Global Gene Expression Profiling of a Population Exposed to a Range of Benzene Levels

Cliona M. McHale,[1] Luoping Zhang,[1] Qing Lan,[2] Roel Vermeulen,[3] Guilan Li,[4] Alan E. Hubbard,[1] Kristin E. Porter,[1] Reuben Thomas,[5] Christopher J. Portier,[5] Min Shen,[2] Stephen M. Rappaport,[1] Songnian Yin,[4] Martyn T. Smith,[1] and Nathaniel Rothman[2]

[1]School of Public Health, University of California–Berkeley, Berkeley, California, USA; [2]Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland, USA; [3]Institute of Risk Assessment Sciences, Utrecht University, Utrecht, the Netherlands; [4]Institute of Occupational Health and Poison Control, Chinese Center for Disease Control and Prevention, Beijing, China; [5]Environmental Systems Biology Group, Laboratory of Molecular Toxicology, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, North Carolina, USA

Background: Benzene, an established cause of acute myeloid leukemia (AML), may also cause one or more lymphoid malignancies in humans. Previously, we identified genes and pathways associated with exposure to high (> 10 ppm) levels of benzene through transcriptomic analyses of blood cells from a small number of occupationally exposed workers.

Objectives: The goals of this study were to identify potential biomarkers of benzene exposure and/or early effects and to elucidate mechanisms relevant to risk of hematotoxicity, leukemia, and lymphoid malignancy in occupationally exposed individuals, many of whom were exposed to benzene levels < 1 ppm, the current U.S. occupational standard.

Methods: We analyzed global gene expression in the peripheral blood mononuclear cells of 125 workers exposed to benzene levels ranging from < 1 ppm to > 10 ppm. Study design and analysis with a mixed-effects model minimized potential confounding and experimental variability.

Results: We observed highly significant widespread perturbation of gene expression at all exposure levels. The AML pathway was among the pathways most significantly associated with benzene exposure. Immune response pathways were associated with most exposure levels, potentially providing biological plausibility for an association between lymphoma and benzene exposure. We identified a 16-gene expression signature associated with all levels of benzene exposure.

Conclusions: Our findings suggest that chronic benzene exposure, even at levels below the current U.S. occupational standard, perturbs many genes, biological processes, and pathways. These findings expand our understanding of the mechanisms by which benzene may induce hematotoxicity, leukemia, and lymphoma and reveal relevant potential biomarkers associated with a range of exposures.

Key words: benzene, biomarker, human, microarray, transcriptomics. *Environ Health Perspect* 119:628–634 (2011). doi:10.1289/ehp.1002546 [Online 13 December 2010]

Benzene is an established cause of acute myeloid leukemia (AML) and myelodysplastic syndromes, and is a probable cause of lymphocytic malignancies (Baan et al. 2009; Vlaanderen et al. 2010), including non-Hodgkin lymphoma (NHL) in humans, as recently reviewed by Smith (2010). Benzene is also hematotoxic, even at relatively low levels of exposure (Lan et al. 2004). Possible mechanisms underlying these pathologies include the generation of free radicals leading to oxidative stress, immune system dysfunction, and decreased immune surveillance (Smith 2010). Studies of global gene expression in the bone marrow of very highly exposed mice have revealed additional potential mechanisms of benzene toxicity (Faiola et al. 2004; Yoon et al. 2003), but their relevance to risk in occupationally exposed individuals is uncertain. Toxicogenomic studies of exposed human populations are an important alternative approach to the human health risk assessment of environmental exposures. Such studies that have examined environmental exposures have identified potential biomarkers of early effects and revealed potential mechanisms underlying associated diseases (McHale et al. 2010). However, these studies have been of limited size, have mainly addressed high levels of exposure, and have often lacked precise, individual estimates of exposure. Further, such studies are limited by confounding effects and laboratory variation, especially at low doses.

We previously compared global gene expression in the peripheral blood mononuclear cell (PBMC) fractions of six to eight pairs of unexposed controls and workers exposed to high levels of benzene (> 10 ppm) and identified potential biomarkers of exposure and mechanisms of toxicity (Forrest et al. 2005; McHale et al. 2009). We chose PBMCs because they are widely used in human toxicogenomic studies. As an extension of these earlier studies, here we sought to identify potential gene expression biomarkers of exposure and early effects, as well as mechanisms of toxicity, in 125 individuals occupationally exposed to a range of benzene levels, including < 1 ppm, the current U.S. occupational standard (Occupational Safety and Health Administration 1987). In the cross-sectional molecular epidemiological study population, which includes the 125 individuals analyzed here, we previously found that white blood cell counts were decreased in workers exposed to < 1 ppm benzene compared with controls and that a highly significant dose–response relationship was present (Lan et al. 2004), with no apparent threshold within the occupational exposure range (0.2–75 ppm benzene) (Lan et al. 2006). We employed a rigorous study design that included randomization of samples across experimental variables, incorporation of precise individual measurements of exposure, and analysis with a mixed-effects model, with the aim of removing sources of biological and experimental variability (nuisance variability).

Materials and Methods

Study subjects and exposure assessment. All subjects were from a molecular epidemiology study of occupational exposure to benzene that comprised 250 benzene-exposed shoe manufacturing workers and 140 unexposed age- and sex-matched controls who worked in three clothes-manufacturing factories in the same region near Tianjin, China (Lan et al. 2004; Vermeulen et al. 2004). This study complied with all applicable requirements of U.S. and Chinese regulations,

Address correspondence to C.M. McHale, School of Public Health, 211 Hildebrand Hall, University of California–Berkeley, Berkeley, CA 94720 USA. Telephone: (510) 643-5349. Fax: (510) 642-0470. E-mail: cmchale@berkeley.edu

Supplemental Material is available online (doi:10.1289/ehp.1002546 via http://dx.doi.org/).

We thank the participants for taking part in this study.

This research was supported by National Institutes of Health (NIH) grants R01ES06721 and P42ES04705 (to M.T.S.), National Institute of Environmental Health Sciences grants P42ES05948 and P30ES10126 (to S.M.R.), and the intramural research program of the National Cancer Institute.

G.L. has received funds from the American Petroleum Institute for consulting on benzene-related health research. S.M.R. has received consulting and expert testimony fees from law firms representing plaintiffs' cases involving exposure to benzene and has received research support from the American Petroleum Institute and the American Chemistry Council. M.T.S. has received consulting and expert testimony fees from law firms representing both plaintiffs and defendants in cases involving exposure to benzene. The other authors declare they have no actual or potential competing financial interests.

