Assessing Gene-Environment Interaction in the Association Between Smoking and Leukocyte Telomere Length

Kelsey Suzanne Kanavel

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2025

Committee:

Sara Lindstroem

Danny E. Miller

Guanghao Qi

Program Authorized to Offer Degree:

Institute for Public Health Genetics

University of Washington

Abstract

Assessing Gene-Environment Interaction in the Association Between Smoking and Leukocyte Telomere Length

Kelsey Suzanne Kanavel

Chair of the Supervisory Committee:

Sara Lindstroem

Department of Epidemiology

Telomere length is influenced by both genetic and environmental factors, including smoking behavior. While smoking is associated with shorter telomeres, this association may vary by genotype. In a subset of the UK Biobank cohort (N = 360,909) with genetically inferred European ancestry, we analyzed 581,069 common single nucleotide polymorphisms (SNPs) and average relative telomere length estimates derived from quantitative PCR (qPCR) to test whether the association between ever/never smoking status and telomere length differs by genotype. One SNP, rs4418881, reached genome-wide significance ($P = 4.49 \times 10^{-8}$) in the 2 degrees of freedom gene-environment interaction joint test but did not reach significance in the overall genome-wide association study ($P = 4.19 \times 10^{-7}$). Stratified analyses revealed that rs4418881 was genome-wide significant in ever smokers ($P = 1.06 \times 10^{-8}$) but not in never smokers ($P = 0.35$), suggesting that this SNP may modify the effect of smoking on telomere length.

I. Introduction

Background and Significance

Telomeres are structures that protect the ends of chromosomes from enzymatic degradation and inappropriate activation of DNA damage response pathways.[1,2] They are characterized by tandem repeats of 5'-$(TTAGGG)_n$-3' that end in a single-stranded 3' overhang that loops back on itself to form the T-loop and which is stabilized with shelterin protein complexes.[2] Telomerase is a reverse transcriptase primarily expressed in stem cells that maintains telomere length as cells divide.[3] Because differentiated human cells express telomerase at low or negligible levels, somatic cell telomeres steadily decrease with each cell division because DNA polymerase cannot replace the final RNA primer on the lagging strand during DNA replication (the "end-replication problem"), and telomerase is not sufficiently present to aid in extending the telomeres.[4,5]

Accordingly, older individuals, whose cells have undergone more mitotic divisions than those in younger people, have shorter telomeres on average than younger individuals.[6–8] However, even among individuals of the same age, telomere length can vary substantially.[9] For example, in newborns, average telomere length has been estimated to range from 7–14 kb in leukocytes.[10,11]

Besides age, variability in telomere length has been found to be explained by genetic and non-genetic factors, such as parental smoking behavior, parental age at one's birth, cell type, sex-related factors (e.g., estrogen concentrations[12,13]), and body mass index (BMI).[6,14] Additionally, environmental factors, such as smoking behavior and exposure to polychlorinated biphenyls and cadmium, have been found to be associated with telomere length.[6,15–17]

Heritability estimates for leukocyte telomere length vary depending on the study design and measurement method. In twin studies using Southern blots of terminal restriction fragments to measure average absolute telomere length, heritability estimates have ranged from 0.34 to 0.82.[18–21] In contrast, a meta-analysis of six large family and twin cohort studies that used quantitative PCR (qPCR) to measure average relative telomere length reported a heritability estimate of 0.70 (0.64–0.76).[16]

Estimates based on genome-wide genotype data are considerably lower. In a genome-wide association study (GWAS) of 19.4 million imputed genetic variants in the UK Biobank cohort, using qPCR-derived average relative telomere length, all genetic variants with a minor allele frequency (MAF) of 0.1%

or greater combined explained 8.1% of telomere length variation.[22] A more recent study in the same cohort used whole genome sequencing (WGS) and a combined average telomere length metric constructed using the first principal component from a principal component analysis of qPCR and TelSeq (an approach to estimating telomere length from whole genome sequencing data[23]).[24] This study found that genetic variants with a MAF of 0.1% or greater explained 9.9% of the variability in average telomere length.[24] Of these, 192 were found to have a genome-wide significant ($P < 5 \times 10^{-8}$) association with the first principal component telomere length metric.[24]

The same study also conducted an exome-wide association study focused on rare genetic variants (MAF < 0.1%) and identified 62 rare exonic variants across 19 genes significantly associated with average telomere length ($P \leq 1 \times 10^{-8}$).[24] While most of these associations occurred at loci previously identified through GWAS of genetic variants with a MAF of 0.1% or greater, one occurred in a novel locus, suggesting that rare variant analysis may reveal associations missed by common variant approaches.[24]

Shorter telomere length is associated with the development of a wide range of age-related diseases, such as cardiovascular disease, Alzheimer's disease, and various cancers, and is considered to be a hallmark of cellular aging.[25,26] However, long telomeres are also associated with an increased risk of several cancers.[27,28] Therefore, there are health risks posed if telomere length is at either end of the spectrum. Because telomere length is influenced by both genetic and environmental factors, identifying modifiable exposures that affect telomere length may have important implications for disease prevention.

One such modifiable exposure is smoking behavior. Multiple studies have found that telomere length is shorter in ever smokers compared to never smokers, and in current smokers compared to former smokers.[29] However, not all individuals respond to environmental exposures in the same way.[30] It is possible that this association differs depending on whether an individual has certain genetic variants. Whether there are gene-by-smoking interaction effects on telomere length has not been investigated to date. Due to the risks of disease at either extreme of telomere length, investigating potential gene-by-smoking interactions may provide valuable insights about the association between smoking and telomere length and may contribute to explaining some of the telomere length missing heritability.

Smoking behavior is the top risk factor for lung cancer, the leading cause of cancer death in the United States, with almost 9 in 10 lung cancers caused, at least in part, by cigarette smoking.[31–34] There is evidence that telomere length is shorter among ever smokers compared to never smokers in both whole blood cells and lung cells.[29,35,36] While some studies have linked shorter leukocyte telomeres to lung cancer development,[37–39] most recent studies have found that longer leukocyte telomeres are associated with the development of lung cancer.[3,36,40–42] Mendelian randomization studies have provided further evidence that longer telomeres are associated with lung cancer.[43,44] Furthermore, long-telomere associated single nucleotide polymorphisms (SNPs) have been found to also be associated with lung adenocarcinoma.[27] Additionally, there is evidence that smoking interacts with some SNPs in lung cancer.[45] This suggests a complex interplay between genetic variation, smoking behavior, and telomere biology in lung carcinogenesis. Investigating whether SNPs that modify the association between smoking and telomere length are also associated with lung cancer in previous lung cancer GWAS may help with the generation of hypotheses about the biological mechanisms connecting these factors.

Specific Aims

Aim 1. To investigate whether there are common genetic variants that modify the association between smoking behavior and telomere length in a large population-based study of middle-aged British individuals (N = 360,909). We will assess if the effect of any variants identified in the gene-environment interaction (GxE) analysis differs according to smoking status (n = 219,570 for ever smokers and n = 141,339 for never smokers).

Aim 2. To assess if any variants found in Aim 1 are associated with lung cancer.

II. Methods/Approach

Study Subjects

The UK Biobank is a prospective cohort that recruited approximately 500,000 individuals across the United Kingdom primarily aged between 40 and 69 years from 2006 to 2010.[46,47] This thesis analyzes a subset of UK Biobank participants with genetically inferred European ancestry (see **Inferring Genetic Ancestry** below). Additionally, UK Biobank cohort participants were excluded if their age at baseline assessment, reported sex, genotyping data, telomere length estimates, or smoking behavior history (ever/never smoked) were missing.

Furthermore, participants were excluded if genetic sex and reported sex did not match. This was done to mitigate any sample issues such as erroneously labeled or mishandled samples, rather than to intentionally exclude intersex or transgender individuals (additional information would be required to make this distinction).

Because telomere length measurements were made from leukocytes,[48] and because cancer cells engage in strategies to maintain telomere length to survive the cellular crisis induced by critically short telomeres (e.g., upregulating *TERT*, or alternative lengthening of telomeres (ALT)),[49] participants were excluded if they had been diagnosed with a hematologic malignancy at any time before study enrollment and up to two years later. This two-year latency period is most directly informed by studies of radiation-induced leukemia, in which a minimum two-year latency period is often used,[50–52] even though radiation is not presumed to be causal in the leukemia cases among UK Biobank participants.

Finally, to avoid test statistic inflation due to relatedness, for every third-degree or closer relative pair (kinship coefficient > 0.044),[53] individuals were randomly removed until no pairs remained.

Following application of these filters, our final dataset comprised 360,909 individuals whose associated data were analyzed in this thesis project ("thesis cohort").

Inferring Genetic Ancestry

Inferred genetic ancestry assignments were provided by Austin Hammermeister Suger. To calculate ancestry assignments, UK Biobank samples were first projected onto a principal component (PC) space defined by 1000 Genomes Project (1000G) reference samples based on genotypes from a set of harmonized LD-pruned variants using FlashPCA2.[54] Each UK Biobank participant was then assigned to the 1000G populations to which they had the minimum sum of Mahalanobis distances across the first 30 PCs.

Telomere Length Estimates

Leukocyte telomere length estimates assessed by quantitative PCR (qPCR) have been estimated for 474,074 UK Biobank participants from blood samples collected at baseline .[55] Average relative telomere length estimates were assessed for variation due to factors such as qPCR run, primer batch, PCR machine, enzyme, pipetting robot, operator, temperature, humidity, extraction method, and time of day, in addition to the potential influence of two-way interactions between these factors.[55] Estimates

adjusted for the influence of technical parameters were then $\log_e$-transformed and Z-standardized to reduce skewness in the telomere length estimate distribution, as recommended by UK Biobank.[22,56] To further adjust for skewness, any observations not within 3 standard deviations of the mean were removed.

Genotype Data

Genotype data were captured for UK Biobank participants on two different genotyping arrays: the Applied Biosystems UK BiLEVE Axiom Array by Affymetrix and the Applied Biosystems UK Biobank Axiom Array.[57–59] 49,950 participants were genotyped using the Applied Biosystems UK BiLEVE Axiom Array, and 438,427 participants were genotyped using the Applied Biosystems UK Biobank Axiom Array.[59] All samples were processed over 18 months, with DNA preparation performed at UK Biobank, genotyping conducted at Affymetrix, and quality control filters applied by the Wellcome Trust Centre for Human Genetics.[58,60] After quality control filters were applied, genotype data for 805,426 markers (markers included on both arrays) for 488,377 UK Biobank participants were released.[59]

GWAS and GxE Analyses

GWAS and GxE analyses were performed in the DNAnexus cloud computing environment using Swiss Army Knife to execute PLINK 2.0 (v2.0.0-a.6.9) using the genotype data previously described.[61–63] Covariates included sex, age at enrollment, ever/never smoking status, array version, and the first 20 genetically inferred principal components (PCs).

