Priority Report

Cancer Research

Mosaic Y Loss Is Moderately Associated with Solid Tumor Risk

Erikka Loftfield[1], Weiyin Zhou[1,2], Meredith Yeager[1,2], Stephen J. Chanock[1], Neal D. Freedman[1], and Mitchell J. Machiela[1]

Abstract

Mosaic loss of the Y chromosome (mLOY) in peripheral leukocytes is a somatic event in which a fraction of leukocytes have lost the entire Y chromosome. The frequency of mLOY increases with age and may reflect poor genomic maintenance as well as clonal imbalances in normal immune function, making mLOY an attractive candidate marker for cancer risk. Here, we investigated the relationship between mLOY and incident cancer in a large sample of 207,603 cancer-free men from the UK Biobank, in which 13,895 men developed an incident solid tumor during follow-up. We identified mLOY by scanning for deviations in genotyping array log R intensity ratios across the male-specific chromosome Y region. Overall, we detected low proportions of cells with mLOY in 3,358 (1.6%) men and high proportions of mLOY in 524 (0.3%) men. We found an association of mLOY with overall solid tumor incidence using both low and high mLOY thresholds [$HR_{low} = 1.18$; 95% confidence interval $(CI)_{low}$, 1.07–1.30; $P_{low} = 0.001$; $HR_{high} = 1.36$; 95% CI_{high}, 1.09–1.71; $P_{high} = 0.007$] and more specifically we observed an association with lung cancer ($HR_{high} = 2.25$; 95% CI_{high}, 1.36–3.71; $P_{high} = 0.002$). Stronger associations were observed without adjustment for smoking, suggesting that smoking is an important confounder of tumor incidence. It is unlikely that mLOY is a major mediator of the effect of cigarette smoking on cancer risk, as mLOY was observed in only a small fraction of smokers who developed cancer. In summary, mLOY was modestly associated with incidence of solid tumors in the UK Biobank, although for some cancer subtypes these findings may reflect residual confounding by smoking.

Significance: Evidence from the UK Biobank indicates mosaic chromosome Y loss in leukocytes is moderately associated with increased incidence of select solid tumors.

Introduction

Genetic mosaicism is a somatically occurring event in which a detectible clonal population of cells harbors a postzygotic mutation that is distinct from inherited germline variation (1, 2). The most commonly detected large, structural mosaic event is mosaic loss of the Y chromosome (mLOY) in circulating leukocytes of males (3, 4). The frequency of mLOY is age-related, with low observed frequencies in men under 50 years of age and then rapidly increasing in frequency to observed frequencies of approximately 10% to 15% in men by age 80 years of age (3, 4). Interestingly, cigarette smoking, a modifiable exposure, has also been established as a risk factor for mLOY (4, 5).

The effect of mLOY on male health has been a topic of recent investigation, and associations with cancer (3, 4, 6, 7), cardiovascular disease (8), Alzheimer's disease (9), mortality (3, 4), and suicide (10) have been explored. However, studies of mLOY and disease have generally been small, and associations have yet to be firmly established. In particular, the relationship between mLOY and cancer risk requires further investigation as some (3, 7), but not all (4), studies suggest mLOY is associated with increased cancer risk. Potential mechanisms linking mLOY to increased disease risk include reduced capacity for genomic maintenance or clonal aberrations in immune cell function (11); although, future functional studies clarifying the impact of mLOY on cell function and homeostasis are needed.

We investigated the relationship between mLOY and incident solid tumor risk in the UK Biobank, a large, population-based cohort study of approximately 500,000 participants from the United Kingdom (12, 13). Our aim was to clarify the association between mLOY and incident cancer risk overall and by solid tumor type in a large, well-characterized population. Furthermore, we performed mediation analyses to better understand how mLOY may mediate the effects of smoking on risk of smoking-related tumors.

Materials and Methods

The UK Biobank is a population-based cohort study of 503,317 individuals representing individuals in the UK's National Health Service, aged 40 to 69 years, who reside near a UK Biobank assessment center. Participants provided baseline information on demographic, lifestyle, and other health-related factors, biological samples, and physical measures. The National Information Governance Board for Health and Social Care and the National Health Service North West Multicentre Research Ethics Committee

[1]Division of Cancer Epidemiology and Genetics, NCI, Bethesda, Maryland. [2]Cancer Genomics Research Laboratory, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc., Frederick, Maryland.

