Association between genetic variants in *VEGF*, *ERCC3* and occupational benzene haematotoxicity

H D Hosgood III, L Zhang, M Shen, et al.

Occup Environ Med 2009 66: 848-853 originally published online September 22, 2009
doi: 10.1136/oem.2008.044024

Association between genetic variants in *VEGF*, *ERCC3* and occupational benzene haematotoxicity

H D Hosgood III,[1] L Zhang,[2] M Shen,[1] S I Berndt,[1] R Vermeulen,[3] G Li,[4] S Yin,[4] M Yeager,[1,2] J Yuenger,[1,2] N Rothman,[1] S Chanock,[1,2] M Smith,[5] Q Lan[1]

► An additional table is published online only at http://oem.bmj.com/content/vol66/issue12

[1] Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland, USA; [2] Center for Cancer Research, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland, USA; [3] Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands; [4] Chinese Center for Disease Control and Prevention, Beijing, China; [5] School of Public Health, University of California, Berkeley, California, USA

Correspondence to: H Dean Hosgood, National Cancer Institute, Division of Cancer Epidemiology and Genetics, Occupational and Environmental Epidemiology Branch, 6120 Executive Blvd, EPS 8118, MCS 7240, Bethesda, MD 20892-7240, USA; hosgoodd@mail.nih.gov

Accepted 27 June 2009
Published Online First 22 September 2009

ABSTRACT

Introduction: Benzene is an established human haematotoxin, with substantial interindividual variation in benzene-induced toxicity.

Methods: To further examine if genetic variation contributes to benzene haematotoxicity, we analysed 1023 tagSNPs in 121 gene regions important for benzene metabolism, haematopoiesis, leukaemia and lymphoma among 250 workers exposed to benzene and 140 unexposed controls in a cross-sectional study carried out in China. Linear regression was used to analyse the relationship between genetic polymorphisms and total white blood cell (WBC) count and its subtypes, adjusting for potential confounders and occupational exposure to benzene and toluene among exposed workers. The minp test assessed the association on the gene region level. The false discovery rate method was used to control for multiple comparisons.

Results: *VEGF* (minp = 0.0030) and *ERCC3* (minp = 0.0042) were the most significantly associated gene regions with altered WBC counts among benzene-exposed workers, after accounting for multiple comparisons. Highly significant changes were also found for WBC subtype counts, including granulocytes, CD4+ T cells and lymphocytes for *VEGF* and granulocytes and NK cells for *ERCC3*. Further, in workers exposed to <1 ppm, a SNP in *VEGF* was associated with changes in WBC and granulocyte counts, and SNPs in *ERCC3* were associated with changes in WBC, NK cell and granulocyte counts.

Discussion: Our findings suggest that genetic variation in *VEGF*, which plays an important role in blood vessel growth, and *ERCC3*, which is a member of the DNA repair pathway and is responsible for repairing bulky DNA adducts formed by chemicals, may contribute to individual susceptibility to benzene-induced haematotoxicity at relatively low levels of benzene exposure.

What this papers adds

- Benzene is a ubiquitous environmental pollutant and a well known leukaemogen.
- Benzene exposure alters white blood cell counts.
- Genetic variation may influence susceptibility to benzene-associated health effects.
- The *VEGF* and *ERCC3* gene regions are associated with altered white blood cell counts among benzene-exposed workers.
- Genetic variation in *VEGF* and *ERCC3* may contribute to individual susceptibility to benzene-induced haematotoxicity.

Benzene is a ubiquitous environmental pollutant found in automobile exhaust, gasoline and cigarette smoke. Chronic benzene exposure is believed to affect the bone marrow and peripheral blood cells and induce human health effects by directly decreasing chromosomal stability.[1–5] Benzene exposure has been associated with aplastic anaemia, myelodysplastic syndrome, leukaemia and non-Hodgkin lymphoma.[6–9] Although the mechanisms have not been fully elucidated, benzene toxicity has been shown to vary among individuals with similar occupational exposures.[8] Interindividual variation to benzene toxicity suggests that genetic variation may explain susceptibility to benzene-associated health effects.

Previous reports of this cross-sectional study in China have suggested that benzene exposure significantly alters blood cell counts,[8] with interindividual variation attributed to genetic variation in genes involved in benzene metabolism,[8 10] DNA double-strand break repair,[11] and cytokine and cellular adhesion.[12] Here, we present the evaluation of 1536 tagged single nucleotide polymorphisms (tagSNPs) in a broad range of candidate genes important for benzene metabolism and haematopoiesis, and of potential relevance for tumours associated with benzene including leukaemia and lymphoma, genotyped with an Illumina GoldenGate assay (Illumina, San Diego, CA) in a cross-sectional study of 250 benzene-exposed workers and 140 unexposed controls in Tianjin, China.

