Environmental Health Sciences, National Institute of Health Grants P42 ES04705 and R01ES06721 to Dr. Martyn Smith and R01ES017452 to Dr. Luoping Zhang.

References

1. IARC. International Agency for Research on Cancer: A Review of Human Carcinogens: Chemical Agents and Related Occupations: Formadehyde. Monogr Eval Carcinog Risks Hum. 2012; 100F: 401–435.
2. NTP. 12th Report on Carcinogens. Health and Human Services; 2011. http://ntp.niehs.nih.gov/?objectid=03C9AF75-E1BF-FF40-DBA9EC0928DF8B15
3. Smith MT, et al. Causal models of leukaemia and lymphoma. IARC Sci Publ. 2004; 157:373–392. [PubMed: 15055307]
4. Zhang L, et al. Occupational exposure to formaldehyde, hematotoxicity, and leukemia-specific chromosome changes in cultured myeloid progenitor cells. Cancer Epidemiol Biomarkers Prev. 2010; 19:80–88. [PubMed: 20056626]
5. McHale CM, Zhang L, Smith MT. Current understanding of the mechanism of benzene-induced leukemia in humans: implications for risk assessment. Carcinogenesis. 2012; 33:240–252. [PubMed: 22166497]

6. Zhang L, et al. Formaldehyde exposure and leukemia: A new meta-analysis and potential mechanisms. Mutat Res. 2009; 681:150–168. [PubMed: 18674636]
7. North M, Vulpe CD. Functional toxicogenomics: mechanism-centered toxicology. Int J Mol Sci. 2010; 11:4796–4813. [PubMed: 21614174]
8. North M, et al. Genome-wide functional profiling reveals genes required for tolerance to benzene metabolites in yeast. PLoS One. 2011; 6:e24205. [PubMed: 21912624]
9. Ren X, et al. The impact of FANCD2 deficiency on formaldehyde-induced toxicity in human lymphoblastoid cell lines. Arch Toxicol. 2013; 87:189–196. [PubMed: 22872141]
10. Galvan N, et al. Depletion of WRN enhances DNA damage in HeLa cells exposed to the benzene metabolite, hydroquinone. Mutat Res. 2008; 649:54–61. [PubMed: 17875398]
11. Ren X, et al. Werner syndrome protein, WRN, protects cells from DNA damage induced by the benzene metabolite hydroquinone. Toxicol Sci. 2009; 107:367–375. [PubMed: 19064679]
12. Ren X, et al. Comparison of proliferation and genomic instability responses to WRN silencing in hematopoietic HL60 and TK6 cells. PLoS ONE. 2011; 6:e14546. doi:14510.11371/journal.pone.0014546. [PubMed: 21267443]
13. Forrest MS, et al. Discovery of novel biomarkers by microarray analysis of peripheral blood mononuclear cell gene expression in benzene-exposed workers. Environ Health Perspect. 2005; 113:801–807. [PubMed: 15929907]
14. McHale CM, et al. Changes in the peripheral blood transcriptome associated with occupational benzene exposure identified by cross-comparison on two microarray platforms. Genomics. 2009; 93:343–349. [PubMed: 19162166]
15. Thomas R, McHale Cliona M, Zhang Luoping, Lan Qing, Hubbard Alan E, Vermeulen Roel, Li Guilan, Rappaport Stephen M, Yin Songnian, Smith Martyn T, Rothman Nathaniel. Global gene expression response of a population exposed to benzene: a pilot study exploring the use of mRNA sequencing technology. Environmental and Molecular Mutagenesis. 2013; 54:566–573. [PubMed: 23907980]
16. Zhang L, et al. Use of OctoChrome fluorescence in situ hybridization to detect specific aneuploidy among all 24 chromosomes in benzene-exposed workers. Chem Biol Interact. 2005; 153–154:117–122.
17. Zhang L, et al. Chromosome-wide aneuploidy study (CWAS) in workers exposed to an established leukemogen, benzene. Carcinogenesis. 2011; 32:605–612. [PubMed: 21216845]
18. Aldridge JE, et al. Heterogeneity of toxicant response: sources of human variability. Toxicol Sci. 2003; 76:3–20. [PubMed: 12883075]
19. Ross D. The role of metabolism and specific metabolites in benzene-induced toxicity: evidence and issues. Journal of toxicology and environmental health Part A. 2000; 61:357–372. [PubMed: 11086940]
20. Lv L, et al. The TNF-α 238A polymorphism is associated with susceptibility to persistent bone marrow dysplasia following chronic exposure to benzene. Leukemia Research. 2007; 31:1479–1485. [PubMed: 17367855]
21. Hosgood HD 3rd, et al. Association between genetic variants in VEGF, ERCC3 and occupational benzene haematotoxicity. Occupational and environmental medicine. 2009; 66:848–853. [PubMed: 19773279]
22. Lan Q, et al. Large-scale evaluation of candidate genes identifies associations between DNA repair and genomic maintenance and development of benzene hematotoxicity. Carcinogenesis. 2009; 30:50–58. [PubMed: 18978339]
23. Lan Q, et al. Polymorphisms in cytokine and cellular adhesion molecule genes and susceptibility to hematotoxicity among workers exposed to benzene. Cancer research. 2005; 65:9574–9581. [PubMed: 16230423]
24. Shen M, et al. Polymorphisms in genes involved in innate immunity and susceptibility to benzene-induced hematotoxicity. Experimental & molecular medicine. 2011; 43:374–378. [PubMed: 21540635]
25. Jiang S, et al. Genomic damages in peripheral blood lymphocytes and association with polymorphisms of three glutathione S-transferases in workers exposed to formaldehyde. Mutat Res. 2010; 695:9–15. [PubMed: 19818869]

