Carcinogenesis, 2019, Vol. 40, No. 8, 975–978

doi:10.1093/carcin/bgz052
Advance Access publication March 12, 2019
Original Article

ORIGINAL ARTICLE

Variation in ribosomal DNA copy number is associated with lung cancer risk in a prospective cohort study

H. Dean Hosgood III[1,*], Wei Hu[2], Nathaniel Rothman[2], Madelyn Klugman[1], Stephanie J. Weinstein[2], Jarmo R. Virtamo[3], Demetrius Albanes[2,], Richard Cawthon[4,†] and Qing Lan[2,†]

[1]Division of Epidemiology, Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY, USA, [2]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA, [3]Department of Public Health Solutions, National Institute for Health and Welfare, Helsinki, Finland and [4]Department of Human Genetics, University of Utah, Salt Lake City, UT, USA

*To whom correspondence should be addressed: Tel: +718 430 2138; Fax: +718 430 8780; Email: dean.hosgood@einstein.yu.edu
†These authors contributed equally to this work.

Abstract

Disruption of ribosomal DNA (rDNA) has been linked to a variety of diseases in humans, including carcinogenesis. To evaluate the associations between rDNA copy number (CN) and risk of lung cancer, we measured 5.8S and 18S rDNA CN in the peripheral blood of 229 incident lung cancer cases and 1:1 matched controls from a nested case–control study within a prospective cohort of male smokers. There was a dose–response relationship between quartiles of both 18S and 5.8S rDNA CN and risk of lung cancer (odds ratio [OR], 95% confidence interval [CI]: 18S: 1.0 [ref]; 1.2 [0.6–2.1]; 1.8 [1.0–3.4]; 2.3 [1.3–4.1]; P_{trend} = 0.0002; 5.8S: 1.0 [ref]; 1.6 [0.8–2.9]; 2.2 [1.1–4.2]; 2.6 [1.3–5.1]; P_{trend} = 0.0001). The associations between rDNA CN and lung cancer risk were similar when excluding cases diagnosed within 5 years of follow-up, and when stratifying by heavy (>20 cigarettes per day) and light smokers (≤20 cigarettes per day).

We are the first to report that rDNA CN may be associated with future risk of lung cancer. To further elucidate the relationship between rDNA and lung cancer, replication studies are needed in additional populations, particularly those that include non-smokers.

Introduction

Ribosomal DNA (rDNA) genes encode ribosomal RNA (rRNA), whose main function is protein production within ribosomes. In ribosomes, four types of rRNAs complex with proteins to form large subunits (5S, 5.8S and 28S rRNA) and small subunits (18S) between which mRNA sits during translation (1). The rRNAs are transcribed from two multicopy DNA arrays, 5S and 45S (1,2). The rDNAs are encoded in tandem repeats and vary substantially in copy number (CN) within and between species (3,4). Modulation of rRNAs during the cell cycle is necessary for cell growth and proliferation (5). CN variations in rDNA are 4–6 orders of magnitude greater than those of single nucleotides, and CN variations of rDNA arrays have been shown to be potentially influenced by environmental exposures, such as bisphenol A (4).

The biological effects of rDNA CN are not well understood (6,7). However, Gibbons *et al.* (7) have demonstrated that rDNA CN is associated with expression of genes that regulate ribosome assembly and rRNA processing, maturation and expression, important processes that serve to maintain genome stability to prevent carcinogenesis (8). Furthermore, it has been demonstrated that increased transcriptionally active rDNA CN is related to improved cell survival in response to oxidative stress, and several diseases have been associated with low and high rDNA CN including Alzheimer's disease (6) and potentially cancer development (9).

Received: December 19, 2018; **Revised:** February 7, 2019; **Accepted:** March 8, 2019

Abbreviations

CI	confidence interval
CN	copy number
OR	odds ratio
rDNA	ribosomal DNA
rRNA	ribosomal RNA

Against this background, we hypothesized that circulating rDNA CN in peripheral blood may be associated with increased risk of lung cancer, particularly in a population of smokers who are exposed to a large amount of oxidative stress.