METHODS

This study population has been previously described in detail.[8] Briefly, 250 workers exposed to benzene and 140 unexposed sex- and age-frequency matched controls from Tianjin, China were enrolled in 2000 and 2001. Exposure assessment for cases and controls included 3M organic vapour air monitors (3M, St Paul, MN), urinary benzene measurements and dermal benzene exposure measurements, with complete organic vapour solvent scans for a subgroup of badges to evaluate potential co-exposures.[8 13 14] The mean (SD) benzene air exposure 1 month prior to phlebotomy was 5.4 ppm (12.1 ppm) among exposed workers.[12] About 40% of the exposed workers were exposed to <1 ppm benzene in the month prior to blood sample collection. A detailed questionnaire on lifetime occupational and environmental exposures, recent flu and respiratory infections in the previous month, medical history, current medication use, tobacco smoking and alcohol intake was administered to all subjects. Interviews, physical

exams and biological sample collection took place in June 2000 (n = 88) and in May and June 2001 (n = 330). Twenty eight subjects were enrolled in both years. Complete blood cell counts and differentials were analysed with a Beckman-Coulter T540 blood counter (Beckman-Coulter, Fullerton, CA). Lymphocyte subsets were measured with a Becton Dickinson FACSCalibur flow cytometer (SimulSET v 3.1) (Becton Dickinson, Franklin Lakes, NJ).[8]

Genomic DNA for genotyping was extracted from peripheral blood or buccal cells using a phenol-chloroform-extraction method.[15] TagSNPs from a broad range of candidate genes, including several genes in which a limited number of candidate SNPs were previously genotyped and reported from this study,[11 12] were chosen from the designable set of common SNPs (minor allele frequency (MAF) <5%) genotyped in the Caucasian population sample of the HapMap Project (Data Release 20/Phase II, NCBI Build 35 assembly, dpSNPb125) using Tagzilla (http://tagzilla.nci.nih.gov/), which implements a tagging algorithm based on the pairwise binning method.[16] For each gene region, tagSNPs located within 20 kb 5′ of the start of transcription (exon 1) and 10 kb 3′ of the end of the last exon were grouped and selected using a binning threshold of $r^2>0.8$. When there were multiple transcripts available for the gene, the primary transcript was assessed. In total, 1536 tagSNPs were genotyped using an Illumina GoldenGate assay at the National Cancer Institute's Core Genotyping Facility (see supplementary table 1). Blinded replicate samples (n = 20) were interspersed throughout the genotyping plates to assess quality control. TagSNPs with a concordance rate <95% (n = 13) and completion rate <90% (n = 45) were excluded. All subjects had a completion rate ⩾90%. Of the 1478 successfully genotyped tagSNPs, 327 were excluded from analysis due to low MAF (<0.05). Hardy-Weinberg equilibrium (HWE) for each tagSNP was tested in controls with a Pearson χ^2 test or a Fischer's exact test if any of the cell counts were less than five. TagSNPs (n = 128) that deviated from HWE ($p \leqslant 0.05$) were removed, leaving 1023 tagSNPs for analysis.

Total white blood cell (WBC) count was used as the main endpoint of this study since altered WBC count is a primary component of benzene poisoning diagnosis in China and has been associated with risk of haematological malignancies and related disorders among benzene-exposed workers.[10] The relationship between each tagSNP and WBC count was evaluated using linear regression adjusting for age (continuous variable), sex, current cigarette smoking status (yes/no), current alcohol consumption (yes/no), recent infections (yes/no) and body mass index (BMI). For analyses restricted to benzene-exposed workers, the model was also adjusted for the natural log (ln) mean air benzene and ln mean air toluene exposure in the month prior to phlebotomy. Tests for trends were conducted assuming a linear dose–response pattern with increasing number of variant alleles (ie, 0, 1 and 2). Homozygotes for the most common allele of each tagSNP were used as the referent group. Gene–benzene interactions were tested by adding an interaction term between the genotype (dominant model) and benzene exposure (yes/no). The effects of tagSNPs on specific WBC types were also tested using the same methods and covariates. Mean cell counts and standard deviations were calculated for granulocytes, lymphocytes, CD4+ T cells, CD8+ T cells, CD4/8 ratio, B cells, NK cells, monocytes and platelets. Data from the 28 exposed workers that were studied in both years were treated independently by using generalised estimating equations to adjust for a potential correlation between the repeated measurements.[17] Results were similar when data from only the first or second year of study were used for these 28 subjects.

To assess the significance of the association between each gene region and WBC count among exposed workers, MatLab was used to perform a minp test that assesses the significance of the minimal p value in each gene region (including all tagSNPs with cell counts ⩾5), using a permutation-based resampling procedure (1000 permutations) that takes into account the number of tagSNPs genotyped within each gene region, as well as the underlying linkage disequilibrium pattern.[18] False discovery rates (FDRs) of the minp gene region results were calculated to control for multiple comparisons.[19] Gene regions with an FDR ⩽0.20 were considered noteworthy. Finally, haplotype blocks were determined using all genotyped tagSNPs (including those with MAF<0.05) by the solid spine LD algorithm in Haploview using data from all subjects. Haplotype frequencies were estimated using the expectation-maximisation algorithm[20] and haplotypes with frequencies less than 1% were collapsed into a single category. The association with WBC count was assessed using a global score test in Haplostats.[21] A two tagSNP sliding window was also performed to identify regions associated with altered WBC counts.[18]

All statistical analyses were performed with SAS unless stated otherwise. This study was approved by the US National Cancer Institute's and the China Center for Disease Control's Institutional Review Boards. Informed written consent was obtained from all study participants.

