

Association between Common Genetic Variation in *Cockayne Syndrome A* and *B* Genes and Nucleotide Excision Repair Capacity among Smokers

Shuguang Leng,¹ Amanda Bernauer,¹ Christine A. Stidley,² Maria A. Picchi,¹ Xin Sheng,⁴ Melissa A. Frasco,⁴ David Van Den Berg,⁴ Frank D. Gilliland,⁴ Richard E. Crowell,^{2,3} and Steven A. Belinsky¹

¹Lung Cancer Program, Lovelace Respiratory Research Institute; ²Department of Internal Medicine, University of New Mexico; ³New Mexico VA Health Care System, Albuquerque, New Mexico and ⁴Keck School of Medicine, University of Southern California, Los Angeles, California

Abstract

Mutagen sensitivity in *in vitro* cultured lymphocytes challenged by benzo[*a*]pyrene diolepoxide (BPDE) has been validated as an intrinsic susceptibility factor for several cancers. Bulky BPDE-DNA adducts are repaired via either transcription-coupled repair or global genome nucleotide excision repair depending on the location of lesions. Cockayne syndrome A (CSA) and B (CSB) play essential roles in integrating the recognition of damage, chromatin remodeling, and the core nucleotide excision repair proteins. This study evaluated the hypothesis that common genetic variation in CSA and CSB is associated with mutagen sensitivity induced by BPDE in 276 cancer-free smokers. Tag single nucleotide polymorphisms (SNP; $n = 37$) selected across the entire coding and putative regulatory regions of CSA and CSB based on a high-density SNP database were

genotyped by the Illumina Golden Gate assay. Major principal components of CSA and CSB that captured the linkage disequilibrium from multiple SNPs were globally associated with the number of breaks per cell at the threshold of 80% ($P \leq 0.02$ for both genes). Haplotype H125 in CSA and H97 in CSB as well as SNPs in high linkage disequilibrium with these two haplotypes were significantly associated with a 13% to 15% reduction in the mean number of chromatid breaks per cell ($P < 0.05$). A resampling-based omnibus test supported the significant association between SNPs and haplotypes in CSA and mutagen sensitivity induced by BPDE ($P = 0.035$). This study implicates transcription-coupled repair in protecting the cell from BPDE-induced DNA damage. (Cancer Epidemiol Biomarkers Prev 2008;17(8):2062–9)

Introduction

Transcription-coupled repair (TCR) is the nucleotide excision repair (NER) subpathway responsible for accelerated removal of transcription-blocking lesions from the transcribed strand of active genes (1). Patients with defective TCR have Cockayne syndrome, which is characterized by pronounced sun sensitivity and neurodevelopmental abnormalities (2). The majority of Cockayne syndrome patients carry mutations in genes that encode for the Cockayne syndrome A (CSA) and/or B (CSB) proteins, although the Cockayne syndrome phenotype can also result from mutations in xeroderma pigmentosum genes (*XPB*, *XPD*, and *XPG*; refs. 3, 4). CSA is a part of an E3-ubiquitin ligase complex that is inactivated by interaction with the COP9 signalosome complex to prevent degradation of the stalled RNA polymerase II in TCR (5). CSB is a DNA-dependent ATPase that shares homology with the SWI2/SNF2 family of ATP-dependent chromatin remodeling enzymes (6). Recently, Foustari et al. (7) found that CSB

recruitment to stalled RNA polymerase II was essential for the subsequent assembly of NER proteins and histone acetyltransferase p300 at the site of DNA damage. CSA (in addition to CSB) is required for recruitment of the nucleosomal binding protein HMG1 (high mobility group N nonhistone nucleosome binding protein), XAB2 (an essential TRP-containing protein), and THII5 (a transcription cleavage factor). Taken together, these two Cockayne syndrome genes play essential but different roles in integrating the damage recognition, chromatin remodeling, and NER through interacting with different genes during TCR.

The mutagen sensitivity assay is an *in vitro* measure of DNA repair capacity expressed as the number of chromatid breaks per cell in short-term cultured lymphocytes challenged by mutagens. Different mutagens are used in the mutagen sensitivity assay that offers information about the involvement of specific DNA damage and repair pathways. Benzo[*a*]pyrene diolepoxide (BPDE) primarily reacts with guanine at the N² amino group to form bulky DNA adducts that distort the DNA double helix (8). Bulky BPDE-DNA adducts are repaired via either TCR or the global genome NER pathway depending on the location of the lesions (8, 9). Mutagen sensitivity in peripheral lymphocytes challenged by BPDE is associated with increased risk for several cancers such as lung cancer and squamous cell

Received 4/7/08; revised 5/16/08; accepted 5/27/08.

Grant support: NIH grant U01 CA097356 and Tobacco Settlement Fund of the State of New Mexico.

Requests for reprints: Steven A. Belinsky, Lung Cancer Program, Lovelace Respiratory Research Institute, 2425 Ridgcrest Drive South East, Albuquerque, NM 87108. Phone: 505-348-9465; Fax: 505-348-4990. E-mail: sbelinsk@LRRRI.org

Copyright © 2008 American Association for Cancer Research.

doi:10.1158/1055-9965.EPI-08-0308

carcinoma of the head and neck (10, 11). Wu et al. (12) evaluated the heritability of mutagen sensitivity to BPDE among 255 twins and found that ~50% of the variability of this phenotype can be explained by genetics, validating the use of mutagen sensitivity as a cancer susceptibility factor. Functional single nucleotide polymorphisms (SNP) in several NER genes, which include *XPA*, *XPC*, and *RAD23B*, modify the mutagen sensitivity induced by BPDE (13, 14). Thus, sequence variations in DNA repair genes that modify protein activity clearly affect mutagen sensitivity, and alterations in the sequence of these genes could cause suboptimal mutagen sensitivity to increase cancer risk.

