OXFORD

JNCI J Natl Cancer Inst (2020) 112(9): djz232

doi: 10.1093/jnci/djz232
First published online January 16, 2020
Article

Extreme Downregulation of Chromosome Y and Cancer Risk in Men

Alejandro Cáceres, PhD,[1,2,*] Aina Jene, MsC,[3] Tonu Esko, PhD,[4] Luis A. Pérez-Jurado, MD, PhD, FRACP,[5,6] Juan R. González, PhD[1,2,7,*]

[1]Barcelona Institute for Global Health (ISGlobal), Barcelona, Spain, [2]Centro de Investigación Biomédica en Red en Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain, [3]Center for Genomics Regulation, Barcelona, Spain, [4]Estonian Genome Centre Science Centre, University of Tartu, Tartu, Estonia, [5]Genetics Unit, Universitat Pompeu Fabra, Institut Hospital del Mar d'Investigacions Mediques (IMIM), Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Barcelona, Spain, [6]Women's and Children's Hospital, South Australian Health and Medical Research Institute & University of Adelaide, Adelaide, Australia and [7]Department of Mathematics, Universitat Autònoma de Barcelona, Bellaterra, Spain

*Correspondence to: Alejandro Cáceres, PhD, Barcelona Institute for Global Health (ISGlobal), Doctor Aiguader 88, Barcelona 08003, Spain (e-mail: alejandro.caceres@isglobal.org) and Juan R. González, PhD, Barcelona Institute for Global Health (ISGlobal), Doctor Aiguader 88, Barcelona 08003, Spain (e-mail: juanr.gonzalez@isglobal.org).

Abstract

Background: Understanding the biological differences between sexes in cancer is essential for personalized treatment and prevention. We hypothesized that the extreme downregulation of chromosome Y gene expression (EDY) is a signature of cancer risk in men and the functional mediator of the reported association between the mosaic loss of chromosome Y (LOY) and cancer.

Methods: We advanced a method to measure EDY from transcriptomic data. We studied EDY across 47 nondiseased tissues from the Genotype Tissue-Expression Project (n = 371) and its association with cancer status across 12 cancer studies from The Cancer Genome Atlas (n = 1774) and seven other studies (n = 7562). Associations of EDY with cancer status and presence of loss-off function mutations in chromosome X were tested with logistic regression models, and a Fisher's test was used to assess genome-wide association of EDY with the proportion of copy number gains. All statistical tests were two-sided.

Results: EDY was likely to occur in multiple nondiseased tissues ($P < .001$) and was statistically significantly associated with the EGFR tyrosine kinase inhibitor resistance pathway (false discovery rate = 0.028). EDY strongly associated with cancer risk in men (odds ratio [OR] = 3.66, 95% confidence interval [CI] = 1.58 to 8.46, $P = .002$), adjusted by LOY and age, and its variability was largely explained by several genes of the nonrecombinant region whose chromosome X homologs showed loss-of-function mutations that co-occurred with EDY during cancer (OR = 2.82, 95% CI = 1.32 to 6.01, $P = .007$). EDY associated with a high proportion of *EGFR* amplifications (OR = 5.64, 95% CI = 3.70 to 8.59, false discovery rate < 0.001) and *EGFR* overexpression along with *SRY* hypomethylation and nonrecombinant region hypermethylation, indicating alternative causes of EDY in cancer other than LOY. EDY associations were independently validated for different cancers and exposure to smoking, and its status was accurately predicted from individual methylation patterns.

Conclusions: EDY is a male-specific signature of cancer susceptibility that supports the escape from X-inactivation tumor suppressor hypothesis for genes that protect women compared with men from cancer risk.

Men are more at risk and less likely to survive cancer than women (1). Besides the different environments to which sexes are exposed, sex-specific molecular processes are also important to explain sexual dimorphism in cancer. For instance, sex hormones are critically involved in cancer development, and numerous loci in sex chromosomes are associated with cancer susceptibility (2). In addition, the complete loss of chromosome Y (LOY) is a frequent event in tumor cells (3–9) and a specific risk factor for cancer in men when found in peripheral blood cells (10–12). Given that LOY is the most common somatic mutation in men, there is a need to understand whether uncontrolled mitosis can lead to LOY, or if LOY, as an age-related condition (13,14), can predispose to cancer (15). A logical consequence of the presence of LOY in a tissue would be the reduction of the overall transcription output of Y across the affected tissue. As such, one should observe an association between the extreme downregulation of chromosome Y (EDY) with cancer that, if stronger than that of LOY, would indicate a directionality from LOY to cancer via EDY.

Studies have shown a strong association between aneuploidies in cancer and gene expression (16). Therefore, gene expression data have been used to identify the functional consequences of aneuploidies (17). However, studies that measure the overall transcription output of an entire chromosome

Received: July 12, 2019; **Revised:** October 31, 2019; **Accepted:** December 11, 2019

© The Author(s) 2020. Published by Oxford University Press. All rights reserved. For permissions, please email: journals.permissions@oup.com

Downloaded from https://academic.oup.com/jnci/article/112/9/913/5697920 by Stanford Lane Medical Library user on 23 September 2020

have not been reported. Therefore, we first proposed a method to measure EDY from transcriptomic data, obtained by either RNA-sequencing or expression microarrays, and then confirmed the biological suitability of the measure by analyzing data from the Genotype Tissue-Expression (GTEx) Project over multiple nondiseased tissues. Using The Cancer Genome Atlas (TCGA) and several microarray studies, we then studied the association between EDY and cancer risk in men. We also investigated whether EDY status in tumors associated with differential methylation across Y and with copy number alterations in autosomes. We thus tested the hypothesis that the novel transcriptomic signature EDY is an important risk factor for cancer in men.