Corresponding Author: Mitchell J. Machiela, NCI, NIH, 9609 Medical Center Drive, Rockville, Bethesda, MD 20892-9774. Phone: 240-760-6518; E-mail: mitchell.machiela@nih.gov

doi: 10.1158/0008-5472.CAN-18-2566

approved the UK Biobank study and all participants provided written informed consent at enrollment, and the study was conducted in accordance with recognized ethical guidelines (e.g., U.S. Common Rule).

Our study analyzed genetic data from either Affymetrix UK BiLEVE or UK Biobank Axiom arrays from 223,507 UK Biobank male participants. We excluded participants with sex discrepancies between self-reported and genomic sex using X-chromosome heterozygosity ($n = 167$) and men with a cancer diagnosis prior to baseline ($n = 15{,}737$), resulting in a final analytic cohort of 207,603 male UK participants. Using 691 markers across male-specific region of chromosome Y (MSY; chrY:2658271-28767492, hg19/GRCh37), we scanned for deviations in median log R ratio (mLRR) to detect evidence for mLOY. We defined mLOY dichotomously using two thresholds reported previously in the literature: mLRR $< -0.15(4)$ and mLRR $< -0.40(3)$. We also calculated a continuous measure of mLOY in which mLRR was scaled by the SD divided by -1 for ease of interpretation.

We tabulated the two definitions of mLOY and mLRR by potential confounding factors including age, smoking status, body mass index (BMI), alcohol drinking status, and education level and used multivariable-logistic or linear regression models to test for an association between each factor of interest and mLOY or mLRR, adjusting for continuous age, age-squared, age-cubed, and the other variables in Table 1.

Follow-up time for cancer incidence was counted from the date of assessment center visit (i.e., between 2006 and 2010), at which time, blood and baseline information were collected until the date of first cancer diagnosis or the date of censor (i.e., January 31, 2016 for England and Wales and November 30, 2015 for Scotland), whichever came first. Only individuals with first cancer diagnoses of solid tumors were included as events in the incident cancer analysis. Participants with first cancer diagnoses of hematologic malignancies during follow-up were censored at their date of diagnosis. HR estimates were generated for all solid tumor cancers combined and for solid tumor types with more than 100 incident cases in our final analytic cohort. In addition, we

