Investigation of leukocyte telomere length and genetic variants in chromosome 5p15.33 as prognostic markers in lung cancer

Linda Kachuri[1,2,3], Jens Helby[4], Stig Egil Bojesen[4,5], David C. Christiani[6,7], Li Su[6], Xifeng Wu[8], Adonina Tardón[9], Guillermo Fernández-Tardón[9], John K. Field[10], Michael P. Davies[10], Chu Chen[11], Gary E. Goodman[11], Frances A. Shepherd[12], Natasha B. Leighl[12], Ming S. Tsao[13,14,15], Yonathan Brhane[1], M. Catherine Brown[13], Kevin Boyd[13], Daniel Shepshelovich[13,16], Lei Sun[17,18], Christopher I. Amos[19], Geoffrey Liu* [2,13,14], Rayjean J. Hung* [1,2]

*G.L and R. J. H. share senior authorship

1. Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, ON, Canada
2. Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
3. Department of Epidemiology & Biostatistics, University of California San Francisco, San Francisco, CA, USA
4. Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Denmark
5. Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
6. Departments of Epidemiology and Environmental Health, Harvard TH Chan School of Public Health, Boston, MA, USA
7. Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA
8. Department of Epidemiology, Division of Cancer Prevention and Population Sciences, The University of Texas MD Anderson Cancer Center, Houston, TX, USA
9. University of Oviedo and CIBERESP, Faculty of Medicine, Campus del Cristo, Oviedo, Spain
10. Roy Castle Lung Cancer Research Programme, Institute of Translational Medicine, Department of Molecular & Clinical Cancer Medicine, University of Liverpool, Liverpool, UK
11. Fred Hutchinson Cancer Research Center, Seattle, WA, USA
12. Cancer Clinical Research Unit, Princess Margaret Cancer Centre, University Health Network, Toronto, ON, Canada
13. Princess Margaret Cancer Center, University Health Network, Toronto, ON, Canada
14. Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada
15. Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada
16. The Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel
17. Department of Statistical Sciences, University of Toronto, Toronto, ON, Canada
18. Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
19. Institute for Clinical and Translational Research, Baylor College of Medicine, Houston, TX, USA

RUNNING TITLE: Telomere length and chromosome 5p15 in lung cancer prognosis

KEY WORDS: lung cancer, survival, prognosis, telomere length, chromosome 5p15.33

CORRESPONDING AUTHOR

Rayjean J. Hung, Ph.D., M.S.

Lunenfeld-Tanenbaum Research Institute of Sinai Health System
Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto
60 Murray St. Room L5-215, Box 18
Toronto ON M5T 3L9 Canada
Tel: (416) 586-4750; Fax: (416) 586-8404
E-mail: rayjean.hung@lunenfeld.ca

CONFLICT OF INTEREST

The authors have no potential conflicts of interest to disclose.

ABSTRACT

Background: Lung cancer remains the leading cause of cancer mortality with relatively few prognostic biomarkers. We investigated associations with overall survival for telomere length (TL) and genetic variation in chromosome 5p15.33, an established telomere maintenance locus.

Methods: Leukocyte TL was measured after diagnosis in 807 non-small cell lung cancer (NSCLC) patients from the Princess Margaret Cancer Center in Toronto and assessed prospectively in 767 NSCLC cases from the Copenhagen City Heart Study and the Copenhagen General Population Study. Associations with all-cause mortality were tested for 723 variants in 5p15.33, genotyped in 4672 NSCLC cases.

Results: Short telomeres ($\leq$10th percentile) were associated with poor prognosis for adenocarcinoma in both populations: TL measured 6 months after diagnosis (hazard ratio (HR) = 1.65, 95% confidence intervals (CI): 1.04-2.64) and for those diagnosed within 5 years after blood sampling (HR=2.42, 1.37-4.28). Short TL was associated with mortality in never smokers with NSCLC (HR=10.29, 1.86-56.86) and adenocarcinoma (HR=11.31, 1.96-65.24). Analyses in 5p15.33 identified statistically significant prognostic associations for rs56266421-G in *LPCAT1* (HR=1.86, 1.38-2.52, P=4.5×10^{-5}) in Stage I-IIIA NSCLC, and for the SLC6A3 gene with OS in females with NSCLC (P=1.6×10^{-3}).

Conclusions: Our findings support the potential clinical utility of TL, particularly for adenocarcinoma patients, while associations in chromosome 5p15.33 warrant further exploration.

Impact: This is the largest lung cancer study of leukocyte TL and OS, and first to examine the impact of the timing of TL measurement. Our findings suggest that extremely short telomeres are indicative of poor prognosis in NSCLC.

INTRODUCTION

Lung cancer is the leading cause of cancer mortality, accounting for 19.4% of all deaths among cancer patients, with an estimated 5-year survival rate of 15% (1,2). Non-small cell lung cancer (NSCLC) accounts for 80-85% of lung cases, and is less aggressive than small cell lung carcinoma (SCLC), which has a 5-year survival rate of 6.3% (3). The prognosis of NSCLC patients with resected tumours varies drastically by stage, with 5-year survival estimates ranging from 73% to 25% for Stage IA and IIIA, respectively (4). However, only 16% of patients present with the most localized disease, and for the majority of NSCLC cases treatment with platinum-based chemotherapy brings modest benefits (3,4). Despite the success of targeted therapies for epidermal growth factor receptor (*EGFR*) mutations and anaplastic lymphoma kinase (*ALK*) rearrangements, the prevalence of actionable molecular targets remains low (5). The present study aims to contribute to the identification of prognostic biomarkers in lung cancer by investigating the association with overall survival for leukocyte telomere length and genetic variation in chromosome 5p15.33, which includes the telomerase reverse transcriptase (*TERT*) gene.

Telomeres consist of tandem repeats of the TTAGGG sequence, which prevent chromosome ends from being perceived as DNA breaks and engaging unwarranted DNA damage repair mechanisms (6,7). Human telomeres lose between 50 and 200 base pairs with each replication, and progressively erode until senescence or apoptosis occurs (6,7). Although telomere attrition can be countered by telomerase, which synthesizes *de novo* TTAGGG repeats, it is not expressed in most normal somatic tissues (8). In contrast, telomerase is overexpressed in 85–90% of tumours, which is an important step for the immortalization and continued proliferation of cancer cells (9,10). In addition to *TERT*, which is critical for the functionality of telomerase and maintenance of telomeres (11,12), the 5p15.33 region also includes *CLPTM1L* (cleft lip and palate-associated transmembrane 1 like protein), which plays a role in cisplatin-induced apoptosis (13). Together these genes denote an established cancer susceptibility locus in 5p15.33 for lung and other cancers (14,15). Beyond its canonical telomere-related functions, *TERT* may also potentiate cell proliferation by activating *EGFR* signaling (16,17).

In contrast to lung cancer susceptibility, consistent evidence linking 5p15.33 variants to patient outcomes has been lacking (18-20). Generally, the overlap between loci influencing lung cancer susceptibility and survival has been limited, and few prognostic variants have been identified in genome-wide scans and candidate gene approaches (19-21). The apparent lack of associations may be due to inadequate statistical power and heterogeneity between patient populations, and it is also possible that less common variants may have been overlooked in previous studies that have selectively examined prognostic associations for common lung cancer susceptibility variants.

Telomere length (TL) has long been an active area of clinical research, with ongoing efforts to develop cancer therapeutics targeting telomerase. However, the findings of epidemiological studies to date have been inconsistent, partly due to differences in patient populations and tissues in which TL was measured. Most studies of lung cancer have examined telomere alterations in tumour tissue, with only one analysis of leukocyte TL observing an inverse relationship with mortality (22). It appears that shorter TL in tumour tissue may be a marker of poor prognosis, but evidence for the clinical utility of leukocyte TL remains sparse (18). In particular, it remains unclear whether telomere status observed at the time of clinical presentation contributes useful prognostic information.

We present two complementary survival analyses of leukocyte TL, its genetic determinants, and genetic variation in the 5p15.33 region. First, we assessed the prognostic potential of leukocyte TL, measured at two time points: at the time of lung cancer diagnosis in a cohort from the Princess Margaret Cancer Center in Toronto, and prior to cancer diagnosis in two population-based cohorts, the Copenhagen City Heart Study (CCHS) and the Copenhagen General Population Study (CGPS) (22,23). Next, we investigated associations with overall survival in chromosome 5p15.33, including rare variants previously identified through deep sequencing (24,25), in 4672 NSCLC cases from the OncoArray collaboration (26). Since mechanisms of cancer development and progression may be distinct, we extended the scope of our investigation beyond established susceptibility genes, *TERT* and *CLPTM1L*, to provide a more comprehensive analysis of the 5p15.33 region. Second, we examined prognostic associations between genetic determinants of TL and mortality in the OncoArray NSCLC population. In addition, to link the two approaches, we investigated whether the TL prognostic effect on lung cancer was modified by 5p15 genetic variants.

MATERIALS & METHODS

Study Populations

Telomere Length

TL was measured in 1127 lung cancer cases from the Mount Sinai Hospital-Princess Margaret Cancer Centre Case Control Study (MSH-PMH). Eligible patients were recruited from the Princess Margaret Cancer Center in Toronto between 2006 and 2013, and included individuals older than 18 years of age, with a histologically confirmed, incident lung cancer diagnosis. Blood samples were collected at enrolment or the next earliest clinic visit (median lag time: 27 days after diagnosis). Information on clinical characteristics and vital status were extracted from electronic medical records and the Ontario Cancer Registry. Since TL can be affected by a number of factors following cancer diagnosis (27,28), patients who provided blood samples more than 1 year after diagnosis were considered ineligible, leaving a total of 888 cases (807 NSCLC). The population for the main association analyses was further restricted to 823 patients (745 NSCLC) with DNA collected within 6 months of diagnosis.

