

## **Changes in the Peripheral Blood Transcriptome Associated with Occupational Benzene Exposure Identified by Cross-Comparison on Two Microarray Platforms**

CLIONA M. MCHALE<sup>1</sup>, LUOPING ZHANG<sup>1</sup>, QING LAN<sup>2</sup>, GUILAN LI<sup>3</sup>, ALAN E. HUBBARD<sup>1</sup>, MATTHEW S. FORREST<sup>1</sup>, ROEL VERMEULEN<sup>4</sup>, JINSONG CHEN<sup>1</sup>, MIN SHEN<sup>2</sup>, STEPHEN M. RAPPAPORT<sup>5</sup>, SONGNIAN YIN<sup>3</sup>, MARTYN T. SMITH<sup>1</sup>, and NATHANIEL ROTHMAN<sup>2</sup>.

<sup>1</sup>School of Public Health, University of California, Berkeley, CA

<sup>2</sup>Division of Cancer Epidemiology and Genetics, NCI, NIH, DHHS, Bethesda, MD

<sup>3</sup>Chinese Center for Disease Control and Prevention, Beijing, China

<sup>4</sup>Institute of Risk assessment Sciences, Utrecht University, Utrecht, the Netherlands.

<sup>5</sup>School of Public Health, University of North Carolina, Chapel Hill, NC

Corresponding Author: Cliona M. McHale, <sup>1</sup>School of Public Health, 140 Warren Hall, University of California, Berkeley, CA 94720. (510) 6435349. [cmchale@berkeley.edu](mailto:cmchale@berkeley.edu)

---

## ABSTRACT

Benzene is an established cause of leukemia and a possible cause of lymphoma in humans but the molecular pathways underlying this remain largely undetermined. This study sought to determine if the use of two different microarray platforms could identify robust global gene expression and pathway changes associated with occupational benzene exposure in the peripheral blood mononuclear cell (PBMC) gene expression of a population of shoe-factory workers with well-characterized occupational exposures to benzene. Microarray data was analyzed by a robust t-test using a Quantile Transformation (QT) approach. Differential expression of 2692 genes using the Affymetrix platform and 1828 genes using the Illumina platform was found. While the overall concordance in genes identified as significantly associated with benzene exposure between the two platforms was 26 % (475 genes), the most significant genes identified by either array were more likely to be ranked as significant by the other platform (Illumina = 64%, Affymetrix = 58 %). Expression ratios were similar among the concordant genes (mean difference in expression ratio = 0.04, standard deviation = 0.17). Four genes (*CXCL16*, *ZNF331*, *JUN* and *PF4*), which we previously identified by microarray and confirmed by real-time PCR, were identified by both platforms in the current study and were among the top 100 genes. Gene Ontology analysis showed overrepresentation of genes involved in apoptosis among the concordant genes while Ingenuity® Pathway Analysis (IPA) identified pathways related to lipid metabolism. Using a two-platform approach allows for robust changes in the PBMC transcriptome of benzene-exposed individuals to be identified.

Key words: benzene exposure, gene expression, human blood

## INTRODUCTION

Benzene is an established cause of leukemia and a possible cause of lymphoma in humans (1). A possible mechanism underlying these pathologies is the induction by benzene of genetic changes leading to chromosome aberrations, translocations, aneuploidy and long-arm deletions (2, 3) along with alterations in cell differentiation and immune surveillance. Benzene is hematotoxic, causing a decrease in total white blood cells, granulocytes and lymphocytes even among workers with relatively low level exposure to benzene (4). Benzene is thought to lower blood cell counts via metabolite effects on hematopoietic progenitor cells (4, 5). Depression of the mitogenic response of B and T lymphocytes, as well as impairment of macrophage activity, also result from benzene exposure (6). Damage to the bone marrow stromal microenvironment is another aspect of benzene associated hematotoxicity (7, 8). Individual susceptibility to the genotoxic and hematotoxic effects of benzene is mediated through polymorphisms in DNA-repair genes (9), cytokine and cell adhesion genes (3), and genes involved in benzene metabolism (4, 10).

While pathological outcome and susceptibility studies have generated some understanding of the mechanisms of action of benzene, global gene expression studies have the ability to inform on a more detailed level the involvement of specific genes and molecular pathways. The p53-dependent nature of benzene toxicity and carcinogenesis was revealed by examination of gene expression changes in mouse bone marrow (BM) in response to a 2-week exposure to inhaled benzene at 300 ppm (11). Gene expression in mouse hematopoietic stem cells (HSC) exposed to inhaled benzene (100 ppm) implicated a number of response pathways including apoptosis, growth control of damaged HSC, repair of damaged DNA, and HSC growth arrest (12). We

previously identified several genes (*ZNF331*, *CXCL16*, *JUN*, and *PF4*) altered by benzene in peripheral blood mononuclear cells (PBMC) from benzene-exposed (>10 ppm) workers compared with unexposed controls. The genes were identified by applying high-throughput microarray analysis to discover potential biomarkers and relatively low-throughput real-time PCR for confirmation (3).

In order to confirm previous findings and to discover more differentially expressed genes associated with benzene exposure, in the current study we analyzed more samples using the Affymetrix microarray platform, and have expanded the study to include a second microarray platform (Illumina). Recent reports have shown good inter-platform reproducibility of gene expression measurements between these two platforms (13). The approach integrates high-throughput confirmation with discovery, helping to further elucidate genetic pathways and mechanisms underlying hematotoxicity induced by benzene exposure.

## MATERIALS AND METHODS

### *Study Subjects*

Eight highly exposed workers (mean air benzene level  $\pm$  SD =  $39.0 \pm 25.5$  ppm) and eight unexposed controls ( $<0.04$  ppm) who were frequency-matched to these subjects on the basis of age and gender, were chosen from a large molecular epidemiology study (4) investigating occupational exposure to benzene. Six of the individuals were included in a previous study (3). The study was approved by institutional review boards at all institutions. Participation was voluntary, written informed consent was obtained, and the participation rate was approximately 95%. Four pairs were male and the other four female. Mean age was  $33.5 \pm 7.0$  years for the eight exposed workers and  $35.4 \pm 7.0$  years for the controls. Exposure assessment, biologic sample collection and RNA isolation were described previously (3, 14). A single RNA isolation was performed from each individual and stored in aliquots.

### *Affymetrix microarray analysis*

The Affymetrix Human U133 GeneChip set containing ~44,000 probes targeting >39,000 unique transcripts derived from approximately 33,000 well-substantiated genes, are included in this chip set. The complete protocol used for Affymetrix microarray analysis was described previously (3).

