Cigarette smoking increases copy number alterations in nonsmall-cell lung cancer

Cigarette smoking increases copy number alterations in nonsmall-cell lung cancer

Yen-Tsung Huang[a,b], Xihong Lin[b], Yan Liu[c], Lucian R. Chirieac[d], Ray McGovern[b], John Wain[e,f], Rebecca Heist[e,g], Vidar Skaug[h], Shanbeh Zienolddiny[h], Aage Haugen[h], Li Su[g], Edward A. Fox[i], Kwok-Kin Wong[c], and David C. Christiani[a,g,j,1]

[a]Department of Epidemiology, [b]Department of Biostatistics, and [g]Department of Environmental Health, Harvard School of Public Health, Boston, MA 02115; [c]Department of Medical Oncology and [i]Molecular Diagnostics Laboratory, The Dana-Farber Cancer Institute, Boston, MA 02115; [d]Department of Pathology, Brigham and Women's Hospital, Boston, MA 02115; [e]Cancer Center, [f]Thoracic Surgery Unit, and [j]Pulmonary and Critical Care Unit, Massachusetts General Hospital, Boston, MA 02114; and [h]Department of Biological and Chemical Working Environment, National Institute of Occupational Health, N-0033 Oslo, Norway

Edited by Wing Hung Wong, Stanford University, Stanford, CA, and approved August 11, 2011 (received for review April 05, 2011)

Cigarette smoking has been a well-established risk factor of lung cancer for decades. How smoking contributes to tumorigenesis in the lung remains not fully understood. Here we report the results of a genome-wide study of DNA copy number and smoking pack-years in a large collection of nonsmall-cell lung cancer (NSCLC) tumors. Genome-wide analyses of DNA copy number and pack-years of cigarette smoking were performed on 264 NSCLC tumors, which were divided into discovery and validation sets. The copy number-smoking associations were investigated in three scales: whole-genome, chromosome/arm, and focal regions. We found that heavy cigarette smokers (>60 pack-years) have significantly more copy number gains than non- or light smokers (≤60 pack-years) ($P = 2.46 \times 10^{-4}$), especially in 8q and 12q. Copy number losses tend to occur away from genes in non/light smokers ($P = 5.15 \times 10^{-5}$) but not in heavy smokers ($P = 0.52$). Focal copy number analyses showed that there are strong associations of copy number and cigarette smoking pack-years in 12q23 ($P = 9.69 \times 10^{-10}$) where *IGF1* (insulin-like growth factor 1) is located. All of the above analyses were tested in the discovery set and confirmed in the validation set. DNA double-strand break assays using human bronchial epithelial cell lines treated with cigarette smoke condensate were also performed, and indicated that cigarette smoke condensate leads to genome instability in human bronchial epithelial cells. We conclude that cigarette smoking leads to more copy number alterations, which may be mediated by the genome instability.

tumor genome | multimarker analyses

Lung cancer, of which 85% is nonsmall-cell lung carcinoma (NSCLC), is the second most common cancer and the leading cause of cancer-related death in the United States (1). The epidemiologic evidence supporting that cigarette smoking is an important factor in causing lung cancer was reported almost six decades ago (2–4). Moreover, lung cancer mortality mirrors trends in tobacco use (5). Carcinogens derived from cigarette smoking damage lung epithelium by oxidative stress and direct DNA damage (6). Although there has been progress in our understanding of lung carcinogenesis over the past few decades, the knowledge of mechanisms by which cigarette smoking causes lung cancer remains incomplete.

Profiles of copy number alterations (CNAs) in NSCLC have been studied (7, 8). However, the cause of copy number (CN) changes remains unknown. Several mechanisms of CN changes have been proposed, including homologous recombinations and nonhomologous mechanisms (9, 10). Bacteria, yeast, and human seem to share similar mechanisms (10). In bacteria, CNAs can be induced by environmental stress to enable swifter evolution in response to such stress. In the cell population within a tumor or precancerous lesion, similar stress, such as hypoxia, may induce CN changes. Thus, it is plausible to hypothesize that cigarette smoking serves as an environmental stress on the cells that leads to tumorigenesis by means of CNAs.

Using the tumor cells separated from malignant pleural effusions, it was found that gains of 11p were more frequent in smoking men than nonsmoking men (11). Furthermore, another study identified a CN-based genomic signature in resected lung tumors for current smokers compared with never smokers (12). However, these studies had significant limitations. First, discrete smoking status (smokers vs. nonsmokers) may not be an optimal indicator to capture the dose–response relationship between cigarette smoking and CN changes. Second, smoking may have different implications on CN, depending on whether it induces gains or losses. Third, the conclusions in the previous studies were drawn based on modest sample sizes. Finally, none of previous studies provided a biological explanation on how cigarette smoking causes CNAs. To better investigate the relationship between cigarette smoking and CNAs, we conducted a genome-wide study of CNs and smoking pack-years in a large collection of resected NSCLC tumors. Our analyses cover the association of cigarette smoking with CNs on three different scales: whole-genome, chromosome/arm, and focal CNs. The causal mechanism behind such smoking–CNA association was further explored in a human nontumorigenic bronchial cell line.

Results

A total of 264 subjects were randomly divided into two datasets: discovery and validation sets. The characteristics of the populations are similar (Table 1), indicating the balance of the two sets. Two alternative data splittings were pursued to prevent the possibility that the results presented here are simply because of chance or multiple comparisons. (*SI Appendix*, Tables S1–S3) To account for batch effects, we also performed batch-adjusted analyses by normalization and explicitly adjusting for the batch identity as a covariate in the regression. The batch effect-adjusted analyses showed similar patterns to those without adjustment. (*SI Appendix*, Fig. S1 and Tables S4 and S5) The analyses of smoking vs. CN associations are outlined as three parts: on the genome-wide scale, on the chromosome/arm-specific scale, and on the focal-region scale.