To assess whether pre-diagnostic TL is associated with lung cancer prognosis, 923 incident lung cancer cases were identified in the CCHS, enrolled between 1991 and 1994, and the CGPS, enrolled between 2003 and 2012. Both studies have been described previously (22,23,29). Briefly, participants completed questionnaires, underwent a physical examination, and provided blood samples at enrollment. Diagnoses of invasive cancer, with corresponding dates and disease stage, were obtained from the national Danish Cancer Registry, which identifies >98% of all cancers in Denmark (30). Vital status and date of death was obtained from the Danish Civil Registration system, which is 100% accurate.

In order to contrast TL associations with prognosis, and distinguish from the effects of aging, we conducted a separate analysis of TL and mortality in 2905 controls without lung cancer selected from the CCHS and CGPS, matching on birthdate, date of blood sample collection, sex, and cumulative smoking in pack-years. The median difference between the birthdate in cases and their matched controls was 4.7 months, with 75% of controls having a birthdate within 1 year of the index lung cancer case, and 88% within 2 years. Eligible controls were not diagnosed with lung cancer at enrolment or during follow-up, but other cancers may have occurred.

TL was measured in DNA isolated from peripheral blood leukocytes using previously described quantitative polymerase chain reaction (qPCR) assays performed in Toronto (25) and Copenhagen studies (29). This method expresses relative TL as the T/S ratio between telomere (T) repeat copy numbers and single gene (S) copy numbers (details in Supplementary File).

Genetic variants in chromosome 5p15.33

In addition to Toronto MSH-PMH study, lung cancer patients from the following five studies were included: the Harvard Lung Cancer Susceptibility (HLCS), Cáncer de Pulmón en Asturias (CAPUA), the University of Texas MD Anderson Cancer Center (MDACC) lung cancer study, the Liverpool Lung Project (LLP), and the Carotene and Retinol Efficacy Trial (CARET). All patients were incident cases with histologically confirmed lung cancer. Detailed descriptions of participant recruitment and endpoint ascertainment have been previously published for the HLCS (31), MDACC (21), CAPUA (32), LLP (33), and CARET (34) studies. All studies were approved by research ethics review boards at the involved institutions and informed consent was obtained for all participants.

Descriptions of the genotyping and quality control methods for OncoArray data have been previously published (26). Custom content for the 5p15.33 region included variants identified through a targeted deep sequencing of 288 independent lung cancer case-control pairs (24). Analyses were restricted to individuals of ≥80% European ancestry (26). Variants were filtered to retain those with a minor allele frequency (MAF) of 0.001 or greater in the entire OncoArray lung cancer dataset, even if study-specific MAF was lower. A total of 723 variants (43 insertions/deletions), were analyzed. The overall genotyping rate was 0.9977.

Statistical Analysis

For individuals in the Toronto study, follow-up time was calculated from the date of diagnosis to date of death, or date of last visit if no death occurred prior to January 17, 2017. For individuals in the CGPS and CCHS, follow up time was calculated from the date of diagnosis to date of death, or March 22, 2017, whichever came first. The end point for all analyses was death from any cause, with patients considered censored if no death occurred within 5 years of follow-up time. Cox proportional hazards regression was used to estimate the hazard ratios (HRs) and corresponding 95% confidence intervals (CIs) for all-cause mortality. All analyses were adjusted for the following variables of prognostic significance: disease stage (early: Stage I-IIIA; advanced: Stage IIIB-IV), age at diagnosis, sex, and cigarette pack-year categories: 0 (never smokers with fewer than 100 lifetime cigarettes), >0 to 20, >20 to 40, and >40.

Violation of the proportionality assumption was assessed by examining the association between standardized Schoenfeld residuals and time, and by plotting -ln(-ln(survival) against ln(analysis time). No major violations were detected. Associations were further refined in analyses stratified by lung cancer histology and established prognostic factors.

Leukocyte Telomere Length

TL was analyzed as a continuous and dichotomous variable. To facilitate comparison between studies, T/S ratio values were converted to normalized Z-scores in each study. Short TL was defined as length in the 10th percentile (Z ≤ -1.28) and the comparison group consisted of TL >10th percentile. In addition to standard prognostic covariates, Cox models were also adjusted for body-mass index (BMI) categories corresponding to normal weight (18.5 to <25), underweight (<18.5), overweight (25 to <30), obese (≥30), and unknown. To account for the number of comparisons conducted across strata defined by histology, sex, and smoking status, we also present false discovery rate (FDR) q-values.

Cox regression models in the Copenhagen study population were further adjusted for the time elapsed between blood sample collection and lung cancer diagnosis, and qPCR calibrator lot number, since percentiles were calculated separately for each lot. For comparative analyses of mortality in controls without lung cancer, follow-up time started from the date of lung cancer diagnosis in the matched case. Heterogeneity of associations was assessed using the likelihood ratio test p-value for comparing nested models with and without an interaction term between case-control status and TL.

To further characterize how the timing of TL measurement, analyses of pre-diagnostic TL in the Copenhagen cohort were restricted to lung cancers diagnosed within 5 years of blood sample collection, since only 6% of cases occurred within the first year.

Genetic variants in chromosome 5p15.33

Single-variant analyses and regional, gene-based analyses were carried out to explore predictors of survival in chromosome 5p15.33. For each variant, Cox models were fit to the number of copies of the minor allele, assuming a log-additive genetic model with 1 degree of freedom. To achieve sufficient power for rare variants in subgroup analyses, individual OncoArray studies were pooled. In addition to prognostic factors, study-specific analyses were adjusted for 5 genetic ancestry principal components (PCs), and pooled analyses were adjusted for 10 PCs.

For NSCLC, fixed effect meta-analysis was conducted using the R package *meta* (http://cran.r-project.org/web/packages/meta/). Heterogeneity across studies was assessed with the Cochran Q test (P_{het}) and I^2. Cox models for pooled analyses incorporated a frailty term for study to account for heterogeneity between the study populations and non-independence in survival outcomes between patients from the same center. The estimated number of effective independent tests (35) was applied into the Bonferroni correction formula to determine the significance threshold for single variant analysis. The significance threshold was set at $P<1.6\times10^{-4}$ based on 306 independent tests. The level for suggestive associations was set at one order of magnitude higher ($P<1.6\times10^{-3}$).

Gene-level tests were conducted using Cox SKAT-LRT (36). All variants in the 5p15.33 target region were assigned to mutually exclusive sets, based on RefSeq genes downloaded from the UCSC Genome Browser and GENCODE version 13, implemented using HaploReg v4.1 (37) . Each study was analyzed separately, adjusting for appropriate covariates. Single-variant test statistics were aggregated by gene and combined across studies using Cox SKAT-LRT and meta-analysis implemented within the R package *seqMeta* (http://cran.r-project.org/web/packages/seqMeta/). Associations with $P<2.5\times10^{-3}$ were considered statistically significant after Bonferroni correction for 5 genes tested in four NSCLC subgroups.

RESULTS

Leukocyte Telomere Length

Demographic and clinical characteristics of the Toronto MSH-PMH participants are presented in Table 1. Median survival time (MST) was 26.2 months (95% CI: 22.7-30.1) for the MSH-PMH cases, 29.4 months (95% CI: 25.4 - 36.9) for 807 NSCLC and 12.3 months (95% CI: 10.3-17.3) for 74 SCLC patients. The mean age at diagnosis was 67 years and was negatively correlated with TL (Pearson's r = -0.25, $P=6.4\times10^{-14}$). Over half the patients (57.1%) were diagnosed with early stage disease. Several prognostic variables were associated with TL (Table 1), including stage ($P=4.5\times10^{-3}$), histology ($P=1.8\times10^{-3}$), and smoking status ($P=1.5\times10^{-3}$). Patients with adenocarcinoma and never smokers were characterized by longer leukocyte TL. *EGFR* mutation and *ALK* translocation status was ascertained in 58 NSCLC patients. Mean TL was significantly longer in the patients with *EGFR* mutation positive (n=26) or *ALK* rearranged tumours (n=3), when compared to the 29 non-*EGFR*/non-*ALK* patients ($P=1.6\times10^{-3}$).

Baseline characteristics of the Copenhagen cohorts are presented in Supplementary Table S1. Most of the lung cancer cases (91.1%) were over 60 years of age. Lung cancer cases diagnosed in Copenhagen were predominantly smokers (95.4%), and 50.9% presented with metastatic disease (Stage IV) at diagnosis. Information on somatic mutation status was not available in these cohorts.

Telomere Length Measured At The Time of Diagnosis

In the Toronto MSH-PMH study, survival outcomes were examined in 823 patients with DNA samples collected up to 6 months following diagnosis (Table 3). Among the 745 NSCLC cases, short leukocyte TL (≤10th percentile) was not associated with 5-year all-cause mortality (HR=1.16, 95% CI: 0.83-1.63). Compared to patients with long TL, short TL was associated with a higher mortality risk in females with NSCLC (HR=1.88, 1.11-3.17; n=358), but not in males (HR=0.93, 0.59-1.33). A significant association with poor prognosis was observed among never smokers with short TL (HR=10.29, 1.86-56.86; q_{FDR}<0.05).

In histology-specific analyses, short TL was associated with an increased mortality risk in 459 patients with adenocarcinoma (HR=1.65, 1.04-2.64). The magnitude of this association was larger in females (HR=2.54, 1.20-5.38, n=248) and never smokers with short TL (HR=11.31, 1.96-65.24; q_{FDR}<0.05, n=77). No associations with prognosis were observed for patients with squamous cell carcinoma (HR=0.69, 0.33-1.45) and SCLC (HR=0.62, 0.26-1.45). Association estimates for short TL could not be reliably estimated in never smokers with squamous carcinoma (8 cases) and in strata defined by *EGFR* or *ALK* mutation status (short TL: n=1).

When analyzing TL values as continuous, we observed decreased mortality in females with NSCLC (HR=0.81, 0.67-0.98) and adenocarcinoma (HR=0.72, 0.56-0.93) per unit increase in the standardized Z-score (Supplementary Table S2). Relative TL was not associated with survival in strata defined by *EGFR* or *ALK* status. Deviations from the proportionality assumption (p<0.05) were not observed for either TL metric.

