Original research

Elevated Alu retroelement copy number among workers exposed to diesel engine exhaust

Jason Y.Y. Wong ,[1] Richard Cawthon,[2] Yufei Dai,[3] Roel Vermeulen ,[4] Bryan A. Bassig,[1] Wei Hu,[1] Huawei Duan ,[3] Yong Niu,[3] George S. Downward ,[4] Shuguang Leng,[5] Bu-Tian Ji,[1] Wei Fu,[6] Jun Xu,[7] Kees Meliefste,[4] Baosen Zhou,[8] Jufang Yang,[6] Dianzhi Ren,[6] Meng Ye,[3] Xiaowei Jia,[3] Tao Meng,[3] Ping Bin,[3] H. Dean Hosgood, III,[9] Debra T. Silverman,[1] Nathaniel Rothman,[1] Yuxin Zheng,[3] Qing Lan[1]

For numbered affiliations see end of article.

Correspondence to
Dr Jason Y.Y. Wong, Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland, USA; jason.wong@nih.gov

Received 6 February 2021
Revised 3 May 2021
Accepted 7 May 2021
Published Online First 26 May 2021

ABSTRACT

Background Millions of workers worldwide are exposed to diesel engine exhaust (DEE), a known genotoxic carcinogen. Alu retroelements are repetitive DNA sequences that can multiply and compromise genomic stability. There is some evidence linking altered Alu repeats to cancer and elevated mortality risks. However, whether Alu repeats are influenced by environmental pollutants is unexplored. In an occupational setting with high DEE exposure levels, we investigated associations with Alu repeat copy number.

Methods A cross-sectional study of 54 male DEE-exposed workers from an engine testing facility and a comparison group of 55 male unexposed controls was conducted in China. Personal air samples were assessed for elemental carbon, a DEE surrogate, using NIOSH Method 5040. Quantitative PCR (qPCR) was used to measure Alu repeat copy number relative to albumin (Alb) single-gene copy number in leucocyte DNA. The unitless Alu/Alb ratio reflects the average quantity of Alu repeats per cell. Linear regression models adjusted for age and smoking status were used to estimate relations between DEE-exposed workers versus unexposed controls, DEE tertiles (6.1–39.0, 39.1–54.5 and 54.6–107.7 μg/m^3) and Alu/Alb ratio.

Results DEE-exposed workers had a higher average Alu/Alb ratio than the unexposed controls (p=0.03). Further, we found a positive exposure–response relationship (p=0.02). The Alu/Alb ratio was highest among workers exposed to the top tertile of DEE versus the unexposed controls (1.12±0.08 SD vs 1.06±0.07 SD, p=0.01).

Conclusion Our findings suggest that DEE exposure may contribute to genomic instability. Further investigations of environmental pollutants, Alu copy number and carcinogenesis are warranted.

Key messages

What is already known about this subject?

- Alu retroelements are repetitive mobile DNA sequences that can multiply and 'jump' throughout the human genome, which can potentially influence gene expression, genomic architecture and health outcomes including cancer. DNA damage from ionising radiation and chemotherapy was previously found to promote increased Alu retrotransposition; however, whether genotoxic air pollution can produce similar effects in humans is unknown.

What are the new findings?

- In a cross-sectional study of workers exposed to diesel engine exhaust and unexposed controls, we found that increased diesel exhaust exposure was associated with increased Alu copy number, a marker that could reflect genomic instability.

How might this impact on policy or clinical practice in the foreseeable future?

- Our findings of the detrimental effects of diesel exhaust on genomic stability reinforce the need for policies aimed at limiting occupational exposure.

INTRODUCTION

Diesel engine exhaust (DEE) is a known carcinogen based on considerable evidence in humans.[1] These emissions contain genotoxic constituents such as nitro-polycyclic aromatic hydrocarbons (nitro-PAHs) and fine particulate matter ($PM_{2.5}$) that are produced by diesel internal combustion engines that commonly power automobiles, aircraft, locomotives, mining and construction equipment and electric generators. Due to the widespread use of diesel engines in transportation and industrial settings, millions of workers around the world are exposed to DEE emissions,[2 3] which constitutes a substantial occupational and public health issue. Although DEE is thought to contribute to cancer development by inducing oxidative stress,[4] inflammation[5 6] and DNA damage,[7] the underlying pathogenic mechanism of DEE has not been fully characterised.