Received 10 June 2010; accepted 13 December 2010.

including institutional review board approval. Participation was voluntary, and written informed consent was obtained.

Exposure assessment to benzene was performed as described previously (Vermeulen et al. 2004). For this study, we categorized exposure groups using mean individual air benzene measurements obtained during the 3 months preceding phlebotomy. A subgroup of subjects was selected from each benzene exposure category as follows: 13 workers with very high exposure (> 10 ppm), 11 workers with high exposure (5–10 ppm), 30 workers with low exposure (< 1 ppm; average < 1 ppm), and 29 workers with very low exposure (<< 1 ppm; average < 1 ppm, with most individual measurements < 1 ppm) (Table 1). We previously reported that urinary benzene and mean individual air levels of benzene were strongly correlated (Spearman $r = 0.88$, $p < 0.0001$) in the epidemiological study population (Lan et al. 2004). Among the individuals with occupational exposure to benzene in the present study for which urinary benzene levels were available ($n = 82$), a similar correlation was noted (Spearman $r = 0.76$, $p < 0.0001$). A group of 42 unexposed controls were frequency matched to the exposed subjects on the basis of age and sex. Mean age (± SD) was 29.5 ± 8.7 years for the 83 exposed workers and 29.5 ± 8.2 years for the controls.

Biological sample collection was described previously (Forrest et al. 2005; Vermeulen et al. 2004). We transferred field-stabilized samples on dry ice. We isolated RNAs using the *mir*Vana miRNA (microRNA) isolation kit (Applied Biosystems, Austin, TX, USA), stored them in aliquots at –80°C, and thawed them immediately before microarray analysis. All RNA samples analyzed had absorbance ratios for A_{260}:A_{280} and A_{260}:A_{230} between 1.7 and 2.1, and we confirmed integrity by the presence of sharp 28S and 18S rRNA bands and a ratio of 28S:18S intensity of approximately 2:1 after denaturing gel electrophoresis.

Microarray study design and analysis. We randomized samples, and thus exposure groups, across labeling and hybridization reactions and across chips as uniformly as possible [see Supplemental Material, Table 1 (doi:10.1289/ehp.1002546)]. Technical replicates ($n = 19$), randomly chosen from among the 125 study subject samples, were included in the study to assess variability in the labeling, hybridization, and chip steps of the microarray procedure. We labeled samples (200 ng) in batches of 24 using the Illumina RNA Amplification kit (Ambion, Austin, TX, USA) and hybridized them to Illumina HumanRef-8 V2 BeadChips in batches of 32 (four chips) following the manufacturer's protocol. All sample processing was performed in a blinded manner.

Data analysis. We conducted variance components analysis using a linear mixed model (Laird and Ware 1982) to assess the proportion of total variation due to variation between subjects, hybridizations, labels, and chips, both before and after normalization [quantile normalization in the affy package (Gautier et al. 2004) in R (R Development Core Team 2010)]. For each probe, we estimated the association between exposure level and expression level using a mixed-effects model with random intercepts that accounted for clustering by subject, hybridization, and label. The fixed effects in our model, in addition to benzene exposure level, included sex (1 = male, 0 = female), current smoking status (1 = yes, 0 = no), and age (in years, linear term) as potential confounders of associations between gene expression and benzene exposure. We fitted the mixed-effects model in R with the lmer function in the lme4 package (Bates and Maechler 2010). We identified differentially expressed probes as those with a statistically significant log-fold change (based on likelihood ratio tests). We computed *p*-values adjusted for multiple testing by controlling the false discovery rate (FDR) with the Benjamini-Hochberg procedure (Benjamini and Hochberg 1995), using the multtest package in R. These values are FDR-adjusted *p*-values and were considered significant if they were ≤ 0.05, the traditional experiment-wise type I error rate. The raw data discussed here have been deposited in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) (Edgar et al. 2002) and are accessible through the GEO database (accession number GSE21862; NCBI 2002).

Pathway analysis. We imported microarray probe IDs into Pathway Studio software (Ariadne Genomics, Rockville, MD, USA), and queried the ResNet 7.0 database (Ariadne Genomics) for interactions among genes and gene products derived from the current literature (Nikitin et al. 2003). We also used a method known as "structurally enhanced pathway enrichment analysis" (SEPEA_NT3) (Thomas et al. 2009), which incorporates the associated network information of KEGG (Kyoto Encyclopedia of Genes and Genomes) biochemical pathways (Kanehisa and Goto 2000; Kyoto Encyclopedia of Genes and Genomes 2000). KEGG pathways are manually drawn pathway maps representing current knowledge on the molecular interaction and reaction networks involved in cellular processes such as metabolism and the cell cycle.

Gene Ontology (GO) analysis. The GO project (The Gene Ontology Consortium 2000) provides an ontology of defined terms representing gene product properties in the domains, cellular components, molecular functions, and biological processes. GO has a hierarchical structure that forms a directed acyclic graph in which each term has defined relationships to one or more other terms in the same domain, which can be described as parent–child relationships. Every GO term is represented by a node in this graph, and the nodes are annotated with a set of genes. We used TopGO (topology-based GO scoring; Bioconductor 2010) to calculate the significance of biological terms from gene expression data taking the GO structure into account (Alexa et al. 2006). We used the "elim" algorithm, which differs from standard GO analyses in that it eliminates genes from parent nodes that are members of "significant" child nodes. The elim score is the *p*-value returned by Fisher's exact test, and a node is marked as significant if the *p*-value is smaller than a previously defined threshold (Alexa et al. 2006). Typically this threshold is set to be 0.01 divided by the number of nodes in the GO graph with at least one annotated gene. This corresponds to a Bonferroni adjustment of the *p*-values. The most highly significant nodes thus derived are denoted as key nodes.

Both TopGO and SEPEA_NT3 have limitations (Barry et al. 2005; Nettleton et al. 2008). They assume independence between expressions of the genes, violation of which can lead to greater false positives than allowed by the nominal threshold set. These methods were chosen over more computationally intensive permutation-based subject sampling approaches.

Hierarchical clustering. We performed simple supervised clustering based on complete linkage (Murtagh 1985) in order to make heat maps [hierarchical agglomerative clustering with complete linkage; implemented

Table 1. Characteristics of study subjects.