26. Jiang SF, et al. Association between XRCC1 gene polymorphisms and DNA damage of workers exposed to formaldehyde. Wei Sheng Yan Jiu. 2006; 35:675–677. [PubMed: 17290738]
27. Cheng XM, et al. Relationship between susceptibility of formaldehyde metabolism and genetic polymorphisms of ALDH2 and cytochrome P4502E1. Zhonghua Yu Fang Yi Xue Za Zhi. 2008; 42:582–587. [PubMed: 19115622]
28. Hieter P, Boguski M. Functional genomics: it's all how you read it. Science. 1997; 278:601–602. [PubMed: 9381168]
29. Yamazoe M, et al. Reverse genetic studies of the DNA damage response in the chicken B lymphocyte line DT40. DNA Repair (Amst). 2004; 3:1175–1185. [PubMed: 15279806]
30. Evans TJ, et al. Mutant cells defective in DNA repair pathways provide a sensitive high-throughput assay for genotoxicity. DNA Repair (Amst). 2010; 9:1292–1298. [PubMed: 21030320]
31. Carette JE, et al. Haploid genetic screens in human cells identify host factors used by pathogens. Science. 2009; 326:1231–1235. [PubMed: 19965467]
32. Baba TW, Giroir BP, Humphries EH. Cell lines derived from avian lymphomas exhibit two distinct phenotypes. Virology. 1985; 144:139–151. [PubMed: 2998040]
33. Buerstedde JM, Takeda S. Increased ratio of targeted to random integration after transfection of chicken B cell lines. Cell. 1991; 67:179–188. [PubMed: 1913816]
34. Kennedy RD, D'Andrea AD. The Fanconi Anemia/BRCA pathway: new faces in the crowd. Genes Dev. 2005; 19:2925–2940. [PubMed: 16357213]
35. Ridpath JR, et al. Cells deficient in the FANC/BRCA pathway are hypersensitive to plasma levels of formaldehyde. Cancer Res. 2007; 67:11117–11122. [PubMed: 18056434]
36. Jacquemont C, Taniguchi T. The Fanconi anemia pathway and ubiquitin. BMC Biochem. 2007; 8(Suppl 1):S10. [PubMed: 18047734]
37. Ji K, et al. A novel approach using DNA-repair-deficient chicken DT40 cell lines for screening and characterizing the genotoxicity of environmental contaminants. Environ Health Perspect. 2009; 117:1737–1744. [PubMed: 20049126]
38. Giaever G, et al. Functional profiling of the Saccharomyces cerevisiae genome. Nature. 2002; 418:387–391. [PubMed: 12140549]
39. Pierce SE, et al. Genome-wide analysis of barcoded Saccharomyces cerevisiae gene-deletion mutants in pooled cultures. Nature protocols. 2007; 2:2958–2974.
40. Pierce SE, et al. Chemogenomic approaches to elucidation of gene function and genetic pathways. Methods in molecular biology. 2009; 548:115–143. [PubMed: 19521822]
41. Giaever G, et al. Genomic profiling of drug sensitivities via induced haploinsufficiency. Nature genetics. 1999; 21:278–283. [PubMed: 10080179]
42. Shoemaker DD, et al. Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. Nature genetics. 1996; 14:450–456. [PubMed: 8944025]
43. Steinmetz LM, et al. Systematic screen for human disease genes in yeast. Nature genetics. 2002; 31:400–404. [PubMed: 12134146]
44. Foury F. Human genetic diseases: a cross-talk between man and yeast. Gene. 1997; 195:1–10. [PubMed: 9300813]
45. Bassett DE Jr, et al. Genome cross-referencing and XREFdb: implications for the identification and analysis of genes mutated in human disease. Nature genetics. 1997; 15:339–344. [PubMed: 9090377]
46. Shen Y, et al. Benzene metabolites enhance reactive oxygen species generation in HL60 human leukemia cells. Human and Experimental Toxicology. 1996; 15:422–427. [PubMed: 8735467]
47. Badham HJ, Winn LM. In utero exposure to benzene disrupts fetal hematopoietic progenitor cell growth via reactive oxygen species. Toxicological sciences: an official journal of the Society of Toxicology. 2010; 113:207–215. [PubMed: 19812361]
48. Barreto G, et al. The role of catechols and free radicals in benzene toxicity: an oxidative DNA damage pathway. Environmental and molecular mutagenesis. 2009; 50:771–780. [PubMed: 19449395]