Alu retroelements are repetitive mobile DNA sequences approximately 300 base pairs in length that are ubiquitous throughout the human genome.[8] With over one million copies, Alu retroelements compose nearly 11% of the genome and are a major source of genetic variation.[9] There are

To cite: Wong JY.Y, Cawthon R, Dai Y, *et al*. *Occup Environ Med* 2021;**78**:823–828.

three major families of Alu retroelements (ie, AluJ (Jurka), AluS (Smith) and AluY ('Young')), which arose at different periods of hominid evolution.[10] These families exhibit different degrees of active retrotransposition, which is the ability for Alu sequences to multiply, insert ('jump') into different genomic locations and increase in copy number throughout the genome. Originating nearly 65 million years ago, AluJ is the most ancient family and currently exhibits negligible active retrotransposition among humans.[10] The AluS family arose nearly 30 million years ago and still has several known active elements.[10] AluY, which comprises approximately 1% of all Alu elements, is the youngest of the three families and exhibits the greatest tendency for active retrotransposition.[11]

The ability of AluY and AluS elements to jump throughout the genome has notable biological and health consequences. Although rare, Alu insertions into exons could alter RNA splicing and shift reading frames, which can lead to disruption of gene expression, as well as alter protein expression and function.[12 13] Additionally, given that Alu sequences are often enriched with CpG sequences that are targets for DNA methylation, insertion of Alu elements near regulatory regions may alter epigenetic regulation of crucial genes.[14] Furthermore, an overall change in Alu retroelement copy number throughout the genome may contribute to increased non-allelic homologous recombination[15] and disruption of DNA architecture.[16] Taken together, increased Alu retroelement copy number could be a potential marker of genomic instability.

Alu repeats and insertions have been linked to various cancers, chronic diseases and mortality in human studies[16–19]; however, this evidence is largely preliminary. An analysis of catalogued polymorphic young Alu elements near previously identified genome-wide association signals found that Alu insertions were enriched around and were in strong linkage disequilibrium with genetic variants associated with a number of disease phenotypes, including breast and prostate cancer.[16] However, it is unclear whether those Alu insertions were functional or 'passengers' in the relationship.

DNA-damaging agents such as cisplatin, etoposide or gamma radiation have been found to promote Alu jumping throughout the genome.[20 21] However, whether genotoxic environmental pollutants promote similar Alu activity as those agents and influence Alu copy number is unexplored. As a first step towards understanding the impact of environmental exposures on this potential marker of genomic instability, we investigated associations with Alu copy number in an occupational environment with high levels of ambient DEE. Findings from this study may provide valuable insight into the pathogenic mechanism of DEE's contribution to cancer development and mortality.

METHODS

Study design and sample population

The study design, sample population and exposure assessment were previously described.[6] In brief, we conducted a cross-sectional study of 54 male DEE-exposed workers and a comparison group of 55 male unexposed controls in China in March 2013. DEE-exposed workers were selected from an engine testing facility of a factory that manufactures diesel engines for light and heavy trucks. The DEE-exposed workers spent most of their shift in direct proximity to the engines being tested during work hours; therefore, they had high levels of DEE exposure.[6] In the same local region of China, 55 unexposed men were recruited from workplaces that did not use diesel equipment or have processes that resulted in exposure to any known or suspected genotoxic, haematotoxic or immunotoxic chemicals or above-background particulate levels based on assessments from detailed walkthrough surveys.[6] The selected control facilities included the bottling department of a brewery (n=24), a water treatment plant (n=18), a meat packing facility (n=8) and an administrative facility (n=5). The unexposed participants were frequency-matched to DEE-exposed workers by age (±5 years) and smoking status (ie, never, former and current). The participation rates for DEE-exposed workers and unexposed controls were approximately 90% and 80%, respectively.

Peripheral blood samples were collected for complete blood cell count, analysis of the major lymphocyte subsets via flow cytometry (FACSCalibur, BD Biosciences) and extraction of leucocyte genomic DNA via commercial kits (QIAsymphony, Qiagen) by the National Cancer Institute's Cancer Genomics Research Laboratory. This study was integrated into a regular health exam administered by the local Chinese Center for Disease Control (CDC). The study was approved by Institutional Review Boards at the US National Cancer Institute and the National Institute of Occupational Health and Poison Control, China CDC. All volunteer participants provided written informed consent.

Exposure assessment

The exposure assessment was conducted from October 2012 to March 2013 in the diesel engine testing facility. Repeated full-shift personal air samples for elemental carbon (EC; a surrogate measurement for DEE exposure[22]) and $PM_{2.5}$ were collected using a cyclone attached to the lapel near the breathing zone, with an aerodynamic cut-off of 2.5 µm ($PM_{2.5}$) at a flow rate of 3.5 L/min using quartz or Teflon filters, respectively.[6] $PM_{2.5}$ was measured by preweighing and postweighing of the Teflon filters in an environmentally controlled weighing room using a microbalance. Each measurement was performed in duplicate. EC was measured on the quartz filters using NIOSH Method 5040.[23] Weights were divided by the volume of air drawn through the filters to provide exposure concentrations (µg/m^3). Individual exposure levels were estimated using mixed-effect models.[6]

Alu retroelement copy number assay

Our field study and sample collection were conducted in 2012–2013; however, many new assays were developed over the past decade. The Alu repeat qPCR assay was recently developed by Professor Richard Cawthon in 2018–2019. Given that this novel assay assesses a form of genomic instability and DEE is a genotoxic carcinogen, we applied this assay to genomic DNA extracted from stored samples as soon as possible. Monochrome multiplex qPCR was used to measure Alu repeat copy number and single-copy albumin (Alb) gene copy number in extracted leucocyte genomic DNA. Examination of an alignment of known Alu family sequences permitted targeted qPCR amplification of a region of sequence that was identical across all major Alu subfamilies (J, S and Y).