Methods

Detection of EDY From Transcriptome Data

We analyzed expression data from chromosome Y in the form of count data for RNA-sequencing and signal intensity for microarray experiments. For each individual, we measured the relative expression of the entire chromosome with respect to the autosomes. Having N exons in chromosome Y, with x_e read count for the e-th exon, we computed

$$y = \Sigma_{e=1,..N} \log_2(x_e + 1)/N$$

as a measure of the average expression of Y. Likewise, we obtained the mean expression in autosomes

$$a = \Sigma_{e=1,..M} \log_2(x_e + 1)/M,$$

where M is the number exons with count data in the autosomes. The relative amount of an individual's Y expression with respect to the individual's autosomes was then defined as

$$Ry = y - a.$$

We considered EDY as the extreme phenotype of Ry given by values lower than the 0.05 sample quantile, as has been done for other extreme phenotypes (18). The adequacy to treat EDY as a discontinuous extreme phenotype that is the consequence of LOY is supported by the observation that treating LOY itself as a continuous variable is suboptimal (19). In a study with K patients, we then classified individual j as having EDY if

$$Ry_j < median(Ry) - 1.2 \times IQR(Ry),$$

where IQR is the usual definition for the interquartile range of Ry values over patients. The cutting threshold given by the expression above corresponds to the lower 5% of the data for different types of unimodal distributions. Given that the interquartile range is robust for different distributions, in the case of array intensity data, we used similar definitions for EDY, computing the relative expression Ry from x_e as the intensity value at probe e.

Discovery and Validation Studies

We studied the frequency of EDY in 47 nondiseased tissues using the version-6 RNA-sequencing data from the GTEx project (https://www.gtexportal.org/). Genome-wide single nucleotide polymorphism (SNP) data were available for 298 men for whom we could determine their EDY status. We studied the association between EDY and cancer using the multiomic data for 28 TCGA cancer studies. We downloaded data from 10 642 samples: 5329 were from normal tissues and 5313 were tumorous tissues. For validation, we downloaded from the ArrayExpress Archive (www.ebi.ac.uk/arrayexpress) a large expression matrix of 27 871 arrays with accession number E-MTAB-3732, the largest systematically annotated gene expression dataset of its kind. Gene expression data were searched in the GEO repository (www.ncbi.nlm.nih.gov/geo) for case-control studies of renal clear cell carcinoma and colorectal cancer. Their accession numbers are GSE36895 and GSE44076. We also downloaded transcriptomic data from two additional studies (GSE4573 and GSE5123) on lung squamous cell carcinoma with exposure to smoking. We downloaded normalized expression data and used female samples to check the lower limit of EDY detection in men. Probe annotation was made with the Bioconductor biomaRt package.

We downloaded methylomic data from a case-control study on kidney cancer with accession number GSE61441. Given the strong pattern of methylation associated with EDY, we fitted an elastic-net model to build a subject-wise predictor of EDY from methylomic data using glmnet and caret R packages. The model was trained in 90% of the TCGA cancer samples (n = 1174) and validated in the other 10% of samples (n = 292). The model hyperparameters (mixing and smoothing) were estimated using 10-fold cross-validation. The list of CpGs and coefficients to build the predictive and an R function to get EDY prediction are available at http://github.com/isglobalbrge/EDY. Further details are in the Supplementary Methods (available online).

Statistical Analyses

All analyses were performed using packages from Bioconductor version 3.8 and R version 3.5.2. Logistic regression models were fitted for testing the association between EDY with different outcomes, such as case-control status of the individuals or tumor status of biological samples of cancer patients. We used Bayesian regression models from the arm R package that gave consistent estimates for low frequencies of patients with EDY. Random effects meta-analyses were performed with the rma package where heterogeneity between studies was tested with a χ^2 test. All models were adjusted by age and cancer type when available or needed. Main effect P values were two-sided and, if needed, corrected for multiple comparisons by false discovery rate (FDR). FDR and single-test P values less than .05 were considered statistically significant. The processed data and entire computer code needed to completely reproduce our findings have been made public in the figshare repository at https://figshare.com/projects/Extreme_down-regulation_of_chromosome_Y_and_male_disease/58514.

Results

EDY in Nondiseased Tissues

We first studied EDY in nondiseased tissues analyzing RNA-sequencing data of 371 men across 47 tissues from the GTEx Project. The average number of tissues per man was 12. We detected 140 individuals with EDY in at least one tissue (Supplementary Figure 1, available online). There was large variability of EDY frequency between tissues (mean [SD] = 6.1% [3.7%]) (Supplementary Table 1, available online). We found high rates of individuals with EDY in more than one tissue and, therefore, hypothesized whether EDY was likely to appear in multiple tissues in a single individual, suggesting a genetic

Downloaded from https://academic.oup.com/jnci/article/112/9/913/5697920 by Stang Mongkolsuk Library user on 23 September 2020

Table 1. EDY and LOY status in 12 cancer studies of TCGA.*

TCGA cancer study	No.	EDY, %	LOY, %	Agreement between EDY and LOY, %
Bladder urothelial carcinoma (BLCA)	246	31.3	41.5	85.0
Colon adenocarcinoma (COAD)	108	31.5	55.6	74.1
Esophageal carcinoma (ESCA)	98	49.0	45.9	80.6
Kidney chromophobe (KICH)	46	45.7	45.7	95.7
Kidney renal clear cell carcinoma (KIRC)	347	44.4	46.1	89.6
Kidney renal papillary cell carcinoma (KIRP)	165	77.0	77.6	95.8
Liver hepatocellular carcinoma (LIHC)	114	14.9	16.7	94.7
Lung adenocarcinoma (LUAD)	190	30.0	40.5	89.5
Lung squamous cell carcinoma (LUSC)	268	38.4	54.5	82.5
Prostate adenocarcinoma (PRAD)	413	11.1	7.7	91.8
Rectum adenocarcinoma (READ)	46	37.0	50.0	87.0
Thyroid carcinoma (THCA)	97	7.2	18.6	86.6

*EDY and LOY were estimated from transcriptomic and genomic data, respectively, obtained in both cancer and normal tissues. EDY was computed with respect to samples with no gains or losses of chromosome Y. The proportion of agreement between the measures was high but substantial differences were also observed. BLCA = bladder urothelial carcinoma; COAD = colon adenocarcinoma; EDY = extreme downregulation of chromosome Y gene expression; ESCA = esophageal carcinoma; KICH = kidney chromophobe; KIRC = kidney renal clear cell carcinoma; KIRP = kidney renal papillary cell carcinoma; LIHC = liver hepatocellular carcinoma; LOY = loss of chromosome Y; LUAD = lung adenocarcinoma; LUSC = lung squamous cell carcinoma; PRAD = prostate adenocarcinoma; READ = rectum adenocarcinoma; THCA = Thyroid carcinoma; TCGA = The Cancer Genome Atlas.