49. Shen M, et al. Polymorphisms in genes involved in DNA double-strand break repair pathway and susceptibility to benzene-induced hematotoxicity. Carcinogenesis. 2006; 27:2083–2089. [PubMed: 16728435]
50. North M, et al. Identification of novel biomarkers of formaldehyde toxicity in humans using functional genomics in yeast. The Toxicologist, Supplement to Toxicological Sciences. 2011; 120:548.
51. Iborra FJ, et al. Immunocytochemical and biochemical demonstration of formaldhyde dehydrogenase (class III alcohol dehydrogenase) in the nucleus. J Histochem Cytochem. 1992; 40:1865–1878. [PubMed: 1453005]
52. de Graaf B, Clore A, McCullough AK. Cellular pathways for DNA repair and damage tolerance of formaldehyde-induced DNA-protein crosslinks. DNA Repair (Amst). 2009; 8:1207–1214. [PubMed: 19625222]
53. Jo WJ, et al. Comparative Functional Genomic Analysis Identifies Distinct and Overlapping Sets of Genes Required for Resistance to Monomethylarsonous Acid (MMAIII) and Arsenite (AsIII) in Yeast. Toxicol Sci. 2009; 111:424–436. [PubMed: 19635755]
54. Jo WJ, et al. Acetylated H4K16 by MYST1 protects UROtsa cells from arsenic toxicity and is decreased following chronic arsenic exposure. Toxicol Appl Pharmacol. 2009; 241:294–302. [PubMed: 19732783]
55. Shen H, Smith MT, Zhang L. Using a novel functional genetic screen in human haploid cells to identify genes that modify susceptibility to chemical exposure. Environmental and molecular mutagenesis. 2013; 54:S33.
56. Elling U, et al. Forward and reverse genetics through derivation of haploid mouse embryonic stem cells. Cell Stem Cell. 2011; 9:563–574. [PubMed: 22136931]
57. Zhang S, Teng Y. Powering mammalian genetic screens with mouse haploid embryonic stem cells. Mutat Res. 2013; 741–742:44–50.
58. Li W, et al. Androgenetic haploid embryonic stem cells produce live transgenic mice. Nature. 2012; 490:407–411. [PubMed: 23023130]
59. Leeb M, et al. Germline potential of parthenogenetic haploid mouse embryonic stem cells. Development. 2012; 139:3301–3305. [PubMed: 22912412]
60. Yang H, et al. Generation of genetically modified mice by oocyte injection of androgenetic haploid embryonic stem cells. Cell. 2012; 149:605–617. [PubMed: 22541431]
61. Leeb M, Wutz A. Derivation of haploid embryonic stem cells from mouse embryos. Nature. 2011; 479:131–134. [PubMed: 21900896]
62. Zhang L, Eastmond DA, Smith MT. The nature of chromosomal aberrations detected in humans exposed to benzene. Crit Rev Toxicol. 2002; 32:1–42. [PubMed: 11846214]
63. Smith MT, et al. Increased translocations and aneusomy in chromosomes 8 and 21 among workers exposed to benzene. Cancer research. 1998; 58:2176–2181. [PubMed: 9605763]
64. Zhang L, et al. Aberrations in chromosomes associated with lymphoma and therapy-related leukemia in benzene-exposed workers. Environmental and molecular mutagenesis. 2007; 48:467–474. [PubMed: 17584886]
65. Zhang L, et al. Increased aneusomy and long arm deletion of chromosomes 5 and 7 in the lymphocytes of Chinese workers exposed to benzene. Carcinogenesis. 1998; 19:1955–1961. [PubMed: 9855009]
66. Zhang L, et al. Benzene metabolites induce the loss and long arm deletion of chromosomes 5 and 7 in human lymphocytes. Leuk Res. 1998; 22:105–113. [PubMed: 9593466]
67. Zhang L, et al. Nonrandom aneuploidy of chromosomes 1, 5, 6, 7, 8, 9, 11, 12, and 21 induced by the benzene metabolites hydroquinone and benzenetriol. Environmental and molecular mutagenesis. 2005; 45:388–396. [PubMed: 15662717]
68. Ji Z, Zhang L. Chromosomics: detection of numerical and structural alterations in all 24 human chromosomes simultaneously using a novel OctoChrome FISH assay. Journal of visualized experiments: JoVE. 2012; 60:3619. pii. [PubMed: 22331009]
69. King RW. When 2+2=5: the origins and fates of aneuploid and tetraploid cells. Biochimica et biophysica acta. 2008; 1786:4–14. [PubMed: 18703117]