The qPCR components in a 10 µL reaction volume were 1X Titanium Taq DNA Polymerase (Takara Bio), 200 µM of each deoxynucleoside triphosphate (dNTP) (Invitrogen), 1X SYBR Green I dye (Thermo Fisher), 1 M betaine (Affymetrix, USB), 1X CutSmart Buffer (New England Biolabs) (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate and 100 µg/mL bovine serum Alb), MspI (≥20 units per µg of input DNA) and MboI (≥20 units per µg of input DNA) (New England Biolabs) and 200 nM of each of the following oligonucleotide primers: AluXu: 5'-ATTCACGCCTGTAATCCCAGGAC-3' and AluXd: 5'-ATTGCCTCGGCCTCCCAATGTC-3', to amplify

Table 1 Thermocycling profile for Alu and Alb qPCR amplification

	Step							
	1	2	3	4	5	6	7	8
Temperature (°C)	37	94	96	64	96	69	72	86
Time (MM:SS)	10:00	3:00	0:10	0:40	0:10	0:40	0:15	0:20
Repeated cycles	1	1	2		30			
Signal read							Read	Read

The initial 10 min incubation at 37°C allows restriction endonuclease digestion of the genomic DNA with the restriction enzymes MspI and MboI. The DNA is then denatured and the hot start polymerase is activated. Purposely introduced mutations in the Alu primers require that they are initially annealed at a relatively low temperature (steps 3 and 4, two cycles), before beginning the final stage of thermal cycling (steps 5–8, 30 cycles). This strategy overcomes the problem of Alu amplification curves rising above baseline so early in the run that the software cannot effectively determine a baseline fluorescence level. Alu signal acquisition is at 72°C, and Alb signal acquisition is at 86°C.

Alu retroelements, and Albugc: 5'-CCCCCGCCGCCGCCGC CGCCGCCGCCCTGAAATGCATGGTCGCCTGTT-3' and Albdgc: 5'-CCGCCGCCCGCGCCCGCCGCGCCCCGCT ATGCTGCACAGAATCCTTG-3', to amplify the single copy Alb gene. The qPCR reaction mixture was pH 7.9 at 25°C.

The qPCR reactions were performed on 384-well plates using a Bio-Rad CFX384 Real-Time PCR Detection System with a thermocycling profile shown in table 1. Relative quantification of the amplicons was performed using the standard curve method with Bio-Rad software. The input amounts of the reference DNA for the standard curve were 0.33, 1, 3, 9 and 27 ng. Both the Alu and Alb assays were measured in triplicate for each subject. A unitless Alu/Alb ratio was calculated between the average Alu retroelement copy number to Alb gene copy number, relative to that of a pooled reference DNA sample from the Cawthon laboratory at the University of Utah consisting of six women and two men. The relative Alu/Alb ratio is reflective of the average copy number of Alu retroelements in the leucocyte genome of each participant. For example, an Alu/Alb ratio of 1.05 for a participant represents a 5% greater Alu copy number compared with a common pooled reference sample that was run on all plates.

The average coefficient of variation of five quality control samples that were performed in duplicate was 4%. Furthermore, there was no correlation between input amount of DNA and Alu/Alb ratio (p=0.78 for a linear regression model).

Statistical analyses

First, we compared DEE-exposed workers to the unexposed controls. Second, DEE-exposed workers were classified into low, medium and high DEE exposure groups to evaluate exposure–response relationships using EC as previously described.[6] The median (range) EC levels (μg/m^3) for the three categories were first tertile, 23.8 (6.1–39.0); second tertile, 49.7 (39.1–54.5); and third tertile, 69.4 (54.6–107.7). One DEE-exposed worker was missing outcome data on Alu repeats and was excluded from the analyses. Between the exposure categories, Wilcoxon tests were used to compare continuous variables, while χ^2 tests were used to compare categorical variables. Spearman correlations were used to assess trends between continuous and ordinal variables.