RESULTS

As previously reported in this study population,[12] exposed workers and unexposed workers had similar distributions of

Table 1 Peripheral blood cell counts stratified by benzene exposure status

	Controls (n = 140)	Exposed (n = 250)	p Value*
Total white blood cells	6480 (1710)	5490 (1350)	<0.001
Granulocytes	4110 (1410)	3330 (1050)	<0.001
Lymphocytes	2130 (577)	1940 (520)	0.0014
CD4+ T cells	742 (262)	622 (183)	<0.001
CD8+ T cells	553 (208)	553 (213)	0.88
B cells	218 (94)	173 (88.5)	<0.001
NK cells	586 (318)	542 (277)	0.30
Monocytes	241 (92)	215 (93.2)	0.002

Values are mean (SD).
*Linear regression was used to test for differences between exposed and control subjects, adjusting for age, sex, current smoking, current alcohol drinking, BMI and recent infections.

Table 2 Most highly statistically significant gene region associations with peripheral white blood cell count among benzene-exposed workers in Tianjin, China

Gene region*	Gene name	Gene (and number of SNPs) included	Total SNPs	Minp‡	FDR values
VEGF	Vascular endothelial growth factor	*VEGF* (8)	8	0.003	0.15
ERCC3	Excision repair cross-complementing 3	*ERCC3* (2), *MAP3K2* (2)	4	0.004	0.15
BLM†	Bloom syndrome	*BLM* (14)	14	0.006	0.15
GPX3†	Glutathione peroxidase 3	*GPX3* (8), *TNIP1* (1)	9	0.007	0.15
IL8RB/IL8RA	Interleukin 8 receptor, α and β	*IL8RB* (1), *IL8RA* (2)	3	0.008	0.15
RIPK2	Receptor-interacting serine-threonine kinase 2	*RIPK2* (2)	1	0.010	0.15
IL6	Interleukin 6	*IL6* (2)	2	0.012	0.15
IL6R	Interleukin 6 receptor	*IL6R* (4), *MRPS33P1* (1)	5	0.012	0.15
IL10/IL19†	Interleukin 10 and 19	*IL10* (4), *IL19* (3)	7	0.013	0.15
IL12RB1	Interleukin 12 receptor, β 1	*IL12RB1* (2)	2	0.014	0.15
WRN†	Werner syndrome	*WRN* (8)	8	0.014	0.15
IFNAR2	Interferon receptor 2	*IFNAR2* (3)	3	0.017	0.18

*Defined as SNPs located within 20 kb 5′ of the start of transcription (exon 1) and 10 kb 3′ of the end of the last exon which were grouped and selected using a binning threshold of $r^2>0.8$.
†Previously genotyped and reported in Lan *et al*[12], Shen *et al*[11] and Lan *et al*.[45]
‡Adjusting for age, sex, current cigarette smoking status, current alcohol consumption, recent infections, body mass index and ln mean air benzene and ln mean air toluene exposure in the month prior to phlebotomy.

gender (p = 0.59), age (p = 0.40), BMI (p = 0.94), alcohol use (p = 0.41) and smoking status (p = 0.11). Further, exposed workers had significantly lower total WBC (p<0.001), granulocyte (p<0.001), lymphocyte (p = 0.0014), CD4+ T cell (p<0.001), B cell (p<0.001) and monocyte (p = 0.002) counts compared to unexposed workers (table 1).

Gene region analysis identified 11 regions that were associated with altered WBC counts among benzene-exposed workers, after accounting for multiple comparisons (FDR ⩽ 0.20) (table 2). *VEGF* (minp = 0.0030) and *ERCC3* (minp = 0.0042) were the most significant gene regions associated with altered WBC count among exposed workers, after accounting for multiple comparisons, and are the focus of this report.