Sensitivity analyses using less restrictive selection criteria, including cases with DNA obtained up to 12 months following diagnosis (807 NSCLC), resulted in a similar pattern of associations, although the HR estimates were attenuated (Supplementary Table S3).

Pre-Diagnostic Telomere Length

Association between TL measured prior to cancer diagnosis and all-cause mortality in the CCHS and CGPS are presented in Table 4 (continuous TL in Supplementary Table S4). Among 767 NSCLC cases, short TL (≤10th percentile) at cohort entry (baseline) was associated with an increased 5-year mortality risk (HR=1.52, 1.18-1.95; q_{FDR}<0.05), particularly in females (HR=1.61, 1.11-2.35, q_{FDR}<0.05; n=361). In histology-specific analyses short TL was significantly associated with mortality for adenocarcinoma (HR=1.48, 1.01-2.16). No significant associations were found for

squamous carcinoma (HR=1.45, 0.94-2.25) and SCLC (HR=1.42, 0.74-2.74). In matched controls without lung cancer short TL was not associated with all-cause mortality.

Next, we examined associations with TL measured more proximally to lung cancer diagnosis (Table 4). Among lung cancers diagnosed within 5 years after blood sampling, short telomeres were associated with poor prognosis for lung cancer overall (HR=1.89, 1.34-2.66) and all NSCLC subtypes, including adenocarcinoma (HR=2.42, 1.37-4.28, q_{FDR}<0.05). A statistically significant association between short TL and survival observed in females (HR=4.06, 2.08-7.92; q_{FDR}<0.05), but not in males with NSCLC (HR=1.37, 0.82-2.28). In contrast, short TL was not associated with mortality in controls matched to cases with NSCLC (HR=0.68, 0.40-1.16) or adenocarcinoma (HR=0.70, 0.30-1.63). Significant interactions between TL and case-control status on overall survival was observed for NSCLC (*P*=0.01) and adenocarcinoma (*P*=0.02), suggesting the presence of effect heterogeneity in these subgroups. Analyses stratified by smoking status were not pursued due to the high prevalence of current or former smoking (95.4%) in this population.

Genetic variants in chromosome 5p15.33

Genetic association analyses were based on a pooled sample of 4672 NSCLC cases from six studies, including 1149 from the Toronto MSH-PMH study (Table 2; Supplementary Figure 2). Adenocarcinoma was the most common histology, with 2717 cases (58.2%), followed by squamous carcinoma with 1406 cases (30.1%). Of the 723 variants analyzed in chromosome 5p15.33, none were associated with overall survival in NSCLC at *P* value of <1.6×10⁻⁴, but an indel in *LPCAT1*, Chr5:1463124-TA, was suggestively associated (HR=2.19, 95% CI: 1.46-3.32, *P*=1.8×10⁻⁴; MAF=0.004; I^2=0.45, P_{het}=0.06; Figure 1A, Supplementary Figure 3). A region of high linkage disequilibrium (LD) between *TERT* and *CLPTM1L* (Figure 1B) included a variant suggestively associated with 5-year survival in adenocarcinoma patients (rs4975616-G: HR=1.14, 1.06-1.22, *P*=6.6×10⁻⁴; MAF=0.38), with another suggestive signal observed in *SLC6A3* (rs2617606-A: HR=1.66, 1.27-2.18, *P*=2.1×10⁻⁴; MAF=0.016). For squamous carcinoma (Figure 1C), a significant association was observed for Chr5:1346940-C (HR=1.92, 1.37-2.70, *P*=1.6×10⁻⁴; MAF: 0.015).

Stratifying by disease stage revealed a significant association with mortality for an *LPCAT1* variant in patients with early stage disease (Figure 1D: rs56266421-G HR=1.86, 1.38-2.52, *P*=4.5×10⁻⁵; MAF=0.015). Sex-stratified analyses identified a suggestive signal in *SLC6A3* for females (Figure 1F), denoted by Chr5:1439260-G (HR=4.82, 2.12-10.95, *P*=1.8×10⁻⁴; MAF=0.001). Gene-based Cox SKAT-LRT analyses supported the association pattern emerging in the single-variant analysis (Table 5). Variants in *SLC6A3* were significantly associated with overall survival in females with NSCLC (*P*=1.6×10⁻³) and remained significant after restricting to rare variants with MAF≤0.01 (*P*=1.4×10⁻³).

Integrative Analysis of 5p15.33 and Telomere Length

To integrate telomere and genetic association analyses, we investigate whether any 5p15.33 variants modify the prognostic association between TL and NSCLC survival in a subset of 705 MSH-PMH patients that have both TL and OncoArray data. We examined interaction between short TL and rs2736100 (*TERT*), a variant known to be associated with leukocyte TL in cancer-free controls. There was some evidence of effect modification observed for rs2736100 (interaction *P*=0.037). In stratified analyses short TL was associated with increased NSCLC mortality, and reduced median survival (Supplementary Figure 4), in the rs2736100-CC group (adjusted HR=2.28, 1.20-4.37), but not in patients with rs2736100-CA/AA genotypes (adjusted HR=1.16, 0.74-1.68).

In addition to effect modification, we investigated whether established telomere loci were associated with mortality: *ACYP2* (rs10165485-G), *PXK* (rs6772228-G), *TERC* (rs10936599-A), *TERT* (rs2736100-A), *NAF1* (rs11100479-G), *OBFC1* (rs9420907-A), *ZNF676* (rs10419926-C), and *RTEL1* (rs755017-G). Adjusted HRs were estimated for the allele associated with decreasing TL in 6 OncoArray populations. Generally, regions involved in telomere maintenance did not seem to have prognostic significance (Supplementary Figure 5). A linear combination of all TL-associated loci was not significantly associated with mortality for NSCLC overall (HR=1.01. 0.99–1.04, p=0.34) or any histological subtypes.

DISCUSSION

This comprehensive analysis of two large, well-characterized lung cancer populations suggests that short leukocyte TL may serve as a marker of poor prognosis in NSCLC, particularly for adenocarcinoma. This association was confirmed in two independent study populations, with telomere length assessed at different time points. Other notable findings include the increased mortality risk observed in females and never smokers with short telomeres. However, as adenocarcinoma is the predominant NSCLC subtype in females and never smokers, these findings may primarily reflect a histology-specific relationship.

Other findings of interest include an observation that was previously reported (38) of longer leukocyte TL in patients with *EGFR* or *ALK* mutations. Although mutation status did not modify the association between TL and survival, statistical power was limited in this subgroup. Our integrative analyses suggest that the prognostic associations observed for leukocyte TL may be modified by rs2736100 (*TERT*) genotype. The adverse impact of short TL on mortality was significant only in patients with a genetic predisposition to longer telomeres (rs2736100-CC). However, genetic variants involved in telomere maintenance in 5p15.33 and other regions do not appear to have a direct effect on survival.

The present study offers novel insight into the role of telomeres in lung cancer survival by examining associations with TL observed at the time of diagnosis and clinical presentation. Our findings are consistent with an earlier analysis of the Copenhagen cohorts by Weischer et al. (22), which reported that shortened leukocyte TL, measured at cohort entry, was associated with an increased mortality risk in lung cancer patients. In addition to evaluating associations with TL at clinically relevant time points, our analysis extends previous work by including a 40% increase in the number of lung cancer cases from the Copenhagen population. We also refined prognostic associations by clinically distinct subgroups, such as histology, sex, and smoking status. The consistency of associations observed in the Copenhagen cohorts and Toronto patients suggests that leukocyte TL may capture patient characteristics indicative of prognosis both at the time of disease detection and years prior to NSCLC diagnosis. Short TL may be a marker of systemic factors, such as impaired immune competence (29,39), inflammation (40), and reduced regenerative capacity (41). Therefore, leukocyte TL serves as an integrative biomarker of heritable intrinsic factors, in addition to DNA damage accumulated due to hemodynamic, metabolic, and oxidative stressors.

Short telomeres have also been associated with increased all-cause mortality in the general population (23,42). However, the findings of our comparative analysis conducted in matched controls demonstrate that associations with survival in lung cancer patients are not indicative of a general survival advantage, since controls were carefully selected matching on relevant prognostic characteristics. These differences are also unlikely to be caused by selection bias due to the exclusion of individuals in poor health, since the control group included participants diagnosed with other cancers and medical conditions.

The results of the genetic association analyses suggest that in contrast to lung cancer risk, few prognostic variants are found in the 5p15.33 region. A significant association with survival in early stage NSCLC was observed for rs56266421 (*LPCAT1*), located in a regulatory region enriched for enhancer and promoter histone marks in several tissues, including lung (37). Another suggestive association was observed in adenocarcinoma patients for rs4975616 (*MIR4457*), an established lung cancer susceptibility locus (14) previously linked to overall and progression-free survival in advanced NSCLC (67% adenocarcinoma) (19). Gene-based analyses identified an association with mortality in females with NSCLC for the dopamine transporter *SLC6A3* gene, which contains a number of expression lung quantitative trait loci (43). Dopamine transporters are expressed in lung tumour cell lines (44), and functions related to tumour growth inhibition and regulation of angiogenesis have been proposed (45,46). Overall, our findings suggest that 5p15.33 genes previously implicated in lung cancer susceptibility, such as *TERT* and *CLPTM1L*, are less relevant for clinical endpoints. Similarly, genetic determinants of TL that exhibit strong associations with lung cancer risk, particularly adenocarcinoma (47,48), were not associated with survival outcomes in these patients. However, all-cause mortality is a composite outcome that may obscure

associations with more specific endpoints, such as a progression free survival and treatment response.

In evaluating the potential clinical implications of this work, several limitations should be acknowledged. Firstly, information on cause of death was not available for all studies, which precluded an analysis of lung cancer-specific mortality, treating other causes of death as competing risks. However, deaths in the majority of patients, particularly with advanced disease, can be directly or indirectly attributed to lung cancer. The accuracy of cause of death determination can also vary (49), making all-cause mortality an informative, and potentially less biased endpoint.