Separate unadjusted linear regression models were used to assess associations between demographic, environmental and lifestyle factors and Alu/Alb ratio. Multivariable linear regression models were used to estimate the associations between DEE-exposed workers versus the unexposed controls, categories of DEE exposure (unexposed, low, medium and high) and Alu/Alb ratio, adjusted for age (continuous) and smoking status (never, former and current). We also considered smoking duration (years), current smoking intensity (cigarettes/day), smoking pack-years, body mass index (BMI, kg/m^2, continuous), work years (continuous), current alcohol use, self-reported respiratory infections in the past 4 weeks (yes/no; (1) a cold, (2) sinusitis or sinus problems, (3) influenza and (4) pneumonia) and white blood cell count and subtypes as covariates. However, these factors were not associated with the Alu/Alb ratio, and/or their inclusion did not appreciably change the estimates; therefore, the parsimonious models are presented. We explored interactions between diesel exposure status (exposed workers vs unexposed controls) and smoking status (never, former and current) using multiplicative interaction terms in the linear regression models, as well as stratified analyses by smoking status.

P values <0.05 were considered statistically significant. All analyses were carried out using SAS V.9.4 software.

RESULTS

Study population characteristics

The demographic characteristics of the study population are shown in table 2. The mean age was 42 years, which was similar for both unexposed and DEE-exposed workers. Most of the study population were current smokers (63%) and drinkers (79%), which did not differ between the DEE-exposed and unexposed subjects. Furthermore, 26.6% of the participants reported recent respiratory infections, which did not differ by exposure group.

Associations with Alu/Alb ratio

Age (β=−8.5E-05, 95% CI: −2.2E-03–2.0E-03 and p=0.94), BMI (β=−1.5E-03, 95% CI: −5.6E-03–2.5E-03 and p=0.46), smoking status (ever vs never: β=−0.021, 95% CI: −0.062–0.019 and p=0.30) and current alcohol use (β=1.3E-03, 95% CI:−3.4E-02–3.7E-02 and p=0.94) were not significantly

Table 2 Characteristics of diesel engine exhaust (DEE)-exposed workers and an unexposed comparison group

Characteristic	Unexposed (n=55)	DEE-exposed (n=54)	P value
Age, mean (SD)	42.1 (7.4)	42.0 (6.8)	0.99*
Body mass index, kg/m^2, mean (SD)	25.2 (3.8)	24.7 (3.4)	0.57*
Smoking status			
Current, n (%)	35 (63.6)	34 (63)	0.95†
Former, n (%)	12 (21.8)	11 (20.4)	
Never, n (%)	8 (14.5)	9 (16.7)	
Smoking duration, years, mean (SD)	16.3 (13.1)	16.2 (10.8)	0.96*
Smoking intensity among current smokers, average cigs/day, mean, SD	7.8 (8.3)	8.3 (8.2)	0.75*
Smoking pack-years, mean, SD	9.6 (11.8)	11.3 (10.4)	0.30*
Current alcohol use			
Non-drinker, n (%)	8 (14.5)	15 (27.8)	0.09†
Drinker, n (%)	47 (85.5)	39 (72.2)	
Recent respiratory infection‡			
No, n (%)	41 (74.5)	39 (72.2)	0.78†
Yes, n (%)	14 (25.5)	15 (27.8)	
Elemental carbon, μg/m^3, mean (SD)	–	48.5 (22.1)	–

P values <0.05 were considered statistically significant.
*P value of Wilcoxon test.
†P value of χ^2 test.
‡Defined by the following self-reported conditions in the past 4 weeks (yes/no): (a) cold, (b) sinusitis or sinus problems, (c) influenza and (d) pneumonia.