Table 1. Baseline characteristics and mosaic loss of the Y chromosome

	Entire cohort	mLOY (mLRR < -0.15)		mLOY (mLRR < -0.40)		mLRR	
	N (%)	*N* (%)	*P*[a]	*N* (%)	*P*[a]	Median $\pm$ SD	*P*[b]
Age							
<65 years	176,987 (85.3)	1,733 (51.6)	Ref	234 (44.7)	Ref	0.007 $\pm$ 0.052	Ref
$\geq$65 years	30,616 (14.8)	1,625 (48.4)	$<4.9 \times 10^{-324}$	290 (55.3)	1.0×10^{-97}	-0.008 ± 0.085	$<4.9 \times 10^{-324}$
Smoking status[c]							
Never smoker	101,909 (49.1)	904 (26.9)	Ref	116 (22.1)	Ref	0.007 $\pm$ 0.050	Ref
Former smoker	78,439 (37.8)	1,594 (47.5)	6.9×10^{-23}	244 (46.6)	3.9×10^{-7}	0.003 $\pm$ 0.062	2.3×10^{-17}
Current smoker	26,151 (12.6)	838 (25.0)	1.59×10^{-171}	162 (30.9)	6.9×10^{-46}	0.001 $\pm$ 0.074	3.6×10^{-261}
Race/ethnicity[d]							
White	194,879 (93.9)	3,254 (96.9)	Ref	514 (98.1)	Ref	0.005 $\pm$ 0.059	Ref
Mixed race	1,07035 (0.5)	4 (0.1)	0.05	1 (0.2)	0.60	0.009 $\pm$ 0.051	0.32
Asian	5,543 (2.7)	45 (1.3)	0.09	6 (1.2)	0.32	0.010 $\pm$ 0.050	8.5×10^{-5}
Black	3,150 (1.5)	12 (0.4)	0.01	0 (0)	0.96	0.017 $\pm$ 0.044	1.0×10^{-16}
Other	1,831 (0.9)	18 (0.5)	0.73	2 (0.4)	0.70	0.013 $\pm$ 0.048	0.006
BMI[e]							
<18.5 kg/m^2	437 (0.2)	6 (0.2)	0.33	1 (0.2)	0.60	0.004 $\pm$ 0.053	0.52
18.5 to <25 kg/m^2	49,712 (24.0)	806 (24.0)	Ref	135 (25.8)	Ref	0.004 $\pm$ 0.059	Ref
25 to <30 kg/m^2	101,157 (48.7)	1,718 (51.2)	0.54	268 (51.2)	0.42	0.005 $\pm$ 0.059	0.002
30 to <35 kg/m^2	40,558 (19.5)	600 (17.9)	0.001	88 (16.8)	0.03	0.006 $\pm$ 0.058	1.4×10^{-12}
$\geq$35 kg/m^2	11,793 (5.7)	146 (4.4)	0.004	21 (4.0)	0.10	0.008 $\pm$ 0.056	3.9×10^{-18}
Alcohol drinking status[f]							
Never drinker	5,862 (2.8)	71 (2.1)	0.61	10 (1.9)	0.88	0.008 $\pm$ 0.054	0.30
Former drinker	7,248 (3.5)	121 (3.6)	0.52	19 (3.6)	0.91	0.004 $\pm$ 0.061	0.54
Current drinker (<1 drink/week)	33,650 (16.2)	484 (14.4)	0.38	77 (14.8)	0.87	0.006 $\pm$ 0.058	0.40
Current drinker ($\geq$1 drink/week >7)	41,623 (20.1)	610 (18.2)	0.43	100 (19.1)	0.46	0.005 $\pm$ 0.057	0.65
Current drinker (1–3 drinks/day)	89,159 (43.0)	1,446 (43.1)	Ref	214 (40.8)	Ref	0.005 $\pm$ 0.058	Ref
Current drinker (>3 drinks/day)	29,468 (14.2)	617 (18.4)	0.004	104 (19.9)	0.06	0.005 $\pm$ 0.064	0.26
Education level[g]							
College or university degree	70,395 (33.9)	860 (25.6)	Ref	120 (22.9)	Ref	0.006 $\pm$ 0.054	Ref
A levels/AS levels or equivalent	21,472 (10.3)	289 (8.6)	0.36	57 (10.9)	0.02	0.006 $\pm$ 0.058	0.35
O levels/GCSEs or equivalent	38,569 (18.6)	594 (17.7)	0.41	86 (16.4)	0.84	0.006 $\pm$ 0.056	0.80
CSEs or equivalent	11,546 (5.6)	74 (2.2)	0.77	13 (2.5)	0.40	0.009 $\pm$ 0.049	0.89
NVQ or HND or HNC or equivalent	18,649 (9.0)	336 (10.0)	0.78	53 (10.1)	0.55	0.004 $\pm$ 0.060	0.15
Other qualifications	9,153 (4.4)	186 (5.5)	0.59	31 (5.9)	0.68	0.002 $\pm$ 0.064	0.55

[a]*P* values from multivariable logistic regression models are adjusted for all other covariates in Table 1. In this model, we adjusted for continuous age, age-squared, and age-cubed.

[b]*P* values from multivariable linear regression models are adjusted for all other covariates in Table 1. In this model, we adjusted for continuous age, age-squared, and age-cubed.

[c]$n = 1{,}104$ men missing data on smoking status.

[d]$n = 1{,}165$ men missing data on race/ethnicity.

[e]$n = 3{,}946$ men missing data on BMI.

[f]$n = 593$ men missing data on alcohol drinking.

[g]$n = 37{,}819$ men missing data on education.

considered two major histologic subtypes of lung cancer, squamous cell (ICD-O-3: 8050, 8052, 8070-8075, 8083, 8084), and adenocarcinoma (ICD-O-3: 8140, 8200, 8231, 8250-8255, 8260, 8290, 8310, 8323, 8430, 8480, 8481, 8490, 8550, 8574) with more than 100 cases each.

We used multivariable Cox proportional hazards regression models to estimate HRs and 95% confidence intervals (CI) for cancer incidence, with age used as the underlying time metric because we expected the hazard to change more as a function of age than with time enrolled in the UK Biobank (14). For smoking adjustment, we created a 25-level detailed smoking history variable by combining data on smoking status, lifetime smoking, smoking intensity, time since quitting for former smokers, and type of tobacco smoked. For alcohol adjustment, we created a 6-level variable by combining data on drinking status and amount of alcohol consumed per week, calculated as the sum of all alcoholic beverages consumed on average per day. Categories of BMI were defined according to the World Health Organization (15). Indicator variables were used to account for missing data in regression models. To account for multiple comparisons in our main analysis of continuous mLRR and solid tumor cancers, we considered a Bonferroni-corrected P value of 0.004 (i.e., 0.15/13) to be statistically significant.