For the GWAS, we tested the association between each SNP and telomere length using linear regression. The model fit for each SNP assessed was the following:

$$Y = \beta_0 + \beta_1 \mathbf{G} + \beta_2 \mathbf{X}_1 + \beta_3 \mathbf{X_2} + \ldots + \beta_{26} \mathbf{X_{24}},$$

where Y is telomere length, β_0 is the model intercept, β_1 is the genetic effect, **G** is genotype (additive model coded as effect allele dosages [0, 1, 2]), and $\mathbf{X}_1 - \mathbf{X_{24}}$ are additional covariates (ever smoked status, age at enrollment, sex, genotype array version, and the first 20 genetically inferred PCs).

In the GWAS, the primary parameter of interest was β_1, the genetic effect after adjusting for covariates. The null hypothesis tested was that $\beta_1 = 0$.

For the GxE analysis, we tested whether the genetic association with telomere length differed by smoking status using linear regression with an interaction term. The model fit for each SNP assessed was the following:

$$Y = \beta_0 + \beta_1 \mathbf{G} + \beta_2 \mathbf{E} + \beta_3 \mathbf{GE} + \beta_4 \mathbf{X_1} \ldots + \beta_{26} \mathbf{X_{23}},$$

where Y is telomere length, β_0 is the model intercept, **G** is genotype (additive model coded as effect allele dosages [0, 1, 2]), and **E** is ever smoked status (0 for nonsmokers and 1 for smokers). β_1 represents the genetic effect among never smokers ($\mathbf{E} = \mathbf{0}$), β_2 is the smoking effect among individuals with $\mathbf{G} = \mathbf{0}$ (an effect allele dosage of 0), and β_3 is the gene-environment interaction effect. Covariates $\mathbf{X}_1 - \mathbf{X_{23}}$ include age at enrollment, sex, genotype array version, and the first 20 genetically inferred PCs.

In the GxE analysis, the primary parameter of interest was β_3, the gene-environment interaction effect. The null hypothesis was that $\beta_3 = 0$, meaning there is no interaction between the effect allele and smoking status on telomere length (no gene-environment interaction). This is the 1 degree of freedom interaction test ("1 d.f. interaction test"). Additionally, a 2 degrees of freedom joint test ("2 d.f. joint test") was performed to detect SNPs for which there was either a main genetic effect, a gene-environment interaction effect, or both. The 2 d.f. joint test simultaneously tests the null hypotheses that $\beta_1 = 0$ and $\beta_3 = 0$.

For the GWAS and GxE analyses conducted on the thesis cohort, of the 805,426 markers released by UK Biobank, data from 581,069 markers were included in the analyses after excluding markers with > 2% missing data across the thesis cohort and excluding markers with a minor allele frequency < 1% within the thesis cohort.[64,65]

For the GWAS analysis conducted on the subset of the thesis cohort who had ever smoked, of the 805,426 markers released by UK Biobank, data from 581,076 markers were included in the analyses after excluding markers with > 2% missing data across the thesis cohort and excluding markers with a minor allele frequency < 1% within the thesis cohort.[64,65]

For the GWAS analysis conducted on the subset of the thesis cohort who had never smoked, of the 805,426 markers released by UK Biobank, data from 581,533 markers were included in the analyses after excluding markers with > 2% missing data across the thesis cohort and excluding markers with a minor allele frequency < 1% within the thesis cohort.[64,65]

Identification of Lead Single Nucleotide Polymorphisms (SNPs) with FUMA

Lead SNPs were identified using the FUMA SNP2GENE web application (v1.5.2; https://fuma.ctglab.nl/snp2gene) using default configurations as described in Watanabe et al., 2017.[66] Using the 1000 Genomes Phase 3 European (EUR) reference panel, variants were grouped into genomic risk loci by merging linkage disequilibrium (LD) blocks within 250 kb. Independent significant SNPs were defined as genome-wide significant variants ($P < 5 \times 10^{-8}$) not in high LD with each other ($r^2 < 0.6$), and independent lead SNPs were identified as the SNPs with the smallest p-value within each locus and in low LD ($r^2 < 0.1$) with other SNPs with a lower p-value. Independent lead SNPs were annotated and mapped to protein-coding genes defined by Ensembl v102 using a 10 kb window flanking gene boundaries.

LDTrait: Identifying Variants in Linkage Disequilibrium with Lead SNPs

The LDTrait tool (https://ldlink.nih.gov/?tab=ldtrait) in the LDlink suite of web-based applications was used to find variants in linkage disequilibrium with lead SNPs .[67,68] Search parameters included an $r^2 = 0.1$ threshold, a ± 500 kb window, and the reference population selected was "(EUR) European."

III. Results

The GxE 1 d.f. interaction analysis did not identify any loci reaching genome-wide significance ($P < 5 \times 10^{-8}$) (see Figures 7 and 8). However, the GxE 2 d.f. joint test identified one SNP, rs4418881 (located on chromosome 12 at position 132,481,480 (GRCh38)[69]), that was genome-wide significant in the GxE 2 d.f. joint test ($P < 4.49 \times 10^{-8}$), but not genome-wide significant in the GWAS conducted on the overall thesis cohort ($P < 4.19 \times 10^{-7}$) (see Table 2 and Figures 9, 10, and 11). This SNP, rs4418881, was found to be genome-wide significant ($P < 1.06 \times 10^{-8}$) in the GWAS conducted solely on individuals who had ever smoked but was not significant ($P = 0.35$) in the GWAS conducted solely on individuals who had never smoked (see Table 2 and Figure 11).

Independent lead SNPs were annotated and mapped to protein-coding genes with FUMA (using the parameters described in **Methods**), and rs4418881 was mapped between *MUC8* and *FBRSL1*, neither of which have been previously identified to be relevant to telomeres, based on PubMed searches for '("MUC8" or "mucin 8") AND "telo*"' and '("FBRSL1" or "fibrosin like 1" or "fibrosin-like protein 1") AND "telo*"' on June 2, 2025.

Using the LDTrait tool, rs4418881 was identified to be in modest linkage disequilibrium (r^2 = 0.214) with rs79228077, a variant identified in a recent WGS GWAS of telomere length in the UK Biobank cohort.[24] This variant was significantly associated ($P < 2 \times 10^{-12}$) with the first principal component of telomere length, which was derived from a linear combination of qPCR and TelSeq estimates.[24]

To investigate whether rs4418881 had been previously found to be associated with lung cancer, summary statistics from several lung cancer GWAS were evaluated. In a 2017 lung cancer GWAS based on 29,266 lung cancer cases and 56,450 controls of European ancestry, rs4418881 was not associated with lung cancer in the overall analysis of both ever and never smokers ($P = 0.68$), nor was it associated with lung cancer in their GWAS restricted to only ever smokers ($P = 0.60$).[70] Additionally, rs4418881 was not among those reported with $P < 5 \times 10^{-5}$ in analyses limited to individuals with lung squamous cell carcinoma or lung adenocarcinoma.[70] Similarly, in a 2020 pan-cancer GWAS meta-analysis of individuals with European ancestry from the UK Biobank and Kaiser Permanente Genetic Epidemiology Research on Adult Health and Aging cohorts, rs4418881 did not reach genome-wide significance for lung cancer (P = 0.02).[71] In a 2022 lung cancer GWAS across European, East Asian, and African populations, rs4418881 again did not appear among SNPs with $P < 5 \times 10^{-5}$ for individuals diagnosed with lung adenocarcinoma, lung squamous cell carcinoma, or small cell lung carcinoma.[72] Additionally, a 2023 lung adenocarcinoma GWAS conducted in an East Asian population did not report rs4418881 among its genome-wide significant findings.[73] Finally, in a 2024 lung cancer GWAS based on 16,336 cases and 612,076 controls of Hispanic/Latino American, European American, and African American participants of the Million Veteran Program, rs4418881 was not significant in the overall meta-analysis across groups ($P = 0.38$), the Hispanic/Latino American group analysis ($P = 0.48$; 624 cases and 59,042 controls), the European American group analysis ($P = 0.32$; 13,065 cases and 434,597 controls), or the African American group analysis ($P = 0.66$; 2,647 cases and 118,437 controls).[74]

IV. Discussion/Conclusions

Summary of Findings

The SNP rs4418881 reached genome-wide significance only in the GxE 2 d.f. joint test and the GWAS restricted to ever smokers, but not in the GWAS including both ever and never smokers, nor in the GWAS limited to never smokers. This pattern suggests a potential interaction between rs4418881 and

smoking in the association between smoking and average telomere length, with an association observed only among individuals who have ever smoked.

Although no SNPs reached genome-wide significance in the 1 d.f. interaction test, it is worth noting that this test evaluates whether the difference in effect due to a SNP between strata (e.g., ever smokers versus never smokers) is statistically significant. In this study, rs4418881 had a strong association with telomere length among ever smokers and a non-significant association with telomere length among never smokers, but the difference between the two effects may not have been large enough relative to their standard errors to result in a genome-wide significant interaction in the 1 d.f. test. By contrast, the 2 d.f. test detects SNPs with either a strong main effect *or* interaction effect, which likely explains why rs4418881 was genome-wide significant in that test.

The relationship between smoking, telomere length, and lung cancer is not straightforward. While smoking is known to be causal for lung cancer, having ever smoked is associated with shorter telomeres than those of never smokers, yet lung cancer is associated with longer telomeres.[3,29,33,36] Therefore, it is possible that the potential telomere length shortening effect of smoking does not play a primary role in the mechanisms by which smoking increases risk of lung cancer. Furthermore, while rs4418881 is associated with a difference in telomere length between smokers with and without an effect allele, rs4418881 was not found in any of the previous lung cancer GWAS surveyed to be associated with lung cancer. This suggests that while it is possible that rs4418881 may play a role in smokers having longer telomeres, the magnitude of this increase is possibly insufficient to influence lung cancer risk. In other words, rs4418881 may lead to a measurable difference in telomere length among smokers without producing a biologically meaningful variation in cancer susceptibility, especially if the potential telomere length-modulating effects of smoking play a minor role in the causal role of smoking on lung cancer. Additionally, although leukocyte average relative telomere length has been found to be positively correlated with lung average relative telomere length ($r = 0.127$), this correlation is modest.[35] As such, variants that modify the association between smoking and telomere length in leukocytes may not also modify this association in lung cells.

Limitations

The UK Biobank cohort has limitations due to selection bias. Of individuals invited to participate across the United Kingdom, 5.5% joined the cohort.[75] Additionally, there is evidence of a healthy volunteer selection bias.[75] Compared to the general population in the United Kingdom, participants have fewer self-reported health problems and are less likely to drink alcohol, to smoke, or to be obese.[75] For individuals aged 70-74, all-cause mortality rates and total cancer incidence were found to be lower in the cohort than in the general population (46.2% and 11.8% lower, respectively, in men, and 55.5% and 18.1% lower, respectively, in women).[75] Additionally, participants are more likely to be older, female, and to live in less socioeconomically deprived areas.[75] Because participants were recruited from the ages of 40 and 69 from 2006 to 2010, and most participants are white and British, the results of studies using data from this cohort may not be fully representative of the general population.[75] Because every participant included in the subset of the overall UK Biobank cohort used in this analysis has primarily European genetically inferred ancestry, the results of this analysis may not be applicable to individuals who do not have primarily European ancestry.