Because TCR is an important component of NER and both *CSA* and *CSB* are polymorphic, we hypothesize that common genetic variation in *CSA* and *CSB* may influence the mutagen sensitivity induced by BPDE. A tag SNP-based approach was employed to identify the common genetic variation in *CSA* and *CSB* and to assess the association between both these tag SNPs and haplotypes and mutagen sensitivity induced by BPDE in 276 cancer-free smokers.

Materials and Methods

Study Population and Sample Collection. Subjects in this study were selected from participants of two smoker cohorts, the Lovelace Smokers Cohort and the Veteran Smokers Cohort, which were established in 2001 and 2000, respectively, to conduct longitudinal studies on molecular markers of respiratory carcinogenesis in biological fluids such as sputum and blood from people at risk for lung cancer. These two cohorts have been described previously (15, 16). In brief, cohort participants were all current or former smokers between 40 and 75 years old and were mainly residents of the Albuquerque metropolitan area. All participants signed the consent form. Whole blood was processed within 2 h after blood draw to isolate lymphocytes and plasma. Cryopreservation of lymphocytes began in 2005. Cryopreserved lymphocytes were available from 278 subjects that included 215 subjects from the Lovelace Smokers Cohort and 63 subjects from the Veteran Smokers Cohort. This study was approved by the institutional review boards of the Lovelace Respiratory Research Institute and the University of New Mexico.

Mutagen Sensitivity Assay. Samples from the Lovelace Smokers Cohort and the Veteran Smokers Cohort were randomly mixed and assayed as a batch. Phytohemagglutinin (M form)-stimulated lymphocytes were treated with BPDE to evaluate the generation of chromosome aberrations as an index of mutagen sensitivity. Briefly, cryopreserved lymphocytes were thawed and cultured in RPMI 1640 supplemented with fetal bovine serum (20%) and phytohemagglutinin (1.5%). Cell density was adjusted to $<0.5 \times 10^6$ /mL, and 72 h after phytohemagglutinin stimulation, the culture was split into two T25 flasks. Cells were treated with vehicle or BPDE for 24 h because BPDE is a S-phase-dependent clastogen. The final concentration for BPDE in culture medium was 0.3 $\mu\text{mol/L}$, a concentration defined through dose-response studies using isolated lymphocytes from cohort subjects and three lymphoblastoid cell lines: GM02345 (mutant XPA), GM16024 (mutant XPG),

and GM00131 (wild-type XPA and XPG). The dose selected was within the linear dose-response range and caused obvious genotoxicity but minimal cytotoxicity. One hour before harvest, colcemid was added to the cultures at a final concentration of 0.06 mg/mL. Slides were prepared according to conventional procedures and 100 well-spread metaphases were examined for chromatid breaks or exchanges. All experiments were carried out with reagents with the same lot number. One person who was blinded to the genotype of *CSA* and *CSB* scored the slides. Each simple chromatid break was scored as one break, whereas each isochromatid break set and each exchange event (including interstitial deletions) were considered as two breaks (17). Mutagen sensitivity was expressed as the mean number of chromatid breaks per cell. Thirteen samples were randomly selected for evaluating assay repeatability during the entire experiment. First and second tests of a sample were carried out in different batches of assayed samples. The average of the percent change between the two test results for these 13 samples was 10% with range of 0% to 25%. A XPA-defective lymphoblastoid cell line (GM02345) treated with 0.15 $\mu\text{mol/L}$ BPDE for 24 h was included in each batch for quality control. The average chromatid breaks per cell in GM02345 induced by 0.15 $\mu\text{mol/L}$ BPDE was 0.55 with a SD of 0.07.

Tag SNP Selection and Genotyping by the Illumina Golden Gate Assay. Tag SNPs ($n = 37$) were selected across the entire coding and putative regulatory regions of *CSA* and *CSB* (20 kb upstream and 10 kb downstream) based on a high-density SNP database of Hispanics and non-Hispanic Whites generated by Haiman et al. (18) plus Phase I HapMap data for Caucasians.⁵ The pairwise r^2 method was used to select tag SNPs with $r^2 \geq 0.8$ and minor allele frequency ≥ 0.05 . Five nonsynonymous SNPs that include rs2228527, rs2228529, rs2228526, rs4253211, and rs2228528 in *CSB* were either selected as tag SNPs or were in perfect linkage disequilibrium (LD; $r^2 = 1$) with selected tag SNPs. rs2228527 and rs2228529 were tagged by rs2228526. rs4253211 and rs2228528 were tagged by rs4253126 and rs4253226, respectively. One additional SNP for bins with at least six SNPs was selected as a redundant SNP in case of genotyping failure. The redundancy SNP was in perfect LD with the tag SNP in the same bin. Genomic DNA was isolated from the mononuclear blood cells by phenol-chloroform extraction followed by precipitation with ethanol. These SNPs were genotyped by the Illumina Golden Gate Assay⁶ for the 278 DNA samples. R_h^2 measures were also calculated to assess how well the chosen tag SNPs could predict all common haplotypes in *CSA* and *CSB* with an estimated frequency of >5% in non-Hispanic Whites and Hispanics, the two major ethnicities. The minimal R_h^2 for these tag SNPs in prediction of the common haplotypes is >0.9 for both *CSA* and *CSB* in non-Hispanic Whites and Hispanics in the reference panel for tag SNP selection.