Three of the eight and two of the four tagSNPs in *VEGF* and *ERCC3*, respectively, were associated with altered WBC counts among exposed workers ($p_{trend} \leqslant 0.05$) (table 3). The variant allele *C* at *VEGF* rs3025030 was associated with a significantly increasing trend of WBC counts in exposed workers ($p_{trend}<0.001$), compared to a significantly decreasing trend in unexposed workers ($p_{trend} = 0.013$) ($p_{interaction}<0.001$). The variant allele *C* at *VEGF* rs833058 was associated with a significantly increasing trend of WBC counts in exposed workers ($p_{trend} = 0.0011$). WBC counts increased about 10% in the homozygote variant carriers of *VEGF* rs3025030 and rs833058, compared to homozygote wildtype carriers. In the *ERCC3* gene region, the variant allele *T* at rs4150441

Table 3 Effect on total WBC counts in *VEGF and ERCC3* SNPs among unexposed controls and benzene-exposed subjects

						Workers exposed to benzene						
Gene region	SNP	Genotypes	Controls, n	WBC, mean (SD)	p*	All, n	WBC, mean (SD)	p*	p For interaction	<1 ppm, n	WBC, mean (SD)	p*
VEGF	rs3025030	GG	78	6785 (1786)		146	5284 (1217)			59	5219 (994)	
		CG	31	6052 (1438)	0.081	79	5771 (1271)	<0.001		35	5863 (1366)	0.034
		CC	5	5560 (1016)	0.026	12	5942 (1405)	0.083		8	5988 (1675)	0.17
		Trend			0.013			<0.001	<0.001			0.035
	rs833058	TT	55	6618 (1803)		115	5275 (1192)			44	5473 (1232)	
		CT	45	6407 (1533)	0.89	107	5644 (1309)	0.0061		53	5523 (1235)	0.68
		CC	14	6593 (1928)	0.74	15	5880 (1303)	0.014		5	5500 (1323)	0.32
		Trend			0.84			0.0011	0.084			0.41
	rs699946	AA	41	6346 (1789		77	5522 (1112)			31	5406 (1070)	
		AG	50	6734 (1693)	0.66	107	5583 (1394)	0.88		48	5696 (1372)	0.43
		GG	22	6505 (1588)	0.67	53	5209 (1179)	0.0092		23	5217 (1062)	0.17
		Trend			0.79			0.016	0.36			0.23
ERCC3	rs4150441	CC	46	6724 (1501)		72	5272 (1284)			33	5115 (1108)	
		CT	43	6381 (1803)	0.24	116	5531 (1218)	0.017		50	5612 (1155)	0.0056
		TT	25	6436 (1911)	0.15	49	5663 (1330)	0.013		19	5874 (1468)	0.17
		Trend			0.12			0.0086	0.011			0.010
	rs6731176	TT	46	6724 (1501)		73	5260 (1279)			34	5094 (1098)	
		CT	43	6381 (1803)	0.24	116	5541 (1213)	0.013		49	5637 (1154)	0.0036
		CC	25	6436 (1911)	0.15	48	5665 (1344)	0.014		19	5874 (1468)	0.17
		Trend			0.12			0.0087	0.010			0.0068

*Adjusted for age, sex, current smoking, current alcohol drinking, BMI, recent infections, and among exposed workers ln air benzene exposure and ln air toluene exposure in the month before phlebotomy

($p_{trend} = 0.0086$) and the variant allele *C* at rs6731176 ($p_{trend} = 0.0087$) were associated with significantly increasing trends of WBC counts in exposed workers. The association was not observed among unexposed workers, but the interactions between SNPs and exposure status were significant ($p_{interaction} = 0.011$ and $p_{interaction} = 0.010$, respectively).

WBC subtype analyses among exposed workers found *VEGF* rs3025030 and rs833058 to be associated with increasing granulocyte and CD4+ T cell ($p_{trend} \leqslant 0.05$) counts (table 4). *VEGF* rs3025030 and rs833058 were also associated with border-line significant ($p_{trend} = 0.061$) and significant ($p_{trend} = 0.011$) increased lymphocyte counts, respectively. *ERCC3* rs4150441 and rs6731176 were associated with altered ($p_{trend} \leqslant 0.05$) granulocyte and NK cell counts in exposed workers.

The variant allele *C* at *VEGF* rs3025030 was significantly associated with increased WBC ($p_{trend} = 0.0035$) and granulocyte ($p_{trend} = 0.025$) counts in workers exposed to <1 ppm of benzene. In *ERCC3*, the variant allele *T* at rs4150441 and the variant allele *C* at rs6731176 were significantly associated with increased WBC ($p_{trend} = 0.010$; $p_{trend} = 0.0068$, respectively), NK cell ($p_{trend} = 0.044$; $p_{trend} = 0.038$, respectively) and granulocyte ($p_{trend} = 0.0087$; $p_{trend} = 0.0052$, respectively) counts in workers exposed to <1 ppm of benzene.

Haplotype analyses for *VEGF* and *ERCC3* did not provide any additional insights into the associations beyond those observed in the individual SNP analyses.

DISCUSSION

Through an exploratory analysis of 121 gene regions, *VEGF* and *ERCC3* were most significantly associated with WBC count change among exposed workers. The variant allele *C* at two *VEGF* SNPs was associated with altered WBC counts among benzene-exposed workers: *VEGF* rs3025030 and rs833058. Altered cell counts were also seen in granulocytes, CD4+ T cells and lymphocytes, in benzene-exposed carriers of the variant allele *C* at *VEGF* rs3025030 and at rs833058. In the *ERCC3* gene region, rs4150441 and rs6731176 were both associated with altered WBC counts, as well as granulocyte and NK cell counts, in exposed workers. Similar associations from WBC and WBC subtype counts were observed when evaluating workers with <1 ppm of exposure.