Quantitative PCR can be used to estimate mean relative telomere length by comparing the amplification of telomeric TTAGGG repeats to that of a single reference gene.[76] This ratio is proportional to total telomere length.[76] There is evidence that it is the shortest telomere rather than average telomere length in a cell that triggers cellular senescence, so potentially useful information may be lost when using average telomere lengths.[77,78] qPCR has been found to be a sensitive and valid tool to estimate mean telomere length[79], but qPCR telomere length estimates are very sensitive to variation in DNA isolation method and PCR conditions, which can lead to inconsistent results even within the same laboratory.[80] However, UK Biobank samples were processed in the same laboratory at the University of Leicester, and telomere length estimates have been adjusted for technical variation and have been found to have good reproducibility.[81] Additionally, qPCR estimates of telomere length have not been found to be consistently sensitive enough to identify patients with dyskeratosis congenita when their telomeres are in the first percentile for their age.[82]

Analyses were limited to less than 600,000 SNPs (MAF > 1%) present on the Applied Biosystems UK BiLEVE Axiom Array and the Applied Biosystems UK Biobank Axiom Array. As such, this analysis of

common variants is not a comprehensive assessment of all possible variants that may play a role in the determination of telomere length.

Finally, a two-year latency period was selected because it is frequently used for radiation-related hematologic malignancies.[50–52] While some reported hematologic malignancies within the UK Biobank cohort may be partially attributable to radiation exposure, there is no reason to assume that all cases are radiation-related. Therefore, this latency threshold may represent an overly conservative approach to reducing the risk of reverse causation.

Conclusions

This study identified a potential gene-by-smoking interaction on telomere length in an analysis of a limited set of 581,069 common SNPs among a cohort of individuals with primarily genetically inferred European ancestry. It is possible that additional or stronger interaction signals could be uncovered by expanding the analysis to include a broader range of variants, including common SNPs not included on the genotyping arrays and rare variants. Additionally, the results of this thesis may only be applicable to individuals with primarily European genetic ancestry, so it would be ideal to expand this investigation such that results are applicable to individuals with diverse genetic ancestries. Several planned analyses remain for future research including the full spectrum of genetic ancestries included in the overall UK Biobank cohort.

First, this project focused only on common SNPs (MAF ≥ 1%) measured on genotyping arrays. Future analyses should incorporate both common SNPs not included on the arrays and rare variants in analyses of whole genome sequencing. Such variants may interact with smoking and account for additional unexplained variability in telomere length.

Second, any variants identified as smoking-interaction candidates for telomere length should be evaluated for whether they also interact with smoking in the development of lung cancer. While rs4418881 was not found to be directly associated with lung cancer in any prior lung cancer GWAS reviewed, including in GWAS conducted for ever smokers alone, it is possible that other variants identified may play a role in the smoking-lung cancer relationship. Because smoking is causal for lung cancer and associated with telomere length, and telomere length itself has a complex relationship with cancer (increased risk of cancer at both very short and very long telomeres), exploring whether the same

variants modify both outcomes could clarify underlying biological relationships and potentially contribute to a better understanding of how smoking causes lung cancer.

Finally, future work should include the creation and evaluation of polygenic scores for telomere length. This approach could complement single-variant analyses by capturing the cumulative effects of many variants in the investigation of potential gene-environment interactions. By stratifying individuals by their genetically predicted telomere length, we could test whether the association between smoking and observed telomere length differs. It is possible that the magnitude of interaction, if present at all, may differ between single variants and aggregations of variants.

Together, these analyses may shed light on the interplay between genetics, smoking, telomere length, and telomere-related disease risk.

V. References

1. Blasco MA. Telomeres and human disease: ageing, cancer and beyond. *Nat Rev Genet*. 2005;6(8):611-622. doi:10.1038/nrg1656
2. Lange T de. Shelterin-Mediated Telomere Protection. *Annu Rev Genet*. 2018;52(Volume 52, 2018):223-247. doi:10.1146/annurev-genet-032918-021921
3. Savage SA. Telomere length and cancer risk: finding Goldilocks. *Biogerontology*. 2024;25(2):265-278. doi:10.1007/s10522-023-10080-9
4. Chakravarti D, LaBella KA, DePinho RA. Telomeres: history, health, and hallmarks of aging. *Cell*. 2021;184(2):306-322. doi:10.1016/j.cell.2020.12.028
5. Chan SRWL, Blackburn EH. Telomeres and telomerase. Sherratt DJ, West SC, eds. *Philos Trans R Soc Lond B Biol Sci*. 2004;359(1441):109-122. doi:10.1098/rstb.2003.1370
6. Andreu-Sánchez S, Aubert G, Ripoll-Cladellas A, et al. Genetic, parental and lifestyle factors influence telomere length. *Commun Biol*. 2022;5(1):1-14. doi:10.1038/s42003-022-03521-7
7. Allsopp RC, Vaziri H, Patterson C, et al. Telomere length predicts replicative capacity of human fibroblasts. *Proc Natl Acad Sci*. 1992;89(21):10114-10118. doi:10.1073/pnas.89.21.10114
8. Karimian K, Groot A, Huso V, et al. Human telomere length is chromosome end–specific and conserved across individuals. *Science*. 2024;384(6695):533-539. doi:10.1126/science.ado0431
9. Benetos A, Kark JD, Susser E, et al. Tracking and fixed ranking of leukocyte telomere length across the adult life course. *Aging Cell*. 2013;12(4):615-621. doi:10.1111/acel.12086
10. Factor-Litvak P, Susser E, Kezios K, et al. Leukocyte Telomere Length in Newborns: Implications for the Role of Telomeres in Human Disease. *Pediatrics*. 2016;137(4):e20153927. doi:10.1542/peds.2015-3927
11. Okuda K, Bardeguez A, Gardner JP, et al. Telomere Length in the Newborn. *Pediatr Res*. 2002;52(3):377-381. doi:10.1203/00006450-200209000-00012
12. Lee DC, Im JA, Kim JH, Lee HR, Shim JY. Effect of Long-Term Hormone Therapy on Telomere Length in Postmenopausal Women. *Yonsei Med J*. 2005;46(4):471-479. doi:10.3349/ymj.2005.46.4.471
13. Lin J, Kroenke CH, Epel E, et al. Greater endogenous estrogen exposure is associated with longer telomeres in postmenopausal women at risk for cognitive decline. *Brain Res*. 2011;1379:224-231. doi:10.1016/j.brainres.2010.10.033
14. Broer L, Codd V, Nyholt DR, et al. Meta-analysis of telomere length in 19 713 subjects reveals high heritability, stronger maternal inheritance and a paternal age effect. *Eur J Hum Genet*. 2013;21(10):1163-1168. doi:10.1038/ejhg.2012.303
15. Patel CJ, Manrai AK, Corona E, Kohane IS. Systematic correlation of environmental exposure and physiological and self-reported behaviour factors with leukocyte telomere length. *Int J Epidemiol*. 2017;46(1):44-56. doi:10.1093/ije/dyw043
16. Broer L, Codd V, Nyholt DR, et al. Meta-analysis of telomere length in 19 713 subjects reveals high heritability, stronger maternal inheritance and a paternal age effect. *Eur J Hum Genet*. 2013;21(10):1163-1168. doi:10.1038/ejhg.2012.303
17. Bountziouka V, Musicha C, Allara E, et al. Modifiable traits, healthy behaviours, and leukocyte telomere length: a population-based study in UK Biobank. *Lancet Healthy Longev*. 2022;3(5):e321-e331. doi:10.1016/S2666-7568(22)00072-1
18. Andrew T, Aviv A, Falchi M, et al. Mapping Genetic Loci That Determine Leukocyte Telomere Length in a Large Sample of Unselected Female Sibling Pairs. *Am J Hum Genet*. 2006;78(3):480-486. doi:10.1086/500052
19. Vasa-Nicotera M, Brouilette S, Mangino M, et al. Mapping of a Major Locus that Determines Telomere Length in Humans. *Am J Hum Genet*. 2005;76(1):147-151. doi:10.1086/426734
20. Bischoff C, Graakjaer J, Petersen HC, et al. The Heritability of Telomere Length Among the Elderly and Oldest-Old. *Twin Res Hum Genet*. 2005;8(5):433-439. doi:10.1375/twin.8.5.433
21. Slagboom PE, Droog S, Boomsma DI. Genetic determination of telomere size in humans: a twin study of three age groups. *Am J Hum Genet*. 1994;55(5):876-882.
22. Codd V, Wang Q, Allara E, et al. Polygenic basis and biomedical consequences of telomere length variation. *Nat Genet*. 2021;53(10):1425-1433. doi:10.1038/s41588-021-00944-6
23. Ding Z, Mangino M, Aviv A, Consortium U, Spector T, Durbin R. Estimating telomere length from whole genome sequence data. *Nucleic Acids Res*. 2014;42(9):e75-e75. doi:10.1093/nar/gku181