Secondly, treatment information was not available for all patient populations. Adjustment for stage may partly account for treatment patterns, since early-stage patients are typically eligible for surgery, whereas those with locally advanced and metastatic disease tend to receive chemotherapy. Thirdly, this analysis did not account for the impact of comorbid conditions on survival. Certain chronic conditions, such as congestive heart failure and chronic obstructive pulmonary disease, may have an independent effect on survival and contribute to adverse outcomes by preventing patients from receiving aggressive lung cancer treatment (50). However, this is somewhat mitigated by adjustment for disease stage, since the impact of these conditions may predominate in early-stage disease.

Despite these limitations, this work provides further insight into the potential clinical utility of telomere length and genetic variants in 5p15.33. In this large-scale study of well-characterized clinical populations, we demonstrated that leukocyte TL predicts survival in NSCLC and adenocarcinoma patients, at multiple time points, and this association is unlikely to simply reflect a relationship with longevity. Although associations between TL and mortality were also observed in never smokers and patients with other NSCLC histology, we urge caution in interpreting these associations, particularly findings observed in strata with few individuals.

The large size of the OncoArray study population enabled a thorough exploration of rare and novel variants in an important genetic region. However, even with a sample size of 4672 NSCLC cases, this analysis was underpowered for extremely rare variants and heterogeneity between study populations may have also reduced power for detecting associations with survival. In summary, this study supports the potential of leukocyte telomere length to contribute information that may improve prognostic assessment of NSCLC patients. Future studies are needed to validate these associations and inform clinically relevant thresholds of telomere attrition.

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M*, et al.* Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. International journal of cancer Journal international du cancer **2015**;136(5):E359-86 doi 10.1002/ijc.29210.
2. Canadian Cancer Society's Advisory Committee on Cancer Statistics. Canadian Cancer Statistics 2017. Toronto, ON: Canadian Cancer Society; 2017.
3. National Cancer Institute. SEER Cancer Statistics Review (CSR) 1975-2014. Bethesda, MD: National Cancer Institute; 2017.
4. Goldstraw P, Ball D, Jett JR, Le Chevalier T, Lim E, Nicholson AG*, et al.* Non-small-cell lung cancer. Lancet **2011**;378(9804):1727-40 doi 10.1016/S0140-6736(10)62101-0.
5. Johnson DH, Schiller JH, Bunn PA, Jr. Recent clinical advances in lung cancer management. Journal of clinical oncology : official journal of the American Society of Clinical Oncology **2014**;32(10):973-82 doi 10.1200/JCO.2013.53.1228.
6. Blackburn EH. Structure and function of telomeres. Nature **1991**;350(6319):569-73 doi 10.1038/350569a0.
7. de Lange T. Protection of mammalian telomeres. Oncogene **2002**;21(4):532-40 doi 10.1038/sj.onc.1205080.
8. Harley CB, Sherwood SW. Telomerase, checkpoints and cancer. Cancer Surv **1997**;29:263-84.
9. Newbold RF. The significance of telomerase activation and cellular immortalization in human cancer. Mutagenesis **2002**;17(6):539-50.
10. Borah S, Xi L, Zaug AJ, Powell NM, Dancik GM, Cohen SB*, et al.* Cancer. TERT promoter mutations and telomerase reactivation in urothelial cancer. Science **2015**;347(6225):1006-10 doi 10.1126/science.1260200.
11. Codd V, Nelson CP, Albrecht E, Mangino M, Deelen J, Buxton JL*, et al.* Identification of seven loci affecting mean telomere length and their association with disease. Nat Genet **2013**;45(4):422-7, 7e1-2 doi 10.1038/ng.2528.
12. Prescott J, Kraft P, Chasman DI, Savage SA, Mirabello L, Berndt SI*, et al.* Genome-wide association study of relative telomere length. PLoS One **2011**;6(5):e19635 doi 10.1371/journal.pone.0019635.
13. Yamamoto K, Okamoto A, Isonishi S, Ochiai K, Ohtake Y. A novel gene, CRR9, which was up-regulated in CDDP-resistant ovarian tumor cell line, was associated with apoptosis. Biochemical and biophysical research communications **2001**;280(4):1148-54 doi 10.1006/bbrc.2001.4250.
14. Landi MT, Chatterjee N, Yu K, Goldin LR, Goldstein AM, Rotunno M*, et al.* A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. Am J Hum Genet **2009**;85(5):679-91 doi 10.1016/j.ajhg.2009.09.012.
15. Rafnar T, Sulem P, Stacey SN, Geller F, Gudmundsson J, Sigurdsson A*, et al.* Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. Nat Genet **2009**;41(2):221-7 doi 10.1038/ng.296.
16. Low KC, Tergaonkar V. Telomerase: central regulator of all of the hallmarks of cancer. Trends Biochem Sci **2013**;38(9):426-34 doi 10.1016/j.tibs.2013.07.001.
17. Smith LL, Coller HA, Roberts JM. Telomerase modulates expression of growth-controlling genes and enhances cell proliferation. Nat Cell Biol **2003**;5(5):474-9 doi 10.1038/ncb985.
18. Kachuri L, Latifovic L, Liu G, Hung RJ. Systematic Review of Genetic Variation in Chromosome 5p15.33 and Telomere Length as Predictive and Prognostic Biomarkers for Lung Cancer. Cancer Epidemiol Biomarkers Prev **2016**;25(12):1537-49 doi 10.1158/1055-9965.EPI-16-0200.
19. Azad AK, Qiu X, Boyd K, Kuang Q, Emami M, Perera N*, et al.* A genetic sequence variant (GSV) at susceptibility loci of 5p15.33 (TERT-CLPTM1L) is associated with survival outcome in locally advanced and metastatic non-small-cell lung cancer (NSCLC). Lung cancer **2014**;84(3):289-94 doi 10.1016/j.lungcan.2014.03.008.
20. Xun WW, Brennan P, Tjonneland A, Vogel U, Overvad K, Kaaks R*, et al.* Single-nucleotide polymorphisms (5p15.33, 15q25.1, 6p22.1, 6q27 and 7p15.3) and lung cancer survival in the

European Prospective Investigation into Cancer and Nutrition (EPIC). Mutagenesis **2011**;26(5):657-66 doi 10.1093/mutage/ger030.

21. Wu X, Ye Y, Rosell R, Amos CI, Stewart DJ, Hildebrandt MA*, et al.* Genome-wide association study of survival in non-small cell lung cancer patients receiving platinum-based chemotherapy. J Natl Cancer Inst **2011**;103(10):817-25 doi 10.1093/jnci/djr075.

22. Weischer M, Nordestgaard BG, Cawthon RM, Freiberg JJ, Tybjaerg-Hansen A, Bojesen SE. Short telomere length, cancer survival, and cancer risk in 47102 individuals. J Natl Cancer Inst **2013**;105(7):459-68 doi 10.1093/jnci/djt016.

23. Rode L, Nordestgaard BG, Bojesen SE. Peripheral blood leukocyte telomere length and mortality among 64,637 individuals from the general population. J Natl Cancer Inst **2015**;107(6):djv074 doi 10.1093/jnci/djv074.

24. Zuzarte PC, Denroche RE, Fehringer G, Katzov-Eckert H, Hung RJ, McPherson JD. A two-dimensional pooling strategy for rare variant detection on next-generation sequencing platforms. PLoS One **2014**;9(4):e93455 doi 10.1371/journal.pone.0093455.

25. Kachuri L, Amos CI, McKay JD, Johansson M, Vineis P, Bueno-de-Mesquita HB*, et al.* Fine mapping of chromosome 5p15.33 based on a targeted deep sequencing and high density genotyping identifies novel lung cancer susceptibility loci. Carcinogenesis **2016**;37(1):96-105 doi 10.1093/carcin/bgv165.

26. Amos CI, Dennis J, Wang Z, Byun J, Schumacher FR, Gayther SA*, et al.* The OncoArray Consortium: a Network for Understanding the Genetic Architecture of Common Cancers. Cancer Epidemiol Biomarkers Prev **2016** doi 10.1158/1055-9965.EPI-16-0106.

27. Benitez-Buelga C, Sanchez-Barroso L, Gallardo M, Apellaniz-Ruiz M, Inglada-Perez L, Yanowski K*, et al.* Impact of chemotherapy on telomere length in sporadic and familial breast cancer patients. Breast Cancer Res Treat **2015**;149(2):385-94 doi 10.1007/s10549-014-3246-6.

28. Epel ES, Blackburn EH, Lin J, Dhabhar FS, Adler NE, Morrow JD*, et al.* Accelerated telomere shortening in response to life stress. Proc Natl Acad Sci U S A **2004**;101(49):17312-5 doi 10.1073/pnas.0407162101.

29. Helby J, Nordestgaard BG, Benfield T, Bojesen SE. Shorter leukocyte telomere length is associated with higher risk of infections: a prospective study of 75,309 individuals from the general population. Haematologica **2017**;102(8):1457-65 doi 10.3324/haematol.2016.161943.

30. Storm HH. Completeness of cancer registration in Denmark 1943-1966 and efficacy of record linkage procedures. International journal of epidemiology **1988**;17(1):44-9.

31. Heist RS, Zhai R, Liu G, Zhou W, Lin X, Su L*, et al.* VEGF polymorphisms and survival in early-stage non-small-cell lung cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology **2008**;26(6):856-62 doi 10.1200/JCO.2007.13.5947.

32. Marin MS, Lopez-Cima MF, Garcia-Castro L, Pascual T, Marron MG, Tardon A. Poly (AT) polymorphism in intron 11 of the XPC DNA repair gene enhances the risk of lung cancer. Cancer Epidemiol Biomarkers Prev **2004**;13(11 Pt 1):1788-93.

33. Field JK, Smith DL, Duffy S, Cassidy A. The Liverpool Lung Project research protocol. International journal of oncology **2005**;27(6):1633-45.