Benzene exposure category (ppm)	Subjects (*n*)	Air benzene (ppm)[a]	WBC count (per µL blood)	Age (years)	Sex [*n* (%)]		Currently smoking [*n* (%)]	
					Male	Female	Yes	No
Control (—)	42	< 0.04[b]	6454.8 ± 1746.5	29.5 ± 8.2	17 (33)	25 (34)	9 (35)	33 (33)
Very low (<< 1)[c]	29	0.3 ± 0.9	5524.1 ± 1369.2	30.3 ± 9.2	8 (16)	21 (28)	6 (23)	23 (23)
Low (< 1)[d]	30	0.8 ± 0.8	5510.0 ± 1170.7	27.9 ± 7.2	19 (37)	11 (15)	5 (19)	25 (25)
High (5–10)	11	7.2 ± 1.3	5418.2 ± 1376.8	29.7 ± 9.1	1 (2)	10 (14)	1 (4)	10 (10)
Very high (> 10)	13	24.7 ± 15.7	5176.9 ± 1326.8	30.9 ± 10.5	6 (12)	7 (9)	5 (19)	8 (8)

WBC, white blood cell. Values for air benzene, WBC count, and age are mean ± SD.
[a]Air benzene level in the 3 months preceding phlebotomy. [b]The limit of detection for benzene was 0.04 ppm (Lan et al. 2004). [c]The average level of benzene was < 1 ppm and dosimetry levels were < 1 ppm at most measurements in the 3 months preceding phlebotomy and at all measurements in the prior month. [d]The average level of benzene was < 1 ppm (in the 3 months preceding phlebotomy) but dosimetry levels were not always < 1 ppm in the previous 3 months.

in the hclust function in R (R Development Core Team 2010), called by the heatmap.2 function available with the gplots library in Bioconductor (Gentleman et al. 2004)]. Input data consisted of the four columns of $\log_2$-adjusted ratios (the coefficients from the linear mixed-effects models adjusted for both random and fixed effects). This provides clusters driven by average responses within dose groups rather than by potential confounding within groups.

Results

Application of a mixed-effects model to analyze gene expression. We applied a mixed model (variance components analysis) to assess the proportion of total variation due to variation among subjects, hybridizations, labels, and chips, among the randomly selected within-subject replicates (n = 19). Plotting the distribution of the contribution of variance across all probes after normalization revealed that the greatest source of variation was between subjects and was therefore consistent with biological causes (Figure 1). We also found substantial variation between labeling reactions. Therefore, for each probe, we estimated the association between exposure level and expression level using a mixed-effects model with (crossed) random intercepts that account for clustering by subject and by label (Laird and Ware 1982). Because the study design included randomization of samples—and thus exposures—across labeling reactions, an inferential procedure was necessary that allowed the existence of nonnested sources of correlation (labeling and subject). Thus, we used mixed models with so-called crossed random effects (Fitzmaurice et al. 2004), with the goal of providing more trustworthy inference than procedures that would have ignored, for instance, the variability caused by the labeling. (Many microarray studies are not designed to partition out the sources of variability and thus, if such sources are important, could provide misleading inference. In addition, it is often assumed that normalization will eliminate these sources of variability, but this assumption cannot be verified unless the study design allows for partitioning of the variance.) In the model, we also adjusted, as simple fixed effects, for biological variation in expression associated with differences in sex, age, and smoking status.

Effects of benzene exposure on gene expression, biological processes, and pathways. Analysis of the overall effect of benzene across the four exposure categories (very high, high, low, and very low) relative to unexposed controls (n = 42) revealed significantly altered expression (FDR-adjusted p-values ≤ 0.05) of 3,007 probes representing 2,846 genes [see Supplemental Material, Table 2 (doi:10.1289/ehp.1002546)]. Immune response (p = 3.78E-07) was the most significant key node among the GO processes associated with exposure (see Supplemental Material, Table 3), as determined by TopGO analysis. Pathway analysis by SEPEA_NT3 (Thomas et al. 2009) revealed highly significant (p < 0.001) impacts on the Toll-like receptor signaling pathway, oxidative phosphorylation, B-cell receptor signaling pathway, apoptosis, AML, and T-cell receptor signaling (see Supplemental Material, Table 4).

Large numbers of genes were significantly differentially expressed (FDR-adjusted p-values ≤ 0.05) in samples from each of the four exposure categories relative to controls [see Supplemental Material, Figure 1 and Tables 5–8 (doi:10.1289/ehp.1002546)]. We identified several GO processes implicated in

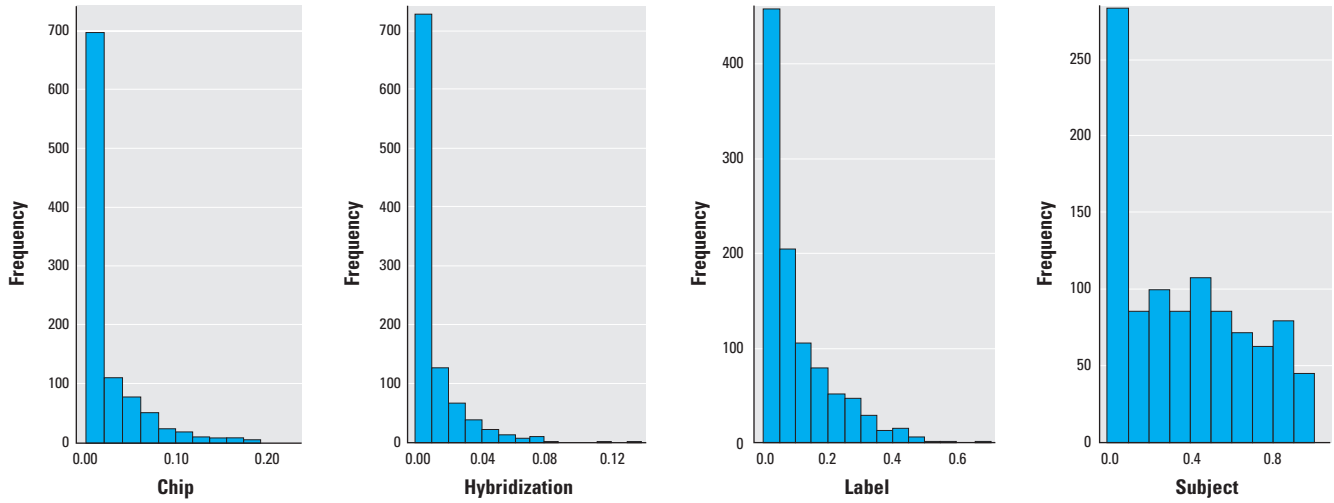

Figure 1. Distribution of the intraclass correlation coefficients (the proportion of variability estimated to come from each source on a probe-by-probe basis) calculated by variance components analysis based on a mixed-effects model allowing assessment of independent contributions of variability from chip, hybridization, label, and biological (subject), as well as residual variability.

