Article:
Telomeres Mendelian Randomization Collaboration, , Haycock, PC, Burgess, S et al. (196 more authors) (2017) Association Between Telomere Length and Risk of Cancer and Non-Neoplastic Diseases: A Mendelian Randomization Study. JAMA Oncology, 3 (5). pp. 636-651. ISSN 2374-2437

https://doi.org/10.1001/jamaoncol.2016.5945

The association between genetically longer telomeres and risk of cancer and non-neoplastic diseases

The Telomere Length Mendelian Randomization Studies Collaboration

Correspondence: Philip Haycock

MRC Integrative Epidemiology Unit

University of Bristol

Bristol

UK

philip.haycock@bristol.ac.uk

Tel: +44 1173 310 088

3185 words [word limit 3000]

3 figures, 2 tables, 130 references; 7 supplementary figures / 6 supplementary tables

ABSTRACT 341 WORDS

Importance Due to the susceptibility of observational studies to confounding and reverse causation, the causal direction and magnitude of the association between telomere length and incidence of cancer and non-neoplastic diseases is uncertain.

Objective To appraise the causal relevance of telomere length for risk of cancer and non-neoplastic diseases using germline genetic variants as instrumental variables.

Data Sources Genome-wide association studies (GWAS) published up to January 15 2015.

Study Selection GWAS of non-communicable diseases that assayed germline genetic variation and did not select cohort or control participants on the basis of pre-existing diseases. Of 163 GWAS of non-communicable diseases identified, 103 shared data for our study.

Data Extraction Summary association statistics for single nucleotide polymorphisms (SNPs) that are strongly associated with telomere length in the general population.

Main Outcomes Odds ratios (ORs) for disease per 1-SD higher telomere length due to germline genetic variation.

Results Summary data were available for 35 cancers and 47 non-neoplastic diseases, corresponding to 409,819 cases (median 2,092 per disease) and 1,404,633 controls (median 7,738 per disease). Increased telomere length due to germline genetic variation was generally associated with increased risk for site-specific cancers. The strongest associations were observed for (ORs per 1-SD change in genetically longer telomeres): glioma 5.27 (3.15, 8.81), serous low malignant potential ovarian cancer 4.35 (2.39-7.94); lung adenocarcinoma 3.19 (2.40-4.22); neuroblastoma 2.98 (1.92-4.62); bladder cancer 2.19 (1.32-3.66); melanoma 1.87 (1.55, 2.26); testicular cancer 1.76 (1.02-3.04); kidney cancer 1.55 (1.08-2.23); and endometrial cancer 1.31 (1.07-1.61). Associations with cancer were stronger for rarer cancers and tissue sites with lower rates of stem cell division ($P<0.05$). There was generally little

evidence of association between genetically longer telomeres and risk of psychiatric, autoimmune, inflammatory, diabetic and other non-neoplastic diseases, except for coronary heart disease (0.78 [0.67-0.90]), abdominal aortic aneurysm (0.63 [0.49-0.81]), celiac disease (0.42 [0.28-0.61]) and interstitial lung disease (0.09 [0.05- 0.15]).

Conclusions Genetically longer telomeres are associated with increased risk for several cancers, but the relative increase in risk is highly heterogeneous across cancer types, and with reduced risk for some non-neoplastic diseases, including cardiovascular diseases.

INTRODUCTION

Telomeres are DNA-protein structures at the end of linear chromosomes that protect the genome from damage, shorten progressively over time in most somatic tissues[1] and are proposed markers of biological ageing. Shorter leukocyte telomeres are correlated with older age, male sex and other known risk factors for non-communicable diseases[2–4] and are generally associated with higher risk of cardiovascular diseases[5,6], type 2 diabetes[7] and non-vascular non-neoplastic causes of mortality.[6] Whether these associations are causal, however, is unknown. Telomere length has also been implicated in risk of cancer but the direction and magnitude of the association is uncertain and contradictory across observational studies.[8–12] The uncertainty reflects the considerable difficulty of designing observational studies of telomere length and cancer incidence that are robust to reverse causation, confounding and measurement error. For example, it is possible to detect changes in rates of telomere attrition in cancer cases 3-4 years prior to diagnosis[12], suggesting that even well designed prospective studies may be susceptible to reverse causation. These limitations undermine the potential clinical application of telomere length as a tool for risk prediction and disease prevention.