34. Omenn GS, Goodman G, Thornquist M, Grizzle J, Rosenstock L, Barnhart S*, et al.* The beta-carotene and retinol efficacy trial (CARET) for chemoprevention of lung cancer in high risk populations: smokers and asbestos-exposed workers. Cancer research **1994**;54(7 Suppl):2038s-43s.

35. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity (Edinb) **2005**;95(3):221-7 doi 10.1038/sj.hdy.6800717.

36. Chen H, Lumley T, Brody J, Heard-Costa NL, Fox CS, Cupples LA*, et al.* Sequence kernel association test for survival traits. Genetic epidemiology **2014**;38(3):191-7 doi 10.1002/gepi.21791.

37. Ward LD, Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic Acids Res **2016**;44(D1):D877-81 doi 10.1093/nar/gkv1340.

38. Wei R, Cao L, Pu H, Wang H, Zheng Y, Niu X*, et al.* TERT Polymorphism rs2736100-C Is Associated with EGFR Mutation-Positive Non-Small Cell Lung Cancer. Clinical cancer

research : an official journal of the American Association for Cancer Research **2015**;21(22):5173-80 doi 10.1158/1078-0432.CCR-15-0009.

39. Gadalla SM, Khincha PP, Katki HA, Giri N, Wong JY, Spellman S*, et al.* The limitations of qPCR telomere length measurement in diagnosing dyskeratosis congenita. Molecular genetics & genomic medicine **2016**;4(4):475-9 doi 10.1002/mgg3.220.

40. Amsellem V, Gary-Bobo G, Marcos E, Maitre B, Chaar V, Validire P*, et al.* Telomere dysfunction causes sustained inflammation in chronic obstructive pulmonary disease. Am J Respir Crit Care Med **2011**;184(12):1358-66 doi 10.1164/rccm.201105-0802OC.

41. Hammadah M, Al Mheid I, Wilmot K, Ramadan R, Abdelhadi N, Alkhoder A*, et al.* Telomere Shortening, Regenerative Capacity, and Cardiovascular Outcomes. Circ Res **2017**;120(7):1130-8 doi 10.1161/CIRCRESAHA.116.309421.

42. Needham BL, Rehkopf D, Adler N, Gregorich S, Lin J, Blackburn EH*, et al.* Leukocyte telomere length and mortality in the National Health and Nutrition Examination Survey, 1999-2002. Epidemiology **2015**;26(4):528-35 doi 10.1097/EDE.0000000000000299.

43. Consortium GT. The Genotype-Tissue Expression (GTEx) project. Nat Genet **2013**;45(6):580-5 doi 10.1038/ng.2653.

44. Sokoloff P, Riou JF, Martres MP, Schwartz JC. Presence of dopamine D-2 receptors in human tumoral cell lines. Biochemical and biophysical research communications **1989**;162(2):575-82.

45. Teunis MA, Kavelaars A, Voest E, Bakker JM, Ellenbroek BA, Cools AR*, et al.* Reduced tumor growth, experimental metastasis formation, and angiogenesis in rats with a hyperreactive dopaminergic system. FASEB J **2002**;16(11):1465-7 doi 10.1096/fj.02-0145fje.

46. Basu S, Sarkar C, Chakroborty D, Nagy J, Mitra RB, Dasgupta PS*, et al.* Ablation of peripheral dopaminergic nerves stimulates malignant tumor growth by inducing vascular permeability factor/vascular endothelial growth factor-mediated angiogenesis. Cancer research **2004**;64(16):5551-5 doi 10.1158/0008-5472.CAN-04-1600.

47. Kachuri L, Saarela O, Bojesen SE, Davey Smith G, Liu G, Landi MT*, et al.* Mendelian Randomization and mediation analysis of leukocyte telomere length and risk of lung and head and neck cancers. International journal of epidemiology **2018** doi 10.1093/ije/dyy140.

48. Zhang C, Doherty JA, Burgess S, Hung RJ, Lindstrom S, Kraft P*, et al.* Genetic determinants of telomere length and risk of common cancers: a Mendelian randomization study. Human molecular genetics **2015**;24(18):5356-66 doi 10.1093/hmg/ddv252.

49. Smith Sehdev AE, Hutchins GM. Problems with proper completion and accuracy of the cause-of-death statement. Arch Intern Med **2001**;161(2):277-84.

50. Iachina M, Green A, Jakobsen E. The direct and indirect impact of comorbidity on the survival of patients with non-small cell lung cancer: a combination of survival, staging and resection models with missing measurements in covariates. BMJ Open **2014**;4(2):e003846 doi 10.1136/bmjopen-2013-003846.

Table 1: Demographic characteristics and clinical features of 888 lung cancer cases with telomere length data, expressed as standardized Z-score, in the Toronto Mount Sinai Hospital - Princess Margaret Hospital (MSH-PMH) study

Characteristic	N	(Deaths)	Mean TL (SD)		P	MST	$P_{\text{Log-rank}}$
Status							
Alive/Censored (%)	420	(47.3)	0.07	(0.99)	0.06		
Deceased (%)	468	(52.7)	-0.06	(1.00)			
Histology							
Adenocarcinoma	501	(219)	0.11	(0.99)	1.8×10^{-3}	42.2	2.2×10^{-12}
Squamous carcinoma	167	(94)	-0.22	(0.99)		19.9	
Other NSCLC[2]	139	(92)	-0.01	(1.01)		18.9	
SCLC	74	(57)	-0.15	(0.99)		12.3	
Other[3]	6	(5)	-0.43	(0.98)		20.7	
Age at diagnosis							
≤50 years	55	(31)	0.49	(0.87)	3.0×10^{-11}	25.4	0.51
51 to 60 years	163	(81)	0.28	(1.00)		29.3	
61 to 70 years	301	(152)	0.05	(0.95)		27.9	
71 to 80 years	270	(147)	-0.67	(1.00)		25.4	
>80 years	99	(57)	-0.45	(0.96)		19.7	
Sex							
Males	463	(285)	-0.01	(1.00)	0.87	19.1	3.9×10^{-8}
Females	425	(183)	0.01	(1.00)		41.2	
Smoking status							
Never	102	(39)	0.30	(0.96)	1.5×10^{-3}	51.2	5.6×10^{-3}
Ever	760	(412)	-0.04	(1.00)		24.4	
Former	347	(190)	-0.06	(0.99)		24.2	
Current	387	(213)	-0.04	(1.00)		22.9	
Missing	9	(17)	0.07	(0.93)		22.2	
Clinical Stage							
NSCLC						29.4	
I	265	(54)	-0.08	(0.95)	4.5×10^{-3}	-	1.2×10^{-64}
II	78	(19)	-0.17	(1.12)		-	
III	200	(123)	0.00	(0.98)		22.9	
IIIA	118	(64)	0.02	(1.00)		27.6	
IIIB	79	(56)	-0.04	(0.96)		16.6	
IV	264	(209)	0.18	(1.00)		10.7	
SCLC						12.3	
Limited	25	(13)	-0.43	(1.05)		22.4	
Extensive	49	(44)	-0.02	(0.95)		10.7	
Mutation status							
Missing or not tested	749	(375)	-0.03	(1.00)		25.8	
$EGFR^-$ and ALK^-	29	(18)	0.08	(0.86)	1.6×10^{-3}	22.9	2.0×10^{-3}
$EGFR^+$ or ALK^+	29	(12)	0.78	(0.74)		41.3	
ECOG performance status							
0 - Fully active	253	(84)	0.04	(0.97)	0.35	58.3	1.7×10^{-23}
1	340	(190)	0.04	(1.02)		22.9	

						MST[1]	
2	169	(112)	-0.06	(1.00)		16.5	
3	67	(51)	-0.04	(1.01)		8.2	
4 - Completely disabled	3	(3)	0.72	(0.83)		17.3	
Missing	56	(28)	-0.20	(0.95)		37.8	
Body mass index (kg/m^2)							
Normal: 18.5 to <25	354	(181)	0.08	(1.00)	0.14	25.9	9.0×10^{-5}
Underweight: <18.5	45	(34)	-0.23	(1.11)		12.2	
Overweight: 25 to <30	262	(140)	-0.07	(1.03)		29.3	
Obese: >30	141	(58)	0.06	(0.98)		39.5	
Missing	86	(55)	-0.10	(0.85)		19.0	
Self-reported ethnicity							
Caucasian	842	(449)	0.00	(1.00)	0.25	25.8	0.20
Other[4]	7	(2)	0.63	(0.81)		-	
Missing	40	(19)	0.01	(0.96)		39.9	

[1] Median survival time (MST) expressed in months
[2] Includes tumours with adeno-squamous or NSCLC not otherwise specified histology
[3] Includes tumours of mixed SCLC and NSCLC histology
[4] Includes individuals with self-reported Asian/Pacific Island, and Latino/Hispanic ethnicity

Abbreviations: MSH-PMH - Mount Sinai Hospital - Princess Margaret Hospital; NSCLC - Non-small cell lung cancer; SCLC - Small cell lung cancer; *EGFR* - Epidermal growth factor receptor; *ALK* - anaplastic lymphoma kinase; ECOG - Eastern Cooperative Oncology Group

Table 2: Demographic characteristics and clinical features non-small cell lung cancer (NSCLC) patients comprising the OncoArray dataset for the analysis of chromosome 5p15.33 variants and overall survival