predisposition to it. Consequently, we first confirmed that individuals with EDY in one tissue were likely to show EDY in any other tissue (permutation test of tissue labels, $P < .001$). Then, because of the low power expected for the number of individuals, we performed enrichment analysis in genome-wide SNP associations for EDY status. Enrichment analyses were performed for a new variable $EDY_{>1\ tissue}$, defined as positive for individuals where EDY was found in more than one tissue and negative otherwise. Although no statistically significant associations were observed for $EDY_{>1\ tissue}$, we found that $EDY_{>2\ tissues}$ (n = 32) was statistically significantly enriched with SNPs in the EGFR tyrosine kinase inhibitor resistance pathway (*FDR* = 0.028). In this case, we also observed suggestive genome-wide associations mapping to susceptibility genes for basophil percentage of granulocytes (*GRIP1*) (21), lung and gastric cancers and smoke-induced emphysema (*MMP12*) (22–24), and high- and low-density cholesterol and triglycerides levels (*GPAM*) (25) (Supplementary Figure 2, available online). Finally, in whole blood, the single genotyped tissue in the GTEx Project where LOY could be called, only one of the three individuals with positive EDY was detected with LOY. Therefore, this novel signature EDY appears to be more common than LOY in nondiseased individuals, can be identified across tissues, and may have a genetic basis linked to several autosomal loci.

EDY in 12 TCGA Cancer Studies

We analyzed genomic and transcriptomic data of 12 cancer studies with normal and tumor samples of cancer patients from the TCGA project to establish whether EDY explained more cancer variability than LOY (Supplementary Figure 3, available online). We called LOY from genotype data and EDY from transcriptomic data within each cancer study in all samples (normal and tumor) (Table 1). EDY was obtained with respect to the *Ry* distribution of samples with no loss and no gains in chromosome Y (Figure 1). As expected, the proportion of agreement between EDY and LOY status, comprising normal and tumor tissues, was high but varied across all 12 cancer studies (87% [6%]). Comparing cancer with normal samples, we observed that the overall magnitude of the age-adjusted effect of EDY on cancer status (odds ratio [OR] = 8.33, 95% confidence interval [CI] = 3.30 to 20.89, $P = 6.9 \times 10^{-6}$) remained statistically significant after adjusting by LOY within each cancer study (OR = 3.66, 95% CI = 1.58 to 8.46, $P = .002$). Because all samples (normal and tumor) were from cancer patients, there was no association between age and cancer status of the samples (OR = 0.99, 95% CI = 0.98 to 1.00, $P = .07$). However, whereas we observed a statistically significant association between LOY and age (OR = 1.009, 95% CI = 1.003 to 1.01, $P = .001$), we did not observe a statistically significant association between EDY and age (OR = 1.003, 95% CI = 0.99 to 1.00, $P = .2$). Consistent with these findings, we observed that the association between EDY and cancer was robust under different age quartiles (age [16–57] years: OR = 4.56, $P < .001$; age [57–64] years: OR = 5.03, $P < .001$; age [64–71] years; OR = 3.76, $P < .001$; age [71–90] years: OR = 17.25, $P = .008$).

Transcriptome-wide analyses in tumor samples revealed that the transcription levels of *DDX3Y*, *EIF1AY*, *KDM5D*, *RPS4Y1*, *UTY*, and *ZFY* were statistically significantly downregulated across all 12 cancers. Their joint downregulation explained 89% of EDY's variability and 88% of LOY's variability (Supplementary Figure 3, available online). Interestingly, these genes have four remarkable features. First, they are located in pairs in three distant regions of the nonrecombinant region of Y (NRY) (Yp11.31: from Mb 2.7 to Mb 2.9 / Yq11.21: from Mb 15.0 to Mb 15.6 / Yq11.22: from Mb 21.8 to Mb 22.9), suggesting that they may share regulatory elements. Second, the genes encode proteins with important functions in cell cycle regulation: helicase (*DDX3Y*), translation initiation (*EIF1AY*), histone demethylation (*KDM5D* and *UTY/KDM6C*), transcriptional activation (*ZFY*), and ribosomal assembly (*RPS4Y1*). Third, these genes have homologs (*DDX3X*, *EIF1AX*, *KDM5C*, *KDM6A/UTX*) on the X chromosome that escape X-inactivation. And fourth, male-biased loss-of-function (LoF) somatic mutations have been found in four of the X chromosome homologs of these genes across many cancers (26). In line with this last feature, we observed that LoF mutations in the four X chromosome homologs co-occurred with LOY (OR = 3.59, 95% CI = 1.57 to 8.18, $P = .002$) and EDY (OR = 2.82, 95% CI = 1.32 to 6.01, $P = .007$) during cancer.

Three further analyses in TCGA consistently showed that EDY provided a stronger signature of cancer than LOY. First, the meta-analysis between cancer status and the EDY derived from the NRY gene signature was substantially more statistically

Downloaded from https://academic.oup.com/jnci/article/112/9/913/5697920 by Stanford Libraries user on 23 September 2020