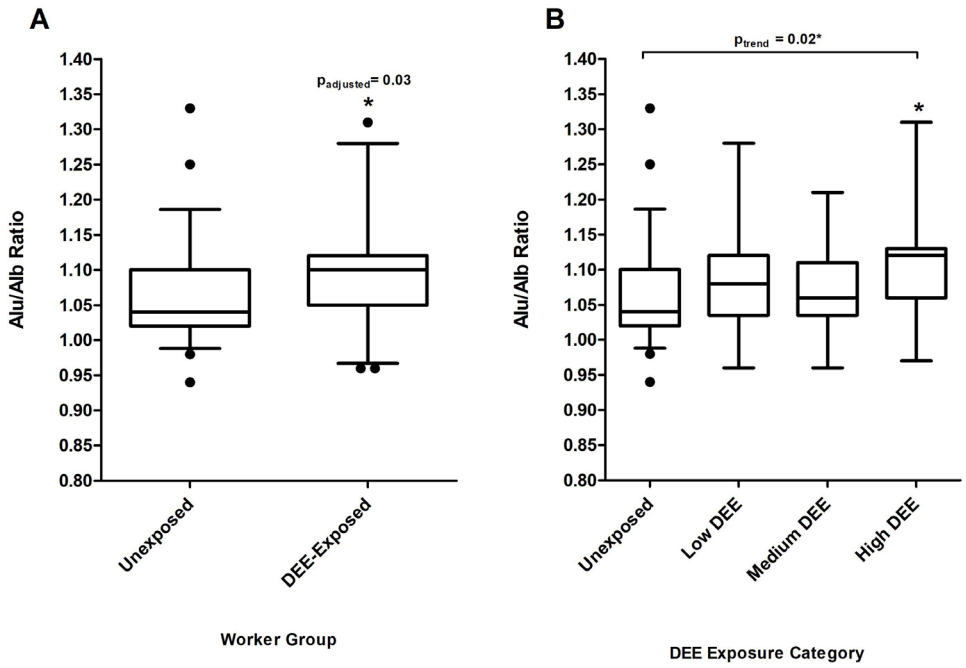

Figure 1 Overall associations between diesel engine exhaust (DEE) exposure and Alu retroelement copy number. The unadjusted Alu/albumin (Alb) ratio, which reflects average Alu retroelement copy number, is the outcome on the Y-axis of the box and whisker plots. The error bars represent the fifth and 95th percentile. (A) Among 55 unexposed and 53 DEE-exposed workers with Alu/Alb ratio outcome data, multivariable linear regression models were performed, adjusted for age and smoking status. (B) Among 55 unexposed and 17 low (elemental carbon (EC): 6.1–39.0 µg/m^3), 17 medium (EC: 39.1–54.5 µg/m^3) and 19 high (EC: 54.6–107.7 µg/m^3) DEE-exposed workers, multivariable linear regression models were conducted to assess trends, adjusted for age and smoking status.

associated with Alu/Alb ratio in univariate analyses among the overall study population, as well as among the DEE-exposed and unexposed control groups.

DEE-exposed factory workers had a statistically significant higher average Alu/Alb ratio than the unexposed controls (1.10±0.08 SD vs 1.06±0.07 SD, $p_{Wilcoxon}$=0.01 and $p_{adjusted}$=0.03) (figure 1A and table 3). Additionally, we found a positive exposure–response relationship between DEE tertiles and Alu/Alb ratio (Spearman correlation=0.28, $p_{Spearman}=3.0\times10^{-3}$ and $p_{trend,adjusted}$=0.02) (figure 1B and table 3). Those in the highest tertile had a statistically significant higher average Alu/Alb ratio than the unexposed controls (1.12±0.08 SD vs 1.06±0.07 SD and $p_{adjusted}$=0.01) (table 3).

We did not find statistically significant or compelling evidence for multiplicative interactions between DEE and smoking status ($\beta_{interaction}$=0.06 and $p_{interaction}$=0.18 for current smoking). When the analyses were stratified by smoking status to assess the consistency of the main findings, the association between DEE exposure and Alu repeats was significant among the current smokers (n=69, β=0.054, 95% CI: 0.019–0.089 and $p=3.2\times10^{-3}$), but not among former smokers (n=23, β=−0.005, 95% CI: −0.064–0.055, p=0.87) or never smokers (n=17, β=−0.001, 95% CI: −0.109–0.106 and p=0.98). However, the data were too sparse for former and never smokers to make any conclusions about heterogeneity by smoking status.

DISCUSSION

We found that DEE-exposed workers had a higher level of Alu retroelement repeats compared with unexposed workers, which potentially reflects increased genomic instability. Additionally, a positive exposure–response relationship was observed, which further supports this relationship. To our knowledge, this is the

Table 3 Associations between exposure to diesel engine exhaust (DEE) and Alu retroelement copy number

Parameter	Average Alu/Alb ratio	% Alteration in Alu/Alb ratio, exposed versus unexposed	β, adjusted difference, Alu/Alb ratio	95% CI lower	95% CI upper	P value
(A)DEE-exposedfactoryworkersversusunexposed						
Unexposed (n=55)	1.06	Reference	Reference			
DEE-exposed (n=53)	1.10	3.8	0.032	0.004	0.061	0.03
(B) DEE exposure response						
Unexposed (n=55)	1.06	Reference	Reference			
Low (n=17)	1.09	2.8	0.032	−0.009	0.073	0.13
Medium (n=17)	1.07	0.9	0.009	−0.032	0.051	0.65
High (n=19)	1.12	5.7	0.053	0.014	0.092	0.01
P-trend	0.02					

P values <0.05 were considered statistically significant. Multivariable linear regression models were adjusted for age and smoking status for analyses A and B. DEE exposure categories: low (6.1–39.0 µg/m^3), medium (39.1–54.5 µg/m^3) and high (54.6–107.7 µg/m^3).

first human study to find a link between genotoxic diesel exhaust and Alu retroelement copy number.

As a subtype of short interspersed nuclear elements, Alu retroelements are one of the most common mobile DNA elements (transposable elements) in the human genome. Studies of the biological function and human health consequences of Alu repeats are in early stages. These transposable elements were once considered 'junk DNA', as the vast majority of Alu repeats are mapped to non-coding intergenic and intronic regions, which makes their biological significance unclear without further experimental and epidemiological investigation.[24] However, there is emerging evidence that Alu repeats could potentially influence genomic architecture, gene expression, biological function and disease phenotypes (including cancer).