In secondary analyses, we further explored potential confounding and mediation by cigarette smoking. First, we stratified analyses for all solid tumors as well as lung cancer by smoking status (i.e., never or ever). Among ever smokers, we ran multivariable Cox proportional hazards regression models with and without detailed adjustment for smoking. Finally, we conducted mediation analyses to estimate how much of the smoking–cancer association, for all solid tumors and lung cancer, is mediated through Y loss (i.e., continuous mLRR) in UK Biobank men using a unified approach developed by Lange and colleagues for estimating natural direct and indirect effects (16). All data analysis was performed in SAS version 9.4 (SAS Institute) on NIH's Biowulf computing cluster.

Data availability statement

All data used in this analysis are available through application to the UK Biobank.

Results

The Affymetrix array intensity data for all included UK Biobank male participants aged 37 to 73 ($N = 207{,}603$; mean = 57, median = 58) were scanned for evidence of mLOY. In total, 3,358 (1.6%) men had detectible mLOY (mLRR < −0.15) and of these 524 (0.3%) had evidence for high proportions of cells affected (mLRR < −0.4, approximately 24% of cells with Y loss). We also observed evidence for mosaic chromosome Y gain (mLOY > 0.15) in 197 (0.09%) men. As observed with previous studies, we observed strong associations between mLOY and increasing age ($P < 4.9 \times 10^{-324}$) as well as former ($P = 6.9 \times 10^{-23}$) and current (1.59×10^{-171}) smoking, indicating these are important covariates to account for in statistical modeling of the association between mLOY and incident tumor risk.

During follow-up, 13,895 men in the UK Biobank developed an incident solid tumor. Skin (excluding basal and squamous, $N = 4{,}519$) and prostate cancer ($N = 4{,}345$) were the most common incident cancers followed by colon ($N = 848$) and lung ($N = 783$) cancer. Other cancer types with more than 100 incident cases are described in Table 2.

An association with incident solid tumors and the continuous measure of mLOY (mLRR; HR = 1.04; 95% CI, 1.02–1.05; $P = 7.5 \times 10^{-7}$) was observed in the multivariable Cox proportional

Table 2. HRs and 95% CIs[a] for mosaic loss of the Y chromosome in incident solid tumor cases in the UK Biobank ($N = 207{,}603$)

		Percent Y loss		mLOY (LRR ≤ 0.15)		mLOY (LRR ≤ 0.40)		SD of continuous LRR [b]	
	Incident cases	LRR	LRR	HR		HR		HR	
Group site	N	≤ 0.15	≤ 0.40	(95% CI)	P	(95% CI)	P	(95% CI)	P
All solid tumors cancer	13,895	3.0%	0.6%	1.18 (1.07–1.30)	0.001	1.36 (1.09–1.71)	0.007	1.04 (1.02–1.05)	7.5×10^{-7}
Oral cavity and pharynx	281	4.3%	1.1%	1.77 (0.98–3.20)	0.061	2.32 (0.73–7.36)	0.153	1.07 (0.97–1.17)	0.164
Esophagus	298	3.7%	0.3%	1.26 (0.68–2.31)	0.465	0.74 (0.10–5.29)	0.764	1.03 (0.94–1.13)	0.562
Stomach	190	3.7%	1.6%	1.30 (0.61–2.80)	0.500	3.58 (1.13–11.34)	0.030	1.11 (1.01–1.22)	0.026
Colon excluding rectum	848	2.5%	0.4%	1.00 (0.65–1.55)	0.992	0.92 (0.30–2.87)	0.886	1.03 (0.97–1.09)	0.335
Rectum & rectosigmoid junction	550	2.4%	0.9%	0.95 (0.54–1.65)	0.846	2.37 (0.98–5.75)	0.057	1.02 (0.95–1.10)	0.598
Liver & intrahepatic bile duct	112	1.8%	0.9%	0.63 (0.15–2.57)	0.518	1.97 (0.27–14.31)	0.503	0.98 (0.83–1.16)	0.788
Pancreas	229	4.8%	0.4%	1.59 (0.86–2.95)	0.140	0.87 (0.12–6.21)	0.866	1.08 (0.99–1.18)	0.089
Lung	783	5.6%	2.0%	1.11 (0.82–1.52)	0.498	2.25 (1.36–3.71)	0.002	1.06 (1.02–1.11)	0.003
Skin excluding basal/squamous	4,519	2.6%	0.4%	1.16 (0.97–1.40)	0.109	1.29 (0.83–2.01)	0.257	1.04 (1.01–1.06)	0.007
Prostate	4,345	2.8%	0.4%	1.13 (0.94–1.36)	0.195	0.98 (0.60–1.60)	0.928	1.03 (1.00–1.05)	0.050
Bladder	369	6.0%	0.8%	1.86 (1.20–2.90)	0.006	1.50 (0.48–4.70)	0.485	1.06 (0.99–1.14)	0.087
Brain & other nervous system	207	1.4%	0.0%	0.70 (0.22–2.21)	0.543	—	—	1.04 (0.92–1.17)	0.587