*Illumina Microarray analysis*

RNA samples were labeled using the Illumina<sup>®</sup> RNA Amplification kit (Ambion, Austin, TX) according to the manufacturer's protocol. RNA samples with  $A_{260}:A_{280}$  ratios between 1.7 and 2.1, and with integrity confirmed by denaturing agarose gel electrophoresis, were reverse transcribed in 20 $\mu$ L reactions comprising 200 ng sample RNA, 1X First Strand Buffer, dNTPs, RNase inhibitor, and ArrayScript enzyme. Reactions were incubated at 37°C for 2 hr after which components of the second strand synthesis reaction including 10X Second Strand Buffer, dNTP mix, DNA polymerase, and RNase H were added to yield a final reaction volume of 100 $\mu$ L. Reactions were incubated at 16°C for 2hr and the resulting cDNAs were purified. cDNA binding buffer (250 $\mu$ L) was added to each reaction which was then mixed and passed through a cDNA filter cartridge by centrifugation at 10,000 x g for 1 min. Filters were washed with wash Buffer (500 $\mu$ L) and dried by centrifugation for an additional minute. cDNA was eluted using 2 x 10 $\mu$ L aliquots of Nuclease-free Water at 55°C. The purified cDNA was dried to completion in a vacuum centrifuge concentrator set to medium heat and resuspended in 10 $\mu$ L in vitro transcription (IVT) reaction mix comprising 1X reaction buffer, dNTP mix, biotin labeled UTP (10 mM; Roche Applied Science, Indianapolis, IN), and T7 enzyme. Reactions were incubated at 37°C for 14 hr after which volumes were adjusted to 100 $\mu$ L by addition of Nuclease-free water. cRNA Binding Buffer (350 $\mu$ L) and 100% ethanol (250 $\mu$ L) were added and mixed by pipetting before passing through a cRNA filter cartridge under centrifugation at 10,000 x g for 1 min. Filters were washed with wash Buffer (650 $\mu$ L) and dried by centrifugation for an additional minute. cDNA was eluted using 100 $\mu$ L of Nuclease-free Water at 55°C. cRNA was quantified using the RiboGreen<sup>®</sup> fluorescence-based assay (Invitrogen, Carlsbad, CA).

Hybridization, washing and detection were performed using the Illumina Gene Expression System Buffer Kit for HumanRef-8 BeadChips (Illumina, San Diego, CA) according to the manufacturer's protocol. An aliquot containing 850 ng cRNA was transferred to a new tube and adjusted to a volume of 11.3  $\mu$ L. Hybridization Mix was prepared by mixing Hyb E1 buffer (125  $\mu$ L), which had been prewarmed in a 55°C oven for 10 min, to formamide (75 $\mu$ L). Hybridization Mix (22.7  $\mu$ L) was added to each cRNA sample. Following sample incubation at 65°C for 5 min, 34  $\mu$ L was dispensed onto the center of each HumanRef-8 BeadChip array. BeadChips were assembled onto Hybridization cartridges, mixed by shaking to ensure bubbles moved freely, and then placed on the BeadChip Hyb Wheel and incubated for 16 hr at 55°C with rotation.

Following hybridization, a manual washing procedure was followed in which BeadChips were placed in a slide rack and washed in supplied solutions in glass staining dishes. Slide racks were plunged in and out of the appropriate solution 5-10 times and then mixing was performed on an orbital shaker (Thermolyne Roto Mix, Type 50800, Barnstead International, Dubuque, IA) at highest possible speed or on a rocker shaker (Rocker II 260350, Boekel Scientific, Feasterville, PA) at medium-high speed, for the times indicated. Following hybridization, slides were washed successively in wash E1BC solution (250 mL), 100% ethanol (250 mL), and fresh Wash E1BC solution (250 mL), with 15 min, 10 min and 2 min orbital shaking, respectively. In order to block the slides, BeadChips were placed face-up in a wash tray (supplied) containing Block E1 buffer (4 mL) and rocked for 10 min. For detection BeadChips were transferred to a fresh wash tray containing Block E1 buffer (2 mL) containing streptavidin-Cy3 (1 $\mu$ g/mL; Amersham Biosciences, Piscataway, NJ) and rocked for 10 min. Slides were then placed in a staining rack

and washed in Wash E1BC solution (250 mL) with 5 minutes of orbital shaking. The slides were then dried by centrifugation at 275 x g for 4 min at 20°C in a Jouan CR4.22 centrifuge (Thermo Electron Corporation, Waltham, MA) and stored in the dark until scanned. Scanning was performed using a BeadArray Reader and BeadScan software (Illumina).

### *Data Analysis*

Raw data files for each microarray experiment have been deposited at GEO, accession number Series GSE9569 (GSM241938 through GSM243811) and access is available at:

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=dhovtgesoygsyng&acc=GSE9569>.

Data was normalized by quantile normalization using Bioconductor (Affymetrix data by RMA and Illumina data by “Affy” package) and two-sample Welch t-statistics (unequal variance) were calculated. To adjust for multiple testing a novel Quantile Transformation (QT) approach (15) was employed.

### *Comparison of array platforms*

In order to determine comparable targets from the Affymetrix (Human U133 GeneChip set; ~45,000 probe sets targeting 39,000 transcripts from 33,000 well-substantiated genes) and Illumina (HumanRef-8 BeadChip; > 23,000 RefSeq-curated gene targets) platforms, probe sequences from each platform and transcript sequences from RefSeq Release 13 (<http://www.ncbi.nlm.nih.gov/RefSeq/>) were compared. For both platforms, probes that were not valid were filtered out. A probe was defined as valid if it perfectly matched a transcript sequence

and did not perfectly match any other transcript sequences with a different gene symbol. For Affymetrix probe sets, individual probes were determined to be valid by applying the definition above. Then probe sets were defined as valid if at least 80% of the probes within the set were valid. It was then determined which transcript sequences contained a valid probe or probe sets. If a transcript sequence contained multiple valid probes or probe sets, the one closest to the 3' end of the transcript was selected. Based on these criteria 14,708 targets were included in the cross-platform analysis.

In a second approach, all the significant genes identified by each platform (2692 by Affymetrix and 1828 by Illumina) were subjected to an ID conversion program called Gene Expression Pattern Analysis Suite v3.1 (<http://www.gepas.org>) (16) and significant gene lists were then compared by gene symbol.

### *Pathway Analysis*

Gene RefSeq accession numbers were imported into Ingenuity Pathway Analysis software (Ingenuity<sup>®</sup> Systems, Redwood City, CA, ([www.ingenuity.com](http://www.ingenuity.com))) a web-based application, which queries the Ingenuity Pathway Knowledge Base (IPKB) for genetic interactions. To evaluate the significance of the association of a particular gene set with the relevant canonical pathway within Ingenuity, a ratio of the number of genes from the data set that map to the pathway divided by the total number of genes that map to the canonical pathway is displayed and Fischer's exact test is used to calculate the corresponding *p*-value. Pathway Express (PE), a tool in the Onto-Tools ensemble (17) (<http://vortex.cs.wayne.edu/Projects.html>), was also employed. After generating a

list of pathways from the input list of genes from the Onto-tools database, PE first calculates a perturbation factor (PF) for each input gene. This PF takes into account the normalized fold-change of the gene and the number and amount of perturbation of genes downstream from it. The impact factor of the entire pathway includes a probabilistic term that takes into consideration the proportion of differentially regulated genes on the pathway and gene perturbation factors of all genes in the pathway.