24. Burren OS, Dhindsa RS, Deevi SVV, et al. Genetic architecture of telomere length in 462,666 UK Biobank whole-genome sequences. *Nat Genet*. 2024;56(9):1832-1840. doi:10.1038/s41588-024-01884-7
25. Rossiello F, Jurk D, Passos JF, d'Adda di Fagagna F. Telomere dysfunction in ageing and age-related diseases. *Nat Cell Biol*. 2022;24(2):135-147. doi:10.1038/s41556-022-00842-x
26. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. Hallmarks of aging: An expanding universe. *Cell*. 2023;186(2):243-278. doi:10.1016/j.cell.2022.11.001
27. McNally EJ, Luncsford PJ, Armanios M. Long telomeres and cancer risk: the price of cellular immortality. *J Clin Invest*. 129(9):3474-3481. doi:10.1172/JCI120851
28. Armanios M. The Role of Telomeres in Human Disease. *Annu Rev Genomics Hum Genet*. 2022;23:363-381. doi:10.1146/annurev-genom-010422-091101
29. Astuti Y, Wardhana A, Watkins J, Wulaningsih W. Cigarette smoking and telomere length: A systematic review of 84 studies and meta-analysis. *Environ Res*. 2017;158:480-489. doi:10.1016/j.envres.2017.06.038
30. Geneletti S, Gallo V, Porta M, Khoury MJ, Vineis P. Assessing causal relationships in genomics: From Bradford-Hill criteria to complex gene-environment interactions and directed acyclic graphs. *Emerg Themes Epidemiol*. 2011;8(1):5. doi:10.1186/1742-7622-8-5
31. Common Cancer Sites - Cancer Stat Facts. SEER. Accessed December 30, 2024. https://seer.cancer.gov/statfacts/html/common.html
32. CDC. Lung Cancer Awareness. Cancer. October 29, 2024. Accessed December 29, 2024. https://www.cdc.gov/cancer/features/lung-cancer.html
33. National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health. *The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General*. Centers for Disease Control and Prevention (US); 2014. Accessed December 29, 2024. http://www.ncbi.nlm.nih.gov/books/NBK179276/
34. CDC. Lung Cancer Risk Factors. Lung Cancer. October 15, 2024. Accessed December 30, 2024. https://www.cdc.gov/lung-cancer/risk-factors/index.html
35. Demanelis K, Jasmine F, Chen LS, et al. Determinants of telomere length across human tissues. *Science*. 2020;369(6509):eaaz6876. doi:10.1126/science.aaz6876
36. Seow WJ, Cawthon RM, Purdue MP, et al. Telomere Length in White Blood Cell DNA and Lung Cancer: A Pooled Analysis of Three Prospective Cohorts. *Cancer Res*. 2014;74(15):4090-4098. doi:10.1158/0008-5472.CAN-14-0459
37. Wu X, Amos CI, Zhu Y, et al. Telomere Dysfunction: A Potential Cancer Predisposition Factor. *JNCI J Natl Cancer Inst*. 2003;95(16):1211-1218. doi:10.1093/jnci/djg011
38. Qian Y, Ding ,Tingting, Wei ,Lijuan, Cao ,Shui, and Yang L. Shorter telomere length of T-cells in peripheral blood of patients with lung cancer. *OncoTargets Ther*. 2016;9:2675-2682. doi:10.2147/OTT.S98488
39. Jang JS, Choi YY, Lee WK, et al. Telomere length and the risk of lung cancer. *Cancer Sci*. 2008;99(7):1385-1389. doi:10.1111/j.1349-7006.2008.00831.x
40. Yuan JM, Beckman KB, Wang R, et al. Leukocyte telomere length in relation to risk of lung adenocarcinoma incidence: Findings from the Singapore Chinese Health Study. *Int J Cancer*. 2018;142(11):2234-2243. doi:10.1002/ijc.31251
41. Lan Q, Cawthon R, Gao Y, et al. Longer Telomere Length in Peripheral White Blood Cells Is Associated with Risk of Lung Cancer and the rs2736100 (CLPTM1L-TERT) Polymorphism in a Prospective Cohort Study among Women in China. *PLOS ONE*. 2013;8(3):e59230. doi:10.1371/journal.pone.0059230
42. Shen M, Cawthon R, Rothman N, et al. A prospective study of telomere length measured by monochrome multiplex quantitative PCR and risk of lung cancer. *Lung Cancer*. 2011;73(2):133-137. doi:10.1016/j.lungcan.2010.11.009
43. Machiela MJ, Hsiung CA, Shu XO, et al. Genetic variants associated with longer telomere length are associated with increased lung cancer risk among never-smoking women in Asia: a report from the female lung cancer consortium in Asia. *Int J Cancer*. 2015;137(2):311-319. doi:10.1002/ijc.29393
44. Cortez Cardoso Penha R, Smith-Byrne K, Atkins JR, et al. Common genetic variations in telomere length genes and lung cancer: a Mendelian randomisation study and its novel application in lung tumour transcriptome. Banovich NE, Franco EL, Voight B, eds. *eLife*. 2023;12:e83118. doi:10.7554/eLife.83118

45. Li Y, Xiao X, Han Y, et al. Genome-wide interaction study of smoking behavior and non-small cell lung cancer risk in Caucasian population. *Carcinogenesis*. 2018;39(3):336-346. doi:10.1093/carcin/bgx113
46. Data-Field 21003. Accessed June 5, 2025. https://biobank.ctsu.ox.ac.uk/ukb/field.cgi?id=21003
47. Baseline assessment. May 27, 2021. Accessed June 5, 2025. https://www.ukbiobank. ac.uk/enable-your-research/about-our-data/baseline-assessment
48. Codd V, Wang Q, Allara E, et al. Polygenic basis and biomedical consequences of telomere length variation. *Nat Genet*. 2021;53(10):1425-1433. doi:10.1038/s41588-021-00944-6
49. Barthel FP, Wei W, Tang M, et al. Systematic analysis of telomere length and somatic alterations in 31 cancer types. *Nat Genet*. 2017;49(3):349-357. doi:10.1038/ng.3781
50. Little MP, Eidemüller M, Kaiser JC, Apostoaei AI. Minimum latency effects for cancer associated with exposures to radiation or other carcinogens. *Br J Cancer*. 2024;130(5):819-829. doi:10.1038/s41416-023-02544-z
51. United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR). *Annex B: Epidemiological Studies of Cancer Risk Due to Low-Dose-Rate Radiation from Environmental Sources*. United Nations; 2017:67-113. https://www.unscear.org/unscear/uploads/documents/publications/UNSCEAR_2017_Annex-B.pdf
52. Ban N, Cléro E, Vaillant L, et al. Radiation detriment calculation methodology: summary of ICRP Publication 152. *J Radiol Prot*. 2022;42(2):023001. doi:https://iopscience.iop.org/article/10.1088/1361-6498/ac670d
53. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics*. 2010;26(22):2867-2873. doi:10.1093/bioinformatics/btq559
54. Abraham G, Qiu Y, Inouye M. FlashPCA2: principal component analysis of Biobank-scale genotype datasets. *Bioinformatics*. 2017;33(17):2776-2778. doi:10.1093/bioinformatics/btx299
55. Codd V, Denniff M, Swinfield C, et al. Measurement and initial characterization of leukocyte telomere length in 474,074 participants in UK Biobank. *Nat Aging*. 2022;2(2):170-179. doi:10.1038/s43587-021-00166-9
56. UK Biobank. *Z-adjusted T/S Log (UK Biobank Field 22192)*. UK Biobank; 2021. Accessed June 8, 2025. https://biobank.ndph.ox.ac.uk/ukb/field.cgi?id=22192
57. Wain LV, Shrine N, Miller S, et al. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *Lancet Respir Med*. 2015;3(10):769-781. doi:10.1016/S2213-2600(15)00283-0
58. Welsh S, Peakman T, Sheard S, Almond R. Comparison of DNA quantification methodology used in the DNA extraction protocol for the UK Biobank cohort. *BMC Genomics*. 2017;18(1):26. doi:10.1186/s12864-016-3391-x
59. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-209. doi:10.1038/s41586-018-0579-z
60. Wellcome Trust Centre for Human Genetics (University of Oxford). *Genotyping and Quality Control of UK Biobank, a Large-Scale, Extensively Phenotyped Prospective Resource*.; 2015. Accessed June 8, 2025. https://biobank.ctsu.ox.ac.uk/crystal/crystal/ docs/genotyping_qc.pdf
61. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. 2015;4(1):s13742-015-0047-0048. doi:10.1186/s13742-015-0047-8
62. Purcell S, Chang C. PLINK 2.0. www.cog-genomics.org/plink/2.0/
63. End-to-end target discovery with GWAS and PheWAS | Research Analysis Platform. September 18, 2024. Accessed June 9, 2025. https://dnanexus.gitbook.io/uk-biobank-rap/science-corner/end-to-end-target-discovery-with-gwas-and-phewas
64. Hemstrom W, Grummer JA, Luikart G, Christie MR. Next-generation data filtering in the genomics era. *Nat Rev Genet*. 2024;25(11):750-767. doi:10.1038/s41576-024-00738-6
65. Turner S, Armstrong LL, Bradford Y, et al. Quality Control Procedures for Genome-Wide Association Studies. *Curr Protoc Hum Genet*. 2011;68(1):1.19.1-1.19.18. doi:10.1002/0471142905.hg0119s68
66. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*. 2017;8(1):1826. doi:10.1038/s41467-017-01261-5

67. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*. 2015;31(21):3555-3557. doi:10.1093/bioinformatics/btv402
68. Lin SH, Brown DW, Machiela MJ. LDtrait: An Online Tool for Identifying Published Phenotype Associations in Linkage Disequilibrium. *Cancer Res*. 2020;80(16):3443-3446. doi:10.1158/0008-5472.CAN-20-0985
69. rs4418881 RefSNP Report - dbSNP - NCBI. Accessed June 11, 2025. https://www.ncbi.nlm.nih.gov/snp/rs4418881
70. McKay JD, Hung RJ, Han Y, et al. Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. *Nat Genet*. 2017;49(7):1126-1132. doi:10.1038/ng.3892
71. Rashkin SR, Graff RE, Kachuri L, et al. Pan-cancer study detects genetic risk variants and shared genetic basis in two large cohorts. *Nat Commun*. 2020;11(1):4423. doi:10.1038/s41467-020-18246-6
72. Byun J, Han Y, Li Y, et al. Cross-ancestry genome-wide meta-analysis of 61,047 cases and 947,237 controls identifies new susceptibility loci contributing to lung cancer. *Nat Genet*. 2022;54(8):1167-1177. doi:10.1038/s41588-022-01115-x
73. Shi J, Shiraishi K, Choi J, et al. Genome-wide association study of lung adenocarcinoma in East Asia and comparison with a European population. *Nat Commun*. 2023;14(1):3043. doi:10.1038/s41467-023-38196-z
74. Verma A, Huffman JE, Rodriguez A, et al. Diversity and scale: Genetic architecture of 2068 traits in the VA Million Veteran Program. *Science*. 2024;385(6706):eadj1182. doi:10.1126/science.adj1182
75. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am J Epidemiol*. 2017;186(9):1026-1034. doi:10.1093/aje/kwx246
76. Cawthon RM. Telomere measurement by quantitative PCR. *Nucleic Acids Res*. 2002;30(10):e47.
77. Hemann MT, Strong MA, Hao LY, Greider CW. The Shortest Telomere, Not Average Telomere Length, Is Critical for Cell Viability and Chromosome Stability. *Cell*. 2001;107(1):67-77. doi:10.1016/S0092-8674(01)00504-9
78. Xu Z, Duc KD, Holcman D, Teixeira MT. The Length of the Shortest Telomere as the Major Determinant of the Onset of Replicative Senescence. *Genetics*. 2013;194(4):847-857. doi:10.1534/genetics.113.152322
79. Barrett JH, Iles MM, Dunning AM, Pooley KA. Telomere length and common disease: study design and analytical challenges. *Hum Genet*. 2015;134(7):679-689. doi:10.1007/s00439-015-1563-4
80. Ferrer A, Stephens ZD, Kocher JPA. Experimental and Computational Approaches to Measure Telomere Length: Recent Advances and Future Directions. *Curr Hematol Malig Rep*. 2023;18(6):284-291. doi:10.1007/s11899-023-00717-4
81. Codd V, Denniff M, Swinfield C, et al. A major population resource of 474,074 participants in UK Biobank to investigate determinants and biomedical consequences of leukocyte telomere length. Published online March 24, 2021:2021.03.18.21253457. doi:10.1101/2021.03.18.21253457
82. Gadalla SM, Khincha PP, Katki HA, et al. The limitations of qPCR telomere length measurement in diagnosing dyskeratosis congenita. *Mol Genet Genomic Med*. 2016;4(4):475-479. doi:10.1002/mgg3.220
83. Allaire P, He J, Mayer J, et al. Genetic and clinical determinants of telomere length. *Hum Genet Genomics Adv*. 2023;4(3):100201. doi:10.1016/j.xhgg.2023.100201
84. Levy D, Neuhausen SL, Hunt SC, et al. Genome-wide association identifies OBFC1 as a locus involved in human leukocyte telomere biology. *Proc Natl Acad Sci*. 2010;107(20): 9293-9298. doi:10.1073/pnas.0911494107
85. Lyu X, Sang PB, Chai W. CST in maintaining genome stability: Beyond telomeres. *DNA Repair*. 2021;102:103104. doi:10.1016/j.dnarep.2021.103104
86. Awad A, Glousker G, Lamm N, et al. Full length RTEL1 is required for the elongation of the single-stranded telomeric overhang by telomerase. *Nucleic Acids Res*. 2020;48(13): 7239-7251. doi:10.1093/nar/gkaa503

VI. Tables

Table 1. Thesis cohort characteristics.