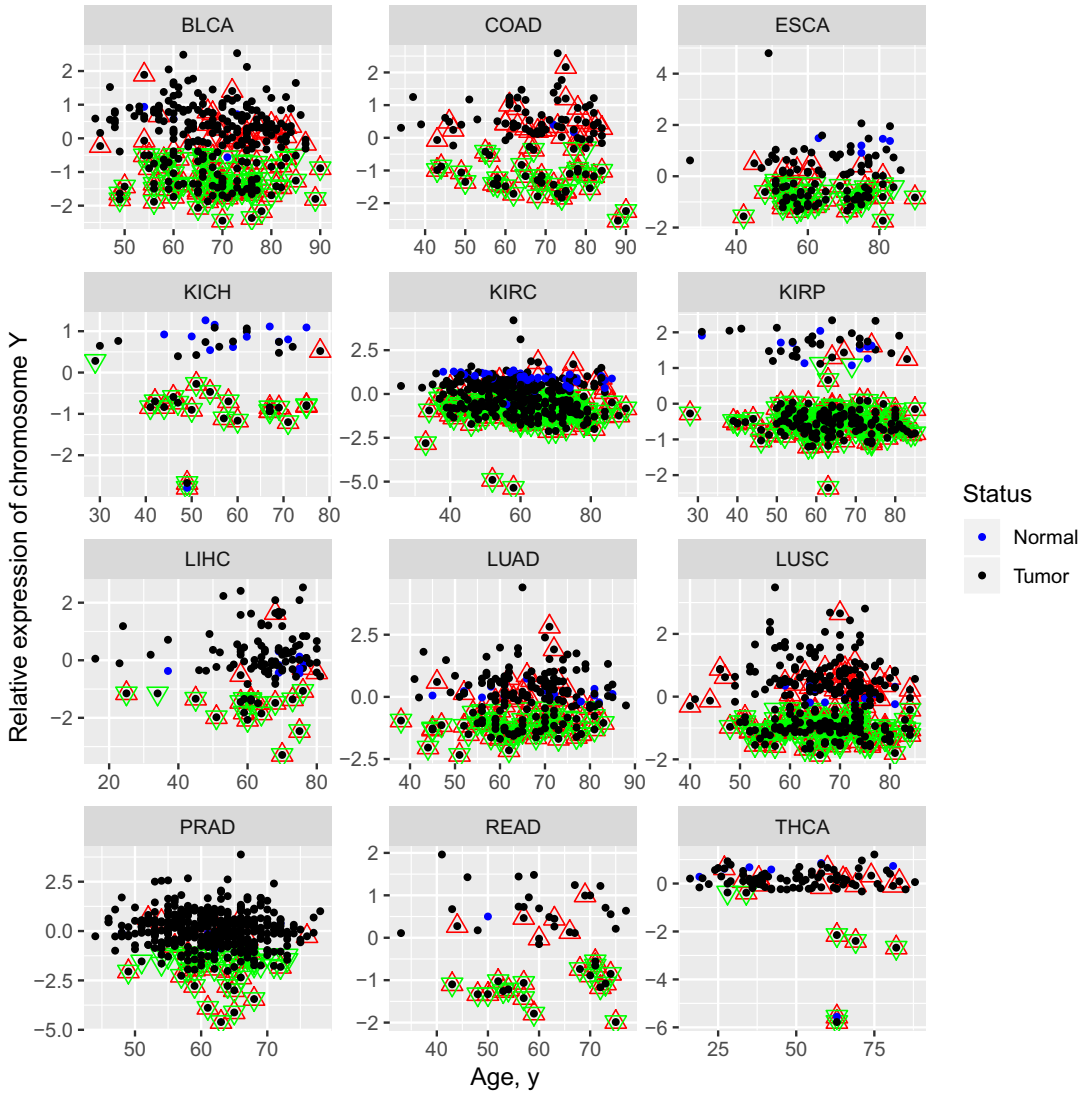

Figure 1. Relative chromosome Y expression (Ry) shown as a function of age for 12 cancer studies from The Cancer Genome Atlas. The figure shows tumor **(black)** and normal samples **(blue)**. Samples with loss of chromosome Y (LOY), obtained from genotype intensity data, are shown in **red triangles**. Samples with extreme downregulation of chromosome Y gene expression (EDY) **(green triangles)** are those with low values of Ry relative to samples with no chromosome Y losses or gains. Although EDY and LOY status overlap, numerous individuals are observed with LOY but no EDY, particularly those with high values of Ry. Normal samples consistently have high Ry values across studies.

significant than previous associations (OR = 8.14, 95% CI = 4.29 to 15.40, $P < .001$), remaining statistically significant after adjusting by LOY (OR = 3.61, 95% CI = 1.51 to 8.63, $P = .003$). Second, Bayesian network analyses indicated that the causal sequence given by aging, LOY, EDY, and cancer was more probable than that given by aging, cancer, LOY, and EDY (Figure 2F). Third, total EDY mediated 48.9% (95% CI = 25.3% to 66.0%) of the age-adjusted association between LOY and the cancer status of the samples. Overall, these observations on TCGA data support a possible cancer mechanism underlying LOY given by the simultaneous inactivation of NRY genes derived by EDY and their functional homologs on chromosome X by LoF mutations.

Validation in Independent Studies

Using data from independent transcriptomic studies, we performed numerous replication and consistency analyses (Figure 3; Supplementary Table 2, available online). We first replicated the EDY association with colorectal (OR = 5.16, 95% CI = 1.30 to 20.45, $P = .01$) and kidney cancer (OR = 20.09, 95% CI = 2.07 to 195.11, $P = .009$) in two independent transcriptomic case-control studies, where tumor tissues were compared with normal tissues of cancer patients. In the kidney study, cancer was not associated with age (OR = 0.99, 95% CI = 0.93 to 1.05, $P = .8$). EDY was not associated with age either (OR = 1.004, 95% CI = 0.94 to 1.06, $P = .8$), and, despite low numbers (12 case patients and 17 control patients), the association between EDY and cancer appeared to be consistent between two age strata (age [35–59] years: OR = 14.9, 95% CI = 1.11 to 201.05, $P = .04$, 5 case patients, 8 control patients; age [59–83] years: OR = 3.31, 95% CI = 0.80 to 13.91, $P = .09$, 8 case patients, 6 control patients). Because LOY in blood is a risk factor for cancer, we then confirmed that EDY in blood associated with cancer diagnosis in a large Estonian population sample (OR = 3.23, 95% CI = 1.24 to 8.40, $P = .01$). We also aimed to determine the range of cancers

ARTICLE

Downloaded from https://academic.oup.com/jnci/article/112/9/913/5697920 by Stang Mongkolsuk Library user on 23 September 2020