Methods

The Alpha-Tocopherol, Beta-Carotene Cancer Prevention (ATBC) Study (10), as well as this lung cancer nested case–control study (11), has been described previously. Briefly, 29 133 male smokers, aged 50–69 years, were recruited from 1985 to 1988 from southwest Finland. Subjects were randomized to receive α-tocopherol and/or β-carotene supplements. All participants provided written informed consent. The Finnish Cancer Registry and the Register of Causes of Death, which provides nearly 100% of cancer case ascertainment in Finland, was used to identify incident lung cancers (n = 229) through 30 April 2002 (12). Fewer than 50% of cases had clear classification as squamous cell carcinoma (n = 74) or adenocarcinoma (n = 34). Controls were selected from the ATBC Study members who were alive, free of cancer at the time of the case diagnosis and were individually matched to cases on date of birth (±5 years). The study was approved by institutional review boards at the United States' National Cancer Institute and the National Public Health Institute of Finland.

DNA was extracted from the whole blood using the phenol–chloroform method (13), and the monochrome multiplex quantitative polymerase chain reaction (MMqPCR) (14) with SYBR Green I as the detecting dye, was used to measure the rDNA CN to single copy gene CN ratio, relative to that in a reference DNA sample. The 18S rDNA to β-globin gene relative ratios were measured in one set of reactions, and the 5.8S rDNA to albumin gene relative ratios were measured in another set of reactions. The MMqPCR methods, including primer sequences and thermal cycling profiles, are described in detail in another manuscript (in review) and are available on the BioRxiv pre-print server at Cold Spring Harbor Laboratory at https://doi.org/10.1101/361840. All qPCR reactions were performed on a Bio-Rad CFX384 Real-Time PCR Detection System. Relative quantification was analyzed by the standard curve method using Bio-Rad's accompanying software. Cases and their matched controls were blindly assayed consecutively within each batch. rDNA assays were run in triplicate. Triplicate samples were interspersed in each batch to evaluate assay reproducibility. The coefficients of variation for the 18S and 5.8S assay were 5.1% and 9.3%, respectively. The intraclass correlation coefficients for the 18S and 5.8S assay were 87.9% and 81.9%, respectively.

The correlation between age at randomization and 18S rDNA and 5.8S rDNA CN was determined by the Spearman correlation coefficient among all subjects and stratified by case status. Each rDNA marker was respectively categorized into quartiles based on the distribution among controls. Odds ratios (ORs) and 95% confidence intervals (95% CIs) were estimated using conditional logistic regression models, adjusting for age at randomization, number of cigarettes smoked per day and number of years smoking. A test for trend was calculated using each rDNA CN as a continuous variable. Further adjustment by other factors, such as body mass index, physical activity and caloric intake, did not change the β-coefficient for rDNA CN ≥ 15%. The 18S rDNA and 5.8S rDNA CN and smoking multiplicative interactions were tested by including the interaction term of the dichotomous variables (based on the median among controls) in the model. The additive interactions were also tested by including a four-level categorical variable (i.e. − −, + −, − +, + +) in the model (15). All statistical analyses were performed using SAS (Cary, NC).

Results

Cases and controls were similar with respect to age, alcohol and calorie consumption (Table 1). As expected, cases smoked more cigarettes per day and smoked for a longer period than controls. The mean CNs of both 5.8S and 18S rDNA CN were higher higher among cases than among controls (Table 1), a pattern that remained when stratifying by supplementation group. Among cases and controls, neither 18S nor 5.8S rDNA CN was associated with age (cases: 18S rDNA: $\rho < 0.01$, $P = 0.99$; 5.8S: $\rho = -0.03$, $P = 0.68$; controls: 18S rDNA: $\rho = -0.07$, $P = 0.28$; 5.8S: $\rho = -0.02$, $P = 0.82$).

Men in the highest quartile of 18S and 5.8S rDNA experienced significantly increased odds of lung cancer (18S: OR = 2.3, 95% CI = 1.3–4.1; 5.8S rDNA: OR=2.6, 95% CI = 1.3–5.1), compared to