The aim of the present report was to conduct a Mendelian randomization study to help clarify the nature of the association between telomere length and risk of cancer and non-neoplastic diseases, using germline genetic variants as instrumental variables for telomere length. The approach, which mimics the random allocation of individuals to the placebo and intervention arms of a randomized controlled trial, allowed us to: (1) estimate the direction and broad magnitude of the association of telomere length with risk of multiple cancer and non-neoplastic diseases; (2) appraise the evidence for causality in the estimated etiological associations; (3) investigate potential sources of heterogeneity in findings for site-specific

cancers; and (4) compare genetic estimates to findings based on directly measured telomere length in prospective observational studies.

METHODS

Study design

The design of our study, illustrated in Figure S1, had three key components: 1) the identification of genetic variants to serve as proxies for telomere length; 2) the acquisition of summary data for the genetic proxies from genome wide association studies (GWASs) of diseases and risk factors; and 3) the classification of diseases and risk factors into primary or secondary outcomes based on a priori statistical power. As a first step, we searched the GWAS catalog[13,14] on the 15 January 2015, to identify single nucleotide polymorphisms (SNPs) associated with telomere length. To supplement the list with additional potential proxies, we also searched the original study reports curated by the GWAS catalog (using a P value threshold of $5x10^{-8}$).[15–23] We acquired summary data for all SNPs identified by our search from a meta-analysis of GWASs of telomere length, involving 9,190 participants of European ancestry.[16] SNPs initially identified as potential proxies for telomere length were subsequently excluded if they lacked strong evidence of association with telomere length. We defined strong evidence of association as a p-value $<5x10^{-8}$ in: i) the discovery stage of at least one published GWAS of telomere length[15–22] or ii) a meta-analysis of summary data from Mangino et al[16] and other GWASs of telomere length,[15,17–22] with any overlapping studies excluded from Mangino et al.[16] We also excluded SNPs with a minor allele frequency <0.05 or showing strong evidence of between-study heterogeneity in associations with telomere length ($P \leq 0.001$).

The second key component of our design strategy involved the acquisition of summary data, corresponding to the selected genetic proxies for telomere length, from GWASs of non-communicable diseases and risk factors (Fig. S1). As part of this step, we invited principal investigators of non-communicable disease studies curated by the GWAS catalog[13,14] to share summary data for our study (see Fig. S1 for further details). We also downloaded summary data for diseases and risk factors from publically available sources, including study-specific websites, dbGAP and the GWAS catalog (Fig. S1).

The third key component of our design strategy was the classification of diseases and risk factors into either primary or secondary outcomes, which we defined on the basis of a priori statistical power to detect associations with telomere length. Primary outcomes were defined as diseases with sufficient cases and controls for >50% power (i.e. moderate-to-high statistical power) and secondary outcomes defined as diseases with <50% power (i.e. low statistical power) to detect odds ratios ≥2.0 per standard deviation increase in telomere length (alpha assumed to be 0.01). All risk factors were defined as secondary outcomes. Risk factors with low statistical power were excluded from all analyses. Further details on the power calculations and the study design are provided in the supplementary methods.

Comparison with prospective observational studies

We searched PubMed for prospective observational studies of the association between telomere length and disease (see Tables S3 and S4 for details of the search strategy and inclusion criteria). Study-specific relative risks for disease per unit change or quantile comparison of telomere length were transformed to a standard deviation scale using previously described methods.[24] Hazard ratios, risk ratios, and odds ratios were assumed to approximate the same measure of relative risk. Where multiple independent studies of the same disease were identified, these were combined by fixed effects meta-analysis, unless

there was strong evidence of between-study heterogeneity ($P_{Cochran's\ Q}<0.001$), in which case they were kept separate.