Cigarette Smoking and Whole-Genome CN Pattern. There is a significant increase in total events of CN gains among heavy smokers (>60 pack-years) ($P = 0.0080$, 0.0095, and 2.5×10^{-4} for

Author contributions: Y.-T.H., X.L., Y.L., J.W., R.H., V.S., S.Z., A.H., K.-K.W., and D.C.C. designed research; Y.-T.H., X.L., Y.L., L.R.C., J.W., R.H., V.S., S.Z., A.H., L.S., E.A.F., K.-K.W., and D.C.C. performed research; Y.-T.H., Y.L., L.S., E.A.F., K.-K.W., and D.C.C. contributed new reagents/analytic tools; Y.-T.H., X.L., Y.L., R.M., K.-K.W., and D.C.C. analyzed data; and Y.-T.H., X.L., Y.L., L.R.C., R.M., J.W., R.H., V.S., S.Z., A.H., L.S., E.A.F., K.-K.W., and D.C.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: dchris@hsph.harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1102769108/-/DCSupplemental.

Table 1. Characteristics of study populations

	Discovery set	Validation set	*P* value*
Sample size	134	130	
Male (%)	65.67	56.92	0.18
Age			
Mean ± SD	67.27 ± 8.17	67.59 ± 8.39	0.75
Cigarette smoking pack-years			
Median ± interquartile range	34.25 ± 39.64	38 ± 35.93	0.28
Clinical stage			0.43
Stage 1 (%)	77.27	70.00	
Stage 2 (%)	15.15	19.23	
Stage 3 or 4 (%)	7.58	10.77	
Cigarette smoking status			0.23
Never smokers (%)	7.46	6.15	
Ex-smokers (%)	43.28	53.85	
Current smokers (%)	49.25	40.00	
Adenocarcinoma (%)	67.91	64.62	0.66

**P* values were calculated with X^2 test for percentage of male (1 degree of freedom, d.f.), adenocarcinoma, patients from MGH (1 d.f.), clinical stage (2 d.f.), and cigarette smoking status (2 d.f.); with *t* test for age, and with Wilcoxon test for cigarette smoking pack-years.

discovery, validation, and both sets, respectively), but no difference in CN losses (Fig. 1 *A* and *B*). No significant difference was observed in age, clinical stage, histology, and sex between heavy and light or nonsmokers.

For CN losses, G/T ratios in light/nonsmokers (≤60 pack-years) are significantly lower than the null ratio (i.e., the ratio when CNAs occur at random with respect to the gene location) ($P = 0.011$, 9.80×10^{-4}, and 5.15×10^{-5} for discovery, validation, and both sets, respectively) but heavy smokers (>60 pack-years) show no difference ($P = 0.78$, 0.31, and 0.52, respectively) (Fig. 1 *C* and *D*). These results suggest that CN losses tend to occur away from genes, but such tendency disappears in heavy smokers. In contrast, there is no consistent pattern for CN gains. Heavy smokers seem to have more genes with CN changes, especially in gains. (*SI Appendix*, Fig. S2).

Cigarette Smoking and CN Pattern by Chromosome/Arm. The chromosome/arm-specific analyses suggest the majority responsible for the genome-wide difference comes from chromosomes 8q ($P = 1.19 \times 10^{-5}$ for total events of CN gains between light and heavy smokers) and 12q ($P = 2.1 \times 10^{-4}$) (*SI Appendix*, Fig. S3*A*), as well as many others (chromosomes 1, 3, 7, 10, 11, 16, and 17) (*SI Appendix*, Fig. S4). Similar results were observed when genomic location was taken into account, especially in 8q and 12q. (*SI Appendix*, Fig. S5) The dose–response relationships between continuous CNs and smoking pack-years are also significant in 8q ($P = 0.015$) and 12q ($P = 0.0025$). (*SI Appendix*, Fig. S3*B*) These two regions also found the most signals in focal CN analyses, as will be shown in the following section.

Cigarette Smoking and Focal CNs. As stated in *Materials and Methods*, we performed single- and multiple-marker analyses to investigate the association of cigarette smoking and focal CNs. In the moving-window 10-marker analyses, we selected the top 50 sets with smallest *P* values in the discovery set and tested the 50 sets using the validation set ($P < 0.05$). Using such criteria, we identified one 10-marker set in 12q23 with *P* values of 9.69×10^{-10}, which reached the genome-wide significance (Fig. 2*A*). The region harbors a gene, *IGF1* (insulin-like growth factor 1), that plays an important role in tumorigenesis (Fig. 2*B*). In the single-marker analyses, the most significant signals are also in the same region of 12q23: two loci are in the intron between the last two exons of *IGF1* and two loci are located downstream of *IGF1*. (*SI Appendix*, Fig. S6 *A* and *B*, and Table S6) The *P* value of the 10-marker set and the corresponding *P* values and R^2 from the single-marker analyses are shown in Table 2. Compared with the single-marker analyses, statistical power was gained from the 10-marker analyses by borrowing information in the neighboring markers, accounting for correlation among the CN variation in the marker set, reducing degrees-of-freedom of the test, and reducing the total number of tests across the genome.