For instance, studies have found that transposable elements (including Alu) contribute to nearly half of the active regulatory elements of the human genome[25] by altering gene promoters and creating alternative promoters and enhancers that influence gene activity.[26–28] In computational analysis of human genome data, Alu elements were associated with upregulated gene expression and had the highest probability of any transposable element of contributing to regulatory regions, with the most recently derived types of transposable element (ie, young Alu) having the greatest overall impact.[29] Furthermore, Alu insertions located in intronic regions could regulate gene function by promoting alternative splicing of genes since Alu elements can have multiple splice donor/accept sites.[12] In addition, Alu repeats within intergenic regions can become embedded in long intergenic non-coding RNA, which have been correlated with human endogenous retrovirus subfamily H transcriptional regulatory signals.[30]

There is evidence linking Alu repeats to human diseases. In an analysis of combined genome-wide association study (GWAS) data, Payer *et al* (2017) identified an enrichment of Alu polymorphisms in regions of the genome associated with human disease risk. They found 44 instances where the trait-associated single nucleotide polymorphism (SNP) is a surrogate for presence or absence of an Alu insertion, which indicates that the Alu could be the variant effecting disease risk.[16] Polymorphic Alu elements were candidate causal variants for malignancies such as acute lymphoblastic leukaemia (10q21.2: ARID5B), breast cancer (8q23.21: CASC8, 12p11.22: PTHLH, 2q35: TCF4 and 6p23: RANBP9) and prostate cancer (8q24.21).[16] Furthermore, prospective nested case–control studies have found associations between global DNA methylation of Alu sequences and risk of cancer development.[31–33] Additionally, Alu insertions have been linked to breast cancer when inserted into the BRCA1/2 genes in both tumour and blood samples.[34 35] Aside from cancers, Alu repeats have been found to be associated with genetic variants related to multiple sclerosis (1p13.1: CD58), obesity (2p25) and psoriasis (12q13.3: STAT2), among others.[16]

No prospective cohort study to our knowledge has examined associations of an environmental exposure with Alu retroelement copy number. DNA methylation levels among Alu sequences (along with long interspersed nuclear elements (LINE-1)) are often used as surrogate measures of global DNA methylation given their ubiquity throughout the genome and their enrichment for CpG target sites.[36] As such, altered copy number of Alu sequences, which are targeted by DNA methylation machinery, may influence the degree of epigenetic modifications to the genome globally and at specific loci. Given that Alu retrotransposition and insertion throughout the genome over time could potentially alter gene expression, protein function and genomic stability, large prospective studies of Alu copy number and cancer risk are warranted.

Alu insertions into the genome require two sequential single-stranded DNA breaks around the target site in order to proceed.[8] Genotoxic agents that nick DNA including cisplatin, etoposide and gamma radiation have been found to promote Alu jumping throughout the genome.[20 21] Notably, Hagan *et al* (2003) found that the chemotherapeutic agent etoposide, which inhibits topoisomerase II and induces DNA strand breaks,[37] activates retrotransposition of human Alu sequences that were introduced into mouse cells.[21] Similar to etoposide, DEE also contains genotoxic constituents that can nick DNA, which could potentially facilitate Alu insertions. For example, benzo[a]pyrene (a prominent surrogate PAH and known carcinogen) has also been shown to induce DNA strand breaks in experiments with mouse oocytes and cumulus cells.[38] Further, studies have reported that 1-nitropyrene, a notable nitro-PAH found in DEE, can induce DNA damage in cell lines even at low concentrations.[39] As such, a mechanistic link between the genotoxic components of DEE and Alu retrotransposition activity is biologically feasible.

The primary strength of this study was that it was conducted in an occupational environment with relatively high DEE exposure levels and a nearly 18-fold range of exposure. DEE levels in our study were considerably higher than those found in a previous study of US trucking industry workers.[40] Additionally, we used personal air monitors to measure individual EC exposure, which allowed for the estimation of an exposure–response relationship for DEE. Our study had some limitations. First, the sample size was limited; however, given the wide difference in exposure levels between the exposed and unexposed participants, we had enough statistical power to detect modest associations. Second, given the modest association observed between DEE and Alu copy number, we could not discount the possible contribution of unmeasured or residual confounding to the findings. However, we evaluated various pertinent covariates, and none were found to be influential. Lastly, the correlation between Alu retrotransposition in the DNA of leucocytes and target tissues such as lung has not been previously evaluated.

In summary, our findings suggest that elevated exposure to carcinogenic DEE may contribute to increased Alu retroelement copy number. Our results are consistent with those from animal and in vitro studies that reported increased Alu retrotransposition in response to treatment with genotoxic compounds. Our findings represent a first step towards understanding the impact of environmental exposures on this potential marker of genomic instability in humans. Further exploration and replication in larger studies are warranted.