### *Gene Ontology Analysis*

Two publicly available tools for assessing enrichment of Gene Ontology (GO) terms over that which would be expected by chance alone, were applied. In GoMiner (18, 19) the two-sided Fisher's exact test generates the enrichment  $p$ -value while in GOstat (4), a  $\chi^2$  test is used to generate a  $p$ -value. Adjustment for multiple comparisons is based on False Discovery Rate (FDR) in both applications.

## RESULTS

The PBMC transcriptome of 8 individuals occupationally exposed to benzene compared with 8 matched controls was examined on two microarray platforms. Distinct processing protocols appropriate to each platform, from labeling through hybridization and detection, were applied. Data was analyzed using a novel Quantile Transformation approach (15).

### *Cross-Comparison of Genes Associated with Benzene Exposure by Affymetrix and Illumina Microarray Platforms*

The microarray data obtained from the Affymetrix and Illumina platforms is summarized in Figure 1. On the Affymetrix platform, 2692 genes (represented by 3549 probes) were differentially expressed (raw  $p \leq 0.05$ ). Considering genes with expression levels altered by 1.5-fold or greater, 65 genes were down-regulated while 180 genes were up-regulated. On the Illumina platform, 1828 genes (1856 probes) were differentially expressed (raw  $p \leq 0.05$ ). Modification of expression levels by 1.5-fold and higher occurred in 171 genes (88 down-regulated and 83 up-regulated). Supplementary Tables S1 and S2 contain lists of all genes identified as significant by Affymetrix and Illumina platforms, respectively, and show multiple test correction values. Among the Affymetrix data, 15 genes remained significant after multiple testing using the Quantile Transformation (QT) approach while 1 gene among the Illumina data remained significant.

The subset of genes that was identified as differentially expressed (based on raw  $p$ -values) in common by both platforms was determined using two approaches. Results are summarized in Figure 1. First, only those genes identified by each platform which were common to a stringent platform comparison reference file (described in Material and Methods), were analyzed by the same approach used to analyze the full complement of genes from each platform. This approach yielded 1345 significant genes by Affymetrix and 1275 genes by Illumina, which were directly comparable based on the reference file (supplementary Tables S3 and S4, respectively). A total of 356 genes agreed in direction of change of expression. This gives a concordance of 28 % based on the 1275 comparable genes identified by Illumina.

In the second, less stringent approach, all the significant genes identified by each platform (2692 by Affymetrix and 1828 by Illumina) were compared by gene symbol. A further 119 genes commonly identified by both platforms as identified using this less stringent approach, showed the same directional change in expression. From both approaches, therefore, the total number of genes cross-validated by the two platforms was 475 (supplementary Table S5) giving a concordance of 26 % (based on all 1828 significant genes from Illumina).

Expression ratios were similar among the concordant genes (mean difference in expression ratio = 0.04, standard deviation = 0.17). Among the common genes, 37 genes were down-regulated while 89 genes were up-regulated, by 1.5-fold or greater. Four genes (*CXCL16*, *ZNF331*, *JUN* and *PF4*), which we previously identified by microarray and confirmed by real-time PCR, were identified by both platforms in the current study. *JUN* and *ZNF331* are among the top 20 common genes which are listed in Table 1. Both of these genes remained significant after

correction for multiple testing (Quantile Transformation) of the Affy data set (*JUN* QT- $p = 0.044$ ; *ZNF331* QT- $p = 0.042$ ), while *ZNF331* remained significant (QT- $p = 0.039$ ) and *JUN* (QT- $p = 0.072$ ) approached significance upon correction of the Illumina data set. Other genes of note among the top 20 are *HSPA1A* and *HSPA1B*, members of the heat-shock 70 (HSP70) multigene family.

### *Classification of Genes by Gene Ontology and Pathway Analyses*

Among the 475 genes validated by the two platforms, Gostat analysis showed significant association with the GO term apoptosis (GO:0006915;  $p$ -value of 0.0113). The genes are listed in Table 2. Analysis by GoMiner also showed enrichment of apoptosis-related terms. Other notable functional categories among the Affymetrix and Illumina datasets identified by Gostat were immune response, defense response, and response to stress, suggesting concordance between the two platforms at the pathway level. The genes with overlapping classification from the two datasets are listed in Table 2.

Ingenuity canonical pathway analysis identified significant pathways among the common genes as well as among the significant genes from both platforms. Significant pathways and associated genes are shown in Table 3. Lipid metabolism was a key theme among the common genes with involvement of ganglioside biosynthesis, glycerolipid metabolism, glycerphospholipid metabolism and sterol biosynthesis pathways. From the Affymetrix dataset protein ubiquitination was strongly impacted with 23 genes up-regulated and 4 genes down-regulated. Among the Illumina dataset multiple pathways were involved as shown in Table 3.

Using Pathway Express software, no significant pathways were identified among the common genes. The cell adhesion molecules pathway (Impact factor (IF) = 273.4,  $p = 0.04$ ) was significant in the Illumina dataset, with 3 genes down-regulated, while T cell receptor signaling (IF = 6.57,  $p = 0.001$ , N = 6); MAP kinase signaling (IF = 5.722,  $p = 0.003$ , N = 12) and B cell receptor signaling (IF = 5.248,  $p = 0.005$ , N = 5) were significantly associated with the Affymetrix dataset.

#### *Evaluation of Transcripts Discordant between Platforms*

The top genes exclusively identified by each platform, were examined in an effort to understand the discordance. Among the 50 most significant genes identified by Illumina, 32 genes (64 %) were also ranked by Affymetrix as being significant. Of the 18 discordant genes, 8 were not present in the stringent platform comparison file and 9 represent hypothetical proteins. Among the genes that could be compared directly, but which failed to reach significance with the Affymetrix platform analysis, *ZYMND15* was up-regulated 2-fold ( $p$ -value = 1.56E-04) on Illumina but was unchanged on Affymetrix (expression ratio =  $\sim 1$ ,  $p$ -value = 0.72). *UBE2J1* was significantly down-regulated with the Illumina analysis and although the probe targeting this gene had no directly comparable probe on the Affymetrix platform, 7 Affymetrix probes which target *UBE2J1* did not show differential expression of the gene.