Table 2. Summary of GO categories overrepresented at each benzene exposure category.

GO ID[a]	GO term	Total no. of genes[b]	Very low (n = 29)		Low (n = 30)		High (n = 11)		Very high (n = 13)	
			No. genes	p-Value[c]	No. genes	p-Value[c]	No. genes	p-Value[c]	No. genes	p-Value[c]
GO:0006412	translation	456	64	2.0E-06	93	1.2E-03				
GO:0006512	ubiquitin cycle	480	48	7.5E-04	98	1.6E-05				
GO:0006917	induction of apoptosis	216	27	4.1E-04	49	1.6E-04	19	1.5E-03[d]		
GO:0006955	immune response	653	58	3.7E-03[d]	124	4.6E-05	54	4.9E-06	97	1.1E-04
GO:0015986	ATP synthesis coupled proton transport	40	11	2.2E-05	14	5.0E-04			11	1.8E-03
GO:0006915	apoptosis	804	80	5.6E-03	158	9.2E-04			107	2.7E-03
GO:0030301	cholesterol transport	8	5	4.4E-05	4	1.5E-02[d]			4	5.5E-03[d]
GO:0006954	inflammatory response	318			60	4.6E-03[d]	34	2.8E-05		

[a]GO categories that are significant at ≥ 2 doses. [b]Number of annotated genes included on the chip. [c]p-Values were determined using the elim method in TopGO, which computes the statistical significance of a parent node dependent on the significance of its children by Fisher's exact test; nodes are significant if the p-value is smaller than a previously defined threshold (Alexa et al. 2006), 0.01 divided by the number of nodes in the GO graph with at least one annotated gene. [d]Significantly enriched term in classic analysis (which does not take GO hierarchy into account) but not in elim analysis in TopGO. Complete GO data are available in Supplemental Material, Table 9 (doi:10.1289/ehp.1002546).

the overall analysis as key nodes across three to four dose categories, including immune response, apoptosis, and ATP synthesis–coupled proton transport [Table 2; for complete data, see Supplemental Material, Table 9).

Similarly, multiple pathways found to be highly significant in the overall analysis ($p \le 0.005$), including Toll-like receptor signaling, oxidative phosphorylation, B-cell receptor signaling, apoptosis, AML, and T-cell receptor signaling, were enriched among the differentially expressed genes associated with three (including the very low dose category) or four exposure categories [Table 3; for complete data, see Supplemental Material, Table 10 (doi:10.1289/ehp.1002546)].

Twelve genes were up-regulated ≥ 1.5-fold at all four doses relative to unexposed controls, including five genes [*PTX3* (pentraxin-related gene), *CD44* (CD44 antigen), *PTGS2* (prostaglandin-endoperoxide synthase 2), *IL1A* (interleukin 1, alpha), and *SERPINB2* (serpin peptidase inhibitor, clade B, member 2) with FDR-adjusted *p*-values ≤ 0.005. An additional four genes were up-regulated > 1.5-fold at the top three doses, and > 1.3-fold at the lowest dose (Table 4). Expression of each of the 16 signature genes across the five exposure categories shows a distinct pattern, with the highest expression in the < 1-ppm (low) exposure group [see Supplemental Material, Figure 2 (doi:10.1289/ehp.1002546)]. The 16 genes are involved in immune response, inflammatory response, cell adhesion, cell–matrix adhesion, and blood coagulation (see Supplemental Material, Table 11). Ten of the 16 genes (or their products), 7 of which are involved in inflammatory response (p = 1.4E-12), form a network (Figure 2) with central roles for *IL1A* and *PTGS2*.

Dose-specific effects. We used supervised hierarchical clustering to generate a heat map to allow visualization of patterns of gene expression across exposure categories. One group of genes (~ 100) exhibited reduced expression (ratios < 1) with increasing dose relative to controls, whereas a second group (~ 100) appeared to be elevated at all doses but more so at low-dose exposure (Figure 3).

We also observed dose-dependent effects on biological processes and pathways. For example, nucleosome assembly [see Supplemental Material, Table 9 (doi:10.1289/ehp.1002546)] and the ATP-binding cassette (ABC) transporter pathway (see Supplemental Material, Table 10) appeared to be deregulated only at the very high exposure level. Among 78 genes that were highly significantly (FDR *p*-value ≤ 0.05) associated with a ≥ 1.5-fold increase in expression in the very high exposure group, and not significantly altered at any of the other exposure categories relative to controls, a network involving 19 genes (or their products) was apparent, in which v-src sarcoma viral oncogene homolog (*SRC*) and matrix metallopeptidase 9 (*MMP9*) play central roles (see Supplemental Material, Figure 3). Among 29 genes significantly altered only at low-dose benzene exposure,

Table 3. *p*-Values for pathways altered at each benzene exposure category.

	Benzene exposure category			
Pathway name[a]	Very low (n = 29)	Low (n = 30)	High (n = 11)	Very high (n = 13)
Chronic myeloid leukemia	0.034	0.033		
Pancreatic cancer	0.023	0.007		
Oxidative phosphorylation[b]	< 0.001	0.003	0.001	
Small-cell lung cancer[b]	0.004	0.002	0.027	
B-cell receptor signaling pathway[b]	0.008	0.003	0.004	
Insulin signaling pathway	0.015	0.035	0.052	
Adipocytokine signaling pathway	0.034	0.002	0.019	
Circadian rhythm—mammal	0.04	0.045	0.004	
RNA polymerase	< 0.001		0.048	
Toll-like receptor signaling pathway[b]	< 0.001	0.002	0.001	0.004
Epithelial cell signaling in *Helicobacter pylori* infection[b]	< 0.001	0.003	0.006	0.011
GPI-anchor biosynthesis[b]	< 0.001	0.041	< 0.001	0.007
T-cell receptor signaling pathway[b]	0.005	0.002	0.005	0.018
Apoptosis[b]	0.007	0.002	0.007	0.013
Cytokine–cytokine receptor interaction[b]	0.036	0.011	0.030	0.004
AML[b]	0.037	0.002		0.045
Fatty acid metabolism	0.037		0.049	0.033
Nucleotide excision repair	0.001		0.008	0.005
Renal cell carcinoma		0.024	0.015	
Protein export		0.053	0.024	
Steroid biosynthesis			0.004	0.034
Fc epsilon RI signaling pathway		0.006		0.046
Jak-STAT signaling pathway		0.003		0.048
MAPK signaling pathway		0.009		0.023

[a]KEGG pathways that are significant at ≥ 2 doses. [b]FDR-adjusted *p*-value (Benjamini and Hochberg 1995) < 0.005 in overall analysis. Details of all KEGG pathways are available from Kyoto Encyclopedia of Genes and Genomes (2000).