Statistical Analysis. Two samples were excluded from data analysis because of high genotyping failure rates for

⁵ <http://www.hapmap.org/>

⁶ <http://www.illumina.com>

the 37 SNPs. Before data analysis, these 37 tag SNPs were assessed in the remaining 276 samples for the minor allele frequency (>0.05), genotyping failure rate (<5%), and agreement with Hardy-Weinberg equilibrium. All SNPs passed the above criteria, except rs4647128, which had a minor allele frequency of 0.04.

A global test based on principal components analysis was employed to determine whether the sequence variants in each gene were associated with mutagen sensitivity induced by BPDE (19). The original set of SNPs in each gene was transformed to a set of uncorrelated principal components. Each principal component explains a certain percentage of the total variance in the SNPs. The first principal component explains the most SNP variation, and the second principal component both is uncorrelated with the first principal component and explains the second most SNP variation. The number of principal components needed to describe common variation across a locus was defined *a priori* as the number of principal components needed to explain at least 80% of the sequence variance. These components were used as predictors in a linear regression with natural log mutagen sensitivity as the outcome and globally assessed using a likelihood ratio test. The association between mutagen sensitivity and the six nongenetic factors, gender, age, ethnicity, current smoking status, cryopreservation duration, and lymphocyte seeding count, was assessed using linear regression models. Adjustment for sex, current smoking status, and seeding count of lymphocytes per culture was included in the genetic association analysis because these nongenetic variables associated with mutagen sensitivity.

Multivariate linear regression models assessed the association between individual SNPs or haplotypes and the natural log of the number of breaks per cell with adjustment for gender, current smoking status, and seeding count of lymphocytes per culture. The additive model was tested for each SNP, and common homozygotes, heterozygotes, and rare homozygotes were coded as 0, 1, and 2, respectively. Results are expressed as the ratio of the geometric mean number of breaks per cell for heterozygotes relative to common homozygotes or between rare homozygotes and heterozygotes with 95% confidence intervals (95% CI). Because the additive model was used, these two ratios are the same. We

characterized the haplotype block structures of the expanded region of CSA and CSB in Hispanics, non-Hispanic Whites, Blacks, and Asians by using genotype data generated by Haiman et al. (18), from the HapMap Project, and from the Environmental Genomic Project.⁷ No evidence for historical recombination across the entire expanded region in either CSA or CSB (multiallelic D' between blocks = 1) was found in these four ethnicities. Therefore, the entire expanded region was treated as one block to analyze the association between haplotype and number of breaks per cell in the 276 subjects. A Bayesian statistical method implemented in the program PHASE (version 2.1) was used to reconstruct the haplotypes from the tag SNP data in CSA and CSB in non-Hispanic Whites ($n = 206$), Hispanics ($n = 53$), and other ethnicities ($n = 17$; refs. 20, 21). All possible haplotype pairs with corresponding probabilities >0.01 for each individual were generated and haplotypes with frequency <5% were combined into one group. The probabilities of the common haplotypes in each gene for each individual were used as explanatory variables in multivariate linear regression models with adjustment for nongenetic factors to assess the association between mutagen sensitivity and the haplotypes. The effect for each common haplotype was estimated using all other haplotypes as the reference group and is expressed as the ratio of the geometric mean number of breaks per cell between the common haplotype and all other haplotypes in each gene with 95% CI. The effects for the common haplotypes were also analyzed using the most common haplotype in each gene as the reference group and similar results were obtained (data not shown).

A resampling-based omnibus test assessed whether CSA or CSB as a single genetic locus was associated with mutagen sensitivity induced by BPDE (22). The genotype vectors for each of the 276 subjects were randomly permuted 5,000 times, whereas the trait data and any covariates were left intact. With this approach, the permuted samples simulated the null hypothesis of no association, but they maintained the original correlation between genotypes and between traits and covariates. An association test for each SNP and haplotype was done using the permuted data. A distribution of the minimum P values for the SNP and haplotype effects from each permutation was generated for each gene, and the observed minimal P values for the SNP and haplotype effects in each gene were reevaluated in relation to this distribution. The permutation based P value was calculated as the proportion of the minimum P values from each permutation that were equal to or smaller than the observed minimal P (22).

Haplotypes associated with mutagen sensitivity were identified and the number of protective haplotypes was determined for each individual. The association between the number of protective haplotypes and the natural log of the mean number of chromatid breaks per cell was examined with a multivariate linear regression model with adjustment for the nongenetic factors. Least-squares means were obtained, and the results are expressed as geometric mean number of breaks per cell with 95% CI.