Statistical analysis

We combined summary data across SNPs into a single genetic risk score, using maximum likelihood to estimate the slope of the relationship between β_{GD} and β_{GP} and a variance-covariance matrix to make allowance for linkage disequilibrium between SNPs,[25] where β_{GD} is the change in disease or risk factor per copy of the effect allele and β_{GP} is the standard deviation change in telomere length per copy of the effect allele (see supplementary methods for technical details). The slope from this approach can be interpreted as the log odds ratio for binary outcomes, or the unit change for continuous risk factors, per standard deviation change in genetically longer telomeres. P values for heterogeneity in the estimated associations between telomere length and disease amongst SNPs were estimated by likelihood ratio tests.[25] Associations between genetically longer telomeres and continuous risk factors were transformed into standard deviation units. For six diseases where only a single SNP was available for analysis, we estimated associations using the Wald ratio: $\beta_{GD}/\beta_{GP,}$ with standard errors approximated by the delta method.[26]

Inference of causality in the estimated etiological associations between telomere length and disease depends on satisfaction of Mendelian randomization assumptions.[27,28] The assumptions are: 1) the genetic proxies must be associated with telomere length; 2) the genetic proxies should not be associated with confounders; and 3) the genetic proxies must be associated with disease exclusively through their effect on telomere length. When these assumptions are satisfied, genetic proxies are said to be valid instrumental variables. We modeled the impact of violations of these assumptions through two sets of sensitivity analyses: a weighted median function[29] and MR-Egger regression[27] (see supplementary

methods for technical details). We restricted our sensitivity analyses to diseases showing the strongest evidence of association with genetically longer telomeres (defined as $P_{Bonferroni}<0.05$).

We used meta-regression to appraise potential sources of clinical heterogeneity in our findings for cancer outcomes. The association of genetically longer telomeres with the log odds of cancer was regressed on cancer incidence, survival time and median age at diagnosis, downloaded from the National Cancer Institute Surveillance, Epidemiology, and End Results (SEER) Program,[30] and tissue-specific rates of stem cell division from Tomasetti and Vogelstein.[31] As the downloaded cancer characteristics from SEER correspond to the United States population, 77% of which was of white ancestry in 2015[32], the meta-regression analyses excluded genetic studies conducted in East Asian populations.

All analyses were performed in R version 3.1.2[33] and Stata release 13.1 (StataCorp, College Station, TX). P values were two-sided and evidence of association was declared at $P<0.05$. Where indicated, Bonferroni corrections were used to make allowance for multiple testing, although this is likely to be overly conservative given the non-independence of many of the outcomes tested.

RESULTS

We selected 16 SNPs as genetic proxies for telomere length (Fig. S1 & Table 1). The selected SNPs correspond to 10 independent genomic loci that collectively account for 2-3% of the variance in leukocyte telomere length, which is equivalent to an F statistic of ~18. This indicates that the genetic risk score, constructed from these 10 independent genomic loci, is strongly associated with telomere length (see supplementary discussion for a more detailed consideration).[34] Summary data for the genetic proxies for telomere length were available for 83 non-communicable diseases and 44 risk factors, corresponding to 409,819 cases (median 2,092 per disease) and 1,404,633 controls (median 7,738 per disease) (Fig. S1, Table 2 and Table S1). The median number of SNPs available across disease datasets was 11 (min=1, max=13) and across risk factor datasets was 13 (min=10, max=13). Of the 83 diseases, 55 were classified as primary outcomes and 28 as secondary outcomes (Table 2, Fig. S1 and Table S1).

The results from primary analyses of non-communicable diseases are presented in Figure 1; results from secondary analyses of risk factors and diseases with low a priori power are presented in the supplementary materials (Fig. S2, S5 and S6). Genetically longer telomeres were associated with higher odds of disease for 9 of 22 primary cancer outcomes, including glioma, endometrial cancer, kidney cancer, testicular germ cell cancer, melanoma, bladder cancer, neuroblastoma, lung adenocarcinoma and serous low malignancy potential ovarian cancer ($P<0.05$) (Fig. 1). The associations were, however, highly variable across cancer types, varying from an odds ratio of 0.86 (95% confidence interval: 0.50 to 1.48) for head and