Characteristic	**Total** N = 360,909
Sex	
Female, n(%)	194,950 (54.0%)
Male, n(%)	165,959 (46.0%)
Age	
Median	58
Mean (SD)	57.7 (8.0)
Range	39 – 73
Ever Smoked	
Yes, n(%)	219,570 (61%)
No, n(%)	141,339 (39%)
Technically adjusted, $\log_e$-transformed, Z-standardized telomere length, with values <-3 and >3 standard deviations from the mean removed	
Mean (SD)	-0.01 (0.96)
Range	-3 – 3
Genotyping Array	
UK BiLEVE Axiom, n(%)	38,842 (10.8%)
UK Biobank Axiom, n(%)	322,067 (89.2%)

Table 2. Beta value estimates with 95% confidence intervals (calculated using non-robust standard errors) and associated p-values for rs4418881 across the different models/tests evaluating its association with telomere length. The 2 degrees of freedom joint test does not result in a beta estimate.

Model / Test	N	Beta (95% CI)	P-value
GWAS – Overall	360,909	0.0118 (0.0072, 0.0163)	4.19×10^{-7}
GWAS – Only Smokers	219,570	0.0171 (0.0112, 0.0230)	1.06×10^{-8}
GWAS – Only Nonsmokers	141,339	0.0035 (-0.0038, 0.0102)	0.352
GxE – 1 d.f. Interaction	360,909	0.0137 (0.0043, 0.0230)	0.004
GxE – 2 d.f. Joint Test	360,909	—	4.49×10^{-8}

VII. Figures

GWAS – Entire Thesis Cohort

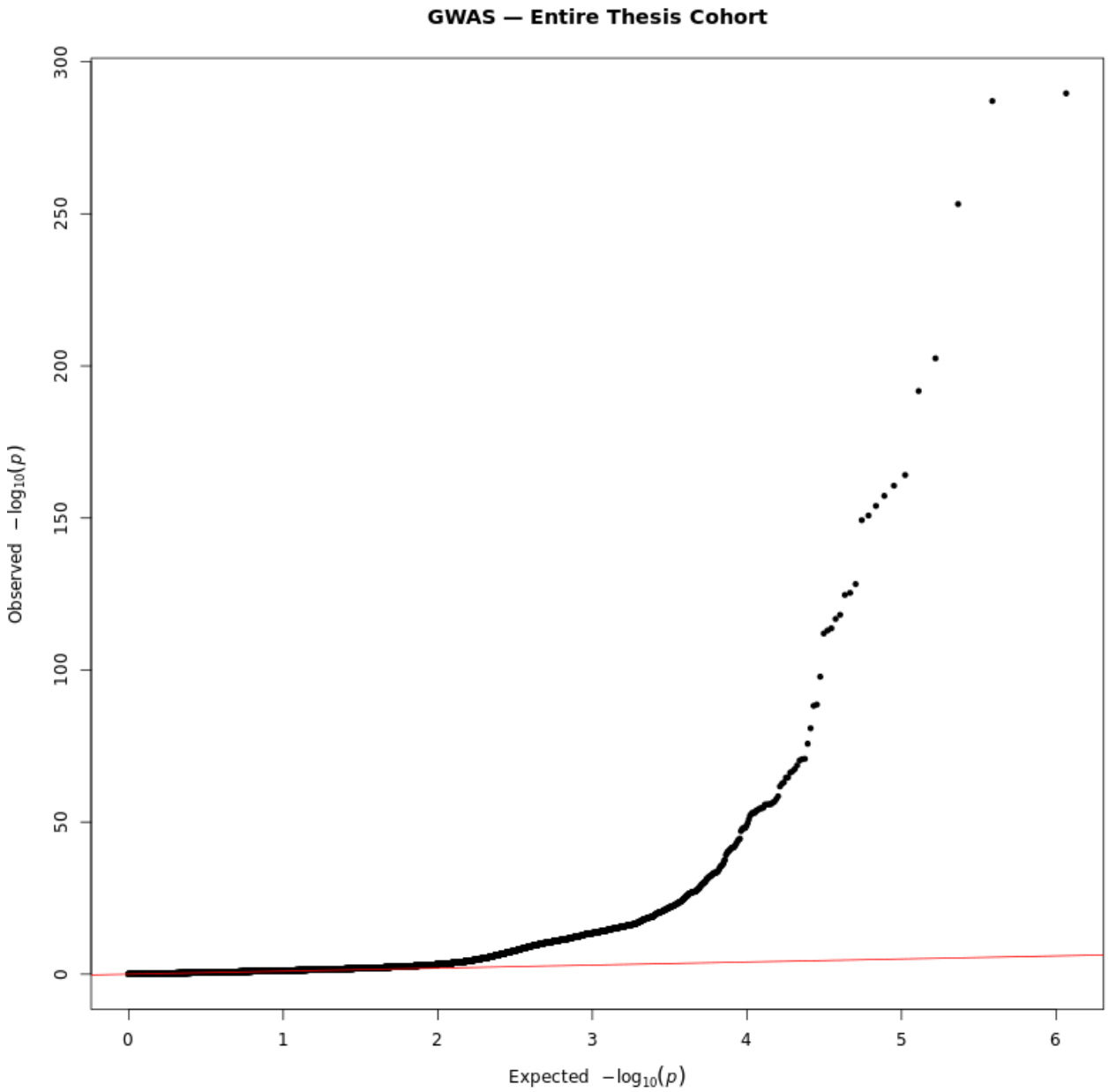

Figure 1. Quantile-quantile (Q-Q) plot for the GWAS conducted on the entire thesis cohort. The red line represents the expected distribution of test statistics under the null hypothesis. The genomic inflation factor (l) was 1.155.

GWAS – Entire Thesis Cohort

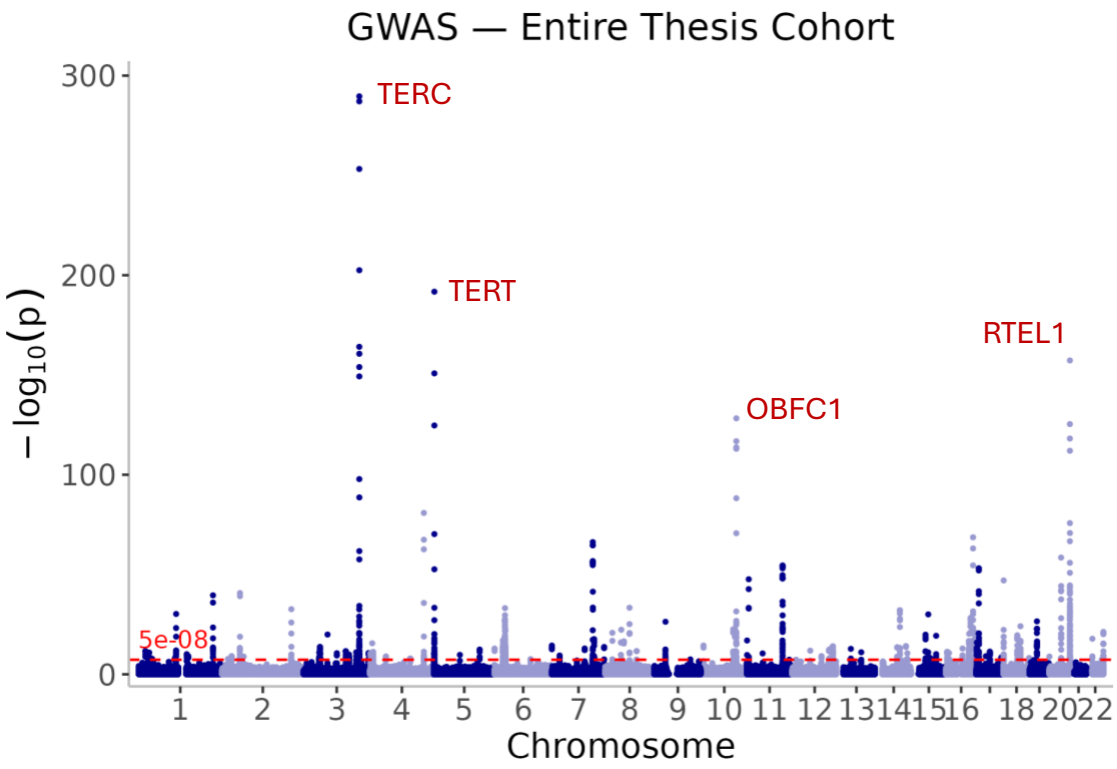

Figure 2. Manhattan plot for the GWAS conducted on the entire thesis cohort. The red dotted line represents the genome-wide significance threshold ($P < 5 \times 10^{-8}$). Independent lead SNPs were annotated and mapped to protein-coding genes with FUMA, and several are labeled. The peaks represent SNP associations with known telomere-related loci. *TERC* is the gene that encodes the RNA component of telomerase.[83] *TERT* is the gene that encodes the catalytic component of telomerase.[83] OBFC1 is a gene that encodes a protein that is involved in binding to single-stranded telomeric DNA.[83–85] RTEL1 encodes a DNA helicase known to play a role in elongating and stabilizing telomeres.[83,86]