Table 1. Demographic characteristics

Variables	Study subjects ($n = 276$)
Age, mean \pm SD	58.0 \pm 9.9
Gender (%)	
Female	52.9
Male	47.1
Race (%)	
Non-Hispanic White	75
Hispanic	19
Other	6
Smoking history	
Current (%)	46
Pack years, mean \pm SD	41.0 \pm 25.5
Duration, mean \pm SD	32.7 \pm 10.7
Cryopreservation duration (d, mean \pm SD)	368 \pm 190
Seeding count ($\times 10^6$ /culture, mean \pm SD)	1.7 \pm 0.6
Mutagen sensitivity (number of breaks per cell, mean \pm SD)	0.28 \pm 0.16

⁷ http://egp.gs.washington.edu/finished_genes.html

Table 2. Association between major principal components of CSA and CSB and number of BPDE-induced chromatid breaks per cell

Gene	Principal component	Effect per each unit of change of principal component		<i>P</i> *	Variance explained (%)	Cumulative variance explained (%)	<i>P</i> of Likelihood ratio test
		Estimate [†]	95% CI				
CSA	PC1	0.92	0.85-0.98	0.021	47.5	90.1	0.02
	PC2	0.94	0.87-1.01	0.074	25.9		
	PC3	1.04	0.97-1.12	0.300	16.7		
CSB	PC1	1.04	0.96-1.12	0.351	30.0	89.7	0.018
	PC2	1.08	1.01-1.17	0.046	22.9		
	PC3	0.94	0.87-1.01	0.105	22.1		
	PC4	0.92	0.85-0.99	0.042	14.7		

*Adjustment for gender, seeding count, and current smoking status was included in the multiple linear regression model.

[†] Estimate is ratio of mean number of breaks per cell for each unit of change of principal component.

Sensitivity analyses were conducted using only non-Hispanic White participants (75%) or only participants from the Lovelace Smokers Cohort (77%) to assess the effect of ethnicity and differences in cohorts on the results. Only results from the analysis including the entire study group of 276 participants are presented because the results were similar. All statistical analyses were done with SAS/STAT 9.1.3 and PHASE (version 2.1).

Results

The demographic characteristics of the 276 subjects are described in Table 1. The average age was 58.0 ± 9.9 years. The 276 subjects included 206 (75.0%) non-Hispanic Whites, 53 (19.0%) Hispanics, and 17 (6.0%) subjects with other ethnicities. Most samples were cryopreserved within 2 years of sample collection and had >95% viability of lymphocytes at the start of culturing. The mean number of breaks per cell was 0.28 with a median of 0.24 (interquartile range, 0.17-0.35). Among the covariates listed in Table 1, being female, being a current smoker, and low seeding count of lymphocytes ($<1.6 \times 10^6$ /culture) were associated with a 24% ($P = 0.003$), 17% ($P = 0.032$), and 31% ($P < 0.001$) increase in mean number of breaks per cell, respectively.

A principal components-based approach was used to assess the association between both CSA and CSB and the mutagen sensitivity in this study by testing whether the sequence variants in each gene were associated globally with mutagen sensitivity (Table 2). Major principal components of CSA and CSB were globally associated with the number of breaks per cell at the threshold of 80% of cumulative variance explained ($P \leq 0.02$ for both genes). Of the three principal components evaluated for CSA, PC1 was significant, whereas two (PC2 and PC4) of four principal components in CSB were significant.

Eleven SNPs in CSA and CSB were significantly associated with chromatid breaks per cell (P values = 0.004-0.039; Table 3; Supplementary Table S1).⁸ Among these 11 SNPs, the minor alleles of two SNPs were associated with an increase in mean number of breaks

per cell (effect per SNP allele, 1.14 for both SNPs), whereas the other 9 SNPs were associated with a reduction in breaks per cell (effect per SNP allele range, 0.85-0.89). SNP rs4647100 in CSA was associated with a 15% increase in the mean number of breaks per cell, the largest effect seen among the 37 SNPs. Three of the four SNPs in CSB (rs2228526, rs4253101, and rs4253049) that were significantly associated with mutagen sensitivity were in high LD with each other and were the primary factors in PC2 in the principal components analysis. Similarly, two SNPs (rs4253079 and rs4253023) were borderline significantly associated with mutagen sensitivity ($0.05 < P < 0.06$) and were the primary factors in PC4.

Haplotypes were constructed across the expanded regions in CSA and CSB in non-Hispanic Whites ($n = 206$), Hispanics ($n = 53$), and others ($n = 17$). The 13 common haplotypes in CSA and CSB with frequency >5% in at least one ethnicity are listed in Table 4. The haplotypes with frequency >5% in both non-Hispanic Whites and Hispanics were analyzed in relation to mutagen sensitivity (Table 5). The haplotype H125 of CSA was associated with a 15% reduction in the mean number of breaks per cell with all other haplotypes as the reference ($P = 0.008$), whereas the haplotype H118 in CSA was associated with a 15% increase in breaks per cell ($P = 0.020$). When the most common haplotype (H231) is used as the reference, only haplotype H125 is associated with a significant change in mutagen sensitivity ($P = 0.030$). H125 and H118 have opposite alleles for the seven significant SNPs associated with mutagen sensitivity, so an association between H125 and mutagen sensitivity may create the appearance of a reverse association for H118 when H125 is included in the reference. The haplotype H97 of CSB was associated with a 13% reduction of the breaks per cell using all other haplotypes as the reference ($P = 0.032$). Furthermore, the joint effect between CSA and CSB was evaluated based on two putative protective haplotypes (H125 and H97). A 42% reduction in the mean number of breaks per cell was seen in lymphocytes from subjects with three or four protective haplotypes compared with persons without protective haplotypes in CSA and CSB (Table 6).