70. Aubert G, Lansdorp PM. Telomeres and aging. Physiological reviews. 2008; 88:557–579. [PubMed: 18391173]
71. Mani RS, Chinnaiyan AM. Triggers for genomic rearrangements: insights into genomic, cellular and environmental influences. Nat Rev Genet. 2010; 11:819–829. [PubMed: 21045868]
72. Chi YH, Jeang KT. Aneuploidy and cancer. J Cell Biochem. 2007; 102:531–538. [PubMed: 17661351]
73. McHale CM, et al. Global gene expression profiling of a population exposed to a range of benzene levels. Environmental health perspectives. 2011; 119:628–634. [PubMed: 21147609]
74. OSHA. Administration, O. S. a. H. Fed Reg. Vol. 52. Washington, D.C., USA: 1987. Occupational exposure to benzene; p. 34460-34579.
75. McHale CM, et al. Analysis of the transcriptome in molecular epidemiology studies. Environmental and molecular mutagenesis. 2013; 54:500–517. [PubMed: 23907930]
76. Marioni JC, et al. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome research. 2008; 18:1509–1517. [PubMed: 18550803]
77. Su Z, et al. Comparing next-generation sequencing and microarray technologies in a toxicological study of the effects of aristolochic acid on rat kidneys. Chemical research in toxicology. 2011; 24:1486–1493. [PubMed: 21834575]
78. Liu S, et al. A comparison of RNA-Seq and high-density exon array for detecting differential gene expression between closely related species. Nucleic acids research. 2011; 39:578–588. [PubMed: 20864445]
79. Raghavachari N, et al. A systematic comparison and evaluation of high density exon arrays and RNA-seq technology used to unravel the peripheral blood transcriptome of sickle cell disease. BMC Medical Genomics. 2012; 5:28. [PubMed: 22747986]
80. Maciejewski JP, Padgett RA. Defects in spliceosomal machinery: a new pathway of leukaemogenesis. British journal of haematology. 2012; 158:165–173. [PubMed: 22594801]
81. Zhang L, et al. Systems biology of human benzene exposure. Chemico-Biological Interactions. 2009; 184:86–93. [PubMed: 20026094]
82. Snyder R. Leukemia and benzene. International journal of environmental research and public health. 2012; 9:2875–2893. [PubMed: 23066403]
83. Gentry PR, et al. Formaldehyde exposure and leukemia: critical review and reevaluation of the results from a study that is the focus for evidence of biological plausibility. Critical reviews in toxicology. 2013; 43:661–670. [PubMed: 23902349]
84. Lu K, et al. Distribution of DNA adducts caused by inhaled formaldehyde is consistent with induction of nasal carcinoma but not leukemia. Toxicol Sci. 2010; 116:441–451. [PubMed: 20176625]
85. Lu K, et al. Molecular dosimetry of N2-hydroxymethyl-dG DNA adducts in rats exposed to formaldehyde. Chem Res Toxicol. 2011; 24:159–161. [PubMed: 21155545]
86. Moeller BC, et al. Determination of N2-hydroxymethyl-dG adducts in the nasal epithelium and bone marrow of nonhuman primates following 13CD2-formaldehyde inhalation exposure. Chem Res Toxicol. 2011; 24:162–164. [PubMed: 21222454]
87. Ye X, et al. Inhaled formaldehyde induces DNA-protein crosslinks and oxidative stress in bone marrow and other distant organs of exposed mice. Environmental and molecular mutagenesis. 2013; 54:705–718. [PubMed: 24136419]
88. Zhang Y, et al. Bone marrow injury induced via oxidative stress in mice by inhalation exposure to formaldehyde. PloS one. 2013; 8:e74974. [PubMed: 24040369]