Among the 41 genes represented by the top 50 Affymetrix targets, 17 genes (42 %) were not identified by Illumina as being differentially expressed. Similar to the finding with Illumina, a

much higher platform concordance (58 %) was observed for the most significant genes, compared with that for all genes. Of the 17 discordant genes, only 5 were directly comparable and, *C3ARI*, which was up-regulated 2-fold on Affymetrix, approached significance on the Illumina platform ( $p$ -value = 0.056, ratio 1.38). For the 12 genes that could not be directly compared, the results from all the Affymetrix probes targeting the RefSeq accession associated with each gene were examined. For example *ANXA1* was down-regulated 5-fold on Affymetrix with one probe but another probe targeting the gene was up-regulated in agreement with the Illumina result (approaching significance). Another gene, *GRAP* was down-regulated with one Affymetrix probe but the other probe targeting the gene did not show significant differential expression, in agreement with Illumina. In the case of *FYN*, which was up-regulated by 1.4-fold and 1.8-fold from two Affymetrix probes, the Illumina platform also detected up-regulation but only approached significance and therefore did not appear in the list of common genes (1.2-fold,  $p$ -value = 0.07).

## DISCUSSION

We identified robust changes in gene expression in response to benzene exposure in 8 occupationally exposed individuals compared with 8 unexposed controls, by cross-comparison using two microarray platforms (Affymetrix and Illumina). This approach enabled identification of a greater number of robust biomarkers than our previous approach of single-platform array analysis in conjunction with quantitative PCR confirmation.

A total of 475 genes were identified in common by our two-platform approach, with 37 genes down-regulated and 89 genes up-regulated, by 1.5-fold or greater. Further validation of our approach was provided by the fact that four genes (*CXCL16*, *ZNF331*, *JUN* and *PF4*), which we previously showed to be highly significantly associated with benzene exposure (3), were present in the cross-validated dataset from the current study. *JUN* was previously shown to be down-regulated by benzene exposure in mouse HSC (12). Expression of *FOSB* expression was also down-regulated (~1.6-fold) by both platforms in the current study. JUN and FOS are basic region-leucine zipper (bZIP) members of the AP-1 transcription complex (20), which modulates the decision of a cell to proliferate, differentiate, or die by apoptosis (21). Since JUN promotes proliferation of many cell types (21), reduced levels of *JUN* could indicate that the PBMCs of benzene-exposed individuals are not proliferating or progressing through the cell cycle as quickly as those of non-exposed individuals. Platelet Factor 4 (PF4), a chemokine secreted from activated platelets (22), activated T cells and mast cells (23), is a chemoattractant for neutrophils and fibroblasts and plays a role in inflammation and wound repair. PF4 was down-regulated in the current study in agreement with previous observations (3).

We used gene ontology and pathway analyses in an attempt to elicit potential underlying biology from the data. GO analysis showed enrichment in genes involved in apoptosis among the 475 common genes, while pathway analysis identified an impact on lipid metabolism. Genes involved in the GO categories of immune response, stress response and defense response were enriched in the separate platform datasets, with overlap of genes between platforms. Therefore, despite the fact that concordance between the microarray platforms was ~25 %, the overall mechanisms (concordance at the pathway level) underlying benzene effects in human PBMC are in close agreement and fit well with the phenotypic effects of benzene including decreased blood cell counts (5), and depression of the immune system (6).

In the current study the Bax:Bcl2 ratio, an indicator of the degree of apoptosis was 1:0.5, suggesting a shift towards apoptosis. Down-regulation of two anti-apoptotic hsp70-encoding genes (24) was also observed in the current study. Increased apoptosis is a mechanism that could potentially underlie benzene-associated leukemia. Chromosomal translocations, signature abnormalities of AML, have been proposed to arise from abortive apoptosis and subsequent misrepair of cleaved genes (25, 26). Surviving apoptosis as a possible mechanism of benzene-induced leukemia has been discussed (27) and induction of apoptosis in hematopoietic progenitor cells (12, 28) and cell lines (29) by benzene metabolites has been previously demonstrated. Removal of cells predestined to die by apoptosis is facilitated, at least in some tissues by macrophages (30, 31), and dysfunction of macrophages may lead to survival of cells which would otherwise have been removed. Poisoning of the BM stromal environment (7), particularly macrophages (8), is a hematotoxic effect of benzene.

Fatty acid and cholesterol metabolism processes were impacted in the current study. This is in agreement with a previous study of gene expression in rat liver following a 28-day oral benzene exposure in which plasma cholesterol (and phospholipids) were found to be slightly elevated in benzene-treated groups (32). The fact that PBMC profiling reflects liver gene expression is not surprising in view of a recent study showing that the peripheral blood transcriptome dynamically reflects systems wide biology with 83% of liver genes also expressed in blood (33). Some of our findings also mirror the toxicogenomic profile of hematopoietic stem cells (HSC) of mice exposed to benzene (100 ppm for 2 weeks) (12). In particular, the thrombopoietin receptor or myeloproliferative leukemia virus oncogene (*MPL*) was down-regulated 1.5-fold in HSC and was also down-regulated in our study (1.4 - >2-fold). It was demonstrated that MPL is a selective surface marker for human hematopoietic stem cells in adult bone marrow and cytokine-mobilized peripheral blood stem cells (34) and perhaps the reduced expression in our study reflects reduced HSC.

Our study undertook the discovery of global benzene-induced differential expression using two microarray platforms and we found ~25% concordance between the platforms. Strikingly, the most significant genes (represented by the top 50 probes) identified by either array were much more likely to be ranked as significant by the other platform (Illumina = 64%, Affymetrix = 58%). While very high concordance levels have been reported (~90%) between the Affymetrix and Illumina platforms, these were based on extremely different biological samples with large fold changes in expression (35). Smaller concordance levels were seen when comparing less biologically similar samples (35) or analyzing more similar rat toxicogenomic data (36). Our

study was based on occupationally-exposed individuals with inherent inter-individual variability in baseline expression, determined by factors such as blood count, blood type, genotype, presence of subclinical infection (37), which are less easy to control for by study design as are factors such as age and gender. Other explanations of microarray data discordance have been discussed (35). Our inter-platform concordance is consistent with a study using a similar approach to ours, comparing lists of differentially expressed genes, which showed an average concordance of 22.8% (38). Despite the relatively low concordance observed in our study, similar mechanisms of benzene effect such as response to stress and immune response were found. However the results potentially caution against use of a single platform to identify all relevant biomarkers in human exposure studies. Application of two platforms to a biomarker discovery study could be cost-prohibitive but our data also suggest the need to be very stringent in the selection of potential biomarkers based on a single platform (as platform concordance was much higher among the most highly significant genes).

Challenges are inherent to this type of molecular epidemiology study. While many potential biomarkers of benzene exposure were generated, few genes remained significant after multiple testing (QT  $p$ -value  $\leq 0.05$ ). Another challenge is the biological relevance of small fold-changes in gene expression. Increasing the number of individuals studied is one way to increase the power to select true biomarkers. Cross-comparison by two platforms increases the chances that the genes identified in our study represent true potential biomarkers, but validation of biomarkers in a larger population is also necessary, at both the RNA and protein levels. While limited sample material precluded the validation of the microarray findings by QPCR in the current study, use of two microarray platforms offers a type of inherent validation in that distinct

processing protocols appropriate to each platform, from labeling through hybridization and detection, were applied.