VEGF, or vascular endothelial growth factor, is a key regulator of blood vessel growth.[22] VEGF is primarily involved in the promotion of endothelial cells from arteries, veins and lymphatics.[23] VEGF is also responsible for the survival of endothelial cells by regulation of the AKT signalling pathway,[24] which is an important regulator of apoptosis and is essential for helping cells manage apoptotic stimuli. Further, VEGF is also involved in immune function by playing a role in cytokine production via the nuclear factor kappa B (NFκB) pathway.[25] NFκB expression in cancer is an important regulator of pro-angiogenic and pro-metastatic cytokines, including VEGF, IL-6 and IL-8.[26 27] VEGF has been seen to affect bone marrow-derived cells. For example, VEGF promotes monocyte chemotaxis and induces colony formation of granulocyte-macrophage progenitor cells.[28 29] Further, VEGF has been reported to control the survival of haematopoietic stem cells.[30] Lethality seen in mice embryos with inactivated VEGF results from a lack of vascular structures and the lack of endothelial and haematopoietic stem cells.[31–33] Therefore, VEGF is strongly interconnected with WBC and WBC subtype levels in humans. When taken together, our findings that WBC and WBC subtype counts vary with *VEGF* polymorphisms are biologically plausible.

VEGF expression has been reported in haematologic malignancy cell lines such as multiple myeloma, T cell lymphoma, acute lymophoblastic leukaemia, Burkitt lymphoma, histiocytic lymphoma and chronic myelocytic lymphoma.[34] Further, VEGF receptors have been found to be expressed in acute myelogenous leukaemia, myelodysplastic syndrome, multiple myeloma and chronic myelogenous lymphoma.[35–37] VEGF expression has also been associated with poor prognosis of non-Hodgkin lymphoma.[38] Similarly, benzene exposure has been associated with aplastic anaemia, myelodysplastic syndrome, leukaemia and non-Hodgkin lymphoma.[6–9]

The metabolism of benzene into quinone and hydroquinone can generate free radicals and reactive oxygen species that can damage DNA.[1 39] Benzene-induced haematotoxicity and haematopoietic malignancies are thought to occur through cell transformation and gene mutation.[40] The body's ability to protect the genome from benzene-induced harm is largely dependent on the overlapping DNA repair pathways. Two major DNA repair pathways are the double-strand break repair (DSB) and the nucleotide excision repair (NER). Previous reports in this study population evaluating polymorphisms in the DSB found variant alleles in *WRN* and *TP53* to be associated with altered WBC counts.[11] In this report, we found two SNPs in the *ERCC3* gene region (rs4150441 and rs6731176) to be associated with altered WBC counts. *ERCC3* is a key player in the NER, which is responsible for repairing bulky DNA adducts formed by chemicals, such as benzene. A gene expression profile of 141 DNA repair genes identified *ERCC3* to be associated with benzene poising, among Chinese subjects.[41] Polymorphisms in other NER genes have also been associated with benzene poisoning susceptibility in Chinese workers.[42] Polymorphisms in NER genes, particularly *ERCC5*, have been associated with non-Hodgkin lymphoma susceptibility.[43] It should be noted that rs6731176 is in *MAP3K2*, upstream of *ERCC3*; therefore, this variant may affect *MAP3K2* as well as *ERCC3*. Beyond *ERCC3*, our analyses of NER pathway genes, including *ERCC1*, *ERCC2*, *ERCC5* and *LIG1*, found only one SNP (*ERCC2* rs238415) to be significantly associated ($p_{trend} \leqslant 0.05$) with altered WBC counts among benzene-exposed workers.

The moderate sample size of our study may lead to both false positive and false negative findings.[44] We accounted for possible spurious findings due to multiple comparisons by evaluating FDRs. Although functionality is not known for all genotyped SNPs, our results are biologically plausible given that variants in *VEGF* and *ERCC3* could contribute to benzene-induced haematotoxicity. Our observed interactions between genetic variation in *VEGF* and *ERCC3* with benzene exposure suggest that benzene exposure, in concert with these variants, induces adverse health effects, such as altered WBC and WBC subtype counts. Associations with a particular SNP in this study may be the result of linkage disequilibrium with another functional SNP in the region. Finally, the SNPs genotyped for this study were selected to provide substantial genomic coverage of each candidate gene in Caucasians as this same panel of SNPs has been used in studies of other ethnic groups.

In summary, SNPs in *VEGF* and *ERCC3* were associated with alterations in WBC and WBC subtype counts in workers exposed to benzene, even at relatively low levels of exposure below 1 ppm. These findings lend additional support to the hypothesis that genetic variation plays an important role in individual susceptibility to benzene-induced haematotoxicity.