Author affiliations
[1]Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland, USA
[2]Department of Human Genetics, University of Utah, Salt Lake City, Utah, USA
[3]National Institute for Occupational Health and Poison Control, Chinese Center for Disease Control and Prevention, Beijing, China
[4]Division of Environmental Epidemiology, Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands
[5]Department of Internal Medicine, School of Medicine, University of New Mexico, Albuquerque, New Mexico, USA
[6]Chaoyang Center for Disease Control and Prevention, Chaoyang, Liaoning, China
[7]Hong Kong University, Hong Kong, China
[8]China Medical University, Shenyang, Liaoning, China
[9]Division of Epidemiology, Albert Einstein College of Medicine, Yeshiva University, New York, New York, USA

Acknowledgements We thank all of the study participants.

Contributors JW: study ideation and design, manuscript composition and statistical analysis. RC: study ideation, manuscript composition and laboratory assays. YD, HD, YN, JX, BZ, DR and DH: data collection and manuscript composition. RV: study ideation and design, data collection, exposure assessment, statistical analysis

and manuscript composition. BB: data management and manuscript composition. WH: exposure assessment, data collection, data management, statistical analysis and manuscript composition. GSD: exposure assessment and manuscript composition. SL, BTJ, WF, JY, MY, XJ, TM and PB: manuscript composition. KM: exposure assessment, data collection and manuscript composition. DTS, NR, YZ and QL: study ideation and design, data collection, manuscript composition, statistical analysis and co-supervision of the study. All authors reviewed and approved the manuscript.

Funding This work was supported by intramural funds from the National Cancer Institute and the National Institutes of Health (Project: 1ZIACP010120-24) and the Key Program of National Natural Science Foundation of China.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval The study was approved by Institutional Review Boards at the US National Cancer Institute and the National Institute of Occupational Health and Poison Control, China CDC. The numbers/IDs of the approvals are not applicable. All volunteer participants provided written informed consent.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available. No data are available.

Author note Bryan A. Bassig is currently employed by the US Centers for Disease Control and Prevention, National Center for Health Statistics. All work and participation in this study was conducted while employed by the US National Cancer Institute.

ORCID iDs
Jason Y.Y. Wong http://orcid.org/0000-0003-2820-2133
Roel Vermeulen http://orcid.org/0000-0003-4082-8163
Huawei Duan http://orcid.org/0000-0002-2717-7936
George S. Downward http://orcid.org/0000-0001-6642-7048