The question of the appropriateness of PBMC as a cell target in which to examine benzene's hematotoxic effects must also be addressed. While some of benzene's immunotoxic effects are thought to involve damage to BM stromal cells (7) and early progenitor cells (4) the knock-on effects of damage to these cells might be expected to be manifest in the transcriptome of their downstream cell targets. As discussed above, the peripheral blood transcriptome dynamically reflects system wide biology (33) and in our study as discussed above, some of our findings correlated with other studies examining different cell targets. However, relevant changes may be masked by looking at heterogenous populations of cells e.g. BM compared with HSC (12). Many of the genes identified in this study are expressed in several cell types and have pleiotropic effects, making it challenging to induce function and mechanism when examining PBMC. However, PBMC are convenient for molecular epidemiology research studies and pathways and mechanisms potentially impacted by benzene exposure such as apoptosis, can be further tested in *in vitro* studies using targeted cell subsets.

In conclusion, we have demonstrated robust changes in the PBMC transcriptome of benzene-exposed individuals, using a two-platform approach. The genes identified contribute to further understanding of the mechanisms underlying benzene-induced hematotoxicity and leukemia.

**ACKNOWLEDGEMENTS**

We thank the participants for taking part in this study. Supported by NIH grants RO1ES06721 and P42ES04705 (to M.T.S.), and P42ES05948 and P30ES10126 (to S.M.R.) from the National Institute of Environmental Health Sciences and Intramural Funds from the National Cancer Institute, and by the U.S. DOE under Contract No. DE-AC02-05CH11231.

### Figure Legends:

Figure 1. Summary of Benzene study. The numbers of significant genes ( $P \leq 0.05$ ) from the four analyses of differential expression performed by t-test/quantile transformation (QT) are shown. Results from analysis of the complete Affymetrix and Illumina datasets are shown on the far-left and far-right, respectively. Results from the analysis of the datasets termed “comparable probes”, derived from the application of stringent criteria described in the methods and discussed in the text, are shown inside the dotted line, for Affymetrix (left) and Illumina (right). Concordance among differentially expressed genes identified from each platform based on complete datasets and comparable probes is shown. \*QT is quantile transformation.

### Table Legends:

Table 1: Top 20 genes associated with benzene exposure cross-validated by Affymetrix and Illumina microarray platforms. The central columns list gene title symbol, symbol and RefSeq ID, while platform-specific IDs,  $p$ -values, and differential expression ratios (relative to control) are detailed on the left (Affymetrix) and right (Illumina). \*Raw  $p$ -values i.e. not adjusted for multiple testing are shown. †Differential Expression ratio.

Table 2: Functional classification of genes modified by benzene exposure. \*GOSTat was used to assess for enrichment of GO terms among the genes with significant differential expression from both the Affymetrix and Illumina analyses as well as the genes identified in common.

†Genes commonly identified by GO analysis of data from both platforms. ‡A  $\chi^2$  test is used to

generate a  $p$ -value and adjustment for multiple comparisons is based on False Discovery Rate (FDR).

Table 3: Canonical pathways impacted by benzene exposure. \*Derived from Ingenuity® Pathway Analysis of Affymetrix, Illumina and common datasets. †Fischer's exact test is used to calculate the  $p$ -value.

## References

1. **Hayes RB, Songnian Y, Dosemeci M, and Linet M.** Benzene and lymphohematopoietic malignancies in humans. *Am J Ind Med* 40: 117-126, 2001.
2. **Smith MT, Zhang L, Wang Y, Hayes RB, Li G, Wiemels J, Dosemeci M, Titenko-Holland N, Xi L, Kolachana P, Yin S, and Rothman N.** Increased translocations and aneusomy in chromosomes 8 and 21 among workers exposed to benzene. *Cancer Res* 58: 2176-2181, 1998.
3. **Forrest MS, Lan Q, Hubbard AE, Zhang L, Vermeulen R, Zhao X, Li G, Wu YY, Shen M, Yin S, Chanock SJ, Rothman N, and Smith MT.** Discovery of novel biomarkers by microarray analysis of peripheral blood mononuclear cell gene expression in benzene-exposed workers. *Environ Health Perspect* 113: 801-807, 2005.
4. **Beissbarth T, and Speed TP.** GStat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics* 20: 1464-1465, 2004.
5. **Yoon BI, Hirabayashi Y, Kawasaki Y, Kodama Y, Kaneko T, Kim DY, and Inoue T.** Mechanism of action of benzene toxicity: cell cycle suppression in hemopoietic progenitor cells (CFU-GM). *Exp Hematol* 29: 278-285, 2001.
6. **Snyder R.** Overview of the toxicology of benzene. *J Toxicol Environ Health A* 61: 339-346, 2000.
7. **Garnett HM, Cronkite EP, and Drew RT.** Effect of in vivo exposure to benzene on the characteristics of bone marrow adherent cells. *Leuk Res* 7: 803-810, 1983.
8. **Thomas DJ, Sadler A, Subrahmanyam VV, Siegel D, Reasor MJ, Wierda D, and Ross D.** Bone marrow stromal cell bioactivation and detoxification of the benzene metabolite

hydroquinone: comparison of macrophages and fibroblastoid cells. *Mol Pharmacol* 37: 255-262, 1990.

9. **Shen M, Lan Q, Zhang L, Chanock S, Li G, Vermeulen R, Rappaport SM, Guo W, Hayes RB, Linet M, Yin S, Yeager M, Welch R, Forrest MS, Rothman N, and Smith MT.** Polymorphisms in genes involved in DNA double-strand break repair pathway and susceptibility to benzene-induced hematotoxicity. *Carcinogenesis* 27: 2083-2089, 2006.
10. **Rothman N, Smith MT, Hayes RB, Traver RD, Hoener B, Campleman S, Li GL, Dosemeci M, Linet M, Zhang L, Xi L, Wacholder S, Lu W, Meyer KB, Titenko-Holland N, Stewart JT, Yin S, and Ross D.** Benzene poisoning, a risk factor for hematological malignancy, is associated with the NQO1 609C-->T mutation and rapid fractional excretion of chlorzoxazone. *Cancer Res* 57: 2839-2842, 1997.
11. **Yoon BI, Li GX, Kitada K, Kawasaki Y, Igarashi K, Kodama Y, Inoue T, Kobayashi K, Kanno J, Kim DY, Inoue T, and Hirabayashi Y.** Mechanisms of benzene-induced hematotoxicity and leukemogenicity: cDNA microarray analyses using mouse bone marrow tissue. *Environ Health Perspect* 111: 1411-1420, 2003.
12. **Abernethy DJ, Kleymenova EV, Rose J, Recio L, and Faiola B.** Human CD34+ hematopoietic progenitor cells are sensitive targets for toxicity induced by 1,4-benzoquinone. *Toxicol Sci* 79: 82-89, 2004.
13. **Shi L, Reid LH, Jones WD, Shippy R, Warrington JA, Baker SC, Collins PJ, de Longueville F, Kawasaki ES, Lee KY, Luo Y, Sun YA, Willey JM, Setterquist RA, Fischer GM, Tong W, Dragan YP, Dix DJ, Frueh FW, Goodsaid FM, Herman D, Jensen RV, Johnson CD, Lobenhofer EK, Puri RK, Schrf U, Thierry-Mieg J, Wang C, Wilson M, Wolber PK, Zhang L, Slikker W, Jr., Shi L, and Reid LH.** The MicroArray Quality Control