Table 4. Potential biomarkers of benzene exposure based on gene expression ratios relative to unexposed controls.

			Benzene exposure category							
			Very low (n = 29)		Low (n = 30)		High (n = 11)		Very high (n = 13)	
Probe ID	Symbol	Definition	Ratio	*p*-Value[a]	Ratio	*p*-Value[a]	Ratio	*p*-Value[a]	Ratio	*p*-Value[a]
5090327	*SERPINB2*[b]	serpin peptidase inhibitor, clade B, member 2	2.47	0.002	5.19	0.000	3.03	0.005	3.39	0.001
2370524	*TNFAIP6*	tumor necrosis factor, alpha-induced protein 6	2.26	0.000	2.94	0.000	1.72	0.030	2.13	0.000
6590338	*IL1A*[b]	interleukin 1, alpha	2.00	0.001	3.03	0.000	2.36	0.000	2.53	0.000
1260746	*KCNJ2*	potassium inwardly-rectifying channel, subfamily J	1.97	0.000	2.54	0.000	2.09	0.000	1.56	0.012
2230131	*PTX3*[b]	pentraxin-related gene, rapidly induced by IL-1 beta	1.80	0.000	2.30	0.000	1.62	0.003	1.81	0.000
5860333	*F3*	coagulation factor III (thromboplastin, tissue factor)	1.73	0.003	2.83	0.000	1.78	0.034	2.41	0.001
1410189	*CD44*[b]	CD44 antigen (Indian blood group)	1.64	0.000	1.76	0.000	1.64	0.005	1.78	0.000
2470100	*CCL20*	chemokine (C-C motif) ligand 20	1.63	0.005	2.30	0.000	1.59	0.041	2.11	0.000
4880717	*ACSL1*	acyl-CoA synthetase long-chain family member 1	1.63	0.001	1.79	0.000	1.59	0.010	1.68	0.002
1470682	*PTGS2*[b]	prostaglandin-endoperoxide synthase 2	1.60	0.000	1.98	0.000	1.68	0.003	1.75	0.000
1770152	*CLEC5A*	C-type lectin domain family 5, member A	1.57	0.009	2.26	0.000	1.78	0.014	2.26	0.000
4060674	*IL1RN*	interleukin 1 receptor antagonist	1.55	0.003	2.26	0.000	1.54	0.020	1.61	0.004
7320646	*PRG2*	proteoglycan 2, bone marrow	1.37	0.011	1.83	0.000	1.5	0.007	1.69	0.000
650709	*SLC2A6*	solute carrier family 2, member 6	1.36	0.005	1.72	0.000	1.5	0.000	1.60	0.000
2900286	*GPR132*	G protein-coupled receptor 132	1.34	0.047	1.87	0.000	1.6	0.003	1.80	0.000
3710379	*PLAUR*	plasminogen activator, urokinase receptor	1.29	0.035	1.80	0.000	1.6	0.002	1.58	0.001

Genes shown are up- or down-regulated ≥ 1.5-fold relative to unexposed controls at three or four doses.
[a]FDR-adjusted *p*-value (Benjamini and Hochberg 1995). [b]Genes that have *p*-values ≤ 0.005 at all four doses.

we identified a network of 15 genes involved in immune response (p = 4E-12), with central roles for interferon gamma (*IFNG*) and tumor necrosis factor (*TNF*) (see Supplemental Material, Figure 4). Together, these data suggest that benzene induces dose-dependent effects, with the caveat that differences in power among the different exposure categories may have influenced the resulting significant gene lists.

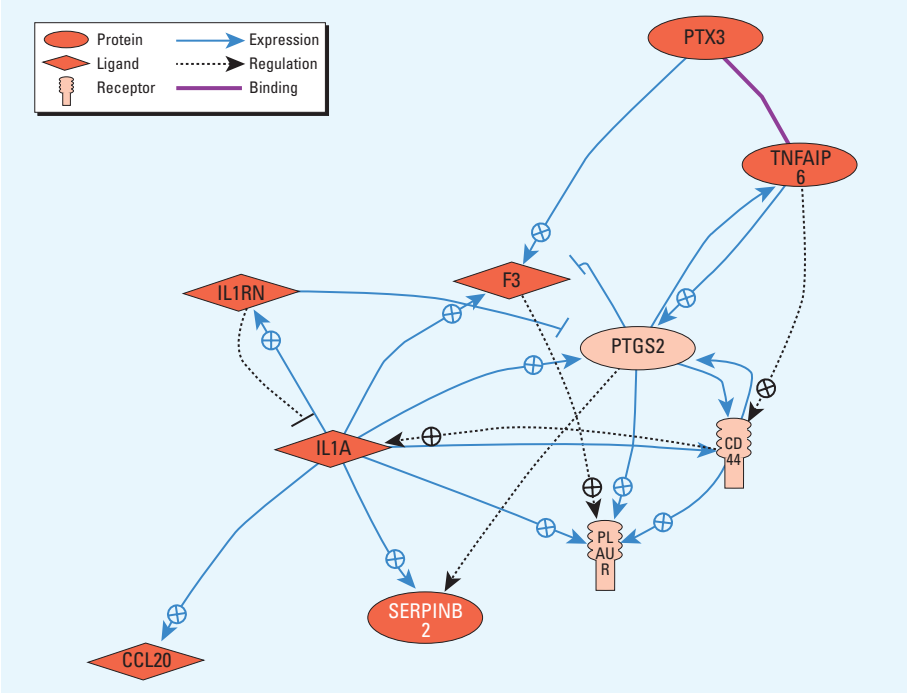

Figure 2. Network interactions among biomarkers of benzene exposure associated with all exposure levels, illustrating a high degree of interrelatedness based on the literature, with central roles for *IL1A* and *PTGS2*. Pathway Studio software identified interactions among 10 of the 16 potential biomarkers of benzene exposure. The interactions are mainly expression, with some regulation (regulator changes the activity of the target) and one binding interaction. Red indicates up-regulation.