The dose–response relationship of CN and smoking pack-years for the four loci in 12q23.2 is shown in *SI Appendix*, Fig. S6 *C–F*, indicating a J-shaped curve. That is, beyond a certain threshold, the higher the smoking pack-years, the more departure from the neutral CN. Notably, the threshold, about 60 pack-years, is consistent with the cut-off used in the above analyses of the whole-genome CNA pattern.

In addition to 12q23, 3q24 and 8q24 are two additional regions that are potentially associated with the pack-years of cigarette smoking from single-marker analyses (*SI Appendix*, Figs. S7 and S8). We also performed the analyses in the dichotomous version, detail of which can be found in *SI Appendix*, Fig. S9 and Table S7.

DNA Double-Strand Break Assay. To investigate further the results of our statistical analyses, we determined whether cigarette smoke could induce DNA double-stand breaks in cultured cells. To mimic longer and heavier cigarette smoking conditions, we treated human nontumorigenic bronchial epithelial cell HBEC 3KT with 0.04 and 0.4 μg/mL cigarette smoke condensate (CSC) for 24 h. Under these conditions, the survival rates are 96.9% and 95.7%, respectively, indicating the dose of CSC and the length of treatment used in this study are not toxic to the cells (Fig. 3*A*). To minimize background DNA double-strand breakage, we treated cells with CSC right after the growth had reached confluence. Under these conditions, ∼5% of non-CSC–treated cells still display double-strand breaks (Fig. 3*C*). When treated with 0.04 μg/mL CSC for 24 h, the percentage of cells with double-strand breaks increased to 15%. This percentage doubled with the application of more concentrated 0.4 μg/mL CSC (Fig. 3*C*). We also treated the cells with 0.4 μg/mL CSC for 2 h, and observed a similar DNA double-strand break ratio as that of the non-CSC–treated control cells, suggesting a DNA double-strand break occurring after a longer time of CSC treatment.

To determine the effects of CSC on induction of cellular apoptosis, which indirectly contributes to DNA double-strand breaks, the same set of cells (as used in Fig. 3 *B* and *C*) were lysed for apoptotic-specific Caspase-3/7 activity. As shown in Fig. 3*D*, there is a basal level of Caspase-3/7 activity in non-CSC–treatment cells. Upon CSC treatment, the value of relative fluorescence unit (RFU) increased in a dose-dependent matter. However, the extent to which the RFU value increased in response to CSC treatment was much less than the corresponding increase in DNA double-strand breaks in Fig. 3*C*. Collectively, these results indicate that higher CSC leads to genome instability in bronchial epithelial cells. As such, theses data provide biological evidence to bridge the associations between CNAs and smoking observed in the above human data.

Discussion

We show that heavy smokers (>60 pack-years) have more CN gains than light/nonsmokers but not CN losses, and that light/nonsmokers (≤60 pack-years) have CN losses away from the gene location, in contrast to heavy smokers. The discrepancy between gains and losses suggests that different mechanisms may exist for the genome impact of cigarette smoking. For gains, smoking executes its oncogenic effect by increasing the event of CN changes. In contrast, for losses smoking does not increase CNA events but increases the proportion of genes being affected. Because losing a fragment of DNA is less favorable than gaining one