- (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* 24: 1151-1161, 2006.
14. **Lan Q, Zhang L, Li G, Vermeulen R, Weinberg RS, Dosemeci M, Rappaport SM, Shen M, Alter BP, Wu Y, Kopp W, Waidyanatha S, Rabkin C, Guo W, Chanock S, Hayes RB, Linet M, Kim S, Yin S, Rothman N, and Smith MT.** Hematotoxicity in workers exposed to low levels of benzene. *Science* 306: 1774-1776, 2004.
  15. **van der Laan MJ, and Hubbard AE.** Quantile-function based null distribution in resampling based multiple testing. *Stat Appl Genet Mol Biol* 5: Article14, 2006.
  16. **Herrero J, Diaz-Uriarte R, and Dopazo J.** Gene expression data preprocessing. *Bioinformatics* 19: 655-656, 2003.
  17. **Khatri P, Sellamuthu S, Malhotra P, Amin K, Done A, and Draghici S.** Recent additions and improvements to the Onto-Tools. *Nucleic Acids Res* 33: W762-765, 2005.
  18. **Zeeberg BR, Feng W, Wang G, Wang MD, Fojo AT, Sunshine M, Narasimhan S, Kane DW, Reinhold WC, Lababidi S, Bussey KJ, Riss J, Barrett JC, and Weinstein JN.** GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol* 4: R28, 2003.
  19. **Zeeberg BR, Qin H, Narasimhan S, Sunshine M, Cao H, Kane DW, Reimers M, Stephens RM, Bryant D, Burt SK, Elnekave E, Hari DM, Wynn TA, Cunningham-Rundles C, Stewart DM, Nelson D, and Weinstein JN.** High-Throughput GoMiner, an 'industrial-strength' integrative gene ontology tool for interpretation of multiple-microarray experiments, with application to studies of Common Variable Immune Deficiency (CVID). *BMC Bioinformatics* 6: 168, 2005.

20. **Shaulian E, and Karin M.** AP-1 as a regulator of cell life and death. *Nat Cell Biol* 4: E131-136, 2002.
21. **Hess J, Angel P, and Schorpp-Kistner M.** AP-1 subunits: quarrel and harmony among siblings. *J Cell Sci* 117: 5965-5973, 2004.
22. **Brandt E, Ludwig A, Petersen F, and Flad HD.** Platelet-derived CXC chemokines: old players in new games. *Immunol Rev* 177: 204-216, 2000.
23. **Boehlen F, and Clemetson KJ.** Platelet chemokines and their receptors: what is their relevance to platelet storage and transfusion practice? *Transfus Med* 11: 403-417, 2001.
24. **Garrido C, Gurbuxani S, Ravagnan L, and Kroemer G.** Heat shock proteins: endogenous modulators of apoptotic cell death. *Biochem Biophys Res Commun* 286: 433-442, 2001.
25. **Khodarev NN, Sokolova IA, and Vaughan AT.** Abortive apoptosis as an initiator of chromosomal translocations. *Med Hypotheses* 52: 373-376, 1999.
26. **McHale CM, and Smith MT.** Prenatal origin of chromosomal translocations in acute childhood leukemia: implications and future directions. *Am J Hematol* 75: 254-257, 2004.
27. **Vaughan AT, Betti CJ, Villalobos MJ, Premkumar K, Cline E, Jiang Q, and Diaz MO.** Surviving apoptosis: a possible mechanism of benzene-induced leukemia. *Chem Biol Interact* 153-154: 179-185, 2005.
28. **Moran JL, Siegel D, Sun XM, and Ross D.** Induction of apoptosis by benzene metabolites in HL60 and CD34+ human bone marrow progenitor cells. *Mol Pharmacol* 50: 610-615, 1996.
29. **Inayat-Hussain SH, and Ross D.** Intrinsic pathway of hydroquinone induced apoptosis occurs via both caspase-dependent and caspase-independent mechanisms. *Chem Res Toxicol* 18: 420-427, 2005.

30. **Lang RA, and Bishop JM.** Macrophages are required for cell death and tissue remodeling in the developing mouse eye. *Cell* 74: 453-462, 1993.
31. **Diez-Roux G, and Lang RA.** Macrophages induce apoptosis in normal cells in vivo. *Development* 124: 3633-3638, 1997.
32. **Heijne WH, Jonker D, Stierum RH, van Ommen B, and Groten JP.** Toxicogenomic analysis of gene expression changes in rat liver after a 28-day oral benzene exposure. *Mutat Res* 575: 85-101, 2005.
33. **Liew CC, Ma J, Tang HC, Zheng R, and Dempsey AA.** The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. *J Lab Clin Med* 147: 126-132, 2006.
34. **Ninos JM, Jefferies LC, Cogle CR, and Kerr WG.** The thrombopoietin receptor, c-Mpl, is a selective surface marker for human hematopoietic stem cells. *J Transl Med* 4: 9, 2006.
35. **Canales RD, Luo Y, Willey JC, Austermiller B, Barbacioru CC, Boysen C, Hunkapiller K, Jensen RV, Knight CR, Lee KY, Ma Y, Maqsoodi B, Papallo A, Peters EH, Poulter K, Ruppel PL, Samaha RR, Shi L, Yang W, Zhang L, and Goodsaid FM.** Evaluation of DNA microarray results with quantitative gene expression platforms. *Nat Biotechnol* 24: 1115-1122, 2006.
36. **Guo L, Lobenhofer EK, Wang C, Shippy R, Harris SC, Zhang L, Mei N, Chen T, Herman D, Goodsaid FM, Hurban P, Phillips KL, Xu J, Deng X, Sun YA, Tong W, Dragan YP, and Shi L.** Rat toxicogenomic study reveals analytical consistency across microarray platforms. *Nat Biotechnol* 24: 1162-1169, 2006.

37. **Whitney AR, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC, Relman DA, and Brown PO.** Individuality and variation in gene expression patterns in human blood. *Proc Natl Acad Sci U S A* 100: 1896-1901, 2003.
38. **Cheadle C, Becker KG, Cho-Chung YS, Nesterova M, Watkins T, Wood W, 3rd, Prabhu V, and Barnes KC.** A rapid method for microarray cross platform comparisons using gene expression signatures. *Mol Cell Probes* 2006.