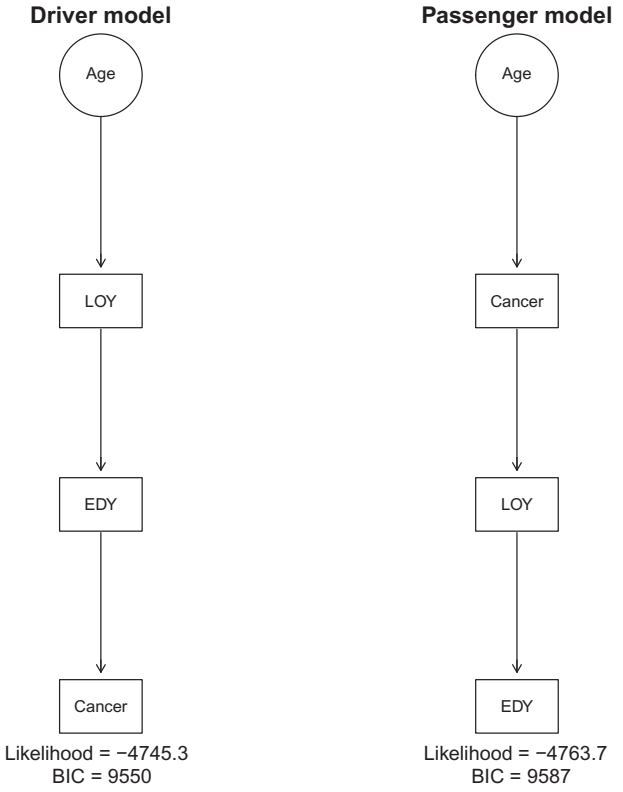

Figure 2. Additive Bayesian network models for age, cancer, loss of chromosome Y (LOY), and extreme downregulation of chromosome Y gene expression (EDY) shown for 12 cancer studies from The Cancer Genome Atlas (TCGA). The **left figure** shows the driver model, where cancer depends on EDY, EDY on LOY, and LOY on age. Maximum likelihood estimate and Bayes information criterion (BIC) are shown on **top**. The **right figure** shows the passenger model, where EDY depends on LOY, LOY on cancer, and cancer on age. In the TCGA studies, the higher likelihood and lower BIC favor the driver model over the passenger model.

associated with EDY using a large collection of multi-disease expression arrays with 3771 diseased tissues and 3127 healthy-male tissues (20). We observed strong positive associations for eight cancer groups; negative associations for myeloma and other types of leukemia; and no association for lymphoma, neuroblastoma, and prostate cancer (Figure 3). The range of cancers associated with EDY largely overlapped with cancer status of samples associated with LOY across all 28 cancer studies from the TCGA. Interestingly, associations of LOY with leukemia and prostate cancers were statistically nonsignificant (Supplementary Table 3, available online). Finally, in line with LOY's association with smoking (27), we observed a statistically significant association of EDY in lung cancer with the number of cigarettes smoked per day (OR = 1.07, 95% CI = 1.02 to 1.12, $P = .003$) in two studies (Figure 3) and confirmed the association with heavy smoking (>1 pack/d, OR = 18.77, 95% CI = 1.02 to 345.32, $P = .04$) in a third study.

EDY Association With Copy Number Variants and Methylation Patterns

To gain further insights on why EDY can be a stronger cancer signature than LOY, we studied whether EDY showed biological correlates in cancer that were independent of LOY. We analyzed the copy number variant differences between EDY statuses in 3034 tumors across all TCGA studies in windows of 1.25 Mb across the genome. Remarkably, we observed that, in individuals with no LOY, EDY was strongly associated with a higher proportion of copy number gains of *EGFR* (OR = 5.64, 95% CI = 3.70 to 8.59, *FDR* < 0.001), whereas in individuals with LOY, EDY associated with a lower proportion of copy number gains in regions containing *SOX4* (OR = 0.29, 95% CI = 0.17 to 0.47, *FDR* < 0.001), *NCOA2*, and the short arm of chromosome 12 (12p) (Figure 4). We also asked whether consistent differences in EDY were associated with methylation across chromosome Y, stratifying by LOY status and adjusting for tumor type. We thus observed a highly reproducible pattern of methylation-probe associations that was independent of LOY (Figure 4; Supplementary Table 4, available online). We found statistically significant methylation changes in the NRY regions deregulated in EDY, the highest association being with hypomethylation surrounding *SRY* (cg04169747, OR = 0.96, 95% CI = 0.95 to 0.97, *FDR* < 0.001) and the highest changes being hypermethylation at *KDM5D* (cg15329860, 95% CI = 1.15 to 1.29, *FDR* < 0.001) (28) and *EIF1AY* (cg08820785, OR = 1.19, 95% CI = 1.12 to 1.24, *FDR* < 0.001). In line with the proportion of gains found in EDY, we observed statistically significant associations of *EGFR* and *SOX4* expression levels with reciprocal hypo- and hypermethylations of the same CpG sites (Supplementary Tables 5–7, available online). In addition to the possible contribution of *SRY* hypomethylation, a gene hypermethylated after sex differentiation early in development (29), our data reinforce the role of NRY genes in cancer sex bias (26). Given the strong pattern of methylation associated with EDY, we used an elastic-net algorithm to build a subject-wise predictor of EDY from methylomic data, trained in the 90% of TCGA cancer samples and validated with 90.7% accuracy in the other 10%. In an independent methylomic study, we externally validated the association between the methylation-inferred EDY with kidney cancer (N = 92, OR = 45.4, 95% CI = 2.11 to 977.32, $P = .01$) (Figure 3).

Discussion

We have provided the first evidence, to our knowledge, of a path in men that leads from LOY and other genetic alterations, such as EGFR pathway activation, to cancer development through EDY. EDY is a male-specific signature of cancer susceptibility that is strongly linked to LOY and chromosome Y methylation patterns as well as to environmental exposures such as smoking (27). The high correlation between EDY and LOY confirmed that EDY is the most likely functional consequence of LOY, yet additional risk was observed for individuals with EDY and no LOY, suggesting that overall decrease of chromosome Y transcript levels is a key element in the susceptibility to disease. We found strong associations with cancer susceptibility comparable with those of smoking. In the population-based Estonian Genome Center of the University of Tartu study, the frequency of EDY in blood in the general population (4.3%) and the fraction of individuals diagnosed with any type of cancer (9.4%) yielded an attributable risk of 16.2%, which is in range of the attributable risk to cancer due to smoking (30), but further studies are needed to refine these risk estimates.