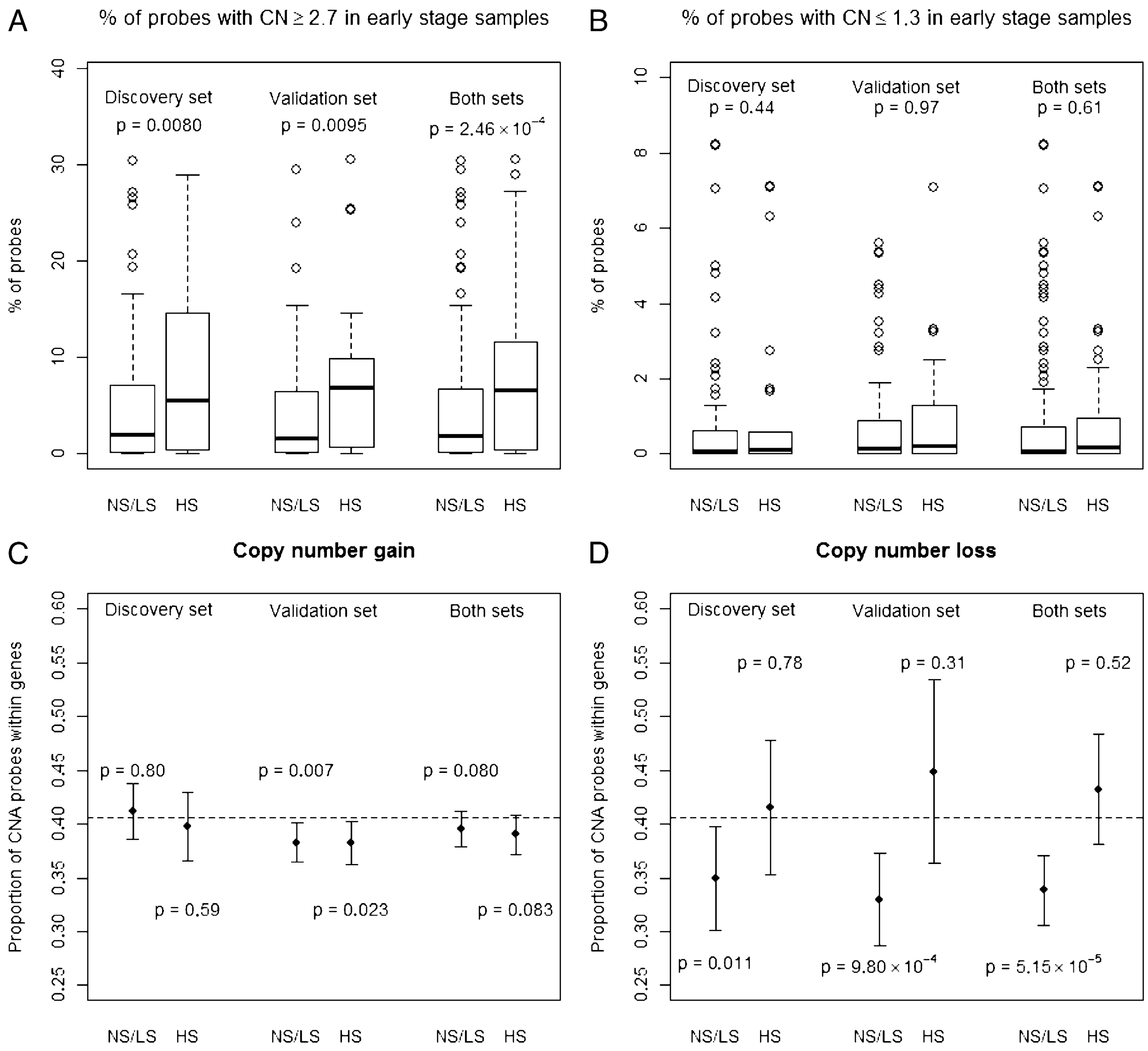

Fig. 1. Association of cigarette smoking and whole-genome CNs. (*A* and *B*) Among the 256,554 total probes, the proportion (%) with CNA (*A*, gains; *B*, losses) events by pack-years of cigarette smoking (NS/LS, non/light smokers; HS, heavy smokers). (*C* and *D*) Mean and its 95% confidence interval of G/T ratios in the HS and NS/LS for CN gains (*C*) and losses (*D*); the dashed lines represent the null G/T ratio on the chip (104,256/256,554 = 40.64%). *P* values were used to test the indicated indices between HS and NS/LS with methods described in *Materials and Methods* and *SI Appendix*.

(13), two separate mechanisms may be developed to hit the genes responsible for tumorigenesis. The phenomenon can be a consequence of selection during cancer development and cell proliferation. Because different cells possess different CNAs, selection by a nutrient-limited environment makes those clones that can grow without regulatory control become dominant.

For heavy smokers, there were more CN gains compared with non- or light smokers and no tendency for CN losses to occur away from the gene location. We have also found that genes with gains are more likely to be oncogenes or to be involved in pathways that are associated with tumor growth, which suggests that lung cancer cells in heavy smokers tend to acquire the growth advantage via CN gains (14). As a result, CN losses within genes have less unfavorable impact on such cells because it is compensated by the fact that they can grow without regulation. This finding explains our observation that the proportion of losses within gene among heavy smokers is not different from that at random.

Previous studies have shown that CNAs are more frequent in smokers than in nonsmokers (11, 12), consistent with our findings based on pack-years. CN-based genomic signature has also been identified to discriminate current smokers and never smokers (12), which, however, does not include *IGF1*. Smoking status may not necessarily reflect the same oncogenic feature as pack-years of smoking, a measure of cumulative exposure. Furthermore, the large sample size and discovery-validation process in this study increases the robustness of the findings.

Smoking causes lung cancer through numerous carcinogens derived from cigarette combustion. There are two parts of the carcinogenic effect: early damage of oxidative stress by reactive oxygen species and late damage by DNA adduct and DNA mutation (6). Both kinds of damage can serve as initiators of CN changes, especially oxidative stress. It has recently been proposed that cellular stress coming from environmental agents can induce CN changes (10).

The most significant region on 12q23 is at the junction of the last two exons and the downstream of *IGF1*. The two loci within *IGF1* are located in the intron between the last two exons of *IGF1*. The protein product of the aberrant genomic DNA can exert its undue influence on the cellular physiology. On the other hand, if the key player is the downstream rather than the coding region of *IGF1*, it