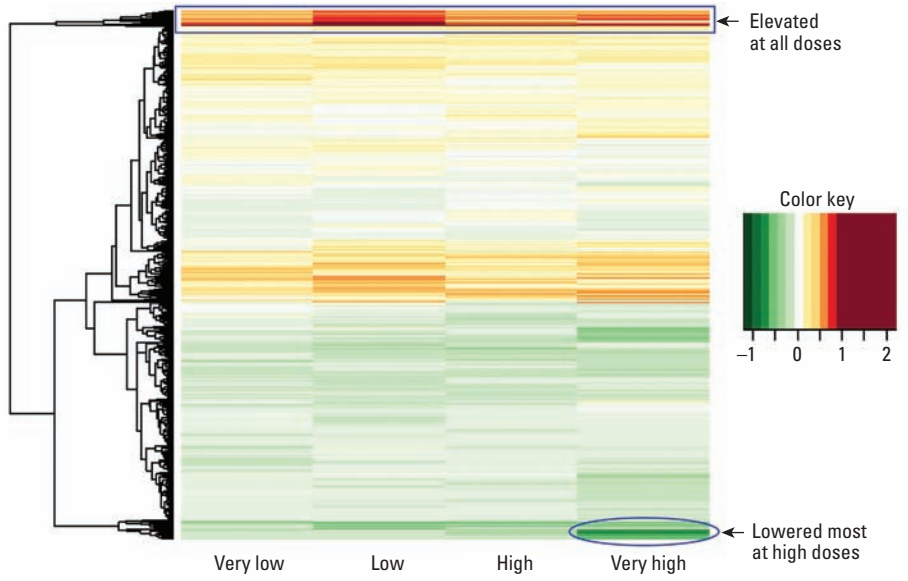

Figure 3. Dose-dependent effects on gene expression. A heat map illustrates simple hierarchical clustering of the differentially expressed 3,007 probes (FDR-adjusted p-value < 0.05) based on the mixed model described in "Materials and Methods." The clustering was done on the four $\log_2$ expression ratios (derived as coefficients returned from the mixed model) all relative to controls. The color key relates to the $\log_2$ ratios observed. Clustering of genes was based on complete linkage (for algorithmic details of algorithms used, see Murtagh 1985), as implemented in the hclust function in R, called by the heatmap.2 function available with the gplots library in Bioconductor (Gentleman et al. 2004). Note that the clustering is based on Euclidean distance.

Discussion

Technical variation is often ignored in human toxicogenomic studies, leading to potential bias in differential expression arising from correlation with technical variation. In the present study, we applied a rigorous study design to assess sources of both potential confounding and experimental variability (nuisance variation) and analyzed the data using statistical techniques that incorporate nonnested sources of variation (i.e., those not eliminated by normalization) and that return estimates of least variability with accurate inference (linear mixed-effects models). This approach increased the power to detect associations between benzene exposure and gene expression, even at low-dose exposure levels.

More genes remained significantly up- or down-regulated compared with controls after multiple test correction in the present study than in an earlier study examining samples from eight pairs of exposed workers and unexposed controls on the Illumina platform (McHale et al. 2009), likely because of the increased number of individuals and the rigorous approach to study design. Nonetheless, we identified 247 genes in both study populations using the Illumina platform. Of 488 significant genes cross-validated on both Illumina and Affymetrix platforms (McHale et al. 2009), 147 genes were significant in the present study. *ZNF331* (zinc finger protein 331), significant after multiple test correction in individuals occupationally exposed to benzene at levels > 10 ppm compared with controls in two earlier studies (Forrest et al. 2005; McHale et al. 2009), was significantly up-regulated at both < 1 ppm and > 10 ppm in the present study.

The finding that genes in the AML pathway were strongly associated with multiple exposure levels of benzene provides support for our approach because epidemiological studies have established that benzene causes AML (Baan et al. 2009; Smith 2010). However, such disease associations must be treated cautiously because the KEGG pathway information, on which the pathway analyses were based, is limited for AML, and a KEGG pathway for NHL has not been defined. Information about altered molecular and cellular processes can provide biological plausibility for probable disease associations. Immune response, previously found to be associated with > 10 ppm benzene exposure in our earlier transcriptomic study of eight high-exposed control pairs (McHale et al. 2009), was one of the major processes significantly altered across multiple exposure levels in the present study, involving both innate (Toll-like receptor signaling) and adaptive (B-cell receptor signaling and T-cell receptor signaling pathway) responses. Additionally, we found central roles for the proinflammatory cytokines *IFNG* and *TNF* among genes uniquely altered

at low-dose exposure in the present study. A single nucleotide polymorphism in *TNF*-α was previously associated with susceptibility to bone marrow dysplasia in chronic benzene poisoning (Lv et al. 2007). Further, genetic variation in *TNF* (Rothman et al. 2006), Toll-like receptor genes (Purdue et al. 2009), and *IFNG* (Colt et al. 2009) has previously been associated with NHL risk. Deregulation of pathways involving these genes through sustained alterations in expression provides biological plausibility for the association of benzene with lymphoid neoplasms.

Findings from the present study are consistent with previous reports of adverse effects of benzene on oxidative stress (Kolachana et al. 1993) and mitochondria (Inayat-Hussain and Ross 2005). Here, we found highly significant associations with ATP synthesis–coupled proton transport and oxidative phosphorylation at all levels of benzene exposure relative to unexposed controls. Expression of superoxide dismutase (*SOD*), a mitochondrial defense against reactive oxygen species, was up-regulated in the present study by 50–100% relative to controls. *HMOX1* [heme oxygenase (decycling) 1], an antioxidant and suppressor of TNF-α signaling (Lee et al. 2009), was down-regulated in the low-dose benzene exposure group. Increased mitochondrial membrane permeability potential induced by benzene metabolites (Inayat-Hussain and Ross 2005) can lead to the initiation of apoptosis. Indeed, apoptosis was associated with all benzene doses in the present study, consistent with our earlier observation of an association with high-dose benzene exposure (> 10 ppm) (McHale et al. 2009).