A resampling-based omnibus test was employed to assess whether the associations seen with individual haplotypes and tag SNPs were the result of multiple testing. The permutation-based P values for the most significant SNPs, rs158923 in CSA and rs4253101 in CSB,

⁸ Supplementary material for this article is available at *Cancer Epidemiology, Biomarkers and Prevention* Online (<http://cebp.aacrjournals.org/>).

were 0.036 and 0.247 respectively, suggesting that the association between SNPs and haplotypes in CSA and mutagen sensitivity induced by BPDE was unlikely a chance finding.

Discussion

This study comprehensively evaluated the common coding and noncoding variation in two essential TCR genes in relation to mutagen sensitivity induced by BPDE. Our results support the conclusion that sequence variants of CSA and CSB may affect an individual's capacity to remove DNA damage induced by BPDE. Two haplotypes in CSA and CSB (H125 and H97, respectively) as well as SNPs in high LD with these two haplotypes were significantly associated with a 13% to 15% reduction in the mean number of chromatid breaks per cell.

This study has several strengths. First, both coding and noncoding common SNPs were captured by the tag

SNPs selected from a high-density SNP database, and the association testing was done based on both haplotypes and individual SNPs. This minimized the probability for false-negative results (22). The significant associations seen were evaluated by a gene-based omnibus test to minimize the family-wise type I error rate when performing multiple testing (22). Second, adjustment for gender, current smoking status, and the seeding count of lymphocytes that were the three significant nongenetic factors affecting the mutagen sensitivity induced by BPDE was included in the genetic association analysis. This allowed for a more accurate estimation of the effect of SNPs and haplotypes on mutagen sensitivity induced by BPDE. The finding of an association between the seeding count of lymphocytes and mutagen sensitivity induced by BPDE was not unexpected because lymphocytes can protect each other from the genotoxicity of BPDE that is highly reactive in culture medium. This premise is supported by the fact that the dose of BPDE added to whole blood culture is ~10-fold greater than that used for the lymphoblastoid

Table 3. Association between individual SNPs in CSA and CSB and number of BPDE-induced chromatid breaks per cell

SNP	Chromosome	Position	Gene	Allele*	Minor allele frequency	Effect per SNP allele		P [†]
						Estimate [‡]	95% CI	
rs12522154	chr5	60217257	CSA	T:C	0.33	0.88	0.79-0.97	0.012
rs4647132	chr5	60218477	CSA	C:T	0.41	1.04	0.94-1.16	0.424
rs4647128	chr5	60219301	CSA	T:C	0.04	0.88	0.67-1.17	0.388
rs4235483	chr5	60223979	CSA	C:T	0.36	0.87	0.79-0.96	0.005
rs929780	chr5	60224783	CSA	G:C	0.10	0.97	0.82-1.15	0.722
rs4647100	chr5	60236422	CSA	A:G	0.23	0.85	0.75-0.96	0.007
rs4647078	chr5	60252795	CSA	G:A	0.22	1.14	1.02-1.28	0.023
rs3797559	chr5	60259498	CSA	A:G	0.42	1.04	0.94-1.15	0.414
rs158937	chr5	60271963	CSA	G:A	0.10	0.97	0.82-1.15	0.727
rs4647028	chr5	60277583	CSA	G:T	0.23	0.85	0.75-0.95	0.006
rs7722373	chr5	60283572	CSA	G:A	0.46	1.03	0.93-1.14	0.594
rs158926	chr5	60286964	CSA	G:T	0.22	1.14	1.02-1.28	0.023
rs158923	chr5	60289598	CSA	T:C	0.36	0.87	0.78-0.96	0.004
rs4253234	chr10	50336488	CSB	G:C	0.39	1.07	0.97-1.18	0.180
rs4253226	chr10	50337451	CSB	T:C	0.18	1.11	0.97-1.26	0.117
rs3810945	chr10	50339175	CSB	C:T	0.09	1.07	0.90-1.26	0.456
rs2228526	chr10	50348723	CSB	T:C	0.19	0.88	0.79-0.99	0.039
rs4240506	chr10	50361015	CSB	C:T	0.50	0.98	0.89-1.09	0.762
rs4253126	chr10	50372196	CSB	G:T	0.10	1.13	0.96-1.33	0.144
rs7903930	chr10	50374003	CSB	G:C	0.13	1.07	0.92-1.24	0.385
rs4253101	chr10	50383065	CSB	T:G	0.19	0.87	0.77-0.98	0.019
rs4253079	chr10	50392831	CSB	T:G	0.11	0.86	0.73-1.00	0.055
rs4253077	chr10	50392975	CSB	C:A	0.11	1.07	0.92-1.26	0.374
rs2281792	chr10	50397362	CSB	T:C	0.41	1.08	0.98-1.19	0.123
rs4253049	chr10	50401891	CSB	G:A	0.18	0.88	0.78-0.99	0.038
rs7903788	chr10	50403985	CSB	G:C	0.13	1.08	0.93-1.26	0.310
rs4253026	chr10	50409122	CSB	G:A	0.09	1.08	0.91-1.28	0.352
rs4253023	chr10	50409269	CSB	C:G	0.12	0.86	0.74-1.01	0.059
rs1012554	chr10	50409563	CSB	T:C	0.29	1.09	0.98-1.22	0.119
rs4253009	chr10	50411435	CSB	C:T	0.47	0.99	0.89-1.09	0.798
rs1917800	chr10	50420923	CSB	C:G	0.11	1.11	0.95-1.28	0.186
rs6537539	chr10	50426040	CSB	T:G	0.23	0.95	0.85-1.07	0.414
rs2377902	chr10	50427510	CSB	A:G	0.41	0.89	0.81-0.98	0.022
rs10776579	chr10	50431373	CSB	C:T	0.19	1.10	0.96-1.27	0.150
rs1917822	chr10	50432753	CSB	A:G	0.43	0.96	0.86-1.07	0.420
rs2222638	chr10	50433801	CSB	G:A	0.20	1.09	0.96-1.23	0.170
rs10776581	chr10	50435466	CSB	T:C	0.44	1.01	0.92-1.11	0.841

*The latter allele is the minor one.