Table 1. Characteristics in lung cancer cases and individually matched controls

Demographics	Cases (n = 229), mean (SD)		Controls (n = 229), mean (SD)		P-value[a]
Age at randomization, years	58.69 (5.00)		58.44 (4.79)		0.57
Number cigarettes per day	20.6 (7.90)		18.7 (8.40)		0.012
Years of smoking	38.53 (7.1)		35.82 (9.21)		0.0005
Pack-years	39.54 (16.31)		34.07 (18.51)		0.0009
Body mass index, kg/m^2	25.61 (3.49)		26.34 (3.89)		0.036
Alcohol consumption[b], g/day	16.99 (18.43)		18.51 (13.07)		0.41
Caloric consumption[b], kcal/day	2705 (737)		2639 (672)		0.33
rDNA copy number by supplementation group	*N*	Mean (SD)	*N*	Mean (SD)	P-value[a]
18S	229	1.01 (0.22)	229	0.94 (0.14)	0.0001
Placebo	53	1.03 (0.22)	64	0.92 (0.15)	0.0031
Alpha-tocopherol (AT)	55	0.97 (0.19)	53	0.94 (0.12)	0.33
Beta-carotene (BC)	58	1.05 (0.28)	54	0.95 (0.14)	0.015
AT and BC	60	0.98 (0.15)	53	0.96 (0.14)	0.43
5.8S	229	0.97 (0.28)	229	0.89 (0.16)	0.0002
Placebo	48	1.00 (0.23)	59	0.88 (0.19)	0.0036
AT	52	0.91 (0.21)	51	0.88 (0.14)	0.32
BC	55	1.04 (0.43)	49	0.91 (0.14)	0.037
AT and BC	56	0.94 (0.15)	50	0.90 (0.15)	0.16

[a]t-test.
[b]Alcohol and caloric consumption data available in only 209 controls and 222 cases.

Table 2. Odds ratio and 95% confidence interval for 18S rDNA and 5.8S rDNA copy number and lung cancer risk

	Case	Control				Excluding cases diagnosed	
Quartile number	*n* (%)	*n* (%)	OR	95% CI[a]	*P*-trend[b]	Within 2 years	Within 5 years
18S rDNA[c]					0.0002		
1	42 (18.6)	56 (25.0)	1.0 (ref)			1.0 (ref)	1.0 (ref)
2	43 (19.0)	56 (25.0)	1.2	0.6–2.1		0.9 (0.5–1.8)	0.9 (0.4–2.2)
3	52 (23.0)	50 (22.3)	1.8	1.0–3.4		1.8 (0.9–3.6)	1.3 (0.6–3.1)
4	89 (39.4)	62 (27.7)	2.3	1.3–4.1		2.5 (1.3–4.9)	2.4 (1.0–5.5)
5.8S rDNA[d]					9.7E-05		
1	33 (15.6)	52 (24.9)	1.0 (ref)			1.0 (ref)	1.0 (ref)
2	49 (23.2)	52 (24.9)	1.6	0.8–2.9		1.4 (0.7–3.0)	1.5 (0.6–3.6)
3	59 (28.0)	52 (24.9)	2.2	1.1–4.2		2.5 (1.2–5.2)	2.1 (0.8–5.2)
4	70 (33.2)	53 (25.4)	2.6	1.3–5.1		3.3 (1.4–7.4)	4.0 (1.4–11.2)

[a]ORs and 95% CIs determined by conditional logistic regression, adjusted for age at randomization, number of years smoking and number of cigarettes per day.
[b]Trend determined by treating log transformed rDNA as a continuous variable, adjusted for age at randomization, number of years smoking and total number of cigarettes.
[c]Quartiles for 18S rDNA: quartile 1 (<0.845), quartile 2 (≥0.845–0.921), quartile 3 (≥0.921–1.030), quartile 4 (≥1.030).
[d]Quartiles for 5.8S rDNA: quartile 1 (<0.790), quartile 2 (≥0.790–0.879), quartile 3 (≥0.879–0.996), quartile 4 (≥0.996).

Table 3. rDNA copy number and risk of lung cancer stratified by smoking level

	Light smokers[a]		Heavy smokers[a]		
rDNA copy number	Cases/controls	OR (95% CI)[b]	Cases/controls	OR (95% CI)[b]	*P*-interaction
18S rDNA					
Low (<0.921)	16/47	1.0 (ref)	69/65	2.9 (1.4–6.1)	0.062
High (≥0.921)	55/50	3.5 (1.6–7.6)	86/62	4.2 (2.0–8.8)	
5.8S rDNA					
Low (<0.879)	17/45	1.0 (ref)	65/59	3.6 (1.6–7.9)	0.0079
High (≥0.879)	48/43	4.6 (1.9–10.8)	81/62	4.0 (1.8–8.9)	