GWAS – Subset of Thesis Cohort Who Have Ever Smoked

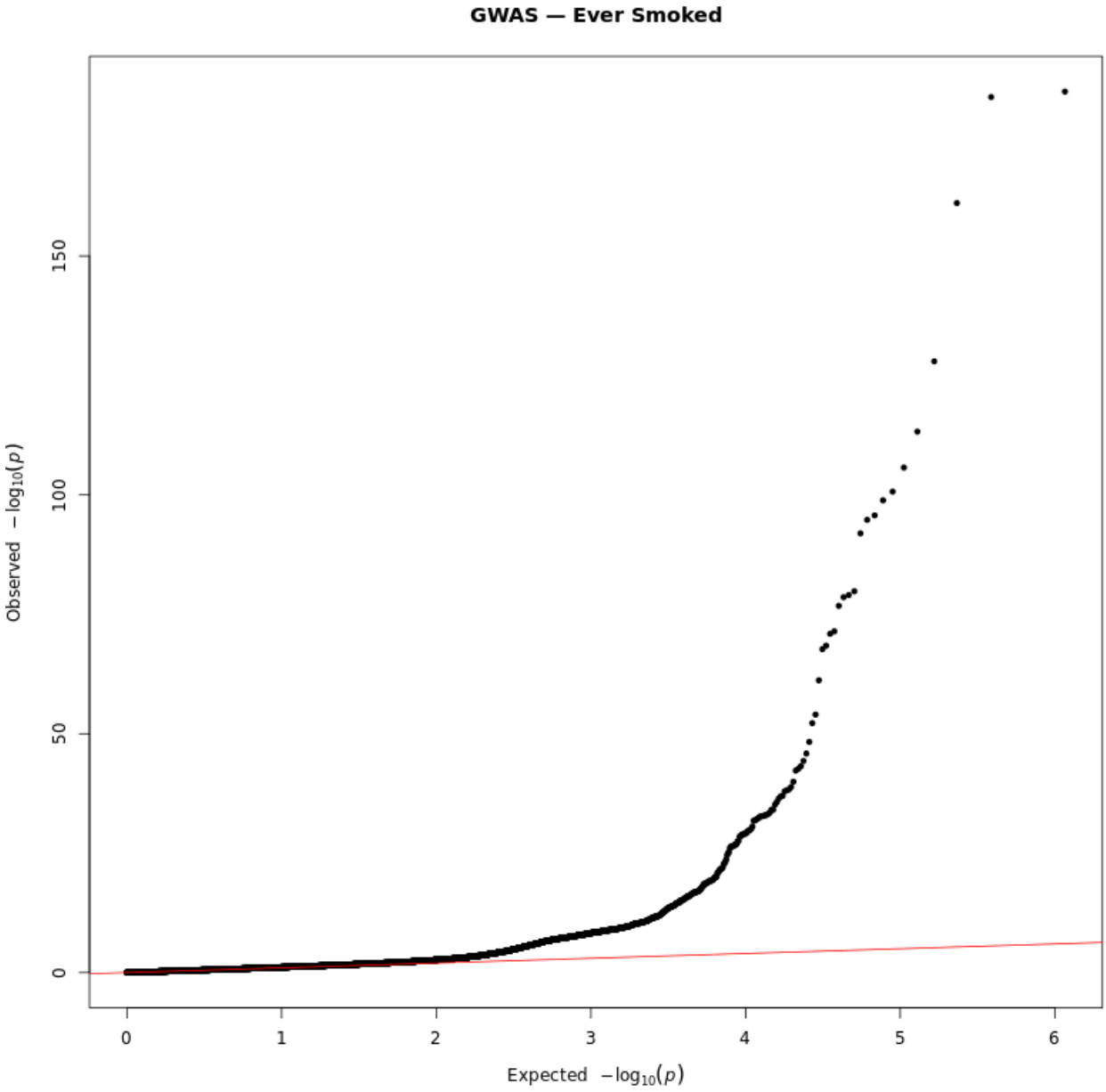

Figure 3. Quantile-quantile (Q–Q) plot for the GWAS conducted on the subset of the thesis cohort who have ever smoked. The red line represents the expected distribution of test statistics under the null hypothesis. The genomic inflation factor (l) was 1.113.

GWAS – Subset of Thesis Cohort Who Have Ever Smoked

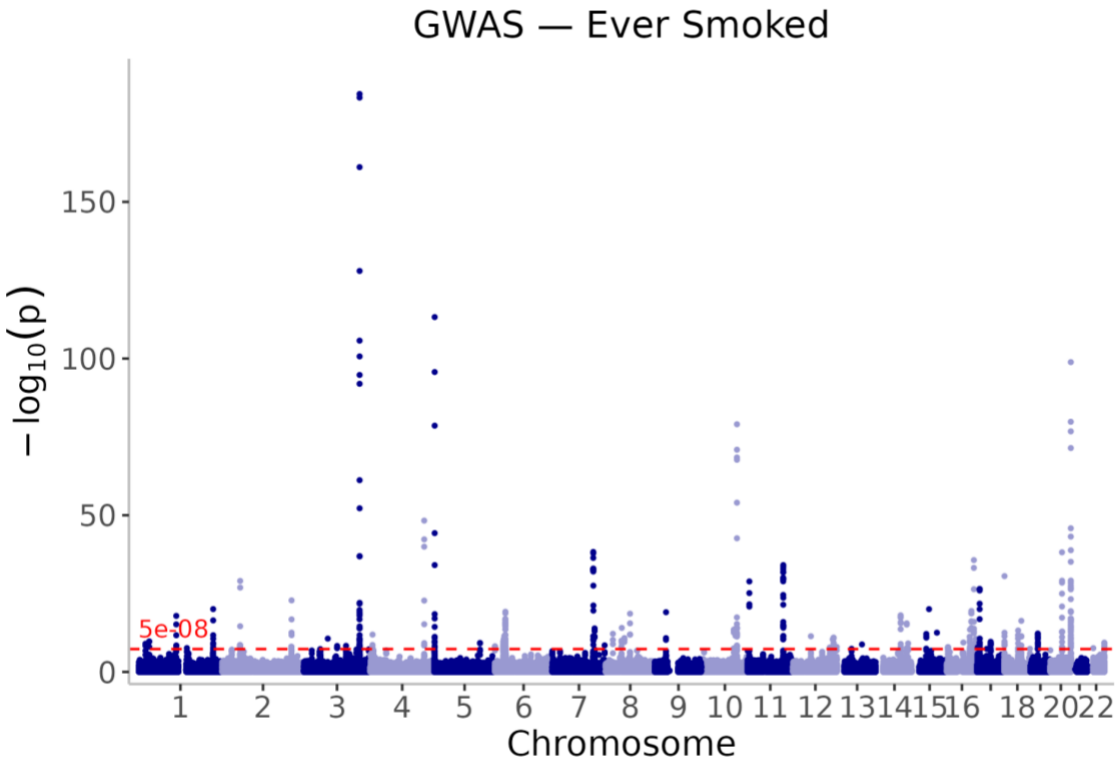

Figure 4. Manhattan plot for the GWAS conducted on the subset of the thesis cohort who have ever smoked. The red dotted line represents the genome-wide significance threshold ($P < 5 \times 10^{-8}$).

GWAS – Subset of Thesis Cohort Who Have Never Smoked

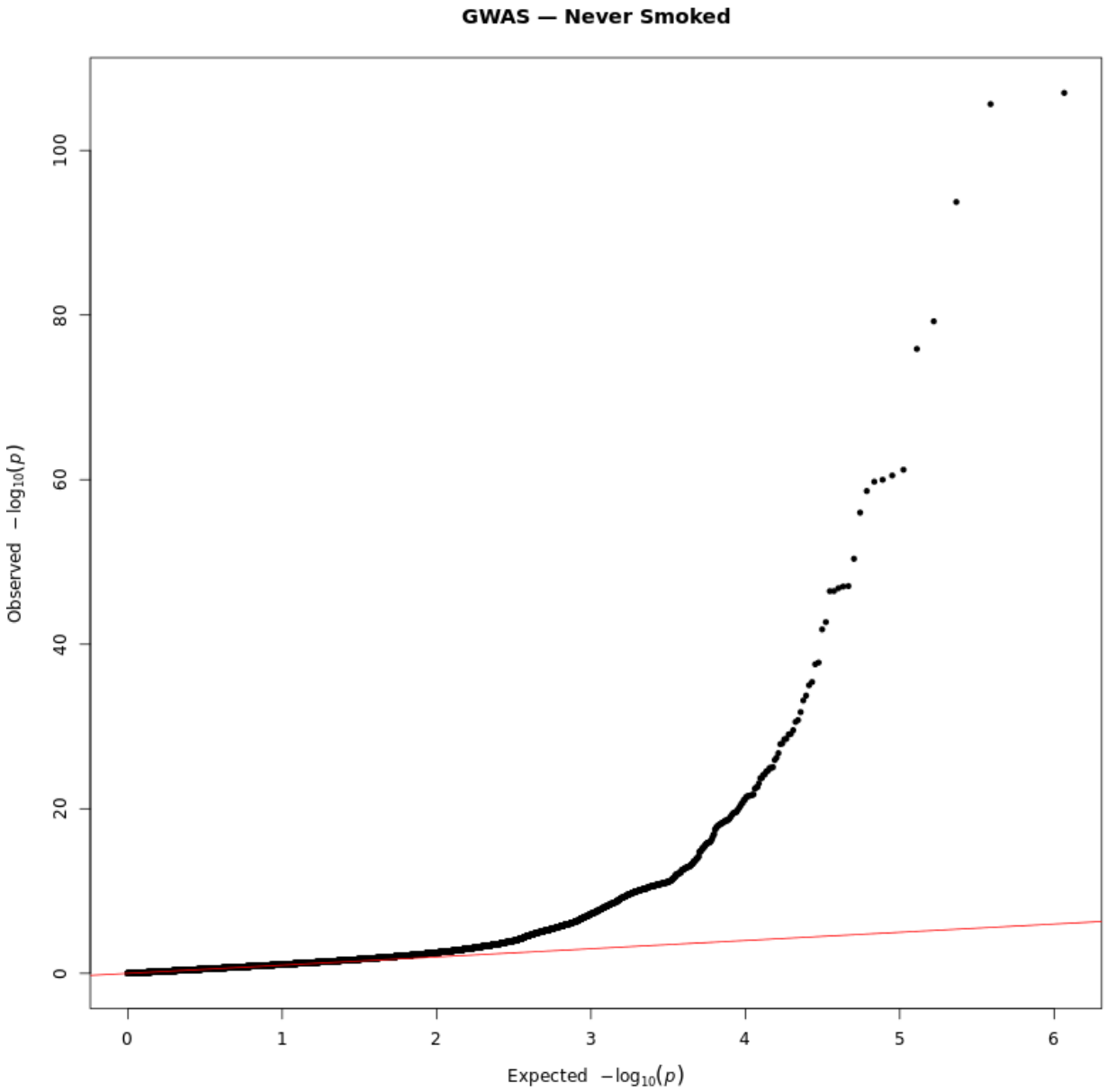

Figure 5. Quantile-quantile (Q–Q) plot for the GWAS conducted on the subset of the thesis cohort who have never smoked. The red line represents the expected distribution of test statistics under the null hypothesis. The genomic inflation factor (l) was 1.08.