† Estimate is the ratio of mean number of breaks per cell between heterozygotes and common homozygotes or between rare homozygotes and heterozygotes. These two ratios were assumed to be the same because the additive model was used in the analysis.

‡ Adjustment for gender, seeding count, and current smoking status was included in the multiple linear regression model.

Table 4. Common (>5%) haplotypes and frequencies in CSA and CSB by ethnicity

Gene/haplotype	Label	Frequency (%)*			All
		Non-Hispanic White (n = 206)	Hispanic (n = 53)	Others † (n = 17)	
CSA					
TTTCGAGGGGAGT	H231	40	43	54	41
CCTTGGGAGTGGC	H125	24	23	6	23
TCTCGAAAGGGTT	H118	23	19	12	22
CCTTCAGAAGGGC	H54	9	8	21	10
TCCTGAGAGGAGC	H18	2	7	1	3
CSB					
CTCTTGGTTCCGGGCTCCTACAGT	H210	38	29	24	35
GTCCCGGGTCTAGGCTTCTGCGGT	H97	18	19	9	18
GTCTCGCTTATGCACCTGGGCAGC	H57	10	8	5	9
GTCTTGGTCTGGGGTCCGGCAGC	H64	10	6	18	10
GCCTCTGTCTGGGCCTCTATGAC	H47	9	5	9	8
GCTTCGGTCTGGGCCTCTATGAC	H39	5	14	14	7
GTCTCGCTTCCGCGCTCCTATGAC	H13	2	2	6	2
GTCCCGGGTCTGGGGTCCGGCAGC	H7	0	7	0	1

*Population haplotype frequencies were estimated from the genotype data in non-Hispanic White, Hispanic, and others separately using Bayesian statistical method implemented in the program PHASE (version 2.1).

†Others included Native American (n = 3), Black (n = 1), Asian (n = 4), and mixed ethnicity (n = 9).

cell lines to induce similar levels of chromatid breaks (10, 23).

A limitation of this study is that the 37 tag SNPs were selected based on a high-density SNPs database but not a resequencing database. The reference panel for tag SNP selection in this study has a SNP density of ~1.0 and 1.3 kb for CSB and CSA, respectively. Resequencing data are now available for CSB in the Environmental Genomic Project in a panel of 90 individuals representative of the U.S. population. Approximately 60% of the coding region in CSB was resequenced with a common SNP density of 1 SNP per 0.6 kb. Fourteen tag SNPs genotyped in the present study were identified in the Environmental Genomic Project resequencing data. These 14 tag SNPs captured 75 of 76 common SNPs identified in the 90 individuals with unknown ethnicities with $r^2 \geq 0.8$, indicating that our approach likely captured all the genetic variation for CSB.

The seven significant SNPs in CSA are located in the introns and 5'-regulatory region. With respect to the 5'-regulatory region, none of these tag SNPs reside in or around putative transcription factor binding sites (24). However, because these SNPs were tag SNPs selected by the pairwise r^2 method and there was no evidence for historical recombination across the entire expanded region of CSA, it is plausible that these SNPs may be in high LD with SNPs in transcriptional regulatory regions or microRNA binding sites in CSA. The Transfac database was searched for SNPs having a potential phenotypic effect on the expression of CSA and identified two SNPs (25). rs158915 (-4,642 with 1 being the translation starting site) and rs158920 (-377) were predicted to locate within several transcription factor binding sites. The rs158915 SNP is in perfect LD with rs158923 and rs4235483, whereas rs158920 is in perfect LD with rs158926 and rs4647078 in both non-Hispanic Whites and Hispanics. The *in silico* functional prediction for these SNPs needs to be tested *in vivo* or *in vitro*.

The finding of no association between CSB and mutagen sensitivity by the permutation test should be interpreted carefully for the following reasons. Chen

et al. (22) and Setiawan et al. (26) found that the resampling-based omnibus test might not be appropriate if multiple SNPs or haplotypes contribute to the significance of the gene. Our test based on principal components analysis that captured the LD from 24 tag SNPs within CSB found that genetic variation in CSB was globally associated with mutagen sensitivity ($P = 0.018$) and that two principal components (PC2 and PC4 in CSB; Table 2) contribute to this global association ($P < 0.05$). Furthermore, the SNP loadings for PC2 and PC4 in CSB were examined. rs2228526, rs4253101, and rs4253049 that are in high LD with each other (average $r^2 = 0.929$) contribute the most to PC2, whereas rs4253079 and rs4253023 that are also in high LD ($r^2 = 0.865$) contribute the most to PC4. These results suggest that at least two independent genetic components in CSB are associated with mutagen sensitivity. rs2228526, one of five

Table 5. Association between haplotypes in CSA and CSB and number of BPDE-induced chromatid breaks per cell

Gene	Label	Frequency	Effect per haplotype allele		
			Estimate*	95% CI	P^\dagger
CSA	H231	41	1.04	0.95-1.15	0.393
	H125	23	0.85	0.76-0.96	0.008
	H118	22	1.15	1.02-1.29	0.020
	H54	10	0.97	0.82-1.15	0.733
	Others	4	0.89	0.72-1.11	0.316
CSB	H210	35	1.08	0.97-1.19	0.152
	H97	18	0.87	0.77-0.99	0.032
	H57	9	1.14	0.96-1.35	0.133
	H64	10	0.87	0.73-1.03	0.107
	H47	8	1.15	0.96-1.38	0.132
	H39	7	1.01	0.83-1.23	0.904
	Others	13	0.95	0.82-1.10	0.484

*Estimate is the ratio of mean number of breaks per cell between any common haplotype and all other haplotypes in each gene.