neck cancer to 5.27 (3.15, 8.81) for glioma. Substantial variability was also observed within tissue sites. For example, the odds ratio for lung adenocarcinoma was 3.19 (2.40 to 4.22) compared to 1.07 (0.82 to 1.39) for squamous cell lung cancer. For serous low malignancy potential ovarian cancer the odds ratio was 4.35 (2.39 to 7.94) compared to odds ratios of 1.21 (0.87 to 1.68) for endometrioid ovarian cancer, 1.12 (0.938 to 1.34) for serous invasive ovarian cancer, 1.04 (0.66 to 1.63) for clear cell ovarian cancer and 1.04 (0.732 to 1.47) for mucinous ovarian cancer. The strongest evidence of association was observed for glioma, lung adenocarcinoma, neuroblastoma and serous low malignancy potential ovarian cancer ($P_{Bonferroni}<0.05$). Results for glioma and bladder cancer showed evidence for replication in independent datasets (independent datasets were not available for other cancers) (Fig. S3).

Genetically longer telomeres were associated with reduced odds of disease for 6 of 32 primary non-neoplastic diseases, including coronary heart disease, abdominal aortic aneurysm, Alzheimer's disease, celiac disease, interstitial lung disease and type 1 diabetes ($P<0.05$) (Figure 1). The strongest evidence of association was observed for coronary heart disease, abdominal aortic aneurysm, celiac disease and interstitial lung disease ($P_{Bonferroni}<0.05$). The associations with coronary heart disease and interstitial lung disease showed evidence for replication in independent datasets (Fig. S3).

Our genetic findings were generally similar in direction and magnitude to estimates based on observational prospective studies of leukocyte telomere length and disease (Figure 3). Our genetic estimates for lung adenocarcinoma, melanoma, kidney cancer and glioma, were, however, stronger in comparison to observational estimates.

In sensitivity analyses, we appraised the potential impact of confounding by pleiotropic pathways on our results. Associations estimated by the weighted median approach were

broadly similar to the main results for glioma, lung adenocarcinoma, serous low malignancy potential ovarian cancer, neuroblastoma, abdominal aortic aneurysm, coronary heart disease, interstitial lung disease and celiac disease (Fig. S4). In the second set of sensitivity analyses, implemented by MR-Egger regression, we found little evidence for the presence of pleiotropy ($P \geq 0.27$) (Fig. S4). The MR-Egger analyses were, however, generally underpowered, as reflected by the wide confidence intervals in the estimated odds ratios.

In meta-regression analyses, we observed that genetically longer telomeres tended to be more strongly associated with rarer cancers ($P=0.02$) and cancers at tissue-sites with lower rates of stem cell division ($P=0.02$) (Figure 2). The associations showed little evidence of varying by percentage survival five years after diagnosis or median age-at-diagnosis ($P=0.4$).

DISCUSSION

Summary of main findings

In this report we show that genetically longer telomeres are associated with increased risk of several cancers and with reduced risk of some non-neoplastic diseases, including coronary heart disease, abdominal aortic aneurysm, celiac disease and interstitial lung disease. The findings for cancer were, however, subject to substantial variation between and within tissue sites, which our results suggest could be partly attributable to differences in cancer incidence and rates of stem cell division. Given the random distribution of genotypes in the general population with respect to lifestyle and other environmental factors, as well as the fixed nature of germline genotypes, these results should be less susceptible to confounding and reverse causation bias in comparison to observational studies. Nevertheless, although compatible with causality, our results could reflect violations of Mendelian randomization assumptions, such as confounding by pleiotropic pathways, population stratification or ancestry.[35] Although we cannot entirely rule out this possibility, the majority of our results persisted in sensitivity analyses that made allowance for violations of Mendelian randomization assumptions. Confounding by population stratification or ancestry is also unlikely, given that the disease GWAS results were generally adjusted for both (see supplementary discussion).