GWAS – Subset of Thesis Cohort Who Have Never Smoked

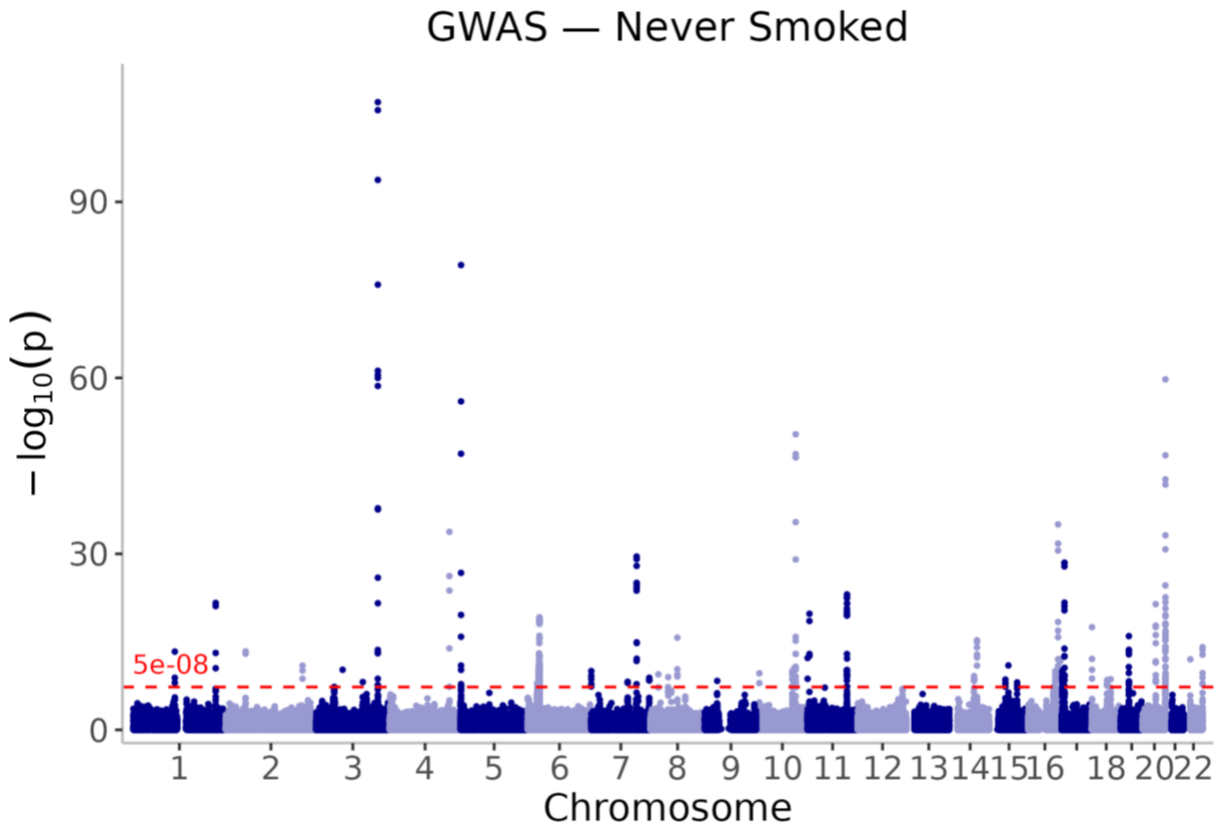

Figure 6. Manhattan plot for the GWAS conducted on the subset of the thesis cohort who have never smoked. The red dotted line represents the genome-wide significance threshold ($P < 5 \times 10^{-8}$).

GxE – Entire Thesis Cohort – 1 Degree of Freedom Interaction Term

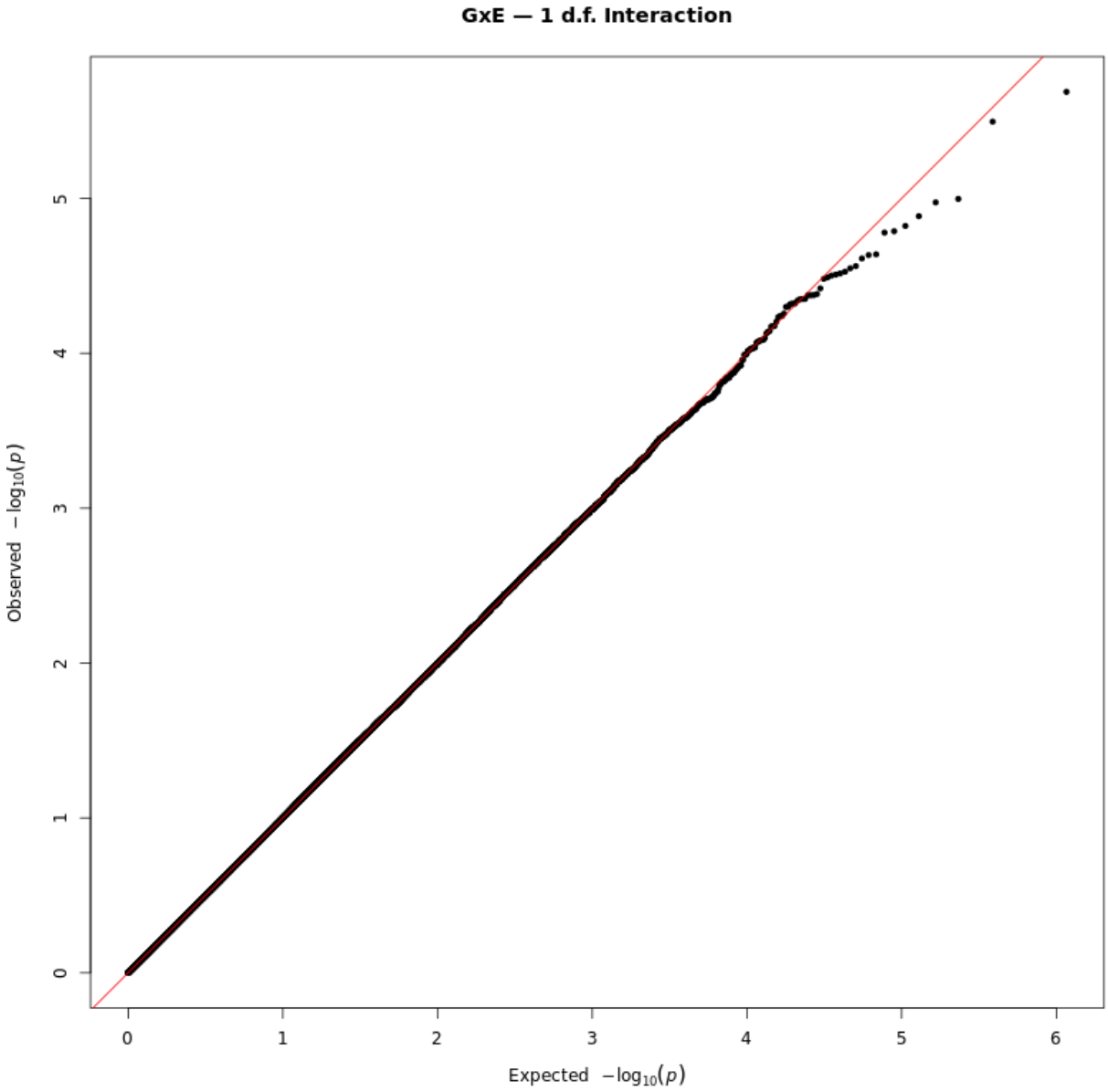

Figure 7. Quantile-quantile (Q-Q) plot for the GxE 1 degree of freedom interaction term. The red line represents the expected distribution of test statistics under the null hypothesis. The genomic inflation factor (I) was 0.999.

GxE – Entire Thesis Cohort – 1 Degree of Freedom Interaction Term

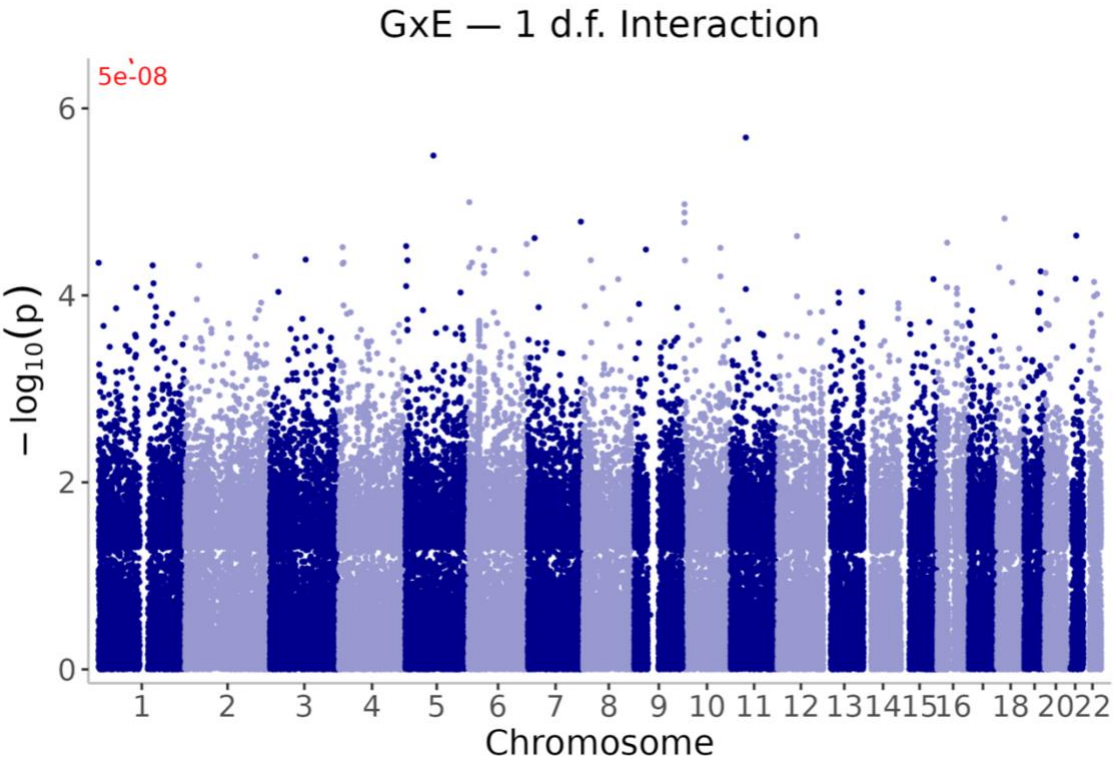

Figure 8. Manhattan plot for the GxE 1 degree of freedom interaction term. The red text indicates that the genome-wide significance threshold ($P < 5 \times 10^{-8}$) is higher on the y-axis than is shown in this plot because no variants have a p-value below that threshold.