† Adjustment for gender, seeding count, and current smoking status was included in the multiple linear regression model.

Table 6. Association between number of protective haplotypes in CSA and CSB and number of BPDE-induced chromatid breaks per cell

No. protective haplotypes	<i>n</i>	Chromatid breaks per cell, geometric mean (95% CI)*	<i>P</i>
0	120	0.26 (0.24-0.29)	Reference
1	103	0.22 (0.20-0.24)	0.02
2	43	0.21 (0.18-0.25)	0.044
3 or 4	10	0.15 (0.11-0.22)	0.005

*Least-squares means (95% CI) were calculated based on natural log-transformed number of chromatid breaks per cell with adjustment for gender, current smoking status, and seeding count; then, they were converted to their exponential form [geometric mean (95% CI)].

nonsynonymous SNPs in *CSB*, causes a change from methionine to valine in codon 1,097 and is in perfect LD with two other nonsynonymous SNPs [rs2228527 (R1213G) and rs2228529 (Q1413R)] in non-Hispanic Whites, Hispanics, and Asians. M1097V, R1213G, and Q1413R surround a putative nucleotide binding domain and a bipartite nuclear location signal domain COOH-terminal to the ATPase domain in *CSB* (4). Functional assessment by searching the PolyPhen database (27) revealed that M1097V and R1213G were predicted to affect the function or structure of CSA protein. Cheng et al. (28) found that the G allele of rs2228526 was associated with a 34% reduction in micronuclei frequency among 140 coke-oven workers in China. Coke-oven workers are mainly exposed to high levels of carcinogenic polycyclic aromatic hydrocarbons, including benzo(a)pyrene, the parent compound of BPDE. Superficial bladder cancer patients with the G allele of rs2228526 had worse clinical outcomes than the A allele carriers after the Bacille Calmette-Guérin treatment, indicating that the G allele may be associated with better DNA repair capacity (29).

The influence of sequence variants in *CSA* and *CSB* on the number of chromatid breaks induced by BPDE is biologically plausible. Bulky BPDE-DNA adducts in the transcribed strands of active genes can lead to blockage of transcription by human RNA polymerase II that in turn provides a recognition signal to initiate TCR, whereas DNA adducts in the nontranscribed strands are repaired via global genome NER (8, 9). In TCR, the Cockayne syndrome proteins act as coupling factors for NER proteins and chromatin remodelers to effectively prevent degradation of components of the stalled RNA polymerase II complex. This allows for elongation of transcription after removal of the damage and cleavage of the extruded mRNA without the need of transcription reinitiation (7). Several studies showed that repair of BPDE adducts in the transcribed DNA strand by TCR is much faster than repair of adducts in the nontranscribed strand by global genome NER (8, 9). Furthermore, slow repair of bulky DNA adducts along the nontranscribed strand of the human *p53* tumor suppressor gene by NER contributes to the mutation hotspots found in this gene in the lung tumors of smokers (9). Therefore, sequence variants that modify the function of *CSA* and/or *CSB* proteins may affect an individual's capacity to remove the BPDE-DNA adducts in the transcribed strand of active genes and increase the risk for genetic and epigenetic changes in tumor suppressor genes causal for lung cancer (9, 16).