Table 4 Effect on white blood cell subtypes in *VEGF* and *ERCC3* SNPs among benzene-exposed subjects

Gene region	SNP	Genotype	n	Granulocytes		Lymphocytes		CD4+ T cells		CD8+ T cells		CD4/8 ratio		B cells		NK cells		Monocytes		Platelets	
				Mean (SD)	p*	Mean (SD)	p*	Mean (SD)	p*	Mean (SD)	p*	Mean (SD)	p*	Mean (SD)	p*	Mean (SD)	p*	Mean (SD)	p*	Mean (SD)	p*
VEGF	rs3025030	GG	146	3169 (953)		1905 (500)		608 (183)		551 (217)		1.206 (0.43)		166 (91)		536 (276)		210 (91)		200 541 (53 345)	
		CG	79	3565 (1074)	<0.001	1984 (437)	0.072	652 (181)	0.058	542 (187)	0.94	1.292 (0.416)	0.12	185 (82)	0.071	544 (231)	0.32	223 (85)	0.31	212 190 (51 508)	0.065
		CC	12	3467 (823)	0.16	2233 (972)	0.26	693 (218)	0.11	637 (328)	0.19	1.188 (0.427)	0.84	199 (116)	0.35	650 (526)	0.97	242 (131)	0.85	200 083 (50 041)	0.83
		Trend			0.0016		0.061		0.024		0.40		0.26		0.081		0.57		0.44		0.17
	rs833058	TT	115	3181 (919)		1885 (495)		608 (187)		545 (189)		1.195 (0.432)		169 (82)		502 (241)		209 (86)		197 922 (51 999)	
		CT	107	3431 (1072)	0.048	1996 (519)	0.015	642 (187)	0.11	564 (244)	0.45	1.257 (0.419)	0.53	172 (83)	0.51	586 (301)	0.0055	217 (88)	0.31	210 280 (50 849)	0.060
		CC	15	3533 (1036)	0.074	2080 (617)	0.10	663 (153)	0.037	520 (175)	0.95	1.365 (0.406)	0.046	218 (166)	0.12	579 (360)	0.71	267 (129)	0.089	212 133 (66 843)	0.20
		Trend			0.016		0.011		0.024		0.57		0.13		0.14		0.069		0.082		0.047
	rs699946	AA	77	3360 (893)		1940 (461)		635 (161)		537 (193)		1.292 (0.426)		175 (99)		541 (289)		222 (106)		196 506 (55 440)	
		AG	107	3406 (1124)	0.78	1964 (541)	0.94	632 (205)	0.35	560 (235)	0.75	1.23 (0.452)	0.27	168 (88)	0.30	551 (276)	0.34	213 (80)	0.81	212 664 (53 547)	0.084
		GG	53	3072 (863)	0.0062	1925 (547)	0.29	604 (178)	0.14	557 (204)	0.86	1.158 (0.359)	0.16	184 (80)	0.88	538 (279)	0.54	213 (90)	0.52	199 189 (44 556)	0.92
		Trend			0.011		0.33		0.13		0.84		0.15		0.98		0.66		0.54		0.75
ERCC3	rs4150441	CC	72	3132 (975)		1929 (574)		607 (192)		553 (239)		1.2 (0.388)		162 (71)		545 (303)		211 (90)		208 472 (56 141)	
		CT	116	3366 (985)	0.022	1946 (510)	0.23	641 (176)	0.099	556 (211)	0.38	1.252 (0.424)	0.62	176 (94)	0.13	531 (285)	0.52	219 (95)	0.33	207 810 (50 978)	0.81
		TT	49	3467 (1064)	0.0064	1980 (441)	0.15	623 (197)	0.64	542 (188)	0.64	1.24 (0.481)	0.98	185 (103)	0.14	578 (226)	0.031	216 (85)	0.53	190 347 (49 884)	0.13
		Trend			0.0044		0.14		0.51		0.57		0.96		0.12		0.044		0.47		0.18
	rs6731176	TT	73	3123 (971)		1926 (571)		610 (192)		550 (238)		1.214 (0.404)		162 (71)		541 (303)		211 (89)		208 644 (55 769)	
		CT	116	3376 (982)	0.015	1947 (510)	0.24	639 (176)	0.13	560 (211)	0.33	1.239 (0.417)	0.79	176 (95)	0.16	532 (284)	0.47	219 (95)	0.36	207 629 (50 968)	0.83
		CC	48	3465 (1076)	0.0074	1983 (445)	0.16	624 (199)	0.68	536 (186)	0.71	1.251 (0.48)	0.99	186 (104)	0.16	582 (227)	0.026	217 (86)	0.55	190 146 (50 392)	0.12
		Trend			0.0045		0.14		0.56		0.62		0.96		0.13		0.038		0.49		0.17

*Adjusted for age, sex, current smoking, current alcohol drinking, BMI, recent infections, and among exposed workers ln air benzene exposure and ln air toluene exposure in the month before phlebotomy

Funding: This project was supported in part by the NIH intramural research program, and by NIH grants RO1ES06721, P42ES04705 and P30ES01896 (to MTS).