Figure 1: Summary of Benzene study

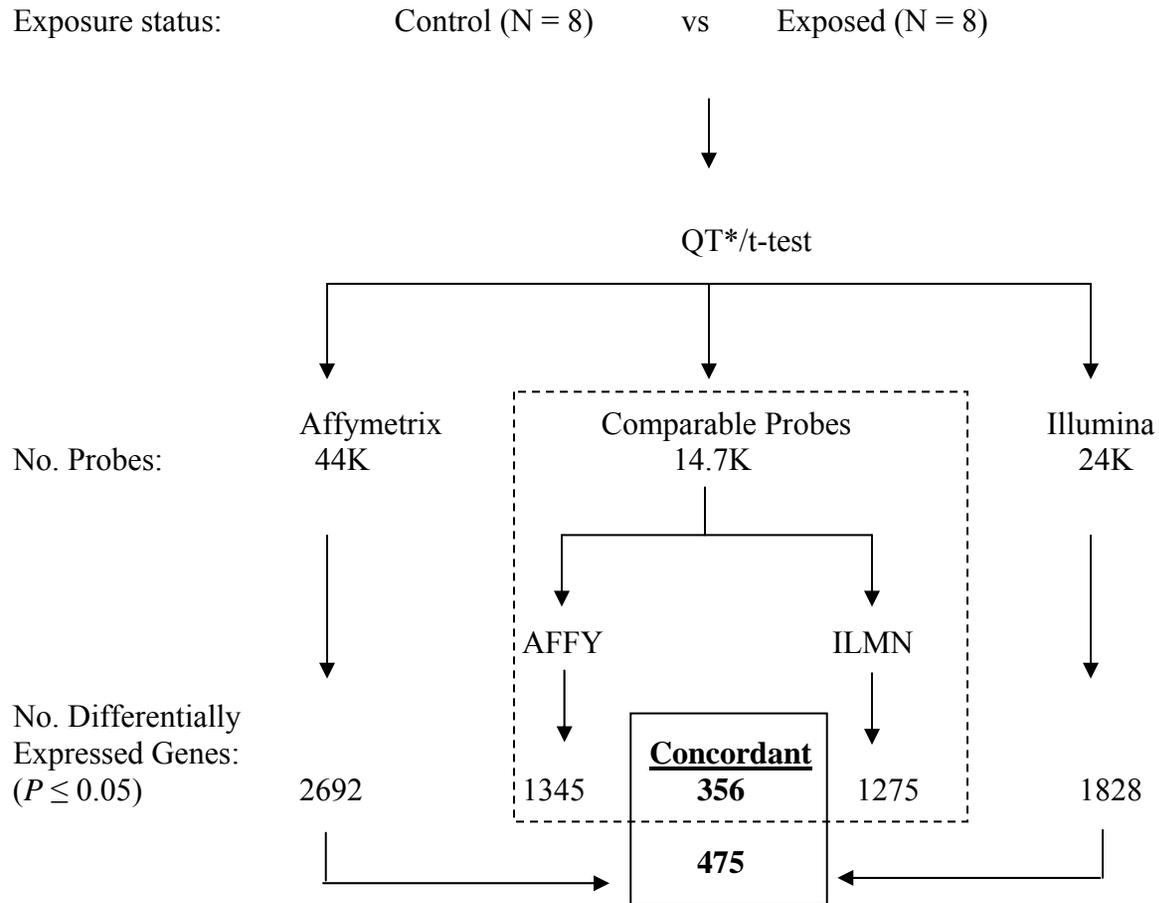


Table 1: Top 20 genes associated with benzene exposure cross-validated by Affymetrix and Illumina microarray platforms

| Affymetrix                 |          |        | Gene Title   | Symbol   | RefSeq ID    | Illumina      |          |       |
|----------------------------|----------|--------|--|----------|--------------|---------------|----------|-------|
| ID                         | p-value* | Ratio† |  |          |              | ID            | p-value* | Ratio |
| <b>Downregulated (N=6)</b> |          |        |  |          |              |               |          |       |
| 202581_at                  | 3.98E-05 | 0.18   | heat shock 70kDa protein 1B                                  | HSPA1B   | NM_005346    | GI_26787974-S | 3.74E-05 | 0.19  |
| 200799_at                  | 6.57E-05 | 0.34   | heat shock 70kDa protein 1A                                  | HSPA1A   | NM_005345    | GI_26787973-S | 1.48E-04 | 0.32  |
| 201466_s_at                | 1.47E-05 | 0.39   | v-jun sarcoma virus 17<br>oncogene homolog (avian)           | JUN      | NM_002228    | GI_44890066-S | 2.43E-05 | 0.33  |
| 208960_s_at                | 1.45E-02 | 0.55   | Kruppel-like factor 6  | KLF6     | NM_001008490 | GI_37655156-S | 1.10E-04 | 0.41  |
| 229054_at                  | 4.37E-07 | 0.60   | FLJ39779 protein   | FLJ39779 | NM_207442    | GI_42660305-S | 1.92E-04 | 0.51  |
| 202014_at                  | 1.33E-03 | 0.62   | protein phosphatase 1,<br>regulatory (inhibitor) subunit 15A | PPP1R15A | NM_014330    | GI_9790902-S  | 4.12E-05 | 0.55  |
| <b>Upregulated (N=14)</b>  |          |        |  |          |              |               |          |       |
| 201939_at                  | 1.05E-03 | 1.82   | polo-like kinase 2 (Drosophila)                              | PLK2     | NM_006622    | GI_5730054-S  | 1.45E-04 | 1.62  |
| 227613_at                  | 8.78E-03 | 1.82   | zinc finger protein 331                                      | ZNF331   | NM_018555    | GI_20127571-S | 6.38E-06 | 1.62  |
| 235568_at                  | 5.09E-04 | 1.81   | mast cell-expressed membrane<br>protein 1                    | MCEMP1   | NM_174918    | GI_27885012-S | 9.11E-05 | 1.95  |
| 202856_s_at                | 5.84E-04 | 1.76   | solute carrier family 16<br>member 3                         | SLC16A3  | NM_004207    | GI_4759111-S  | 3.99E-04 | 1.69  |
| 216248_s_at                | 8.94E-03 | 1.74   | nuclear receptor subfamily 4,<br>group A, member 2           | NR4A2    | NM_173173.1  | GI_27894352   | 3.94E-04 | 1.47  |
| 200768_s_at                | 2.47E-03 | 1.64   | methionine adenosyltransferase<br>II, alpha                  | MAT2A    | NM_005911    | GI_34147493-S | 1.04E-04 | 1.98  |
| 218421_at                  | 6.52E-04 | 1.51   | ceramide kinase  | CERK     | NM_022766    | GI_32967302-A | 3.86E-04 | 1.40  |
| 209272_at                  | 2.24E-02 | 1.44   | NGFI-A binding protein 1<br>(EGR1 binding protein 1)         | NAB1     | NM_005966    | GI_19923347-S | 2.46E-04 | 1.47  |
| 219862_s_at                | 5.74E-04 | 1.38   | nuclear prelamin A recognition<br>factor                     | NARF     | NM_001038618 | GI_14165459-A | 1.84E-05 | 1.35  |
| 217964_at                  | 1.81E-02 | 1.31   | tetratricopeptide repeat domain 19                           | TTC19    | NM_017775    | GI_22547158-S | 2.05E-04 | 1.37  |
| 223093_at                  | 2.27E-02 | 1.26   | ankylosis, progressive homolog<br>(mouse)                    | ANKH     | NM_054027    | GI_34452701-S | 3.99E-04 | 1.32  |
| 36554_at                   | 5.28E-03 | 1.25   | acetylserotonin<br>O-methyltransferase-like                  | ASMTL    | NM_004192    | GI_4757793-S  | 1.02E-04 | 1.29  |