[a]Based on median distribution of number of cigarettes smoked per day in controls (<20 cigarettes per day versus ≥20 cigarettes per day).
[b]OR and 95% CI determined by conditional logistic regression, adjusted for age at randomization and number of years smoking.

those in the lowest quartile (Table 2). For both of the rDNAs, there was evidence that lung cancer risk increased in a dose-dependent relationship with CN (18S rDNA: P_{trend} = 0.0002, 5.8S: P_{trend} < 0.0001). To evaluate if undiagnosed lung cancer cases at the time of blood sample collection may have influenced our findings, cases diagnosed within the 2 and 5 years of follow-up after blood sample collection were excluded from analyses (Table 2). The adjusted association between highest quartiles of 18S and 5.8S rDNA CN and lung cancer were still present and similar when excluding cases diagnosed within 5 years of follow-up (18S: OR = 2.4, 95% CI = 1.0–5.5; 5.8S: OR = 4.0, 95% CI = 1.4–11.2). Significant correlation was evident between 18S and 5.8S rDNA (ρ = 0.81, P < 0.0001), and when both rDNA types were included as covariates in the model, or as the 18S/5.8S ratio, the odds of lung cancer were no longer associated in men with the highest quartile of either rDNA type compared to those in the lowest quartile (18S: OR = 1.83, 95% CI = 0.8–4.4; 5.8S: OR = 1.45, 95% CI = 0.5–3.9; 18S/5.8S ratio: OR = 1.0, 95% CI = 0.5–1.9).

A stratified analysis by smoking intensity was conducted. The association between rDNA CN and lung cancer risk was evident among both heavy smokers (>20 cigarettes per day) and light smokers (≤20 cigarettes per day) (Table 3). The multiplicative interaction between rDNA CN and smoking was significant for 5.8S rDNA CN (P = 0.01), but not for 18S (P = 0.06) (additive interaction: 5.8S, P = 0.0025; 18S, P = 0.0024). Exploratory analyses stratifying the ORs by histology and individual supplementation groups tended to also show increased risk of lung cancer in the high quartiles of 18S and 5.8S rDNA CN; however, these results were non-significant and based on limited sample size (data not shown).

Discussion

We observed an increased risk of developing lung cancer was associated with higher CNs of 18S and 5.8S rDNA. This is the first study, to our knowledge, to evaluate the risk of lung cancer in relation to 18S and 5.8S rDNA CN in a prospective cohort. The major strength of our study is that the biological samples were collected prospectively, before lung cancer diagnosis and that rDNA CN remained significantly associated with lung cancer risk after excluding subjects that were diagnosed within 5 years of blood collection. Further, our laboratory methods had low coefficients of variation and high intraclass correlation coefficients for the 18S and 5.8S assays.

Some caveats must be acknowledged when interpreting our results. First, the 18S rDNA and 5.8S rDNA CNs were highly correlated in our study population. This is consistent with prior studies that have observed rDNA CN to be highly correlated among the subtypes (4). However, this multicollinearity does not allow for the assessment of the independent, direct associations between each unique rDNA and lung cancer risk. Second, we had limited sample size to evaluate our exploratory analyses by histology and treatment arm. Third, our study population

consisted of smoking males, which limits the generalizability of our findings. Interestingly, our results were similar when comparing the effects among light smokers and among heavy smokers, suggesting our findings may be independent of smoking dose. On the basis of these findings, rDNA CN may be a biomarker of the smoking–lung cancer relationship, or potentially even an intermediate-end point marker of lung carcinogenesis independent of smoking exposures. Another potential hypothesis is that rDNA CN may be a marker for cellular aging (16,17) associated with smoking, the carcinogenic progress or both. However, this is unlikely to explain the majority of the variance we observed in rDNA CN given that rDNA CN was not associated with age in our study, and the age range of the subjects was limited (50–69 years). Firm conclusions regarding the mechanistic relationship between rDNA CN and lung cancer risk will remain unknown until studies are specifically conducted in non-smoking lung cancer cases and non-smoking controls. Such a study would be of high scientific value as lung cancer in non-smokers account for ~25% of lung cancer cases, and is a major public health concern in Asia (18), as well as potentially in the USA (19,20).

In conclusion, our results provide the first evidence that rDNA CN may be associated with future development of lung cancer among smokers. Additional research is necessary to explore the generalizability of our findings in additional populations, particularly in women, minorities and non-smokers.