Characteristic	Study												Total	
	HLCS		MSH-PMH		MDACC		CAPUA		LLP		CARET			
	N	%	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)
Status														
Alive/Censored	655	(38.2)	557	(48.5)	410	(58.1)	45	(8.6)	113	(32.2)	43	(18.7)	1823	(39.0)
Deceased	1058	(61.8)	592	(51.5)	296	(41.9)	479	(91.4)	186	(67.8)	186	(81.2)	2849	(61.0)
Histology														
Squamous	398	(23.2)	261	(22.7)	196	(27.8)	288	(55.0)	173	(49.3)	90	(39.3)	1406	(30.1)
Large cell	102	(6.0)	33	(2.9)	5	(0.7)	18	(3.4)	0	(0.0)	20	(8.7)	178	(3.8)
Adenocarcinoma	1027	(60.0)	757	(65.9)	431	(61.0)	218	(41.6)	165	(47.0)	119	(52.0)	2717	(58.2)
NSCLC NOS	186	(10.9)	98	(8.5)	74	(10.5)	0	(0.0)	13	(3.7)	0	(0.0)	371	(7.9)
Age at diagnosis														
Mean (SD)	65.3	(10.8)	66.0	(10.8)	65.3	(9.7)	66.2	(10.6)	67.5	(7.7)	68.7	(5.7)	65.9	(10.2)
≤50 years	159	(9.3)	91	(7.9)	50	(7.1)	49	(9.4)	2	(0.6)	0	(0.0)	351	(7.5)
51 to 60 years	379	(22.1)	231	(20.1)	151	(21.4)	105	(20.0)	63	(17.9)	15	(6.5)	944	(20.2)
61 to 70 years	539	(31.5)	375	(32.6)	284	(40.2)	158	(30.2)	159	(45.3)	119	(52.0)	1634	(35.0)
71 to 80 years	504	(29.4)	349	(30.4)	188	(26.6)	181	(34.5)	115	(32.8)	93	(40.6)	1430	(30.6)
>80 years	132	(7.7)	103	(9.0)	33	(4.7)	31	(5.9)	12	(3.4)	2	(0.9)	313	(6.7)
Sex														
Males	841	(49.1)	576	(50.1)	361	(51.1)	461	(88.0)	203	(57.8)	155	(67.7)	2597	(55.6)
Females	872	(50.9)	573	(49.9)	345	(48.9)	63	(12.0)	148	(42.2)	74	(32.3)	2075	(44.4)
Smoking status														
Never	178	(10.4)	122	(10.6)	70	(9.9)	32	(6.1)	73	(20.8)	0	(0.0)	475	(10.2)
Ever	1517	(88.6)	1014	(88.3)	635	(89.9)	492	(93.9)	278	(79.2)	219	(100)	4165	(89.1)
Former	967	(56.5)	510	(44.4)	374	(53.0)	223	(42.6)	148	(42.2)	45	(19.7)	2267	(48.5)
Current	550	(32.1)	478	(41.6)	261	(37.0)	267	(51.0)	126	(35.9)	184	(80.3)	1866	(39.9)
Missing	18	(1.0)	13	(1.1)	1	(0.1)	0	(0.0)	0	(0.0)	0	(0.0)	32	(0.7)
Clinical Stage														
I	508	(29.7)	404	(35.2)	263	(37.3)	141	(26.9)	120	(34.2)	44	(19.2)	1480	(31.7)
II	192	(11.2)	108	(9.4)	117	(16.6)	45	(8.6)	58	(16.5)	23	(10.0)	543	(11.6)
III	445	(26.0)	295	(25.7)	141	(20.0)	185	(35.3)	67	(19.1)	65	(28.4)	1198	(25.6)
IIIA	201	(11.7)	164	(14.3)	84	(11.9)	70	(13.4)	36	(10.3)	40	(17.5)	595	(12.7)
IIIB	244	(14.2)	127	(11.1)	57	(8.1)	115	(21.9)	31	(8.8)	12	(5.2)	586	(12.5)
IV	537	(31.3)	342	(29.8)	185	(26.2)	152	(29.0)	55	(15.7)	72	(31.4)	1343	(28.7)
Missing	31	(1.8)	0	(0.0)	0	(0.0)	1	(0.2)	51	(14.5)	25	(10.9)	108	(2.3)
Total	1713		1149		706		524		351		229		4672	
MST (months)	29.6 (26.6 - 32.0)		35.6 (31.2 - 40.2)		42.9 (37.5 - 60.0)		9.2 (8.2 - 10.4)		28.2 (21.9 - 37.3)		12.0 (9.6 - 14.4)		26.9 (25.2 - 28.7)	

Abbreviations: HLCS - Harvard Lung Cancer Susceptibility; MSH-PMH - Mount Sinai Hospital - Princess Margaret Hospital; MDACC - MD Anderson Cancer Center; CAPUA - Cáncer de Pulmón en Asturias; LLP - Liverpool Lung Project; CARET - Carotene and Retinol Efficacy Trial; MST - median survival time

Table 3: Adjusted hazard ratios (HR) for 5-year all-cause mortality comparing short (≤10th percentile) with long telomere length (>10th percentile) in lung cancer patients from the Toronto Mount Sinai Hospital - Princess Margaret Hospital (MSH-PMH) study, restricted to samples obtained up to 6 months following diagnosis

	N	(Deaths)	HR[1]	(95% CI)	P	q_{FDR}
All lung cancers	823	(438)	1.09	(0.80, 1.47)	0.61	0.67
Non-small cell lung cancer (NSCLC)	745	(378)	1.16	(0.83, 1.63)	0.38	0.52
Smoking status						
Never	86	(34)	10.29	(1.86, 56.86)	7.5×10^{-3}	0.04
Ever	639	(332)	1.05	(0.74, 1.49)	0.79	0.79
Histology						
Adenocarcinoma	459	(206)	1.65	(1.04, 2.64)	0.03	0.08
Males	211	(117)	1.25	(0.67, 2.33)	0.48	0.59
Females	248	(89)	2.54	(1.20, 5.38)	0.01	0.06
Smoking status						
Never	77	(27)	11.31	(1.96, 65.24)	6.7×10^{-3}	0.04
Ever	371	(174)	1.38	(0.84, 2.29)	0.21	0.46
Squamous cell carcinoma	159	(89)	0.69	(0.33, 1.45)	0.32	0.50
Small cell lung cancer (SCLC)	71	(54)	0.62	(0.26, 1.45)	0.27	0.50

[1] Cox models adjusted for age at diagnosis (years), sex, stage (I-IIIA vs. IIIB-IV or Limited vs. Extensive for SCLC), cigarette pack-years (0, >0-20, 21-40, and >40), and body mass index (<18.5, 18.5 to <25, 25 to <30, ≥30, unknown)

Table 4: Adjusted hazard ratios (HR) for 5-year all-cause mortality comparing short (≤10th percentile) with long telomere length (>10th percentile) in lung cancer cases and controls matched to each lung cancer subgroup from the Copenhagen City Heart Study and Copenhagen General Population Study, stratified by time elapsed between sample collection for telomere length measurement and lung cancer diagnosis

	Lung Cancer Cases					General Population Controls					P for Interaction[3]
	N	(Deaths)	HR[1]	(95% CI)	P	q_{FDR}	N	(Deaths)	HR[2]	(95% CI)	P	
All lung cancers	923	(769)	1.48	(1.18, 1.87)	6.9×10^{-4}	0.002	2905	(429)	1.15	(0.87, 1.52)	0.33	0.73
NSCLC	767	(629)	1.52	(1.18, 1.95)	1.2×10^{-3}	0.003	2437	(346)	0.99	(0.72, 1.37)	0.96	0.39
Histology												
Adenocarcinoma	440	(351)	1.48	(1.01, 2.16)	0.04	0.06	1448	(162)	1.06	(0.66, 1.70)	0.80	0.94
Squamous	211	(175)	1.45	(0.94, 2.25)	0.09	0.11	653	(120)	1.03	(0.58, 1.81)	0.92	0.86
Other NSCLC[4]	116	(103)	2.91	(1.50, 5.65)	1.6×10^{-3}	0.004	336	(64)	0.86	(0.39, 1.87)	0.70	0.01
SCLC	156	(140)	1.42	(0.74, 2.74)	0.29	0.31	468	(83)	1.76	(0.98, 3.14)	0.06	0.83
Diagnosed <5 years after sample collection												
All lung cancers	382	(335)	1.89	(1.34, 2.66)	2.9×10^{-4}	0.001	1313	(189)	0.95	(0.62, 1.45)	0.82	0.12
NSCLC	319	(277)	1.98	(1.37, 2.86)	2.8×10^{-4}	0.001	1093	(149)	0.68	(0.40, 1.16)	0.16	0.01
Males	182	(160)	1.37	(0.82, 2.28)	0.23	0.27	583	(108)	0.62	(0.32, 1.19)	0.15	0.04
Females	137	(117)	4.06	(2.08, 7.92)	4.0×10^{-5}	0.001	510	(41)	0.77	(0.30, 2.01)	0.60	0.33
Histology												
Adenocarcinoma	170	(145)	2.42	(1.37, 4.28)	2.3×10^{-3}	0.005	614	(69)	0.70	(0.30, 1.63)	0.40	0.02
Squamous	86	(75)	2.19	(1.09, 4.41)	0.03	0.05	281	(54)	1.20	(0.53, 2.70)	0.66	0.77
SCLC	63	(58)	1.33	(0.45, 3.89)	0.60	0.60	220	(40)	1.88	(0.85, 4.17)	0.12	0.22

[1] Cox models adjusted for age at diagnosis (years), sex, stage (I-IIIA, IIIB-IV, unknown), cigarette pack-years (0, >0-20, 21-40, >40, unknown), body mass index (<18.5, 18.5 to <25, 25 to <30, ≥30, unknown), calibrator lot for the telomere measurement assay, time elapsed between telomere measurement and lung cancer diagnosis

[2] Cox models adjusted for age at diagnosis (years), sex, cigarette pack-years (0, >0-20, 21-40, >40, unknown), body mass index (<18.5, 18.5 to <25, 25 to <30, ≥30, unknown), calibrator lot for the telomere measurement assay, and time elapsed between blood sample collection and lung cancer diagnosis in the matched case

