Science of the Total Environment 472 (2014) 872–879

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Differential gene expression profiling analysis in workers occupationally exposed to benzene

Ai Gao, Jing Yang, Gengxia Yang, Piye Niu, Lin Tian *

Department of Occupational Health and Environmental Health, School of Public Health, Capital Medical University, Beijing 100069, China
Beijing Key Laboratory of Environmental Toxicology, Capital Medical University, Beijing 100069, China

HIGHLIGHTS

- Benzene toxicity was highlighted by immune-related signaling pathways.
- PIK_3, $GNAI_3$, KRAS and HLA-DMA were key genes of benzene toxicity.

ARTICLE INFO

Article history:
Received 11 October 2013
Received in revised form 17 November 2013
Accepted 18 November 2013
Available online 15 December 2013

Keywords:
Benzene
Gene expression profiling
Microarray analysis
Bioinformatics analysis

ABSTRACT

Benzene is an important industrial chemical and an environmental contaminant, but the pathogenesis of hematotoxicity induced by chronic occupational benzene exposure remains to be elucidated. To gain an insight into the molecular mechanisms and new biomarkers, microarray analysis was used to identify the differentially expressed mRNA critical for benzene hematotoxicity. RNA was extracted from four chronic benzene poisoning patients occupationally exposed to benzene, three benzene-exposed workers without clinical symptoms and three health controls without benzene exposure and mRNA expression profiling was performed using Gene Chip Human Gene 2.0ST Arrays. Analysis of mRNA expression profiles were conducted to identify key genes, biological processes and pathways by the series test of cluster (STC), STC-Gene Ontology analysis (STC-GO), pathway analysis and Signal-net. Principal findings: 1) 1661 differentially expressed mRNAs were identified and assigned to sixteen model profiles. 2) Profiles 14, 10, 11, 1 and 15 are the predominant expression profiles which were involved in immune response, inflammatory response, chemotaxis, defense response, anti-apoptosis and signal transduction. 3) The importance of immune response at benzene hematotoxicity is highlighted by several immune-related signaling pathways such as B/T cell receptor signaling pathway, acute myeloid leukemia, hematopoietic cell lineage and natural killer cell mediated cytotoxicity. 4) Signet analysis showed that PIK3R1, PIK3CG, PIK3R2, GNAI3, KRAS, NRAS, NFKB1, HLA-DMA, and HLA-DMB were key genes involved in benzene hematotoxicity. These genes might be new biomarkers for the prevention and early diagnosis of benzene poisoning. This is a preliminary study that paves the way to further functional study to understand the underlying regulatory mechanisms.

1. Introduction

Benzene is a well-known environmental pollutant which can induce hematotoxicity, aplastic anemia and leukemia (Irons et al., 2013; Lagorio et al., 2013; Li and Yin, 2006; Snyder, 2012). Benzene causes hematotoxicity even at relatively low levels (Glass et al., 2003; Lan et al., 2004; Pesatori et al., 2009; Swaen et al., 2010). A study indicated that white blood cell counts were decreased in workers exposed to less than 1 ppm benzene compared with controls (Lan et al., 2004). However, the mechanisms of benzene-induced hematotoxicity and leukemogenesis remain unclear.

Toxicogenomic studies of the exposed population are an important alternative approach to the human health risk assessment of environmental exposures and have been used to identify potential biomarkers of early effects and underlying mechanisms associated with diseases. Microarray techniques have been used successfully to monitor gene expression profile changes. Forrest and McHale previously compared global gene expression using Illumina HumanRef-8 V2 BeadChips in the peripheral blood mononuclear cell (PBMC) of workers exposed to various levels of benzene (from <1 ppm to >10 ppm) and identified

* Corresponding author at: Department of Occupational Health and Environmental Health, School of Public Health, Capital Medical University, Beijing China, 100069. Tel.: +86 10 83911774; fax: +86 10 83911506.
E-mail address: gjldwsx@163.com (L. Tian).

0048-9697/$ – see front matter.
http://dx.doi.org/10.1016/j.scitotenv.2013.11.089

potential biomarkers of exposure and early effects, as well as mechanisms of toxicity (Forrest et al. 2005; McHale et al., 2009, 2011). However, more evidence on toxic effects of benzene and new bioinformatics analysis method were still required.

To better understand the toxic effect of benzene, peripheral blood mononuclear cells from four chronic benzene poisoning patients occupationally exposed to benzene, three benzene-exposed workers without clinical symptoms and three health controls without benzene exposure were performed gene expression profiling analysis using Affymetrix GeneChip Human Gene 2.0 ST Array. Bioinformatics analysis method such as the series test of cluster (STC), STC-Gene Ontology analysis (STC-GO), pathway analysis and Signal-net were applied in this study. The overall goals are to provide candidate gene biomarkers of exposure and early effect of benzene and to have an in-depth insight into the molecular mechanisms regulating benzene hematotoxicity.