In particular, our data provide additional evidence to support the escape from X-inactivation tumor suppressor hypothesis for genes that protect women compared with men from cancer risk (26), pointing to specific genes that accumulate male-biased mutations in cancer whose chromosome Y homologs on NRY define EDY. These genes (*DDX3Y*, *EIF1AY*, *KDM5D*, *RPS4Y1*, *UTY*,

Downloaded from https://academic.oup.com/jnci/article/112/9/913/5697920 by Stanford Law School Library user on 23 September 2020

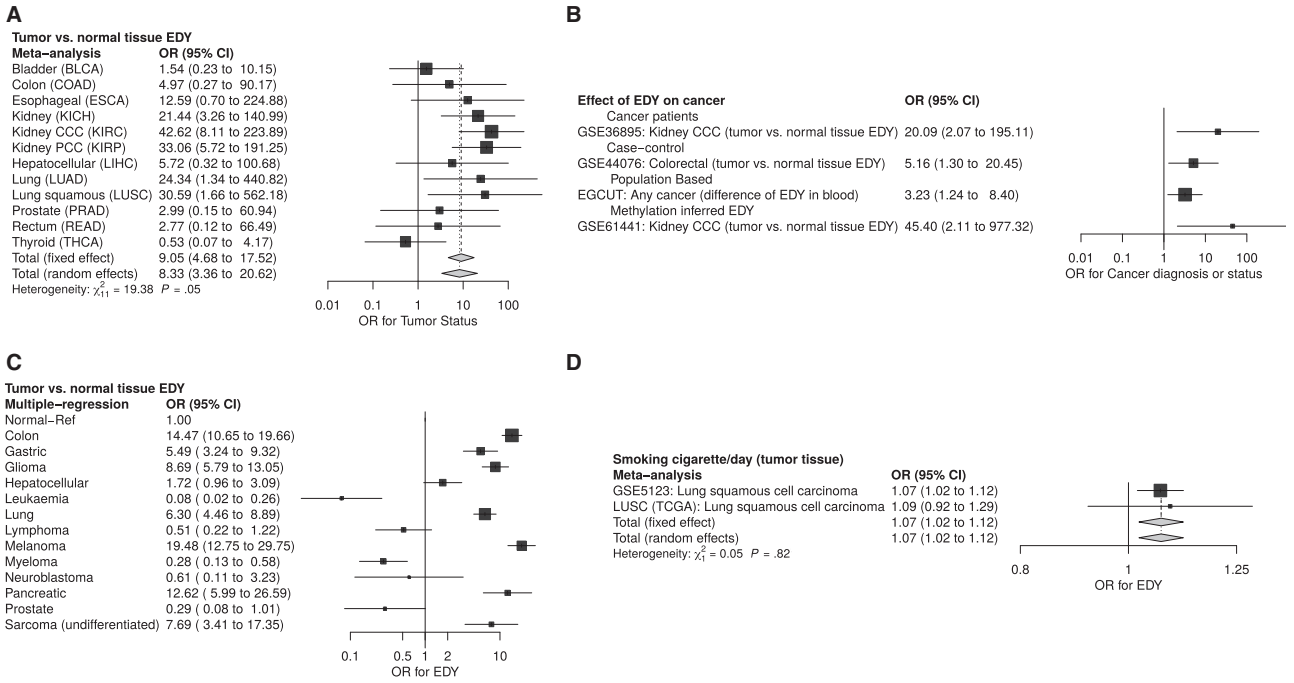

Figure 3. Extreme downregulation of chromosome Y gene expression (EDY) shown as a marker of cancer status of biological samples and individuals. The figure shows the association between EDY and cancer status across different independent studies with publicly available data. **A)** In The Cancer Genome Atlas (TCGA) study (n = 1774), the odds ratio (OR) of EDY for tumor status of the biological samples of cancer patients was obtained from logistic-regression models adjusting by age for 12 different cancers. The overall estimate of the effect of EDY was computed by a random effects meta-analysis and its heterogeneity with a χ^2 test. *P* values are two-sided. **B)** The association between EDY and cancer status of individuals was independently tested in colorectal (n = 142) and kidney (n = 29) cancer case-control studies (GSE66836, GSE36895) and in a population sample of 550 individuals from the Estonian Genome Center of the University of Tartu study. **C)** A large transcriptomic dataset (n = 6898) was used to assess EDY's association with multiple cancer diagnoses (E-MTAB-3732). **D)** Two studies on lung squamous cell carcinoma (GSE5123, LUSC from TCGA, total n = 243) were used to test the association between EDY and cigarettes smoked per day. CI = confidence interval; LUSC = lung squamous cell carcinoma.

and ZFY) regulate the cell cycle through different mechanisms and behave as dosage-sensitive tumor suppressors. In addition to sex-biased LoF mutations on the gene copies of the X chromosome, men would have a higher risk of first or second hits affecting the NRY copies, revealed by EDY derived from LOY and/or other genomic mechanisms associated with NRY hypermethylation. One of the main mechanisms of NRY hypermethylation seems to be related to *EGFR* gene dosage and polymorphisms in the pathway. *EGFR* codes for Epidermal Growth Factor Receptor, one of the four members of the ErbB family of tyrosine kinase receptors whose catalytic activation leads to increase DNA methyltransferase activity, resulting in increased global DNA methylation in some cancers (31,32). Our data also support a role for the EGFR pathway in the process of accumulated DNA methylation affecting the Y chromosome in male cancer progression.

Recent studies indicate that the risk factors of LOY include aging, smoking, and air pollution (10,27,33). Given that they are also risk factors for cancer, they can confound the association between EDY and cancer. More detailed studies into the relationship of EDY with these and other cancer-related factors are needed to determine their role in EDY vs LOY susceptibility. Here, we observed that in tumor vs healthy and cancer vs control studies, EDY was not associated with age or likewise cancer. In these studies, where age was matched, we observed that LOY, however, had a statistically significant association with age, suggesting neutral events deriving in LOY but not EDY. We additionally observed in the TCGA study that adjustment for smoking did not change the statistical significance of the association between EDY and cancer. Although specific studies are needed to characterize these and other risk factors of EDY, our highly reliable predictions from methylation profiles indicate a strong role of environmental exposures. The methylation EDY predictor is available at http://github.com/isglobal-brge/EDY, so its adequacy as a diagnostic or prognostic tool in different male cancers can be further tested in longitudinal studies.