GxE – Entire Thesis Cohort – 2 Degrees of Freedom Joint Test

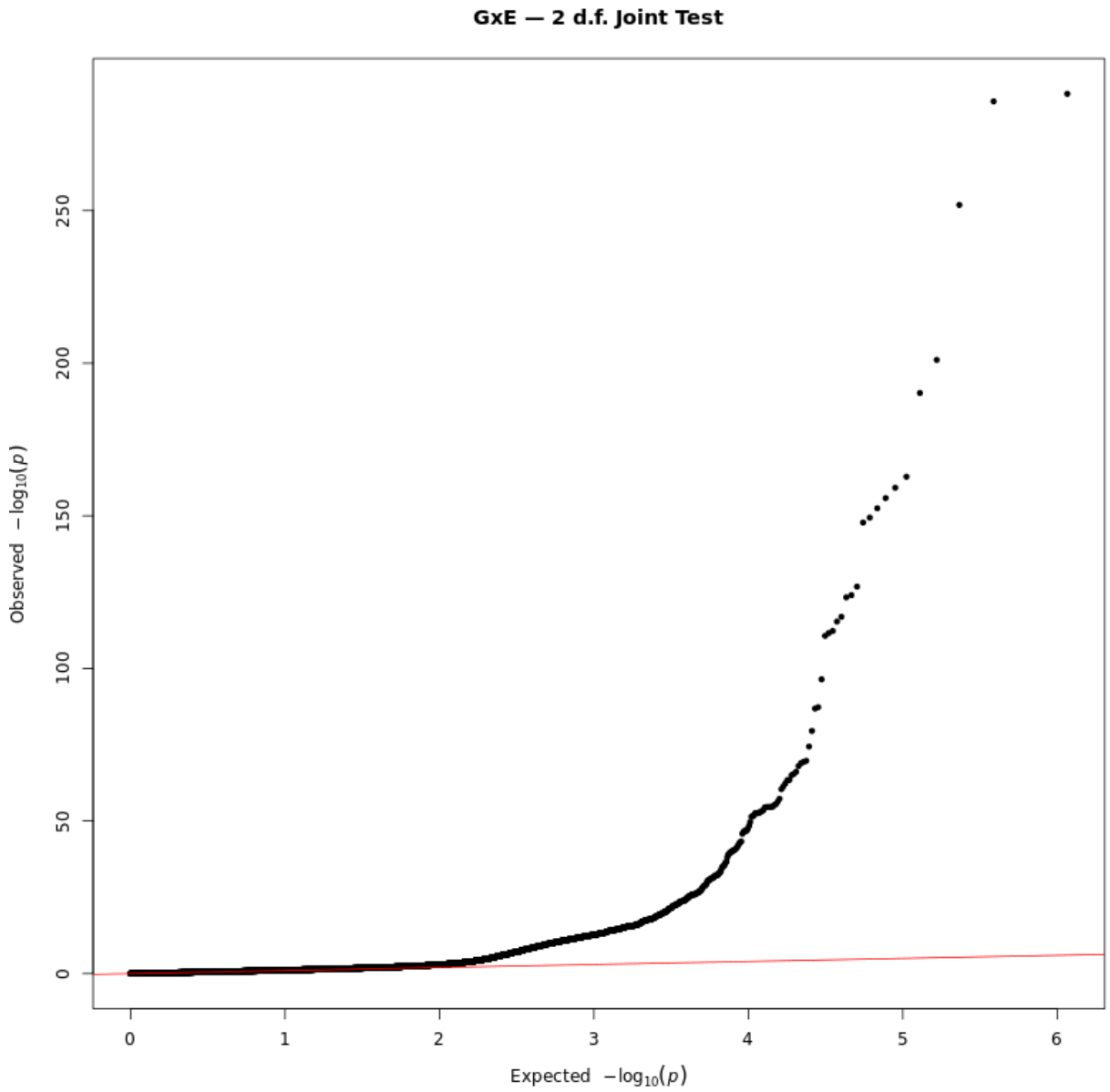

Figure 9. Quantile-quantile (Q–Q) plot for the GxE 2 degrees of freedom joint test. The red line represents the expected distribution of test statistics under the null hypothesis. The genomic inflation factor (l) was 1.082.

GxE – Entire Thesis Cohort – 2 Degrees of Freedom Joint Test

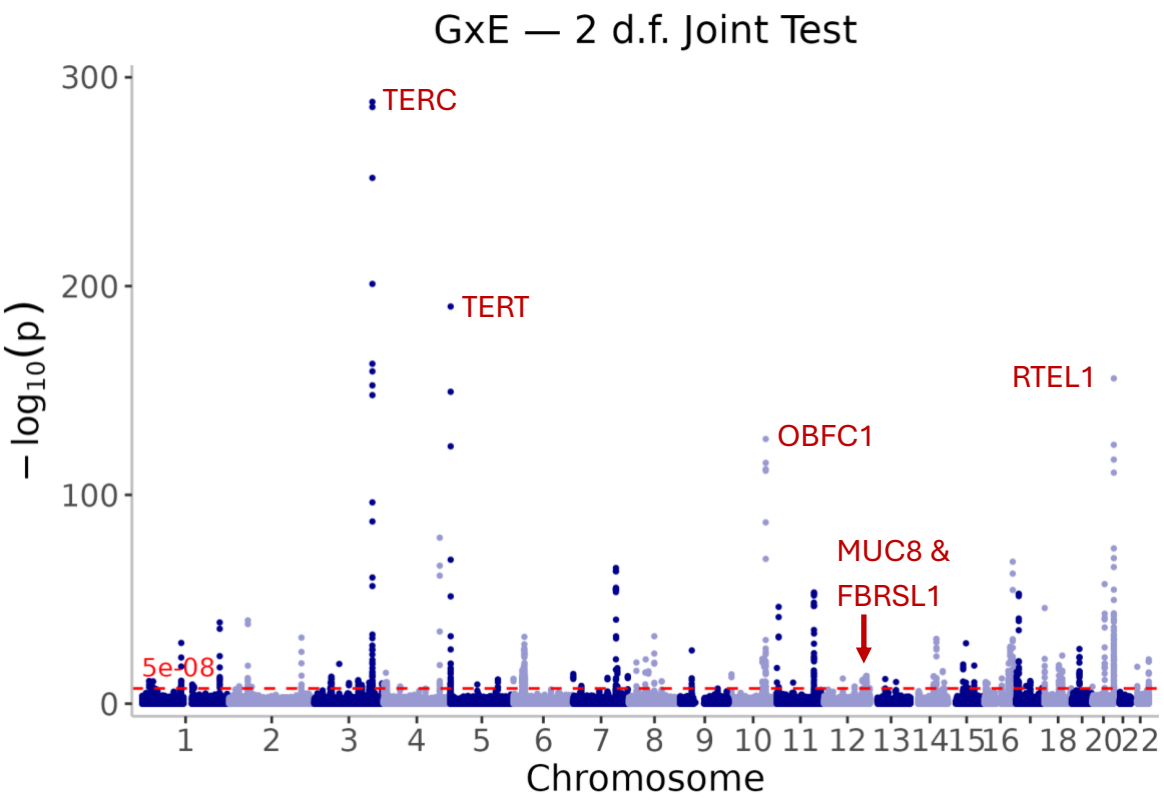

Figure 10. Manhattan plot for the GxE 2 degrees of freedom (d.f.) joint test. The red dotted line represents the genome-wide significance threshold ($P < 5 \times 10^{-8}$). Independent lead SNPs were annotated and mapped to protein-coding genes with FUMA, and several are labeled as described in Figure 2. All independent lead SNPs identified in the 2 d.f. joint test were identified in the GWAS conducted on the entire thesis cohort besides one: rs4418881. This SNP was mapped with FUMA between *MUC8* and *FBRSL1*.

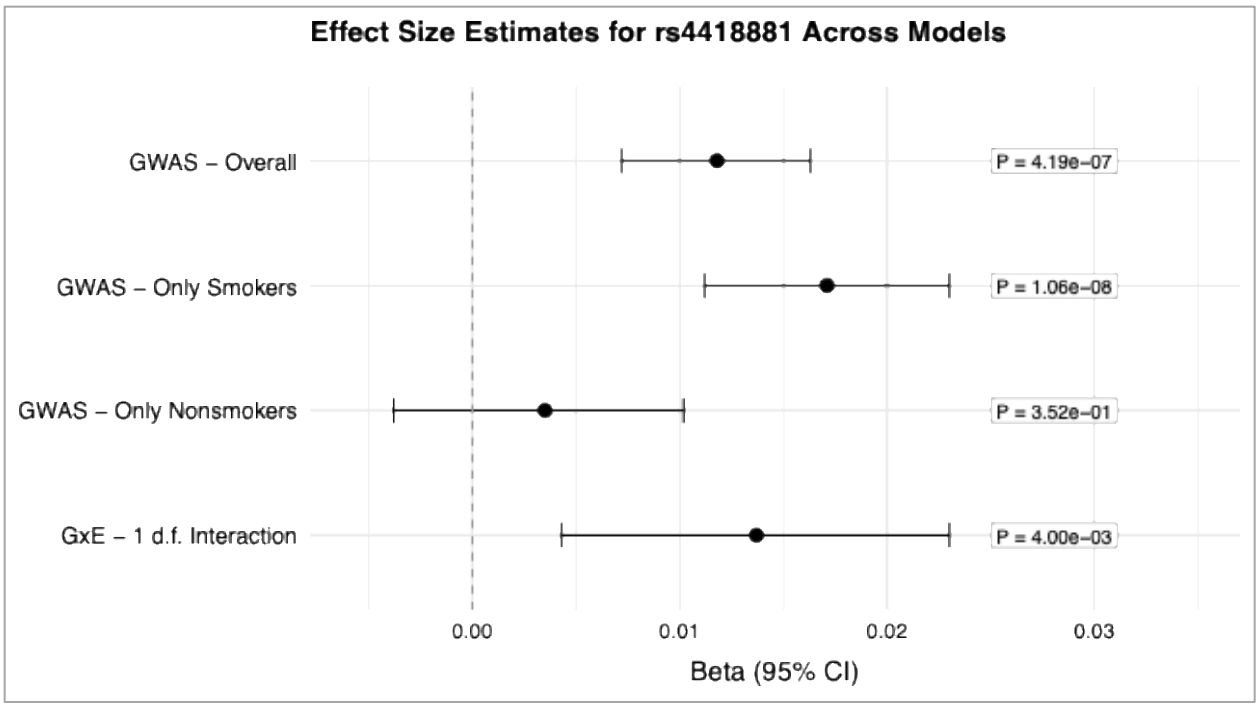

Figure 11. Forest plot of beta value estimates with 95% confidence intervals (calculated using non-robust standard errors) and associated p-values for rs4418881 across different models evaluating its association with telomere length.