[3] P-value for interaction between case-control status and telomere length was calculated using the likelihood ratio test comparing nested models with and without an interaction term
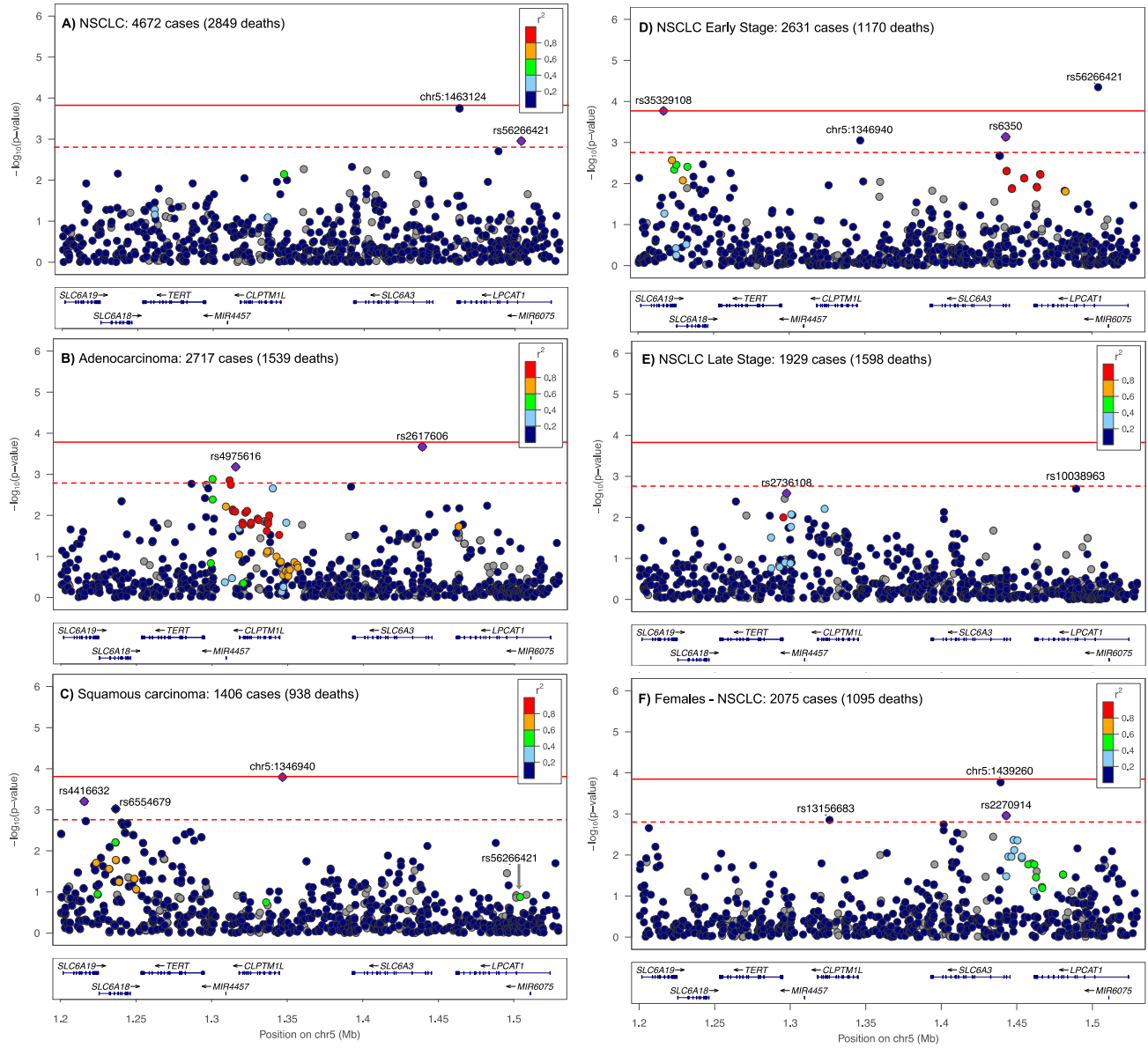
Abbreviations: NSCLC - Non-small cell lung cancer; SCLC - Small cell lung cancer
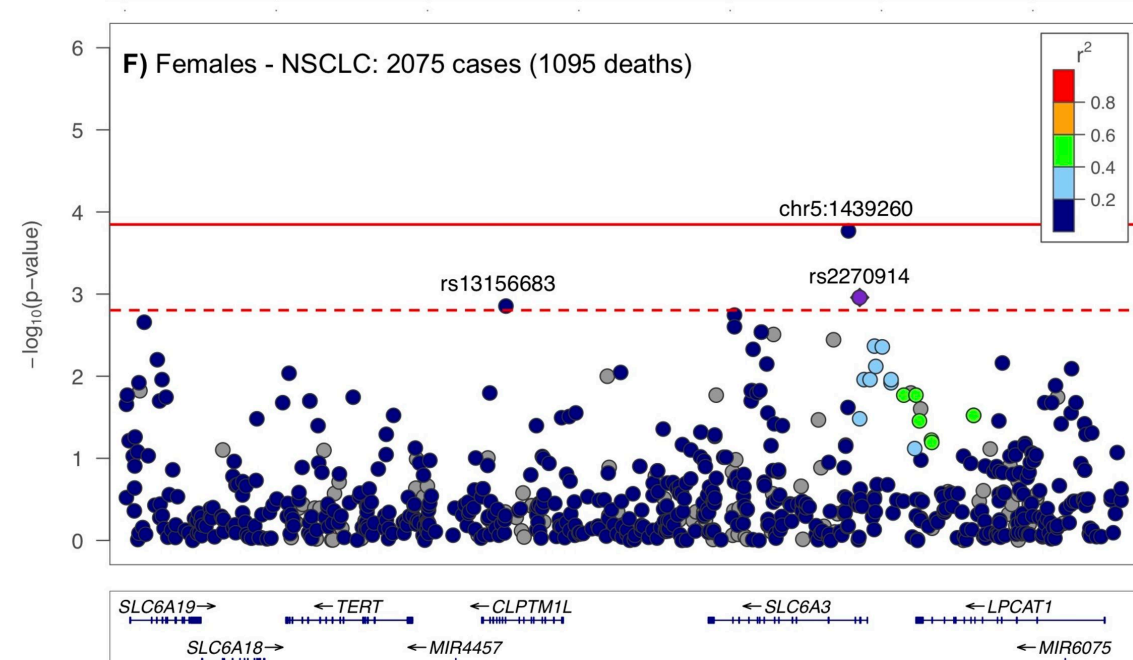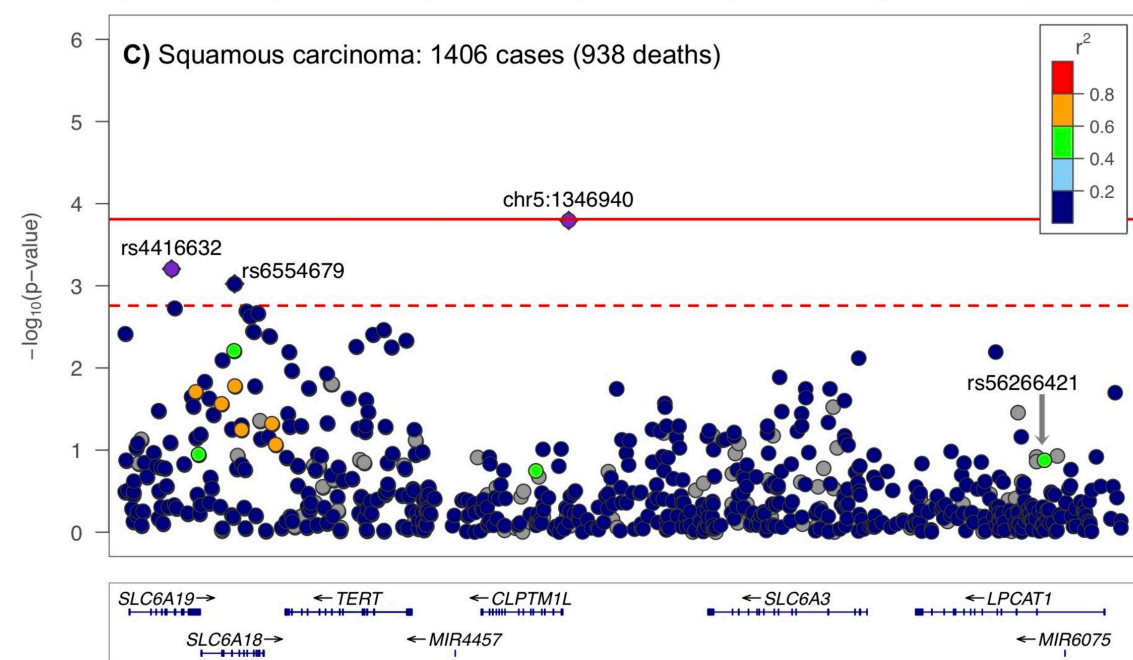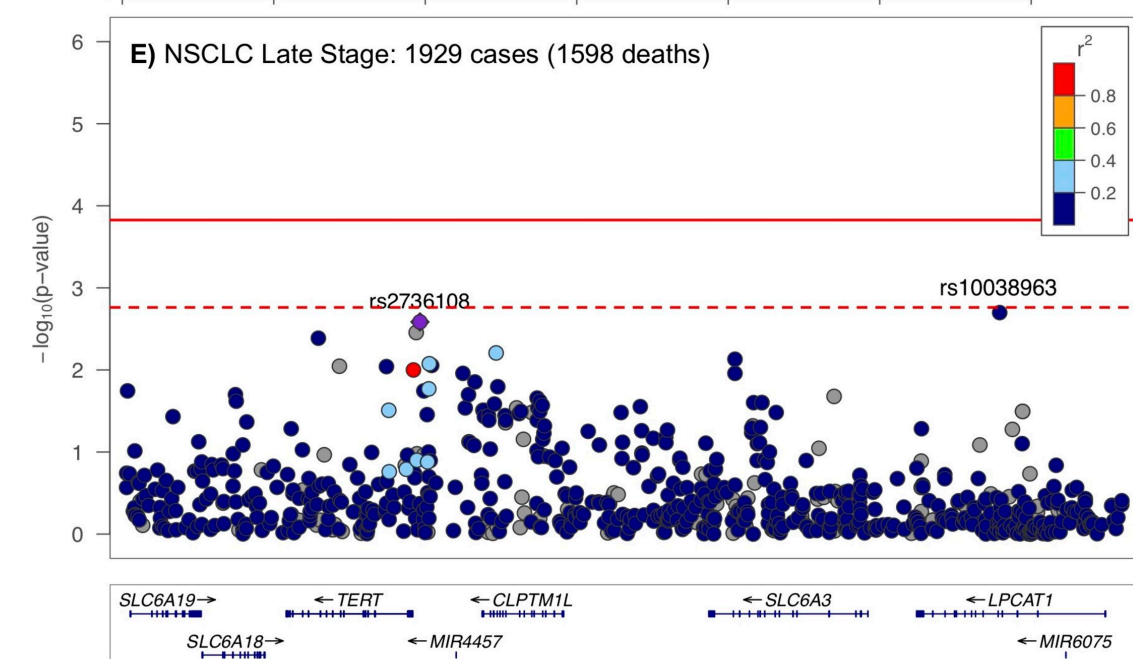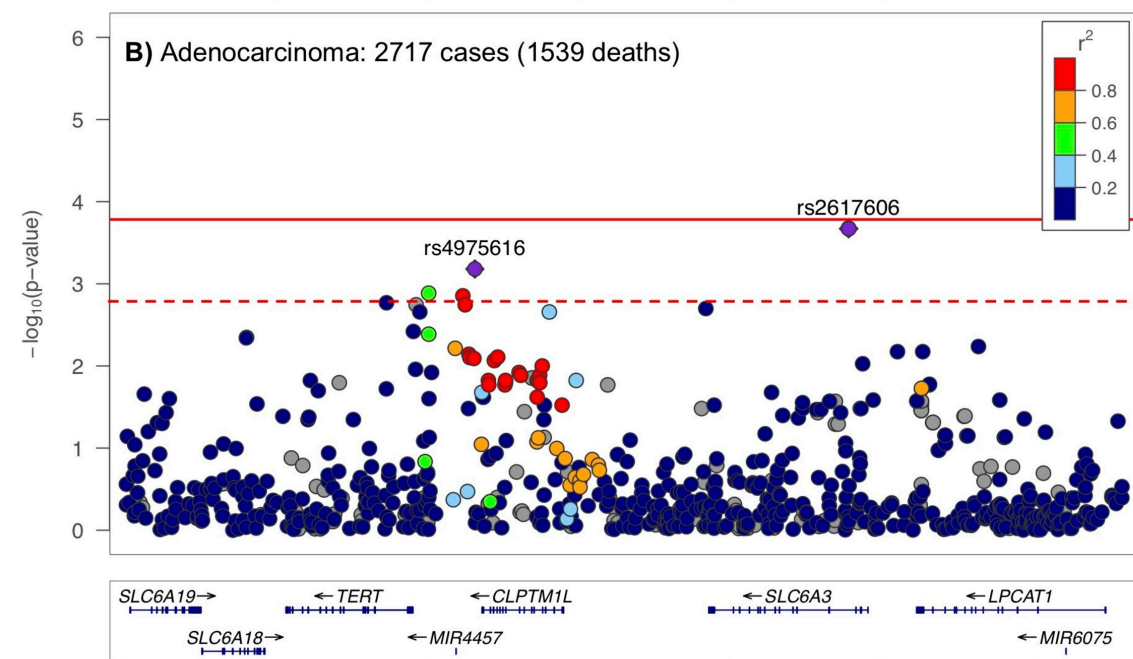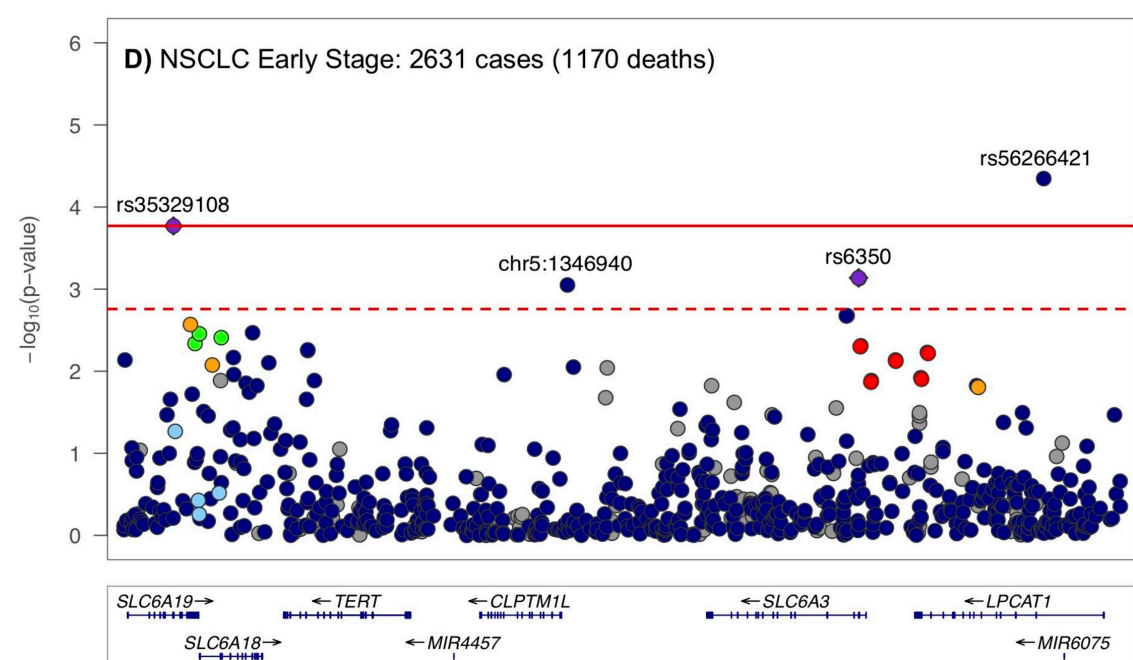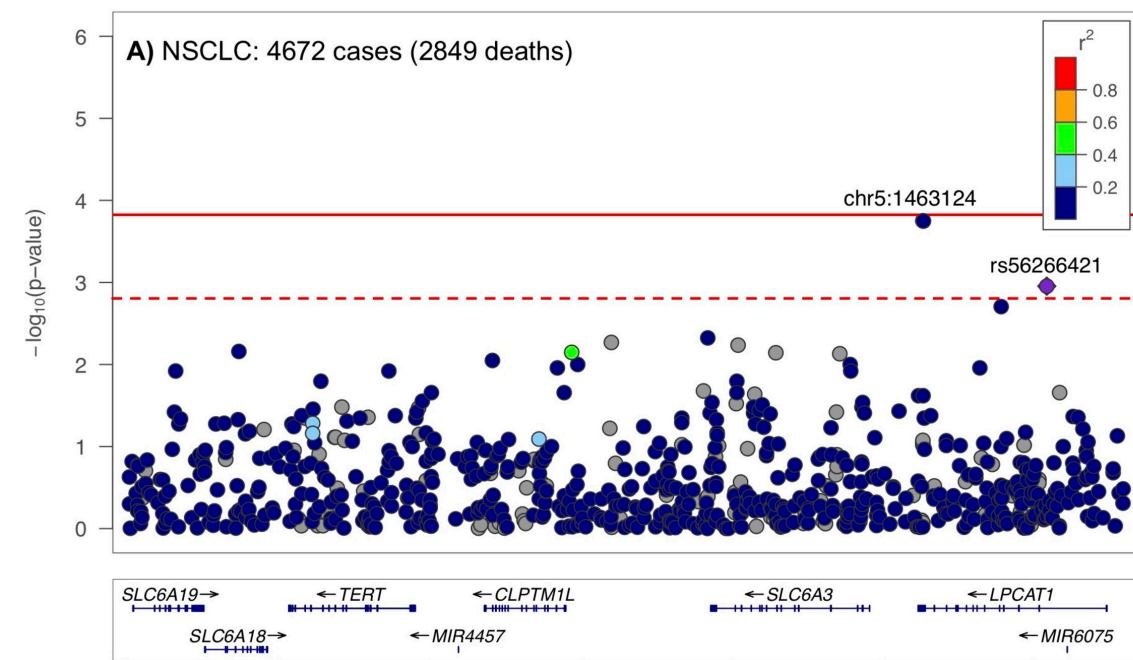
Table 5: Gene-level associations with 5-year overall survival based on the Cox SKAT-LRT meta-analysis of 6 OncoArray studies. Associations with $P < 2.5 \times 10^{-3}$ were considered statistically significant.