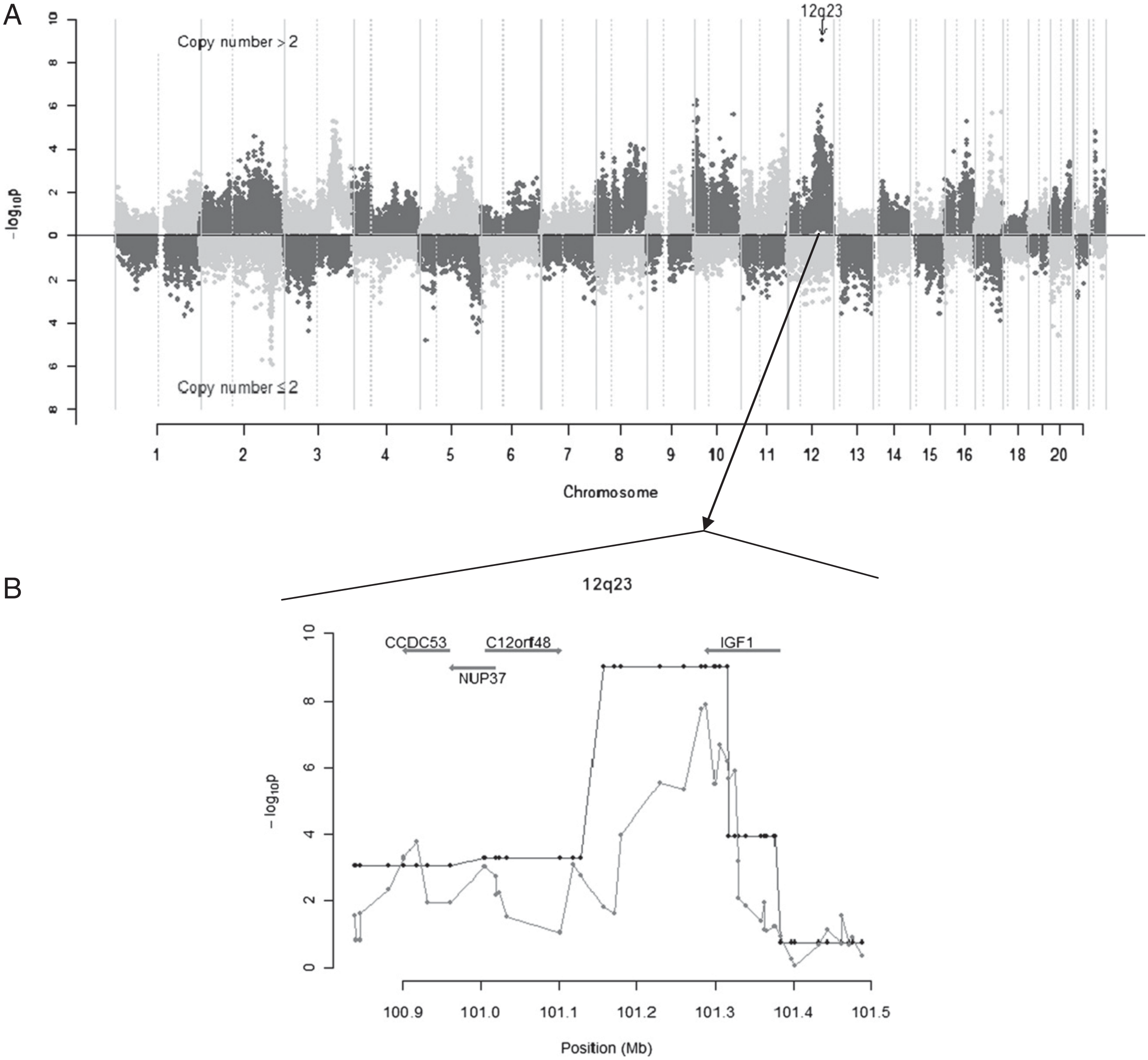

Fig. 2. Association of cigarette smoking and 25,655 moving-window 10-marker focal CNs. (*A*) A plot of -$\log_{10}$P of the association between smoking pack-years and the 10-marker set focal CN, which is analyzed for CN > 2 (*Upper*) and ≤ 2 (*Lower*), separately. (*B*) *P* values of focal CN analyses in 12q23. The black dots and line indicate *P* values from 10-marker analyses, and the superimposed gray dots and line indicate the corresponding ones from single-marker analyses.

is still possible that *IGF1* function is affected because the downstream fragment can serve as a regulatory element of *IGF1* transcription. That is, the alterations of the regulatory element can lead to the abnormal gene expression of *IGF1*.

IGF1/IGF1R signaling pathway can induce many effects, including cell proliferation, differentiation, transformation, and inhibition of apoptosis (15). Because of the overlap with downstream signaling pathways of EGFR signaling, IGF1/IGF1R signaling may modulate the EGFR pathway, a critical pathway in lung tumorigenesis (16), and it may explain, in part, clinical resistances to EGFR inhibitors (17).

Several studies have provided the links among smoking, IGF1, and cancer. For example, it has been reported that smoking may affect IGF1 serum level and its signaling (18, 19). On the other hand, IGF1 and the risk of developing cancer have also been extensively studied in lung cancer (20–23), breast cancer (24), prostate cancer (25), and colorectal cancer (26). Our analysis supports the hypothesis that smoking can act through increasing the CN of *IGF1* to induce its overexpression and subsequent oncogenesis.

Materials and Methods

Study Population, Specimens, and Data Collection. A series of 264 snap-frozen tumor samples from NSCLC patients with complete information on cigarette smoking was collected during surgery or biopsy from the Massachusetts General Hospital (MGH), Boston, MA, and the National Institute of Occupational Health, Oslo, Norway. An additional 50 paired specimens of nonneoplastic lung parenchyma and 63 paired blood samples were included as the reference group for CN estimation. Demographic and smoking information was collected by a trained research assistant using a modified standardized American Thoracic Society respiratory questionnaire (27). A similar approach was used for the Norwegian cohort (28). Written informed consent was obtained from all patients. The study was approved by the institutional review boards of MGH, the Harvard School of Public Health, and the Norwegian Data Inspectorate, and Local Regional Committee for Medical Research Ethics.

DNA Quality, Histopathology, and Genechip. DNA was extracted from tumor and nonneoplastic lung parenchyma after manual microdissection from 5-μm