REFERENCES

1 Silverman DT. Diesel exhaust and lung cancer-aftermath of becoming an IARC group 1 carcinogen. *Am J Epidemiol* 2018;187:1149–52.
2 National Institute of Occupational Safety and Health. *National occupational exposure survey, 1981–1983; estimated total and female employees, actual observation and trade named exposure to products of combustion —diesel fuels*, 1983.
3 Steenland K. Lung cancer and diesel exhaust: a review. *Am J Ind Med* 1986;10:177–89.
4 Hartz AMS, Bauer B, Block ML, *et al*. Diesel exhaust particles induce oxidative stress, proinflammatory signaling, and P-glycoprotein up-regulation at the blood-brain barrier. *Faseb J* 2008;22:2723–33.
5 Bassig BA, Dai Y, Vermeulen R, *et al*. Occupational exposure to diesel engine exhaust and alterations in immune/inflammatory markers: a cross-sectional molecular epidemiology study in China. *Carcinogenesis* 2017;38:1104–11.
6 Lan Q, Vermeulen R, Dai Y, *et al*. Occupational exposure to diesel engine exhaust and alterations in lymphocyte subsets. *Occup Environ Med* 2015;72:354–9.
7 Duan H, Jia X, Zhai Q, *et al*. Long-term exposure to diesel engine exhaust induces primary DNA damage: a population-based study. *Occup Environ Med* 2016;73:83–90.
8 Deininger P. Alu elements: know the SINEs. *Genome Biol* 2011;12:236.
9 Lander ES, Linton LM, Birren B, *et al*. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860–921.
10 Jurka J, Smith T. A fundamental division in the Alu family of repeated sequences. *Proc Natl Acad Sci U S A* 1988;85:4775–8.
11 Liu GE, Alkan C, Jiang L, *et al*. Comparative analysis of Alu repeats in primate genomes. *Genome Res* 2009;19:876–85.
12 Sorek R, Ast G, Graur D. Alu-containing exons are alternatively spliced. *Genome Res* 2002;12:1060–7.
13 Häsler J, Strub K. Alu elements as regulators of gene expression. *Nucleic Acids Res* 2006;34:5491–7.
14 Jordà M, Díez-Villanueva A, Mallona I, *et al*. The epigenetic landscape of Alu repeats delineates the structural and functional genomic architecture of colon cancer cells. *Genome Res* 2017;27:118–32.
15 Ade C, Roy-Engel AM, Deininger PL. Alu elements: an intrinsic source of human genome instability. *Curr Opin Virol* 2013;3:639–45.
16 Payer LM, Steranka JP, Yang WR, *et al*. Structural variants caused by *Alu* insertions are associated with risks for many human diseases. *Proc Natl Acad Sci U S A* 2017;114:E3984–92.
17 Morgan RG, Venturelli M, Gross C, *et al*. Age-associated Alu element instability in white blood cells is linked to lower survival in elderly adults: a preliminary cohort study. *PLoS One* 2017;12:e0169628.
18 Belancio VP, Roy-Engel AM, Deininger PL. All y'all need to know bout retroelements in cancer. *Semin Cancer Biol* 2010;20:200–10.
19 Bakshi A, Herke SW, Batzer MA, *et al*. DNA methylation variation of human-specific Alu repeats. *Epigenetics* 2016;11:163–73.
20 Rudin CM, Thompson CB. Transcriptional activation of short interspersed elements by DNA-damaging agents. *Genes Chromosomes Cancer* 2001;30:64–71.
21 Hagan CR, Sheffield RF, Rudin CM. Human Alu element retrotransposition induced by genotoxic stress. *Nat Genet* 2003;35:219–20.
22 Environment Protection Agency. *Health assessment document for diesel engine exhaust*, 2002.
23 Birch ME. Occupational monitoring of particulate diesel exhaust by NIOSH method 5040. *Appl Occup Environ Hyg* 2002;17:400–5.
24 Martinez-Gomez L, Abascal F, Jungreis I, *et al*. Few SINEs of life: Alu elements have little evidence for biological relevance despite elevated translation. *NAR Genom Bioinform* 2020;2:lqz023.
25 Jacques Pierre-Étienne, Jeyakani J, Bourque G. The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genet* 2013;9:e1003504.
26 Conley AB, Miller WJ, Jordan IK. Human cis natural antisense transcripts initiated by transposable elements. *Trends Genet* 2008;24:53–6.
27 Franchini LF, López-Leal R, Nasif S, *et al*. Convergent evolution of two mammalian neuronal enhancers by sequential exaptation of unrelated retroposons. *Proc Natl Acad Sci U S A* 2011;108:15270–5.
28 Medstrand P, Landry JR, Mager DL. Long terminal repeats are used as alternative promoters for the endothelin B receptor and apolipoprotein C-I genes in humans. *J Biol Chem* 2001;276:1896–903.
29 Zeng L, Pederson SM, Cao D, *et al*. Genome-wide analysis of the association of transposable elements with gene regulation suggests that alu elements have the largest overall regulatory impact. *J Comput Biol* 2018;25:551–62.
30 Kelley D, Rinn J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biol* 2012;13:R107.
31 Gao Y, Baccarelli A, Shu XO, *et al*. Blood leukocyte Alu and LINE-1 methylation and gastric cancer risk in the Shanghai Women's Health Study. *Br J Cancer* 2012;106:585–91.
32 Barry KH, Moore LE, Liao LM, *et al*. Prospective study of DNA methylation at LINE-1 and Alu in peripheral blood and the risk of prostate cancer. *Prostate* 2015;75:1718–25.
33 Joyce BT, Gao T, Zheng Y, *et al*. Prospective changes in global DNA methylation and cancer incidence and mortality. *Br J Cancer* 2016;115:465–72.
34 Teugels E, De Brakeleer S, Goelen G, *et al*. De novo Alu element insertions targeted to a sequence common to the BRCA1 and BRCA2 genes. *Hum Mutat* 2005;26:284.
35 De Brakeleer S, De Grève J, Lissens W, *et al*. Systematic detection of pathogenic Alu element insertions in NGS-based diagnostic screens: the BRCA1/BRCA2 example. *Hum Mutat* 2013;34:785–91.
36 Zheng Y, Joyce BT, Liu L, *et al*. Prediction of genome-wide DNA methylation in repetitive elements. *Nucleic Acids Res* 2017;45:8697–711.
37 Walles SA, Zhou R, Liliemark E. DNA damage induced by etoposide; a comparison of two different methods for determination of strand breaks in DNA. *Cancer Lett* 1996;105:153–9.
38 Einaudi L, Courbiere B, Tassistro V, *et al*. In vivo exposure to benzo(a)pyrene induces significant DNA damage in mouse oocytes and cumulus cells. *Hum Reprod* 2014;29:548–54.
39 Andersson H, Piras E, Demma J, *et al*. Low levels of the air pollutant 1-nitropyrene induce DNA damage, increased levels of reactive oxygen species and endoplasmic reticulum stress in human endothelial cells. *Toxicology* 2009;262:57–64.
40 Smith TJ, Davis ME, Reaser P, *et al*. Overview of particulate exposures in the US trucking industry. *J Environ Monit* 2006;8:711–20.