[a]Adjusted for age using age as the underlying time metric in the Cox proportional hazards regression model as well as the following covariates: detailed smoking history [25-level variable incorporating current smoking status, smoking intensity (current and former smokers); time since quitting (former smokers), and cigar and pipe use (current and former smokers), race/ethnicity (white, black, Asian, mixed, or other race)]; alcohol drinking (never drinker, former drinker, infrequent drinker (<1 drink/week), occasional drinker (>1 drink/week but <1 drink/day), moderate daily drinker (1–3 drinks/day), or heavy daily drinker (>3 drinks/day); education level (college or university degree, A levels/AS levels or equivalent, O levels/GCSEs or equivalent, CSEs or equivalent, NVQ or HND or HNC equivalent, or other professional qualifications); and BMI (<18.5, 18.5–<25, 25–<30, 30–<35, or ≥35 kg/m^2).

[b]Scaled by the − (SD) of mLRR, such that the HR corresponds to a one SD decrease in mLRR.

hazard regression model. Among incident solid tumor cases, a total of 421 (3.0%) men had evidence for mLOY, with 77 (0.6%) men exhibiting high proportions of cells affected, suggesting elevated rates of mLOY in cancer cases (χ^2_{low} = 184.51; df = 1; $P = 5.0 \times 10^{-42}$). Multivariable-adjusted associations between mLOY status and incident solid tumors ($HR_{low} = 1.18$; 95% CI_{low}, 1.07–1.30; $P_{low} = 0.001$; $HR_{high} = 1.36$; 95% CI_{high}, 1.09–1.71; $P_{high} = 0.007$; Table 2) were also observed. The continuous measure mLRR was also associated with lung cancer (HR = 1.06; 95% CI, 1.02–1.11; $P = 0.003$), and those with high proportions of cells with mLOY had an increased risk of lung cancer ($HR_{high} = 2.25$; 95% CI_{high}, 1.36–3.71; $P_{high} = 0.002$; Table 2). Multivariable-adjusted associations of mLRR with two major histologic subtypes of lung cancer, squamous cell (HR = 1.02; 95% CI, 0.93–1.11; $P = 0.71$), and adenocarcinoma (HR = 1.03; 95% CI, 0.95–1.12; $P = 0.46$), did not, however, reach statistical significance.

We also conducted analyses that were stratified by cigarette smoking. For the analysis of all solid tumors, associations with mLOY remained significant among both never and ever smokers (mLRR < −0.15, mLRR < −0.4, and continuous mLRR, Table 3); however, associations appeared stronger among never smokers. For lung cancer, associations were not observed among never smokers, probably due to small numbers, but were observed among ever smokers. Among smokers, associations with mLOY for lung cancer were stronger prior to detailed adjustment for smoking history.

Mediation analyses investigating a potential role of mLOY as a mediator of the effect of smoking on cancer risk found no evidence for an indirect mediated effect through mLOY; the HR estimates for the indirect effects of former and current smoking were 1.00 (95% CI, 0.98–1.03; $P = 0.93$) and 1.01 (95% CI, 0.98–1.04; $P = 0.68$), respectively. We found similar null results for the indirect effects of smoking for lung cancer; the HR estimates for the indirect effects of former and current smoking were 1.01 (95% CI, 0.91–1.11; $P = 0.89$) and 1.07 (95% CI, 0.96–1.19; $P = 0.23$), respectively. In contrast, and as expected, the direct effects of smoking status on risk of solid tumors (HR former vs. never smoker = 1.07, 95% CI, 1.05–1.09; $P < 0.0001$; HR current vs. never smoker = 1.24, 95% CI, 1.21–1.27; $P < 0.0001$) and lung cancer (HR former vs. never smoker = 5.08, 95% CI, 4.91–5.25; $P < 0.0001$; HR current vs. never smoker = 17.36; 95% CI, 17.20–17.53; $P < 0.0001$) were statistically significant.