Previously, we found that chromatin assembly was significantly altered after high-dose benzene exposure (McHale et al. 2009). The finding that nucleosome assembly (a GO category nested within chromatin assembly) was overrepresented in the highest exposure category in the present study confirms and clarifies this potential mechanism of benzene-associated leukemia.

Although significant involvement of the p53 response pathway was previously found in mice exposed to very high levels of benzene (Faiola et al. 2004; Yoon et al. 2003), we did not find such involvement in the present study or in our earlier studies, and the immune and inflammatory effects we found here in humans were not recapitulated in the mouse microarray studies (Faiola et al. 2004; Yoon et al. 2003). These differences suggest that human toxicogenomic studies may be more relevant than animal studies, although differences in exposure levels, tissues examined, and uncontrolled confounding in the human study could also be contributing factors.

Our findings suggest two novel hypotheses regarding benzene toxicity. Glycosylphosphatidylinositol (GPI)-anchor biosynthesis was associated with all doses of benzene exposure in the present study. The GPI anchor is a C-terminal posttranslational modification that anchors the modified protein in the outer leaflet of the cell membrane and putatively plays roles in lipid raft partitioning, signal transduction, and cellular communication (Paulick and Bertozzi 2008). Because epigenetic silencing of genes involved in GPI-anchor biosynthesis may be important in human disease, including lymphomas (Hu et al. 2009), further investigation of its role in benzene-associated disease is warranted.

ABC transporters were associated highly significantly with only the highest (> 10 ppm) benzene dose. In addition to their capacity to extrude cytotoxic drugs, ABC transporters are known to play important roles in the development, differentiation, and maturation of immune cells and are involved in migration of immune effector cells to sites of inflammation (van de Ven et al. 2009).

Our findings also suggest a potential gene expression signature of benzene exposure. In particular, *IL1A* and *PTGS2* played central roles in the interaction network characterizing the gene expression signature associated with benzene in this study. Both molecules are produced by activated macrophages and other cells in inflammatory responses. A single nucleotide polymorphism that increases *IL1A* mRNA expression has been inversely associated with granulocyte count in benzene-exposed individuals (Lan et al. 2005). Overexpression of *PTGS2*, which occurs frequently in premalignant and malignant neoplasms, including hematological malignancies (Bernard et al. 2008), together with overexpression of the prostaglandin cascade, leads to carcinogenesis through a progressive series of highly specific cellular and molecular changes (Harris 2009).

The expression pattern of the signature genes suggests a nonlinear response to benzene. Other biomarkers evaluated in populations exposed to benzene have shown similar patterns, including hematotoxicity (Lan et al. 2004), benzene metabolism (Kim et al. 2006), and the generation of protein adducts (Rappaport et al. 2002, 2005). Further characterization of the expression levels of these genes across a range of benzene exposures in a larger, independent study is necessary to determine the applicability of the signature genes as biomarkers of early effects and to explore more formally the shape of the dose–response curve.

Conclusion

We have identified gene expression biomarkers of early effects across a range of benzene exposures. Our findings support previously reported mechanisms relevant to adverse effects of benzene and suggest potential novel mechanisms for benzene toxicity. Future work should include validation of the potential biomarkers and determining whether the gene expression changes are effected through epigenetic processes such as DNA methylation (Bollati et al. 2007) and miRNA expression.

References

Alexa A, Rahnenfuhrer J, Lengauer T. 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics 22(13):1600–1607.

Baan R, Grosse Y, Straif K, Secretan B, El Ghissassi F, Bouvard V, et al. 2009. A review of human carcinogens—part F: chemical agents and related occupations. Lancet Oncol 10(12):1143–1144.

Barry WT, Nobel AB, Wright FA. 2005. Significance analysis of functional categories in gene expression studies: a structured permutation approach. Bioinformatics 21(9):1943–1949.

Bates D, Maechler M. 2010. lme4: Linear Mixed-Effects Models Using S4 Classes. R Package Version 0.999375-34. Available: http://cran.r-project.org/web/packages/lme4/index.html [accessed 30 September 2010].

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B (57):289–300.

Bernard MP, Bancos S, Sime PJ, Phipps RP. 2008. Targeting cyclooxygenase-2 in hematological malignancies: rationale and promise. Curr Pharm Des 14(21):2051–2060.

Bioconductor. 2010. topGO: Enrichment analysis for Gene Ontology. Available: http://www.bioconductor.org/help/bioc-views/release/bioc/html/topGO.html [accessed 18 March 2011].

Bollati V, Baccarelli A, Hou L, Bonzini M, Fustinoni S, Cavallo D, et al. 2007. Changes in DNA methylation patterns in subjects exposed to low-dose benzene. Cancer Res 67(3):876–880.

Colt JS, Rothman N, Severson RK, Hartge P, Cerhan JR, Chatterjee N, et al. 2009. Organochlorine exposure, immune gene variation, and risk of non-Hodgkin lymphoma. Blood 113(9):1899–1905.

Edgar R, Domrachev M, Lash AE. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30(1):207–210.

Faiola B, Fuller ES, Wong VA, Recio L. 2004. Gene expression profile in bone marrow and hematopoietic stem cells in mice exposed to inhaled benzene. Mutat Res 549(1–2):195–212.

Fitzmaurice GM, Laird NM, Ware JH. 2004. Applied Longitudinal Analysis. New York:John Wiley and Sons.

Forrest MS, Lan Q, Hubbard AE, Zhang L, Vermeulen R, Zhao X, et al. 2005. Discovery of novel biomarkers by microarray analysis of peripheral blood mononuclear cell gene expression in benzene-exposed workers. Environ Health Perspect 113:801–807.

Gautier L, Cope L, Bolstad BM, Irizarry RA. 2004. affy—Analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 20(3):307–315.

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5(10):R80; doi:10.1186/gb-2004-5-10-r80 [Online 15 September 2004].

Harris RE. 2009. Cyclooxygenase-2 (cox-2) blockade in the chemoprevention of cancers of the colon, breast, prostate, and lung. Inflammopharmacology 17(2):55–67.

Hu R, Mukhina GL, Lee SH, Jones RJ, Englund PT, Brown P, et al. 2009. Silencing of genes required for glycosylphosphatidylinositol anchor biosynthesis in Burkitt lymphoma. Exp Hematol 37(4):423–434.

Inayat-Hussain SH, Ross D. 2005. Intrinsic pathway of hydroquinone induced apoptosis occurs via both caspase-dependent and caspase-independent mechanisms. Chem Res Toxicol 18(3):420–427.

Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 28(1):27–30.