Funding

This work was funded in part by the National Institute of Aging (NIA R01 grant: R01 AG038797). The ATBC Study is supported by the Intramural Research Program of the U.S. National Cancer Institute, National Institutes of Health, and by U.S. Public Health Service contract HHSN261201500005C from the National Cancer Institute, Department of Health and Human Services.
Conflict of Interest Statement: None declared.

References

1. Moss, T. *et al.* (2007) A housekeeper with power of attorney: the rRNA genes in ribosome biogenesis. Cell. Mol. Life Sci., 64, 29–49.
2. Yu, S. *et al.* (2018) The long-range interaction map of ribosomal DNA arrays. PLoS Genet., 14, e1007258.
3. Worton, R.G. *et al.* (1988) Human ribosomal RNA genes: orientation of the tandem array and conservation of the 5′ end. Science, 239, 64–68.
4. Gibbons, J.G. *et al.* (2015) Concerted copy number variation balances ribosomal DNA dosage in human and mouse genomes. Proc. Natl. Acad. Sci. USA., 112, 2485–2490.
5. Grummt, I. (1999) Regulation of mammalian ribosomal gene transcription by RNA polymerase I. Prog. Nucleic Acid Res. Mol. Biol., 62, 109–154.
6. Porokhovnik, L.N., Lyapunova, N.A. (2018) Dosage effects of human ribosomal genes (rDNA) in health and disease. Chromosome Res. 27, 5–17. https://www.ncbi.nlm.nih.gov/pubmed/30343462
7. Gibbons, J.G. *et al.* (2014) Ribosomal DNA copy number is coupled with gene expression variation and mitochondrial abundance in humans. Nat. Commun., 5, 4850.
8. Lindström, M.S. *et al.* (2018) Nucleolus as an emerging hub in maintenance of genome stability and cancer pathogenesis. Oncogene, 37, 2351–2366.
9. Deisenroth, C. *et al.* (2010) Ribosome biogenesis surveillance: probing the ribosomal protein-Mdm2-p53 pathway. Oncogene, 29, 4253–4260.
10. ATBC Cancer Prevention Study Group. (1994) The Alpha-Tocopherol, Beta-Carotene Lung Cancer Prevention Study: design, methods, participant characteristics, and compliance. The ATBC Cancer Prevention Study Group. Ann Epidemiol, 4, 1–10. https://www.ncbi.nlm.nih.gov/pubmed/8205268
11. Hosgood, H.D. III *et al.* (2010) Mitochondrial DNA copy number and lung cancer risk in a prospective cohort study. Carcinogenesis, 31, 847–849.
12. Korhonen, P. *et al.* (2002) The Finnish Cancer Registry as follow-up source of a large trial cohort–accuracy and delay. Acta Oncol., 41, 381–388.
13. García-Closas, M. *et al.* (2001) Collection of genomic DNA from adults in epidemiological studies by buccal cytobrush and mouthwash. Cancer Epidemiol. Biomarkers Prev., 10, 687–696.
14. Cawthon, R.M. (2009) Telomere length measurement by a novel monochrome multiplex quantitative PCR method. Nucleic Acids Res., 37, e21.
15. Knol, M.J. *et al.* (2007) Estimating interaction on an additive scale between continuous determinants in a logistic regression model. Int. J. Epidemiol., 36, 1111–1118.
16. Bürkle, A. (2002) In memoriam Bernard Strehler–genomic instability in ageing: a persistent challenge. Mech. Ageing Dev., 123, 899–906.
17. Zafiropoulos, A. *et al.* (2005) Preferential loss of 5S and 28S rDNA genes in human adipose tissue during ageing. Int. J. Biochem. Cell Biol., 37, 409–415.
18. Samet, J.M. *et al.* (2009) Lung cancer in never smokers: clinical epidemiology and environmental risk factors. Clin. Cancer Res., 15, 5626–5645.
19. Samet, J.M. (2017) Is the Incidence of Adenocarcinoma of the Lung Rising in Never Smokers? J. Natl. Cancer. Inst., 109. https://www.ncbi.nlm.nih.gov/pubmed/28376182
20. Pelosof, L. *et al.* (2017) Proportion of never-smoker non-small cell lung cancer patients at three diverse institutions. J. Natl. Cancer. Inst., 109. https://www.ncbi.nlm.nih.gov/pubmed/28132018