Discussion

Previous studies have demonstrated a relationship between mosaicism in peripheral leukocyte DNA and hematologic cancer risk; however, the relationship between mosaicism in peripheral leukocytes and solid tumor risk is unclear. In this large, prospective sample of UK men, we observe evidence for an association of mLOY in peripheral leukocytes and increased risk of incident solid tumors, particularly lung neoplasms. Mechanisms linking mLOY in blood to solid tumor formation are unknown. One possibility is that mLOY in leukocytes may serve as a global biomarker of past exposures to carcinogens (e.g., smoking) and efficient DNA maintenance and repair, and may act as a proxy for these processes in other tissues. Alternatively, clonal imbalances in immune cell composition may impact the ability of immune cells to identify and clear precancerous cells harboring neoantigens from healthy tissues. Future studies on the precise leukocyte populations impacted by mLOY and the phenotypic effects of mLOY on leukocyte function are needed to better understand how mLOY may be related to solid tumor risk.

Our analysis of UK Biobank is the largest prospective study to evaluate the relationship between mLOY and subsequent risk of an incident solid tumor. Other observational studies have provided limited support for such associations. Forsberg and colleagues (3) reported associations for any cancer (HR = 2.47; $P = 0.014$) and nonhematologic cancer (HR = 2.68; $P = 0.008$), although had only modest numbers of cases. Zhou and colleagues (4) reported prospective associations of mLOY with combined cancer (OR = 1.19; 95% CI,1.00–1.42; $P = 0.047$) as well as with bladder (OR = 1.47; 95% CI, 1.09–1.99; $P = 0.011$) and prostate (OR = 1.35; 95% CI, 1.04–1.74; $P = 0.024$) cancers, but not with lung cancer (OR = 0.90; 95% CI, 0.69–1.18; $P = 0.45$). Noveski and colleagues (6) observed lower Y:X ratios for colorectal and prostate cancer ($P = 3.76 \times 10^{-9}$ and 1.15×10^{-4}) and an overall association between LOY percentage and overall cancer (OR = 1.11; 95% CI, 1.09–1.15; $P = 2.0 \times 10^{-9}$), but had modest numbers and only adjusted for age. Machiela and colleagues (7) also found evidence for lower chromosome Y target-to-reference (T/R) ratios in familial cases

Table 3. HRs and 95% CIs[a] for mosaic loss of the Y chromosome in incident solid tumor cases in the UK Biobank stratified by smoking status

	Incident cases	Percent Y loss		mLOY (LRR ≤ 0.15)		mLOY (LRR ≤ 0.40)		SD of continuous LRR[c]	
Group site	N	LRR< −0.15	LRR< −0.40	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P
All solid tumors	1,3895	3.0%	0.6%	1.18 (1.07–1.30)	0.001	1.36 (1.09–1.71)	0.007	1.04 (1.02–1.05)	7.5×10^{-7}
Never smokers	5,830	1.9%	0.3%	1.28 (1.06–1.55)	0.011	1.75 (1.10–2.78)	0.018	1.05 (1.02–1.08)	0.0001
Ever smokers[b]	7,974	3.9%	0.7%	1.15 (1.02–1.29)	0.020	1.29 (1.00–1.67)	0.053	1.03 (1.01–1.05)	0.001
Ever smokers	7,974	3.9%	0.7%	1.20 (1.07–1.35)	0.002	1.36 (1.05–1.75)	0.020	1.04 (1.02–1.06)	9.3×10^{-6}
Lung	783	5.6%	2.0%	1.11 (0.82–1.52)	0.498	2.25 (1.36–3.71)	0.002	1.06 (1.02–1.11)	0.003
Never smokers	67	3.0%	1.5%	1.68 (0.41–6.96)	0.472	6.48 (0.89–47.19)	0.065	1.10 (0.91–1.34)	0.326
Ever smokers[b]	707	5.9%	2.1%	1.08 (0.79–1.48)	0.640	2.13 (1.27–3.58)	0.004	1.06 (1.02–1.11)	0.008
Ever smokers	707	5.9%	2.1%	1.67 (1.22–2.29)	0.001	3.41 (2.04–5.70)	3.1×10^{-6}	1.14 (1.09–1.18)	1.7×10^{-10}