Funding

This research has received funding from Ministerio de Ciencia, Innovación y Universidades de España and Fondo Europeo de Desarrollo, UE (RTI2018-100789-B-I00). The LAP-J laboratory is funded by the Catalan Department of Economy and Knowledge (SGR2014/1468, SGR2017/1974, and ICREA Acadèmia) and also acknowledges support from the Spanish Ministry of Economy and Competitiveness "Programa de Excelencia María de Maeztu" (MDM-2014–0370). We acknowledge support from the Spanish Ministry of Science, Innovation and Universities through the "Centro de Excelencia Severo Ochoa 2019-2023" Program (CEX2018-000806-S), and support from the Generalitat de Catalunya through the CERCA Program

Notes

The funders had no role in the design of the study; the collection, analysis, and interpretation of the data; the writing of the

ARTICLE

Downloaded from https://academic.oup.com/jnci/article/112/9/913/5697920 by Stang Mongkolsuk Library user on 23 September 2020

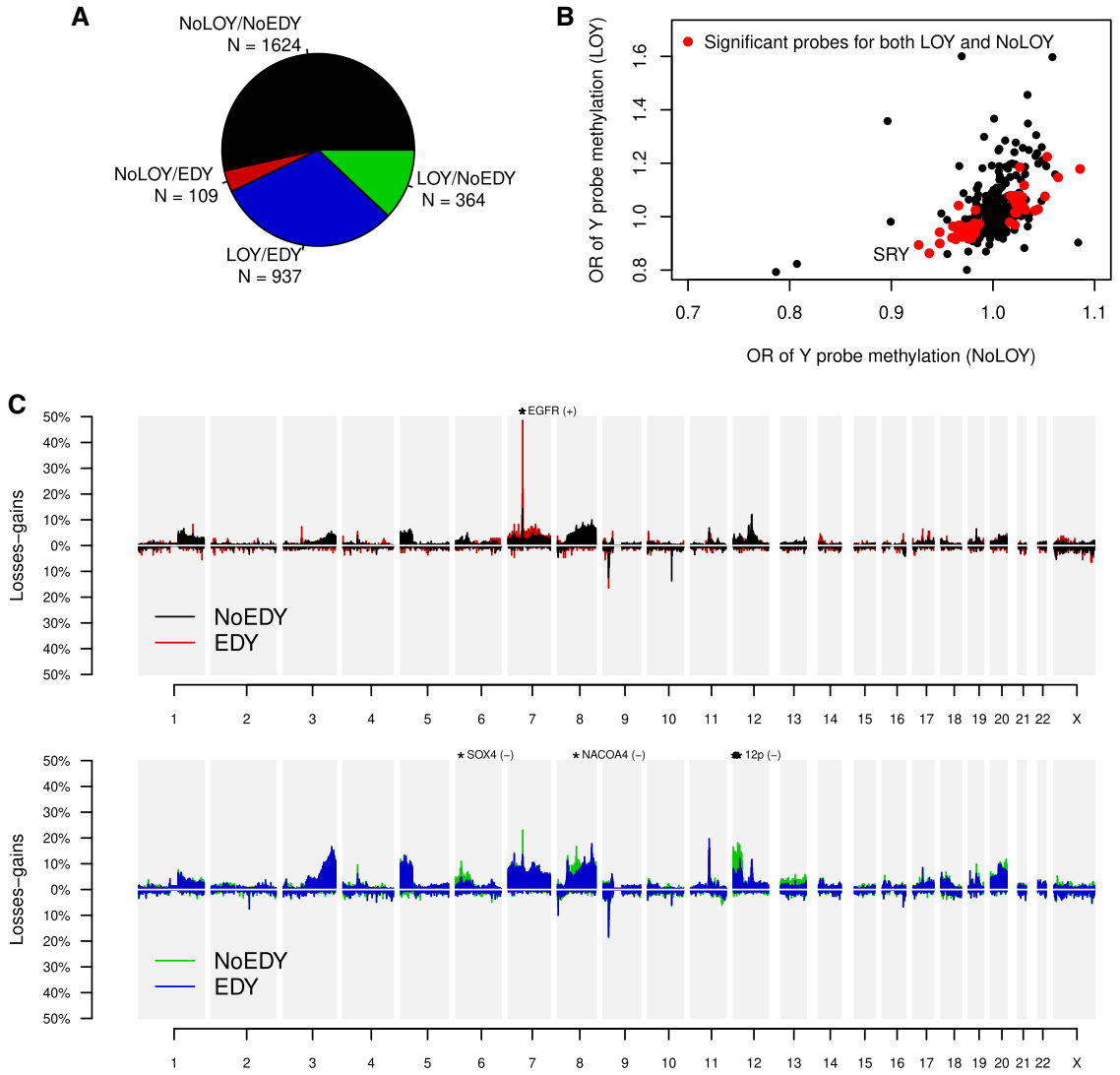

Figure 4. Association of extreme downregulation of chromosome Y gene expression (EDY) with chromosome Y methylation and genome-wide copy number variant proportion for individuals with and without loss of chromosome Y (LOY). **A**) Number of cancer samples in all four LOY and EDY statuses across 12 cancer studies in The Cancer Genome Atlas (BLCA, COAD, ESCA, KICH, KIRC, KIRP, LIHC, LUAD, LUSC, PRAD, READ, and THCA). **B**) Odds ratios (OR) of EDY for methylation sites across Y, stratified by LOY (LOY: **vertical axis**, no LOY: **horizontal axis**). A total of 52 statistically significant associations for both LOY statuses are in **red**. Associations for the SRY (sex determining region Y) gene are shown. **C**) Genome-wide differences in copy number variant (CNV) proportion, positive (+) and negative (−), between no EDY and EDY, and stratified by LOY (no LOY: **top**, LOY: **bottom**). BLCA = bladder urothelial carcinoma; COAD = colon adenocarcinoma; ESCA = esophageal carcinoma; KICH = kidney chromophobe; KIRC = kidney renal clear cell carcinoma; KIRP = kidney renal papillary cell carcinoma; LIHC = liver hepatocellular carcinoma; LUAD = lung adenocarcinoma; LUSC = lung squamous cell carcinoma; PRAD = prostate adenocarcinoma; READ = rectum adenocarcinoma; THCA = thyroid carcinoma.

manuscript; and the decision to submit the manuscript for publication. We would like to thank Francisco Real for his critical reading of the manuscript.