|             |          |      |  |         |             |               |          |      |
|-------------|----------|------|--|---------|-------------|---------------|----------|------|
| 223740_at   | 2.28E-03 | 1.24 | chromosome 6 open reading<br>frame 59            | C6orf59 | NM_024929.1 | GI_13376403   | 3.16E-04 | 1.55 |
| 205791_x_at | 8.96E-03 | 1.23 | zinc finger protein 155,<br>transcript variant 1 | ZNF155  | NM_198089   | GI_37655172-A | 2.63E-04 | 1.29 |

---

Table 2: Functional classification of genes modified by benzene exposure.

| GO Term*           | Genes†   | Illumina  |           | Affymetrix |           | Common  |           |
|--------------------|--|-----------|-----------|------------|-----------|---------|-----------|
|                    |  | p-value‡  | No. Genes | p-value    | No. Genes | p-value | No. Genes |
| Apoptosis          | RABEP1, BCL2, RRAGC, ALOX12, BRE, CASP3, PPM1F, SIPA1, ABL1, DIDO1, F2, CUL4A, TNFSF14, CLU, STK17B, LITAF, PPP1R15A, TRAF3, IFI6, CRADD, CASP9, PTPRC, CTSB, PDCD4, TNFAIP3, MAEA, TLR2, C12ORF22, HSPA1A, TAOK2, TUBB2C, MX1 | -         | -         | -          | -         | 0.0113  | 32        |
| Immune response    | BCL2, CLEC5A, CXCL16, IFI6, IFNG, IL1R2, IL21R, ISG15, KLF6, MX1, PLA2G7, TNFSF14 (N=12)   | 2.58E-05  | 21        | 4.04E-14   | 37        | -       | -         |
| Defense response   | BCL2, CD69, CLEC5A, CXCL16, HIST2H2BE, IFI6, IFNG, IL1R2, IL21R, ISG15, KLF6, MX1, PLA2G7, TNFSF14 (N=14)  | 8.23E-07  | 25        | 1.218E-13  | 39        | -       | -         |
| Response to stress | BCL2, CLEC5A, CXCL16, DNAJB1, DUSP1, HIST2H2BE, HSPA1A, IFI6, IFNG, ISG15, MX1, PLA2G7, PPP1R15A, SRXN1 (N=14)   | 7.26E--06 | 26        | 2.17E-06   | 35        | -       | -         |

Table 3: Canonical pathways impacted by benzene exposure.

| Pathway*                       | Genes  | Affymetrix |           | Illumina |           | Common  |           |
|--------------------------------|--|------------|-----------|----------|-----------|---------|-----------|
|                                |  | p-value†   | No. Genes | p-value  | No. Genes | p-value | No. Genes |
| Death receptor Signaling       | BCL2, CFLAR, CRADD, FASLG, IKBKE, NFKB2, TNFRSF1A                                    | -          | -         | 0.016    | 7         | -       | -         |
| ERK/MAPK signaling             | DUSP1, DUSP4, FOS, MYC, PIK3R1, PPARG, PPP1CB, PPP1R10, PRKAR2B, RPS6KA1, SRC, STAT3 | -          | -         | 0.024    | 12        | -       | -         |
| Ganglioside biosynthesis       | ST3GAL1, ST3GAL4, ST3GAL5, ST6GALNAC2, B3GALT4, ST8SIA4                              | 0.046      | 4         | 0.00034  | 5         | 0.043   | 2         |
| Globoside metabolism           | B3GALT3, HEXA, ST3GAL1, ST8STA4  | -          | -         | 0.022    | 4         | -       | -         |
| Glycerolipid metabolism        | AGPAT4, AGPAT6, CERK, DHRS9, DGAT2, GK, LAC89944, LPL, PPAP2B                        | -          | -         | 0.026    | 9         | 0.019   | 6         |
| Glycerophospholipid metabolism | CERK, PPAP2B, HMOX1, PAFAH1B1, PLAG2G7   | -          | -         | -        | -         | 0.017   | 5         |
| Il-10 signaling                | CCR5, FOS, HMOX1, IKBKE, IL1R2, JUN, NFKB2, RELB, STAT3                              | -          | -         | 0.0031   | 9         | 0.004   | 6         |
| IL-6 signaling                 | ABCB1, FOS, IKBKE, IL1R2, JUN, MAPKAPK2, NFKB2, STAT3,                               | -          | -         | 0.0081   | 9         | -       | -         |

|                              |   |        |    |        |    |       |   |  |
|------------------------------|---|--------|----|--------|----|-------|---|--|
|                              | TNFRSF1A  |        |    |        |    |       |   |  |
| PDGF signaling               | ABL1, FOS, JUN, MYC, PIK3R1, SRC, STAT3   | -      | -  | 0.021  | 7  | -     | - |  |
| PPAR signaling               | FOS, IKBKE, IL1R2, JUN, NCOA1, NFKB2, PPARG, RXRA, STAT5A, TNFRSF1A   | -      | -  | 0.0017 | 10 | -     | - |  |
| Protein ubiquitination       | ANAPC1, BAP1, BIRC4, BTRC, CUL1, IFNG, PSMA3, PSMB4, PSMC2, PSMC4, PSMD2, PSMD4, PSMD11, PSMD12, SMURF2, UBC, UBE2I, UBE2Q1, USP3, USP18, USP24, USP28, USP33, USP36, USP39, USP47, USP9X | 0.0017 | 27 | -      | -  | -     | - |  |
| Sterol biosynthesis          | FDFT1, HMGCR, MVD, SC5DL  | -      | -  | 0.025  | 4  | 0.019 | 3 |  |
| Toll-like receptor signaling | FOS, JUN, NFKB2, RELB, TLR2, TOLLIP   | -      | -  | 0.031  | 6  | -     | - |  |