2. Materials and methods

2.1. Characteristics of study population

These benzene-exposed subjects are all paint sprayers and health controls were office staffs without benzene exposure in the same factory. Airborne benzene concentration among health controls without benzene exposure, benzene-exposed group without clinical symptoms and chronic benzene poisoning group is 0.06 ± 0.01 mg/m^3, 1.82 ± 1.16 mg/m^3, and 6.68 ± 2.28 mg/m^3, respectively. Briefly, peripheral blood samples were randomly obtained from four patients of chronic benzene poisoning (average age 40.5 (32–50) yrs, average white blood cell 3.5 * 10^9/L), three benzene-exposed workers (average age 44.7 (39–51) yrs, average white blood cell 6.4 * 10^9/L) and three health controls without benzene exposure (average age 43.3(34–55) yrs, average white blood cell 6.4 * 10^9/L) with informed consents, and the study was approved by the Committees for Ethical Review of Research involving Human Subjects of Capital Medical University. Chronic benzene poisoning was diagnosed according to diagnostic criteria and principles of occupational benzene poisoning (GBZ 68-2008). The subjects in each group were matched by age, gender and lifestyles. These samples were used for microarray analysis.

2.2. Peripheral blood RNA gene expression profiling analysis

All blood samples from each worker were subjected to RNA extraction. Total RNA was extracted using TRIzol reagent (Invitrogen, USA) and miRNeasy Mini Kit (QIAGEN) according to the manufacturer's protocol. RNA integrity was evaluated by NanoDrop ND-1000 spectrophotometer. RNA was amplified and transcribed into cRNA utilizing a random priming method and cDNA was labeled and hybridized to the GeneChip Human Gene 2.0 ST Array (Affymetrix). 31,650 coding transcripts were collected from many related microarray literatures. GeneChip Human Gene 2.0 ST Array Protocol: Step 1, prepare RNA sample, kit and reagents: TRIzol reagent (Invitrogen, USA) and miRNeasy Mini Kit (QIAGEN); Step 2, total RNA clean-up and RNA QC; Step 3, prepare labeling reaction and reagents: GeneChip WT Terminal Labeling and Controls Kit; Step 4, purify the labeled/amplified RNA and labeled cRNA QC; Step 5, hybridization; Step 6, microarray wash; Step 7, scanning; and Step 8, extract data using Affymetrix Extraction Software. The arrays were scanned by GeneChip® Command Console® Software (AGCC) and the acquired array images were analyzed by Affymetrix GeneChip Operating Software. QC analysis of Gene 2.0 ST Array data was performed using the Affymetrix® Expression Console™ Software. The microarray data discussed in this paper have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE50737 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE50737).

2.3. Differentially expressed mRNAs were identified through random variance model (RVM)

Because of high test fee, only three to four samples at each group were detected. The age, gender, lifestyle such as smoking, drinking, and medical history in each group were matched to reduce the impact of these confounding factors. In addition, the effective statistical method for small samples was adopted. RVM F-test was commonly used to filter the differentially expressed genes because the RVM F-test can effectively increase the degrees of freedom in cases of small samples (Clarke et al., 2008; Wright and Simon, 2003). We computed *P* values adjusted for multiple testing by controlling the false discovery rate (FDR) with the Benjamini–Hochberg procedure (Reiner et al., 2003). These values were FDR-adjusted *P* values and were considered significant if they were <0.05.

2.4. STC and STC-GO analysis

The series test of cluster (STC) algorithm of gene expression dynamics was used to profile the gene expression time series and to identify the most probable set of clusters generating the observed time series. This method explicitly took into account the dynamic nature of the temporal gene expression profiles during clustering and identified the number of distinct clusters. We selected differential expression genes at a logical sequence. In accordance with different signal density change tendencies of genes under different situations, we identified a set of unique model expression tendencies. The raw expression values were converted into $\log_2$ ratio. Using a strategy for clustering short time-series gene expression data, we defined some unique profiles. The expression model profiles were related to the actual or the expected number of genes assigned to each model profile. Significant profiles had higher probability than expected by Fisher's exact test and multiple comparison tests (Xiao et al., 2010; Ramoni et al., 2002; Miller et al., 2002). Gene Ontology (GO) of significant STC cluster profiles was performed using the two-sided Fisher's exact test. $P < 0.05$ was considered as the significant GO categories. Genes in the same temporal expression profiles were more likely to be grouped in certain biological processes.

2.5. Gene ontology analysis and pathway analysis

Gene ontology analysis was applied to analyze the main function of differential expression genes according to the Gene ontology project (Ashburner et al., 2000). Fisher's exact test and χ^2 test were used to classify the GO category, and the FDR was calculated to correct the *P* value (Dupuy et al., 2007). The standard of difference screening was $P < 0.05$. Similarly, pathway analysis was used to find out the significant

Table 1
Significant GO analysis of differentially expressed genes related to benzene hematotoxicity.

GO term	GO_diffgene_count*	*P* value	FDR
Transcription, DNA-dependent	2	3.49E−48	6.32E−45
Inflammatory response	61	2.71E−16	2.45E−13
Cellular protein metabolic process	1	4.72E−09	1.55E−06
I-kappaB kinase/NF-kappaB cascade	15	6.12E−09	1.85E−06
Regulation of transcription, DNA-dependent	72	2.14E−08	4.84E−06
Chemotaxis	25	2.14E−08	4.86E−06
Fibroblast growth factor receptor signaling pathway	1	2.72E−07	4.55E−05
Immune response	80	3.34E−07	5.25E−05
Mitotic cell cycle	1	3.95E−07	5.85E−05
Signal transduction	132	4.21E−07	6.09E−05
Cell cycle	42	2.08E−06	0.0002398
Macrophage activation	5	5.10E−06	0.0005134

* GO_diffgene_count indicates the number of differentially expressed genes corresponding to GO term.