LAP-J is a founding partner and scientific advisor of qGenomics Laboratory. All other authors declare no conflict of interest.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*. 2019; 69(1):7–34.
2. Clocchiatti A, Cora E, Zhang Y, et al. Sexual dimorphism in cancer. *Nat Rev Cancer*. 2016;16(5):330–339.
3. Bianchi NO. Y chromosome structural and functional changes in human malignant diseases. *Mutat Res Rev Mutat Res*. 2009;682(1):21–27.
4. Wright DJ, Day FR, Kerrison ND, et al. Genetic variants associated with mosaic Y chromosome loss highlight cell cycle genes and overlap with cancer susceptibility. *Nat Genet*. 2017;49(5):674–679.
5. Duijf PH, Schultz N, Benezra R. Cancer cells preferentially lose small chromosomes. *Int J Cancer*. 2013;132(10):2316–2326.
6. Dürrbaum M, Storchová Z. Effects of aneuploidy on gene expression: implications for cancer. 2016;283(5):791–802.
7. Klatte T, Rao PN, De Martino M, et al. Cytogenetic profile predicts prognosis of patients with clear cell renal cell carcinoma. *J Clin Oncol*. 2009;27(5):746–753.
8. Nomdedeu M, Pereira A, Calvo X, et al. Clinical and biological significance of isolated y chromosome loss in myelodysplastic syndromes and chronic myelomonocytic leukemia. A report from the Spanish MDS Group. *Leuk Res*. 2017; 63:85–89.
9. Minner S, Kilgué A, Stahl P, et al. Y chromosome loss is a frequent early event in urothelial bladder cancer. *Pathology*. 2010;42(4):356–359.
10. Zhou W, Machiela MJ, Freedman ND, et al. Mosaic loss of chromosome Y is associated with common variation near tcl1a. *Nat Genet*. 2016;48(5):563–568.
11. Forsberg LA, Rasi C, Malmqvist N, et al. Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer. *Nat Genet*. 2014;46(6):624–628.
12. Rodríguez-Santiago B, Malats N, Rothman N, Armengol L, et al. Mosaic uniparental disomies and aneuploidies as large structural variants of the human genome. *Am J Hum Genet*. 2010;87(1):129–138.

Downloaded from https://academic.oup.com/jnci/article/112/9/913/5697920 by Stanford University Libraries user on 23 September 2020

13. Holland AJ, Cleveland DW. Losing balance: the origin and impact of aneuploidy in cancer: 'exploring aneuploidy: the significance of chromosomal imbalance' review series. *EMBO Rep*. 2012;13(6):501–514.
14. Loftfield E, Zhou W, Graubard BI, et al. Predictors of mosaic chromosome y loss and associations with mortality in the UK Biobank. *Sci Rep*. 2018;8(1):12316.
15. Thompson D, Genovese G, Halvardson J, et al. Genetic predisposition to mosaic Y chromosome loss in blood. Nature. *2020;575(7784):652–657*.
16. Fehrmann RS, Karjalainen JM, Krajewska M, et al. Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat Genet*. 2015;47(2):115–125.
17. Carter SL, Eklund AC, Kohane IS, et al. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat Genet*. 2006;38(9):1043–1048.
18. González JR, González-Carpio M, Hernández-Sáez R, et al. FTO risk haplotype among early onset and severe obesity cases in a population of western Spain. *Obesity*. 2012;20(4):909–915.
19. González JR, López-Sánchez M, Cáceres A, et al. A robust estimation of mosaic loss of chromosome Y from genotype-array-intensity data to improve disease risk associations and transcriptional effects. *BioRxiv*. 2019. https://doi.org/10.1101/764845. Accessed February 10, 2020.
20. Torrente A, Lukk M, Xue V, et al. Identification of cancer related genes using a comprehensive map of human gene expression. *PloS One*. 2016;11(6):e0157484.
21. Astle WJ, Elding H, Jiang T, et al. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell*. 2016;167(5):1415–1429.e19.
22. Su L, Zhou W, Asomaning K, et al. Genotypes and haplotypes of matrix metalloproteinase 1, 3 and 12 genes and the risk of lung cancer. *Carcinogenesis*. 2006;27(5):1024–1029.
23. Li Y, Sun DL, Duan YN, et al. Association of functional polymorphisms in MMPs genes with gastric cardias adenocarcinoma and esophageal squamous cell carcinoma in high incidence region of north China. *Mol Biol Rep*. 2010;37(1):197–205.
24. Churg A, Wang RD, Tai H, et al. Macrophage metalloelastase mediates acute cigarette smoke–induced inflammation via tumor necrosis factor-α release. *Am J Respir Crit Care Med*. 2003;167(8):1083–1089.
25. Klarin D, Global Lipids Genetics Consortium, Damrauer SM, et al. Genetics of blood lipids among ~300,000 multi-ethnic participants of the million veteran program. *Nat Genet*. 2018;50(11):1514–1523.
26. Dunford A, Weinstock DM, Savova V, et al. Tumor-suppressor genes that escape from x-inactivation contribute to cancer sex bias. *Nat Genet*. 2017;49(1):10–16.
27. Dumanski JP, Rasi C, Lönn M, et al. Smoking is associated with mosaic loss of chromosome Y. *Science*. 2015;347(6217):81–83.
28. Arseneault M, Monlong J, Vasudev NS, et al. Loss of chromosome y leads to down regulation of KDM5D and KDM6C epigenetic modifiers in clear cell renal cell carcinoma. *Sci Rep*. 2017;7(1):44876.
29. Nishino K, Hattori N, Tanaka S, Shiota K. DNA methylation-mediated control of SRY gene expression in mouse gonadal development. *J Biol Chem*. 2004;279(21):22306–22313.
30. Chyou PH, Nomura AM, Stemmermann GN. A prospective study of the attributable risk of cancer due to cigarette smoking. *Am J Public Health*. 1992;82(1):37–40.
31. Samudio-Ruiz SL, Hudson LG. Increased DNA methyltransferase activity and DNA methylation following epidermal growth factor stimulation in ovarian cancer cells. *Epigenetics*. 2012;7(3):216–224.
32. Bjaanaes MM, Fleischer T, Halvorsen AR, et al. Genome-wide DNA methylation analyses in lung adenocarcinomas: association with EGFR, KRAS and TP53 mutation status, gene expression and prognosis. *Mol Oncol*. 2016;10(2):330–343.
33. Wong JY, Margolis HG, Machiela M, et al. Outdoor air pollution and mosaic loss of chromosome Y in older men from the Cardiovascular Health Study. *Environ Int*. 2018;116:239–247.

Downloaded from https://academic.oup.com/jnci/article/112/9/913/5697920 by Stang Mongkolsuk Library user on 23 September 2020