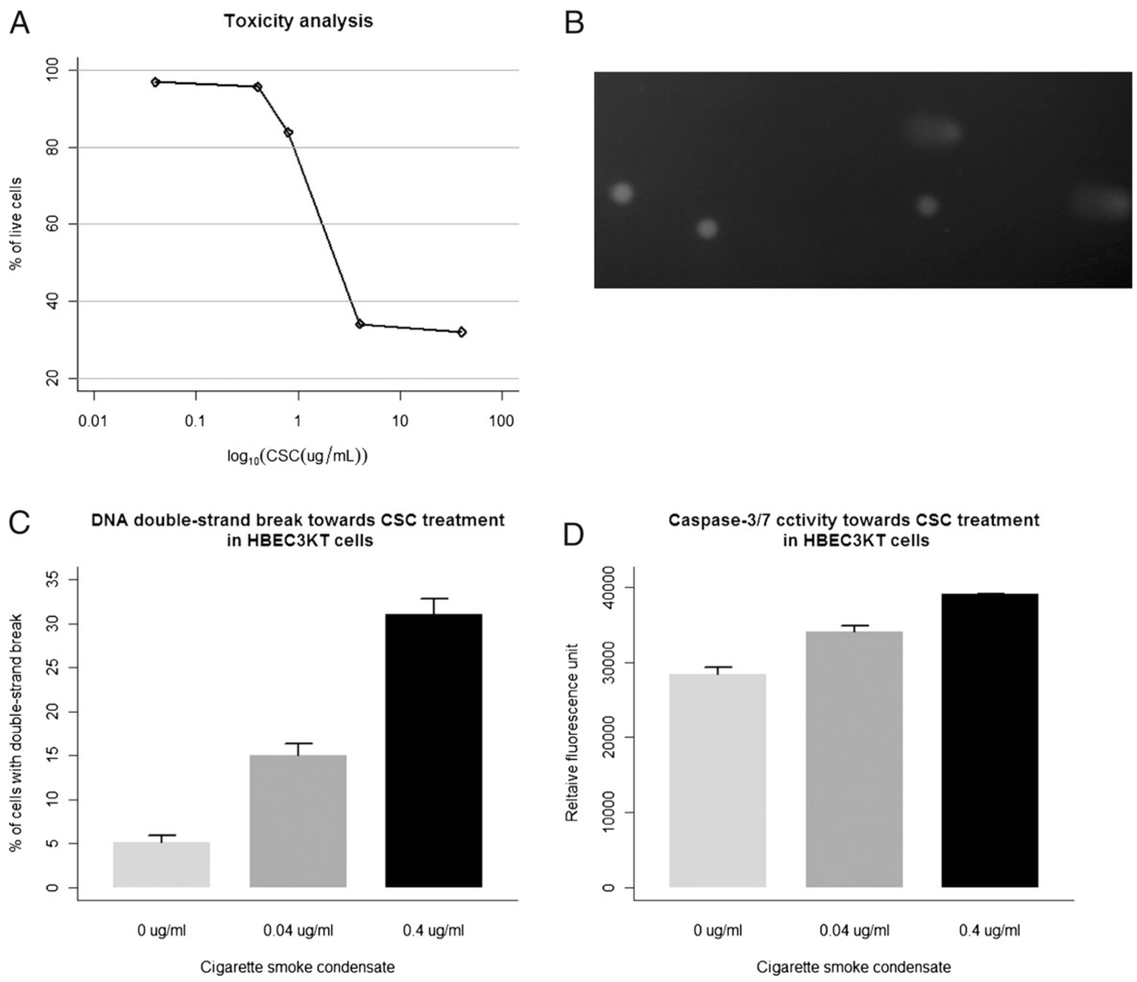

Fig. 3. Effects of CSC treatment on HBEC 3KT DNA double-stand breaks and apoptosis. (*A*) Cytotoxic effect of CSC on HBEC3KT survival. HBEC3KT cells were cultured in 12-well plates to confluence and then incubated with 0, 0.04, 0.4, 0.8, 4, and 40 μg/mL CSC for 24 h. Viable cells were monitored with MTT assay. Live cells treated with 0 μg/mL CSC were defined as 100%. The percentage of live cells vs. CSC concentration was plotted. (*B* and *C*) CSC treatment induces DNA single/double-stand breaks in HBEC 3KT cells. HBEC 3KT cells were cultured to confluence in 100-mm plates and then treated with 0, 0.04, and 0.4 μg/mL CSC for 24 h. Cells were harvested by trypsinization and DNA single/double-stand breaks were analyzed by neutral comet assay. (*B*) A representative photo with undamaged-DNA (bright dot) and DNA with single/double-strand breaks (bright dot with an elongated tail) was shown. (*C*) About 600 to ~800 cells per treatment were viewed, and percentage of DNA single/double-strand breaks vs. CSC dose was plotted. (*D*) CSC treatment induces apoptosis in HBEC 3KT cells. The same set of cells used in *B* and *C* was also analyzed for caspase-3/7 activity. Columns are mean value of RFU. A larger RFU value represents a higher caspase-3/7 activity and thus a stronger apoptotic response. SDs are provided in *C* and *D*.

thick histopathologic sections. Each specimen was evaluated for amount and quality of tumor cells. Tumors were reviewed and classified using the World Health Organization criteria. Specimens with lower than 70% tumor cellularity, inadequate DNA concentration, or not intact genomic DNA were not included for chip hybridization. The platform of genechip is Affymetrix 250K Nsp GeneChip.

Data Preprocessing. CNs were obtained with dChip software by invariant set normalization and median smoothing with the window of 11 loci (29). Only 256,554 probes on somatic chromosomes were analyzed. We further classified the continuous inferred CN into a discrete variable of CNAs: CN gains defined as CNs ≥ 2.7 and CN losses defined as CNs ≤ 1.3, to detect CN ≥ 3 and ≤ 1 by tolerating 30% normal tissue contamination. The probes were mapped to the RefSeq genes with a 2-kb extension both upstream and downstream using the University of California at Santa Cruz Genome Browser. Among the 256,554 probes on somatic chromosomes, 104,256 (40.64%) were mapped to 11,700 genes. The copy number data of the NSCLC tumors reported in this paper are available at http://www.hsph.harvard.edu/~xlin/data.html.