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

References

- de Laat WL, Jaspers NG, Hoeijmakers JH. Molecular mechanism of nucleotide excision repair. *Genes Dev* 1999;13:768–85.
- Bootsma D, Weeda G, Vermeulen W, et al. Nucleotide excision repair syndromes: molecular basis and clinical symptoms. *Philos Trans R Soc Lond B Biol Sci* 1995;347:75–81.
- Henning KA, Li L, Iyer N, et al. The Cockayne syndrome group A gene encodes a WD repeat protein that interacts with CSB protein and a subunit of RNA polymerase II TFIIF. *Cell* 1995;82:555–64.
- Troelstra C, van Gool A, de Wit J, Vermeulen W, Bootsma D, Hoeijmakers JH. ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's syndrome and preferential repair of active genes. *Cell* 1992;71:939–53.
- Groisman R, Polanowska J, Kuraoka I, et al. The ubiquitin ligase activity in the DDB2 and CSA complexes is differentially regulated by the COP9 signalosome in response to DNA damage. *Cell* 2003;113:357–67.
- Licht CL, Stevensner T, Bohr VA. Cockayne syndrome group B cellular and biochemical functions. *Am J Hum Genet* 2003;73:1217–39.
- Fousteri M, Vermeulen W, van Zeeland AA, Mullenders LH. Cockayne syndrome A and B proteins differentially regulate recruitment of chromatin remodeling and repair factors to stalled RNA polymerase II *in vivo*. *Mol Cell* 2006;23:471–82.
- Perlow RA, Kolbanovskii A, Hingerty BE, Geacintov NE, Broyde S, Scicchitano DA. DNA adducts from a tumorigenic metabolite of benzo(a)pyrene block human RNA polymerase II elongation in a sequence- and stereochemistry-dependent manner. *J Mol Biol* 2002;321:29–47.
- Denissenko MF, Pao A, Pfeifer GP, Tang M. Slow repair of bulky DNA adducts along the nontranscribed strand of the human *p53* gene may explain the strand bias of transversion mutations in cancers. *Oncogene* 1998;16:1241–7.
- Wei Q, Gu J, Cheng L, et al. Benzo(a)pyrene diol epoxide-induced chromosomal aberrations and risk of lung cancer. *Cancer Res* 1996;56:3975–9.
- Xiong P, Hu Z, Li C, et al. *In vitro* benzo(a)pyrene diol epoxide-induced DNA damage and chromosomal aberrations in primary lymphocytes, smoking, and risk of squamous cell carcinoma of the head and neck. *Int J Cancer* 2007;121:2735–40.
- Wu X, Spitz MR, Amos CI, et al. Mutagen sensitivity has high heritability: evidence from a twin study. *Cancer Res* 2006;66:5993–6.
- Lin J, Swan GE, Shields PG, et al. Mutagen sensitivity and genetic variants in nucleotide excision repair pathway: genotype-phenotype correlation. *Cancer Epidemiol Biomarkers Prev* 2007;16:2065–71.
- Zhu Y, Yang H, Chen Q, et al. Modulation of DNA damage/DNA repair capacity by XPC polymorphisms. *DNA Repair (Amst)* 2008;7:141–8.
- Tesfaigzi Y, Myers OB, Stidley CA, et al. Genotypes in matrix metalloproteinase 9 are a risk factor for COPD. *Int J Chron Obstruct Pulmon Dis* 2006;1:267–78.
- Leng S, Stidley CA, Willink R, et al. Double-strand break damage and associated DNA repair genes predispose smokers to gene methylation. *Cancer Res* 2008;68:3049–56.
- Hsu TC, Johnston DA, Cherry LM, et al. Sensitivity to genotoxic effects of bleomycin in humans: possible relationship to environmental carcinogenesis. *Int J Cancer* 1989;43:403–9.
- Haiman CA, Hsu C, de Bakker P, et al. Comprehensive association testing of common genetic variation in DNA repair pathway genes in relationship with breast cancer risk in multiple populations. *Hum Mol Genet* 2007;17:825–34.
- Gauderman WJ, Murcray C, Gilliland F, Conti DV. Testing association between disease and multiple SNPs in a candidate gene. *Genet Epidemiol* 2007;31:383–95.
- Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001;68:978–89.

21. Stephens M, Donnelly P. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* 2003;73:1162–9.
22. Chen BE, Sakoda LC, Hsing AW, Rosenberg PS. Resampling-based multiple hypothesis testing procedures for genetic case-control association studies. *Genet Epidemiol* 2006;30:495–507.
23. Wu X, Gu J, Amos CI, Jiang H, Hong WK, Spitz MR. A parallel study of *in vitro* sensitivity to benzo[*a*]pyrene diol epoxide and bleomycin in lung carcinoma cases and controls. *Cancer* 1998;83:1118–27.
24. Conde L, Vaquerizas JM, Dopazo H, et al. PupaSuite: finding functional single nucleotide polymorphisms for large-scale genotyping purposes. *Nucleic Acids Res* 2006;34:W621–5.
25. Wingender E, Chen X, Fricke E, et al. The TRANSFAC system on gene expression regulation. *Nucleic Acids Res* 2001;29:281–3.
26. Setiawan VW, Cheng I, Stram DO, et al. A systematic assessment of common genetic variation in CYP11A and risk of breast cancer. *Cancer Res* 2006;66:12019–25.
27. PolyPhen SNP Data Collection [database available on the Internet]. Cambridge (MA): Harvard University. Available from: <http://genetics.bwh.harvard.edu/pph/data/index.html>.
28. Cheng J, Leng S, Dai Y, et al. Association between nucleotide excision repair gene polymorphisms and chromosomal damage in coke-oven workers. *Biomarkers* 2007;12:76–86.
29. Gu J, Zhao H, Dinney CP, et al. Nucleotide excision repair gene polymorphisms and recurrence after treatment for superficial bladder cancer. *Clin Cancer Res* 2005;11:1408–15.

BLOOD CANCER DISCOVERY

Association between Common Genetic Variation in *Cockayne Syndrome A* and *B* Genes and Nucleotide Excision Repair Capacity among Smokers

Shuguang Leng, Amanda Bernauer, Christine A. Stidley, et al.

Cancer Epidemiol Biomarkers Prev 2008;17:2062-2069.

Updated version Access the most recent version of this article at:
<http://cebp.aacrjournals.org/content/17/8/2062>

Cited articles This article cites 28 articles, 7 of which you can access for free at:
<http://cebp.aacrjournals.org/content/17/8/2062.full#ref-list-1>

Citing articles This article has been cited by 1 HighWire-hosted articles. Access the articles at:
<http://cebp.aacrjournals.org/content/17/8/2062.full#related-urls>

E-mail alerts [Sign up to receive free email-alerts](#) related to this article or journal.

Reprints and Subscriptions To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions To request permission to re-use all or part of this article, use this link
<http://cebp.aacrjournals.org/content/17/8/2062>.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.