Competing interests: M Smith has received consulting and expert testimony fees from law firms representing both plaintiffs and defendants in cases involving exposure to benzene. GL has received funds from the American Petroleum Institute for consulting on benzene-related health research.

Ethics approval: This study was approved by the US National Cancer Institute's and the China Center for Disease Control's Institutional Review Boards.

Provenance and peer review: Not commissioned; externally peer reviewed.

REFERENCES

1. **Ross D.** Metabolic basis of benzene toxicity. *Eur J Haematol Suppl* 1996;**60**:111–18.
2. **Snyder R,** Dimitriadis E, Guy R, *et al*. Studies on the mechanism of benzene toxicity. *Environ Health Perspect* 1989;**82**:31–5.
3. **Irons RD,** Stillman WS. Impact of benzene metabolites on differentiation of bone marrow progenitor cells. *Environ Health Perspect* 1996;**104**(Suppl 6):1247–50.
4. **Smith MT,** Zhang L, Wang Y, *et al*. Increased translocations and aneusomy in chromosomes 8 and 21 among workers exposed to benzene. *Cancer Res* 1998;**58**:2176–81.
5. **Zhang L,** Rothman N, Wang Y, *et al*. Increased aneusomy and long arm deletion of chromosomes 5 and 7 in the lymphocytes of Chinese workers exposed to benzene. *Carcinogenesis* 1998;**19**:1955–61.
6. **Hayes RB,** Songnian Y, Dosemeci M, *et al*. Benzene and lymphohematopoietic malignancies in humans. *Am J Ind Med* 2001;**40**:117–26.
7. **Hayes RB,** Yin SN, Dosemeci M, *et al*. Benzene and the dose-related incidence of hematologic neoplasms in China. Chinese Academy of Preventive Medicine–National Cancer Institute Benzene Study Group. *J Natl Cancer Inst* 1997;**89**:1065–71.
8. **Lan Q,** Zhang L, Li G, *et al*. Hematotoxicity in workers exposed to low levels of benzene. *Science* 2004;**306**:1774–6.
9. **Gist GL,** Burg JR. Benzene—a review of the literature from a health effects perspective. *Toxicol Ind Health* 1997;**13**:661–714.
10. **Rothman N,** Smith MT, Hayes RB, *et al*. Benzene poisoning, a risk factor for hematological malignancy, is associated with the NQO1 609C—>T mutation and rapid fractional excretion of chlorzoxazone. *Cancer Res* 1997;**57**:2839–42.
11. **Shen M,** Lan Q, Zhang L, *et al*. Polymorphisms in genes involved in DNA double-strand break repair pathway and susceptibility to benzene-induced hematotoxicity. *Carcinogenesis* 2006;**27**:2083–9.
12. **Lan Q,** Zhang L, Shen M, *et al*. Polymorphisms in cytokine and cellular adhesion molecule genes and susceptibility to hematotoxicity among workers exposed to benzene. *Cancer Res* 2005;**65**:9574–81.
13. **Vermeulen R,** Li G, Lan Q, *et al*. Detailed exposure assessment for a molecular epidemiology study of benzene in two shoe factories in China. *Ann Occup Hyg* 2004;**48**:105–16.
14. **Vermeulen R,** Lan Q, Li G, *et al*. Assessment of dermal exposure to benzene and toluene in shoe manufacturing by activated carbon cloth patches. *J Environ Monit* 2006;**8**:1143–8.
15. **Garcia-Closas M,** Egan KM, Abruzzo J, *et al*. Collection of genomic DNA from adults in epidemiological studies by buccal cytobrush and mouthwash. *Cancer Epidemiol Biomarkers Prev* 2001;**10**:687–96.
16. **Carlson CS,** Eberle MA, Rieder MJ, *et al*. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* 2004;**74**:106–20.
17. **Zeger SL,** Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 1986;**42**:121–30.
18. **Huang BE,** Amos CI, Lin DY. Detecting haplotype effects in genomewide association studies. *Genet Epidemiol* 2007;**31**:803–12.
19. **Benjamini Y,** Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* 1995;**57**:289–300.
20. **Excoffier L,** Slatkin M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol* 1995;**12**:921–7.
21. **Schaid DJ,** Rowland CM, Tines DE, *et al*. Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet* 2002;**70**:425–34.
22. **Ferrara N,** Gerber HP, LeCouter J. The biology of VEGF and its receptors. *Nat Med* 2003;**9**:669–76.
23. **Ferrara N,** Davis-Smyth T. The biology of vascular endothelial growth factor. *Endocr Rev* 1997;**18**:4–25.
24. **Gerber HP,** McMurtrey A, Kowalski J, *et al*. Vascular endothelial growth factor regulates endothelial cell survival through the phosphatidylinositol 3′-kinase/Akt signal transduction pathway. Requirement for Flk-1/KDR activation. *J Biol Chem* 1998;**273**:30336–43.