Statistical Analysis. Only early-stage tumors were analyzed here because we have found that late-stage tumors have more CNAs. The number of pack-years is defined as the packs of cigarette smoked per day multiplied by the years of smoking. Sixty pack-years of cigarette smoking was chosen as the cut-off for heavy and light/nonsmokers according to the observation of total CNA events by the interval of 10 pack-years in both discovery and validation sets (*SI Appendix*, Fig. S10). Using the cut-off, we had 203 light/nonsmokers and 61 heavy smokers. We developed three methods to test the genome-wide or chromosome/arm-specific CN patterns between heavy and light/nonsmokers and one method to test the association of the chromosome/arm-specific or focal CNs and smoking pack-years (details in *SI Appendix*).

We used another method to investigate gene selection of CNAs between heavy and light/nonsmokers. Both the total probes (T) in which CNAs were detected and the probes locating within genes (G) in which CNAs were detected were calculated for each individual. We proposed a ratio of G vs. T (termed as G/T ratio) to estimate the selection of CNAs with respect to the gene location (14). Under the null hypothesis that CNAs occur randomly relative to where genes locate, we would expect the null ratio of 40.64% (104,256/256,554), where 104,256 is the number of probes located within genes on the chip. By comparing the G/T ratios to the null ratio, 40.64%, with the two-sided *t* test, we were able to test whether CNAs occur preferentially away from genes.

For single-marker analyses, 256,554 regressions for both CN > 2 and ≤ 2 were performed in the discovery set, with continuous CN at each locus as

Table 2. Summary of the 10 candidate loci at 12q23 from both 10-marker and single-marker analyses

					Focal CN-smoking association					
					10-Marker analyses			Single-marker analyses		
Affy ID	dbSNP	Cyto-band	Position (Mb)	Gene	*P* value, discovery set	*P* value, validation set	*P* value, pooled	*P* value, pooled	R^2	*P* value, adjusted*
SNP_A-2002985	rs5011687	12q23	101.157	—	3.17×10^{-8}	0.0291	9.69×10^{-10}	0.0152	0.065	0.0168
SNP_A-2125858	rs17439974	12q23	101.171	—				0.0239	0.060	0.0285
SNP_A-4222341	rs17032384	12q23	101.179	—				0.000110	0.126	0.000263
SNP_A-1899321	rs1520223	12q23	101.229	—				2.92×10^{-6}	0.175	8.07×10^{-6}
SNP_A-4222344	rs4764695	12q23	101.260	—				4.55×10^{-6}	0.167	1.33×10^{-5}
SNP_A-4228436	rs10860860	12q23	101.283	—				1.79×10^{-8}	0.223	9.78×10^{-8}
SNP_A-2106083	rs2946831	12q23	101.289	—				1.29×10^{-8}	0.235	2.63×10^{-8}
SNP_A-2255731	rs10745940	12q23	101.300	*IGF1*				3.26×10^{-6}	0.163	8.20×10^{-6}
SNP_A-2092658	rs9308315	12q23	101.306	*IGF1*				2.10×10^{-7}	0.202	7.74×10^{-7}
SNP_A-2271065	rs2072592	12q23	101.316	*IGF1*				6.17×10^{-7}	0.200	4.42×10^{-6}

**P* values of smoking pack-years were calculated from linear models, with up to quadratic term of square root-transformed smoking pack-years, adjusting for age, sex, clinical stage, and cell type.

a dependent variable and square root of smoking pack-years, and its quadratic term as independent covariates. For the validated candidates ($P < 0.05$ in the validation set), pooled results were generated with linear regressions (with up to the quadratic term of the square root of smoking pack-years), spline regressions (with spline of the square root of smoking pack-years), and locally weighted scatter plot smoothing (LOWESS). Adjusted linear and spline regressions were performed with adjustment of age at diagnosis, sex, two cohorts, clinical stage, and histology.

DNA Double-Strand Break Assay. For the cytotoxicity analysis of cigarette smoke condensate, a human nontumorigenic bronchial epithelial cell line HBEC 3KT was cultured in 12-well plates to confluence and then treated with the indicated concentration of CSCfor 24 h. Viable cells were monitored by MTT assay using the CellTiter 96 AQueous One Solution Cell Proliferation Assay kit (Promega). All assays were performed in triplicate. For the neutral comet assay, HBEC 3KT was cultured in 100-mm plates to confluence and then treated with indicated concentration of CSC for 24 h. Cells having DNA double-stand break were analyzed by Neutral comet assay using CometAssay kit (Trevigen). About 600 to ~800 cells were viewed per treatment. For the apoptosis analysis, HBEC 3KT was cultured in 100-mm plates to confluence and then treated with indicated concentration of CSC for 24 h. The status of cellular apoptosis was determined using SendoLyteTM Homogeneous Rh110 Caspase-3/7 Assay kit (Anaspec). All apoptosis assays were performed in triplicate.

ACKNOWLEDGMENTS. We thank the participants of the Molecular and Genetic Analysis of Lung Cancer Study; Dr. Cheng Li for his advice on data preprocessing; the Harvard School of Public Health Lung Cancer Study Group: Dr. Kofi Asomaning, Dr. Eugene Mark, Dr. Matthew Kulke, Dr. Wei Zhou, Dr. Geoffrey Liu, Marcia Chertok, Andrea Shafer, Lauren Cassidy, Maureen Convery, and Salvatore Mucci; Drs. Panos Fidias and Bruce A. Chabner and the physicians and surgeons of the Massachusetts General Hospital Cancer Center; and Dr. Lodve Stangeland for recruiting patients. This study is supported by National Institutes of Health Grants CA076404 and CA134294 (to Y.-T.H. and X.L.); National Institutes of Health Grants CA092824, CA074386, and CA090578 (to D.C.C.) and CA122794, CA140594, and CA141576 (to K.K.W.); the Raymond P. Lavietes Foundation (D.C.C.); and the Norwegian Cancer Society (A.H.).