[a]Adjusted for age using age as the underlying time metric in the Cox proportional hazards regression model as well as the following covariates: race/ethnicity (white, black, Asian, mixed, or other race); alcohol drinking [never drinker, former drinker, infrequent drinker (<1 drink/week), occasional drinker (>1 drink/week but <1 drink/day), moderate daily drinker (1–3 drinks/day), or heavy daily drinker (>3 drinks/day)]; education level (college or university degree, A levels/AS levels or equivalent, O levels/GCSEs or equivalent, CSEs or equivalent, NVQ or HND or HNC equivalent, or other professional qualifications); and BMI (<18.5, 18.5– <25, 25–<30, 30–<35, or ≥35 kg/m^2).

[b]Additionally adjusted for detailed smoking history [25-level variable incorporating current smoking status, smoking intensity (current and former smokers); time since quitting (former smokers), and cigar and pipe use (current and former smokers)].

[c]Scaled by the − (SD) of mLRR, such that the HR corresponds to a one SD decrease in mLRR.

of testicular germ cell tumors ($P = 0.01$), but not overall testicular germ cell tumors ($P = 0.09$). Finally, in a prior UK Biobank analysis, we found an increased risk of cancer-related death among men with a high proportion of affected cells, but we found no difference in the prevalence of cancer by mLOY status (17). Together, these studies suggest that mLOY is a potential risk factor for select cancers; although, the limited sample size and residual confounding by poorly measured and unmeasured confounders are of concern. Moreover, the estimated effects are small. With more than 200,000 men and detailed information on potential confounders, including smoking, the UK Biobank is the largest and most comprehensive analysis of mLOY and solid tumor risk to date.

A dose–response trend was observed in which the magnitude of associations between mLOY and solid cancer risk increased as the proportions of leukocytes affected by mLOY increased. These findings suggest that men with higher proportions of leukocytes affected by mLOY are at moderately higher risk of developing an incident solid tumor than men with lower proportions. Our analysis did not have the granularity to characterize the shape of the dose–response relationship or establish mLOY thresholds of clinical importance for solid tumor risk. Similarly, a previous report of mLOY and mortality in the UK Biobank also noted an association between higher proportions of cells affected by mLOY and increased risk of mortality (17).

Our analyses identified smoking as a strong confounder of the association between mLOY and incident solid tumor risk. As such, despite careful adjustment, the possibility remains that residual unadjusted effects of smoking may confound some of our observed relationships, particularly for smoking-related tumors. Lung cancer is strongly caused by cigarette smoking (18), and we observed no association of mLOY with lung cancer among never smokers. However, it is unknown whether the lack of an association for lung cancer among never smokers reflects: differences in etiology in never smokers, a small sample set with reduced power to detect an association, or a true lack of an association. For all solid tumors, however, we observed an association with mLOY among never smokers, indicating that mLOY may be associated with the risk of one or more solid tumor subtypes in the absence of smoking. Finally, we investigated the potential for mLOY to mediate the effects of smoking on incident cancer risk. Among ever smokers who developed an incident solid tumor ($n = 6,152$), only 3.3% had evidence of mosaic Y loss (mLRR $\leq$ 0.15), suggesting that if mLOY is a mediator of the effect of smoking, it mediates only a small proportion of the effect of smoking on tumor risk. Likewise, we found no evidence supporting mLOY as a mediator of the effect of smoking on cancer risk in formal mediation analyses.

In conclusion, an investigation of mLOY on incident solid tumor risk in a large, prospective study of UK men found evidence that both low and high levels of mLOY increase overall risk of solid tumors. The estimated risk appears to increase with higher fraction of mLOY. This relationship between higher percentage of leukocytes affected by mLOY and increased cancer risk warrants further characterization through functional studies investigating the leukocyte lineages affected by mLOY and the impact of mLOY on leukocyte function.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: E. Loftfield, M. Yeager, S.J. Chanock, N.D. Freedman, M.J. Machiela
Development of methodology: E. Loftfield, W. Zhou, M.J. Machiela
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): E. Loftfield
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): E. Loftfield, W. Zhou, S.J. Chanock, N.D. Freedman, M.J. Machiela
Writing, review, and/or revision of the manuscript: E. Loftfield, W. Zhou, M. Yeager, S.J. Chanock, N.D. Freedman, M.J. Machiela
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): N.D. Freedman
Study supervision: M.J. Machiela