25. **Crisostomo PR,** Wang Y, Markel TA, *et al*. Human mesenchymal stem cells stimulated by TNF-{alpha}, LPS, or hypoxia produce growth factors by an NF{kappa}B- but not JNK-dependent mechanism. *Am J Physiol Cell Physiol* 2008;**294**:C675–82.
26. **Novotny NM,** Markel TA, Crisostomo PR, *et al*. Differential IL-6 and VEGF secretion in adult and neonatal mesenchymal stem cells: role of NFkB. *Cytokine* 2008;**43**:215–19.
27. **Crisostomo PR,** Wang M, Herring CM, *et al*. Gender differences in injury induced mesenchymal stem cell apoptosis and VEGF, TNF, IL-6 expression: role of the 55 kDa TNF receptor (TNFR1). *J Mol Cell Cardiol* 2007;**42**:142–9.
28. **Clauss M,** Gerlach M, Gerlach H, *et al*. Vascular permeability factor: a tumor-derived polypeptide that induces endothelial cell and monocyte procoagulant activity, and promotes monocyte migration. *J Exp Med* 1990;**172**:1535–45.
29. **Broxmeyer HE,** Cooper S, Li ZH, *et al*. Myeloid progenitor cell regulatory effects of vascular endothelial cell growth factor. *Int J Hematol* 1995;**62**:203–15.
30. **Gerber HP,** Malik AK, Solar GP, *et al*. VEGF regulates haematopoietic stem cell survival by an internal autocrine loop mechanism. *Nature* 2002;**417**:954–8.
31. **Ferrara N,** Carver-Moore K, Chen H, *et al*. Heterozygous embryonic lethality induced by targeted inactivation of the VEGF gene. *Nature* 1996;**380**:439–42.
32. **Carmeliet P,** Ferreira V, Breier G, *et al*. Abnormal blood vessel development and lethality in embryos lacking a single VEGF allele. *Nature* 1996;**380**:435–9.
33. **Shalaby F,** Rossant J, Yamaguchi TP, *et al*. Failure of blood-island formation and vasculogenesis in Flk-1-deficient mice. *Nature* 1995;**376**:62–6.
34. **Gerber HP,** Ferrara N. The role of VEGF in normal and neoplastic hematopoiesis. *J Mol Med* 2003;**81**:20–31.
35. **Bellamy WT,** Richter L, Sirjani D, *et al*. Vascular endothelial cell growth factor is an autocrine promoter of abnormal localized immature myeloid precursors and leukemia progenitor formation in myelodysplastic syndromes. *Blood* 2001;**97**:1427–34.
36. **Fiedler W,** Graeven U, Ergun S, *et al*. Vascular endothelial growth factor, a possible paracrine growth factor in human acute myeloid leukemia. *Blood* 1997;**89**:1870–5.
37. **Verstovsek S,** Estey E, Manshouri T, *et al*. Clinical relevance of vascular endothelial growth factor receptors 1 and 2 in acute myeloid leukaemia and myelodysplastic syndrome. *Br J Haematol* 2002;**118**:151–6.
38. **Salven P,** Teerenhovi L, Joensuu H. A high pretreatment serum vascular endothelial growth factor concentration is associated with poor outcome in non-Hodgkin's lymphoma. *Blood* 1997;**90**:3167–72.
39. **Ross D.** The role of metabolism and specific metabolites in benzene-induced toxicity: evidence and issues. *J Toxicol Environ Health A* 2000;**61**:357–72.
40. **Tsutsui T,** Hayashi N, Maizumi H, *et al*. Benzene-, catechol-, hydroquinone- and phenol-induced cell transformation, gene mutations, chromosome aberrations, aneuploidy, sister chromatid exchanges and unscheduled DNA synthesis in Syrian hamster embryo cells. *Mutat Res* 1997;**373**:113–23.
41. **Chen L,** Bi YY, Tao N, *et al*. [cDNA microarray to identify the significance of DNA replication and damage repair genes associated with benzene poisoning]. *Zhonghua Lao Dong Wei Sheng Zhi Ye Bing Za Zhi* 2005;**23**:248–51.
42. **Huang HL,** Xu JN, Wang QK, *et al*. [Association between polymorphisms of XPD gene and susceptibility to chronic benzene poisoning]. *Zhonghua Lao Dong Wei Sheng Zhi Ye Bing Za Zhi* 2006;**24**:390–3.
43. **Shen M,** Zheng T, Lan Q, *et al*. Polymorphisms in DNA repair genes and risk of non-Hodgkin lymphoma among women in Connecticut. *Hum Genet* 2006;**119**:659–68.
44. **Wacholder S,** Chanock S, Garcia-Closas M, *et al*. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J Natl Cancer Inst* 2004;**96**:434–42.
45. **Lan Q,** Zhang L, Shen M, *et al*. Large-scale evaluation of candidate genes identifies associations between DNA repair and genomic maintenance and development of benzene hematotoxicity. *Carcinogenesis* 2009;**30**:50–8.