Table 1

Functional genomics and omics approaches to study benzene and formaldehyde toxicity

	Organism	Functional genomic/omic screens	Cell type	Genes covered	Advantages	Limitations	Tested
in vitro	Yeast	Parallel Deletion Analysis (PDA)[38–42]	*Saccharomyes cervisiae*	All non-essential genes	Yeast genome well annotated, a good model for human/higher eukaryote disease. Quantitative, high-throughput.	Not all yeast genes have human homologs	BZ (HQ, BT, CAT)[8] FA[50,52]
	Chicken	DT-40 cells—avian leukosis virus transformed[32,33]	B lymphocytes	DNA repair–pathway genes	Highly effective gene targeting, rapid gowth rate	Few pathways covered	FA[29,35]
	Mouse	Mouse haploid embryonic stem cells (ESC)[56,57–61]	ESC	All non-lethal genes	Mouse more relevant to human	Not high-throughput	BZ, FA not tested
	Human	Human haploid cell line (KBM 7)[31]	CML bone marrow cells	All genes except those on chromosome 8	Genes whose insertional disruption allows cells to survive/proliferate can be identified	Chromosome 8 genes not disrupted	FA, HQ[55]
in vivo	Human	Chromosome-wide association study (CWAS)[68]	Human lymphocytes, circulating HSC/HPC	All human chromosomes	Detects aneuploidy in all 24 chromosomes on a single 8-square slide		BZ,[16,17]
		Transcriptomics					
		-Gene expression microarrays	Human PMBC	Illumina HumanRef-8 V2 BeadChip targets	Study design with replicates can minimize potential confounding and experimental variability.	Limited by probes on chips	BZ[13, 14,73]
		-RNA sequencing	Human PMBC	All poly-A-enriched transcripts	Wider dynamic range than microarrays. Can detect coding and non-coding transcripts and splice variants	Expensive. Variability arises in experimental steps.	BZ[15]

BT, 1,2,4-benzenetriol; BZ, benzene; CML, chronic myeloid leukemia; CWAS, chromosome-wide association study; ESC, embryonic stem cell; FA, formaldehyde; HPC, hematopoietic progenitor cell; HQ, hydroquinone; HSC, hematopoietic stem cell; PBMC, peripheral blood mononuclear cells; PDA, parallel deletion analysis.