Acknowledgments

The authors thank Michael Stagner at Information Management Services who provided technical support and Barry I. Graubard and Joshua Sampson at the Division of Cancer Epidemiology and Genetics Biostatistics Branch for statistical input. This work used the computational resources of the NIH's High-Performance Computing Biowulf cluster and was supported by the Intramural Research Program of the NIH; Division of Cancer Epidemiology and Genetics, NCI, NIH; and the Department of Health and Human Services. This research was conducted using the UK Biobank resource (application 21552). The UK Biobank was established by the Wellcome Trust, the Medical Research Council, the United Kingdom Department of Health, and the Scottish Government. The UK Biobank has also received funding from the Welsh Assembly Government, the British Heart Foundation, and Diabetes UK.

Received August 16, 2018; revised October 18, 2018; accepted November 28, 2018; published first December 3, 2018.

References

1. Strachan T, Read AP, Strachan T. Human molecular genetics. Fourth edition. New York, NY: Garland Science; 2011. p. 781.
2. Machiela MJ, Chanock SJ. Detectable clonal mosaicism in the human genome. Semin Hematol 2013;50:348–59.
3. Forsberg LA, Rasi C, Malmqvist N, Davies H, Pasupulati S, Pakalapati G, et al. Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer. Nat Genet 2014;46:624–8.
4. Zhou W, Machiela MJ, Freedman ND, Rothman N, Malats N, Dagnall C, et al. Mosaic loss of chromosome Y is associated with common variation near TCL1A. Nat Genet 2016;48:563–8.
5. Dumanski JP, Rasi C, Lonn M, Davies H, Ingelsson M, Giedraitis V, et al. Mutagenesis. Smoking is associated with mosaic loss of chromosome Y. Science 2015;347:81–3.
6. Noveski P, Madjunkova S, Sukarova Stefanovska E, Matevska Geshkovska N, Kuzmanovska M, Dimovski A, et al. Loss of Y chromosome in peripheral blood of colorectal and prostate cancer patients. PLoS One 2016;11:e0146264.
7. Machiela MJ, Dagnall CL, Pathak A, Loud JT, Chanock SJ, Greene MH, et al. Mosaic chromosome Y loss and testicular germ cell tumor risk. J Hum Genet 2017;62:637–40.
8. Dumanski JP, Sundstrom J, Forsberg LA. Loss of chromosome Y in leukocytes and major cardiovascular events. Circ Cardiovasc Genet 2017;10:e001820.
9. Dumanski JP, Lambert JC, Rasi C, Giedraitis V, Davies H, Grenier-Boley B, et al. Mosaic loss of chromosome Y in blood is associated with Alzheimer disease. Am J Hum Genet 2016;98:1208–19.

10. Kimura A, Hishimoto A, Otsuka I, Okazaki S, Boku S, Horai T, et al. Loss of chromosome Y in blood, but not in brain, of suicide completers. PLoS One 2018;13:e0190667.
11. Machiela MJ, Chanock SJ. The ageing genome, clonal mosaicism and chronic disease. Curr Opin Genet Dev 2017;42:8–13.
12. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 2015;12:e1001779.
13. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with the general population. Am J Epidemiol 2017; 186:1026–34.
14. Korn EL, Graubard BI, Midthune D. Time-to-event analysis of longitudinal follow-up of a survey: choice of the time-scale. Am J Epidemiol 1997;145:72–80.
15. WHO Consultation. Obesity: preventing and managing the global epidemic - introduction. WHO Tech Rep Ser 2000;894: 1–253.
16. Lange T, Vansteelandt S, Bekaert M. A simple unified approach for estimating natural direct and indirect effects. Am J Epidemiol 2012; 176:190–5.
17. Loftfield E, Zhou W, Graubard BI, Yeager M, Chanock SJ, Freedman ND, et al. Predictors of mosaic chromosome Y loss and associations with mortality in the UK Biobank. Sci Rep 2018;8:12316.
18. Freedman ND, Abnet CC, Caporaso NE, Fraumeni JF, Murphy G, Hartge P, et al. Impact of changing US cigarette smoking patterns on incident cancer: risks of 20 smoking-related cancers among the women and men of the NIH-AARP cohort. Int J Epidemiol 2016; 45:846–56.