Gene Set	All variants (N=723)					Rare variants: MAF ≤ 0.01 (N=182)				
	N	P_{NSCLC}	P_{Adeno}	P_{SQC}	$P_{Females}$	N	P_{NSCLC}	P_{Adeno}	P_{SQC}	$P_{Females}$
TERT	119	0.06	0.14	0.03	0.22	48	0.03	0.05	0.02	0.07
CLPTM1L	58	0.54	0.13	0.14	0.05	15	0.51	0.53	0.12	0.03
LINC01511	40	0.60	0.91	0.39	0.02	8	0.51	0.88	0.31	0.10
SLC6A3	144	0.10	4.1×10^{-3}	0.56	1.6×10^{-3}	50	0.09	3.3×10^{-3}	0.93	1.4×10^{-3}
LPCAT1	172	0.09	0.41	0.02	0.13	28	0.05	0.08	0.02	0.31

Abbreviations: NSCLC - Non-small cell lung cancer; MAF - minor allele frequency; SQC - squamous cell carcinoma

Figure 1: Regional plots of associations with 5-year survival for NSCLC overall, based on a meta-analysis of 6 OncoArray studies, and pooled Cox regression analyses stratified by histology, sex, and disease stage. Linkage disequilibrium is illustrated with respect to the lead variant denoted in purple. The threshold for declaring statistical significance (P = 1.6×10^{-4}) is indicated by the solid red line, with suggestive associations denoted by the dotted red line (P = 1.6×10^{-3}).