Kim S, Vermeulen R, Waidyanatha S, Johnson BA, Lan Q, Smith MT, et al. 2006. Modeling human metabolism of benzene following occupational and environmental exposures. Cancer Epidemiol Biomarkers Prev 15(11):2246–2252.

Kolachana P, Subrahmanyam VV, Meyer KB, Zhang L, Smith MT. 1993. Benzene and its phenolic metabolites produce oxidative DNA damage in HL60 cells in vitro and in the bone marrow in vivo. Cancer Res 53(5):1023–1026.

Kyoto Encyclopedia of Genes and Genomes. 2000. KEGG

Pathway Database. Available: http://www.genome.jp/kegg/pathway.html [accessed 18 March 2011].
Laird NM, Ware JH. 1982. Random-effects models for longitudinal data. Biometrics 38(4):963–974.
Lan Q, Vermeulen R, Zhang L, Li G, Rosenberg PS, Alter BP, et al. 2006. Benzene exposure and hematotoxicity: response [Letter]. Science 312(5776):998–999.
Lan Q, Zhang L, Li G, Vermeulen R, Weinberg RS, Dosemeci M, et al. 2004. Hematotoxicity in workers exposed to low levels of benzene. Science 306(5702):1774–1776.
Lan Q, Zhang L, Shen M, Smith MT, Li G, Vermeulen R, et al. 2005. Polymorphisms in cytokine and cellular adhesion molecule genes and susceptibility to hematotoxicity among workers exposed to benzene. Cancer Res 65(20):9574–9581.
Lee IT, Luo SF, Lee CW, Wang SW, Lin CC, Chang CC, et al. 2009. Overexpression of HO-1 protects against TNF-α-mediated airway inflammation by down-regulation of TNFR1-dependent oxidative stress. Am J Pathol 175(2):519–532.
Lv L, Kerzic P, Lin G, Schnatter AR, Bao L, Yang Y, et al. 2007. The TNF-α 238A polymorphism is associated with susceptibility to persistent bone marrow dysplasia following chronic exposure to benzene. Leuk Res 31(11):1479–1485.
McHale CM, Zhang L, Hubbard AE, Smith MT. 2010. Toxicogenomic profiling of chemically exposed humans in risk assessment. Mutat Res 705(3):172–183.
McHale CM, Zhang L, Lan Q, Li G, Hubbard AE, Forrest MS, et al. 2009. Changes in the peripheral blood transcriptome associated with occupational benzene exposure identified by cross-comparison on two microarray platforms. Genomics 93:343–349.
Murtagh F. 1985. Multidimensional Clustering Algorithms. Compstat Lectures 4. Vienna:Physica-Verlag.
NCBI (National Center for Biotechnology Information). 2002. Gene Expression Omnibus (GEO). Available: http://www.ncbi.nlm.nih.gov/geo/ [accessed 22 March 2011].
Nettleton D, Recknor J, Reecy JM. 2008. Identification of differentially expressed gene categories in microarray studies using nonparametric multivariate analysis. Bioinformatics 24(2):192–201.
Nikitin A, Egorov S, Daraselia N, Mazo I. 2003. Pathway studio—the analysis and navigation of molecular networks. Bioinformatics 19(16):2155–2157.
Occupational Safety and Health Administration. 1987. Occupational exposure to benzene. Fed Reg 52:34460–34579.
Paulick MG, Bertozzi CR. 2008. The glycosylphosphatidylinositol anchor: a complex membrane-anchoring structure for proteins. Biochemistry 47(27):6991–7000.
Purdue MP, Lan Q, Wang SS, Kricker A, Menashe I, Zheng TZ, et al. 2009. A pooled investigation of Toll-like receptor gene variants and risk of non-Hodgkin lymphoma. Carcinogenesis 30(2):275–281.
R Development Core Team. 2010. R: A Language and Environment for Statistical Computing. Available: http://www.r-project.org [accessed 25 March 2011].
Rappaport SM, Waidyanatha S, Yeowell-O'Connell K, Rothman N, Smith MT, Zhang L, et al. 2005. Protein adducts as biomarkers of human benzene metabolism. Chem Biol Interact 153–154:103–109.
Rappaport SM, Yeowell-O'Connell K, Smith MT, Dosemeci M, Hayes RB, Zhang L, et al. 2002. Non-linear production of benzene oxide-albumin adducts with human exposure to benzene. J Chromatogr B Analyt Technol Biomed Life Sci 778(1–2):367–374.
Rothman N, Skibola CF, Wang SS, Morgan G, Lan Q, Smith MT, et al. 2006. Genetic variation in TNF and IL10 and risk of non-Hodgkin lymphoma: a report from the InterLymph Consortium. Lancet Oncol 7(1):27–38.
Smith MT. 2010. Advances in understanding benzene health effects and susceptibility. Annu Rev Public Health 31:133–148.
The Gene Ontology Consortium 2000. Gene ontology: tool for the unification of biology. Nat Genet 25(1):25–29. Available: http://www.geneontology.org/GO_nature_genetics_2000.pdf [accessed 18 March 2011].
Thomas R, Gohlke JM, Stopper GF, Parham FM, Portier CJ. 2009. Choosing the right path: enhancement of biologically relevant sets of genes or proteins using pathway structure. Genome Biol 10(4):R44; doi:10.1186/gb-2009-10-4-r44 [Online 24 April 2009].
van de Ven R, Oerlemans R, van der Heijden JW, Scheffer GL, de Gruijl TD, Jansen G, et al. 2009. ABC drug transporters and immunity: novel therapeutic targets in autoimmunity and cancer. J Leukoc Biol 86(5):1075–1087.
Vermeulen R, Li G, Lan Q, Dosemeci M, Rappaport SM, Bohong X, et al. 2004. Detailed exposure assessment for a molecular epidemiology study of benzene in two shoe factories in China. Ann Occup Hyg 48(2):105–116.
Vlaanderen J, Lan Q, Krombout H, Rothman N, Vermeulen R. 2010. Occupational benzene exposure and the risk of lymphoma subtypes: a meta-analysis of cohort studies incorporating three study quality dimensions. Environmental Health Perspect 119:159–167.
Yoon BI, Li GX, Kitada K, Kawasaki Y, Igarashi K, Kodama Y, et al. 2003. Mechanisms of benzene-induced hematotoxicity and leukemogenicity: cDNA microarray analyses using mouse bone marrow tissue. Environ Health Perspect 111:1411–1420.