1. Jemal A, et al. (2009) Cancer statistics, 2009. *CA Cancer J Clin* 59:225–249.
2. Bach PB (2009) Smoking as a factor in causing lung cancer. *JAMA* 301:539–541.
3. Wynder EL, Graham EA (1950) Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma; A study of 684 proved cases. *J Am Med Assoc* 143:329–336.
4. Doll R, Hill AB (1954) The mortality of doctors in relation to their smoking habits; a preliminary report. *BMJ* 1:1451–1455.
5. Dubey S, Powell CA (2009) Update in lung cancer 2008. *Am J Respir Crit Care Med* 179: 860–868.
6. Alavanja MC (2002) Biologic damage resulting from exposure to tobacco smoke and from radon: Implication for preventive interventions. *Oncogene* 21:7365–7375.
7. Kim TM, et al. (2005) Genome-wide screening of genomic alterations and their clinicopathologic implications in non-small cell lung cancers. *Clin Cancer Res* 11: 8235–8242.
8. Weir BA, et al. (2007) Characterizing the cancer genome in lung adenocarcinoma. *Nature* 450:893–898.
9. van Gent DC, van der Burg M (2007) Non-homologous end-joining, a sticky affair. *Oncogene* 26:7731–7740.
10. Hastings PJ, Lupski JR, Rosenberg SM, Ira G (2009) Mechanisms of change in gene copy number. *Nat Rev Genet* 10:551–564.
11. Yen CC, et al. (2007) Chromosomal aberrations of malignant pleural effusions of lung adenocarcinoma: Different cytogenetic changes are correlated with genders and smoking habits. *Lung Cancer* 57:292–301.
12. Massion PP, et al. (2008) Smoking-related genomic signatures in non-small cell lung cancer. *Am J Respir Crit Care Med* 178:1164–1172.
13. Emerson JJ, Cardoso-Moreira M, Borevitz JO, Long M (2008) Natural selection shapes genome-wide patterns of copy-number polymorphism in *Drosophila* melanogaster. *Science* 320:1629–1631.
14. Huang Y-T, et al. (2011) Impact on disease development, genomic location and biological function of copy number alterations in non-small cell lung cancer. *PLoS ONE* 6:e22961.
15. Fürstenberger G, Senn HJ (2002) Insulin-like growth factors and cancer. *Lancet Oncol* 3:298–302.
16. Herbst RS, Heymach JV, Lippman SM (2008) Lung cancer. *N Engl J Med* 359:1367–1380.
17. Morgillo F, et al. (2007) Implication of the insulin-like growth factor-IR pathway in the resistance of non-small cell lung cancer cells to treatment with gefitinib. *Clin Cancer Res* 13:2795–2803.
18. Kaklamani VG, Linos A, Kaklamani E, Markaki I, Mantzoros C (1999) Age, sex, and smoking are predictors of circulating insulin-like growth factor 1 and insulin-like growth factor-binding protein 3. *J Clin Oncol* 17:813–817.
19. Tannheimer SL, Ethier SP, Caldwell KK, Burchiel SW (1998) Benzo[a]pyrene- and TCDD-induced alterations in tyrosine phosphorylation and insulin-like growth factor signaling pathways in the MCF-10A human mammary epithelial cell line. *Carcinogenesis* 19:1291–1297.
20. Yu H, et al. (1999) Plasma levels of insulin-like growth factor-I and lung cancer risk: A case-control analysis. *J Natl Cancer Inst* 91:151–156.
21. Wu X, Yu H, Amos CI, Hong WK, Spitz MR (2000) Joint effect of insulin-like growth factors and mutagen sensitivity in lung cancer risk. *J Natl Cancer Inst* 92:737–743.
22. Lukanova A, et al. (2001) A prospective study of insulin-like growth factor-I, IGF-binding proteins-1, -2 and -3 and lung cancer risk in women. *Int J Cancer* 92:888–892.
23. Spitz MR, et al. (2002) Serum insulin-like growth factor (IGF) and IGF-binding protein levels and risk of lung cancer: A case-control study nested in the beta-Carotene and Retinol Efficacy Trial Cohort. *Cancer Epidemiol Biomarkers Prev* 11:1413–1418.
24. Hankinson SE, et al. (1998) Circulating concentrations of insulin-like growth factor-I and risk of breast cancer. *Lancet* 351:1393–1396.
25. Chan JM, et al. (1998) Plasma insulin-like growth factor-I and prostate cancer risk: A prospective study. *Science* 279:563–566.
26. Ma J, et al. (1999) Prospective study of colorectal cancer risk in men and plasma levels of insulin-like growth factor (IGF)-I and IGF-binding protein-3. *J Natl Cancer Inst* 91: 620–625.
27. Zhou W, et al. (2006) Second hand smoke exposure and survival in early-stage non-small-cell lung cancer patients. *Clin Cancer Res* 12:7187–7193.
28. Zienolddiny S, et al. (2008) A comprehensive analysis of phase I and phase II metabolism gene polymorphisms and risk of non-small cell lung cancer in smokers. *Carcinogenesis* 29:1164–1169.
29. Zhao X, et al. (2004) An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. *Cancer Res* 64: 3060–3071.