Comparison with previous studies

Our findings for cancer are generally contradictory to those based on retrospective studies, which tend to report increased risk for cancer in individuals with shorter telomeres.[9,10,36–39] The contradictory findings may reflect reverse causation bias in the retrospective studies,

whereby shorter telomeres arise as a result of disease, or of confounding effects, e.g. due to cases being slightly older than controls even in age-matched analyses. Our findings for cancer are generally more consistent with those based on prospective observational studies, which tend to report weak or null associations of longer leukocyte telomeres with overall and site-specific risk of cancer[8–11,38,40–59] with some exceptions.[60] Our results are also similar to previously reported Mendelian randomization studies of telomere length and risk of melanoma, lung cancer, chronic lymphocytic leukemia and glioma.[61–64] The shape of the association with cancer may not, however, be linear over the entire telomere length distribution. For example, individuals with dyskeratosis congenita, a disease caused by germline loss-of-function mutations in the telomerase component genes TERC and TERT, have chronically short telomeres and are at increased risk of some cancers, particularly acute myeloid leukemia and squamous cell carcinomas arising at sites of leukoplakia,[65,66] suggesting that the association could be "J" or "U" shaped.[41,54] Our results should therefore be interpreted as reflecting the average association at the population level and may not be generalizable to the extreme ends of the distribution.

Mechanisms of association

Our cancer findings are compatible with known biology.[67] By limiting the proliferative potential of cells, telomere shortening may serve as a tumour suppressor; and individuals with longer telomeres may be more likely to acquire somatic mutations owing to increased proliferative potential.[67] Rates of cell division are, however, highly variable amongst tissues[31] and thus the relative gain in cell proliferative potential, conferred by having longer telomeres, may also be highly variable across tissues. This could explain the almost 9-fold variation in odds ratios observed across cancer types in the present study, as well as the tendency of our results to be stronger at tissue sites with lower rates of stem cell division. For example, the

association was strongest for glioma (OR=5.27) and comparatively weak for colorectal cancer (OR=1.09) and the rates of stem cell division in the tissues giving rise to these cancers differ by several orders of magnitude. In neural stem cells, which give rise to gliomas, the number of divisions is ~270 million and for colorectal stem cells is ~1.2 trillion over the average lifetime of an individual.[31] The observation that genetically longer telomeres were more strongly associated with rarer cancers potentially reflects the same mechanism, since rarer cancers also tend to show lower rates of stem cell division.[31] For example, the incidence of glioma is 0.4 and for colorectal cancer is 42.4 per 100,000 per year in the United States.[30] On the other hand, individuals with chronically short telomeres, such as those with dyskeratosis congenita, could be more susceptible to genome instability and chromosomal end-to-end fusions, which could underlie their increased susceptibility to cancer.[65–67]

The inverse associations observed for some non-neoplastic diseases may reflect the impact of telomere shortening on tissue degeneration and an evolutionary trade-off for greater resistance to cancer at the cost of greater susceptibility to degenerative diseases, particularly cardiovascular diseases.[68,69]

Study limitations

Our study is subject to some limitations, in addition to the Mendelian randomization assumptions already considered above. First, our method assumes that the magnitude of the association between SNPs and telomere length is consistent across tissues. Second, our study assumed a linear shape of association between telomere length and disease risk, whereas the shape could be "J" or "U" shaped.[41,54,65] Third, our results assume that the samples used to define the genetic proxies for telomere length[16] and the various samples used to estimate the SNP-disease associations are representative of the same general population, practically

defined as being of similar ethnicity, age and sex distribution.[70] This assumption would, for example, not apply in the case of the SNP-disease associations derived from East Asian or pediatric populations. Generally speaking, violation of the aforementioned assumptions would potentially bias the magnitude of the estimated association between genetically longer telomeres and disease; but would be unlikely to increase the likelihood of false positives (i.e. incorrectly inferring an association when none exists).[71] Our results should therefore remain informative for the direction and broad magnitude of the average association at the population level, even in the presence of such violations. Fourth, we cannot rule out chance in explaining some of the weaker findings. Fifth, our results may not be fully representative of non-communicable diseases (since not all studies shared data and our analyses were underpowered for the secondary disease outcomes). The diseases represented in our primary analyses probably account for >60% of all causes of death in American adults.[72]

Clinical relevance of findings

Our findings suggest that any potential clinical applications of telomere length, e.g. as a tool for risk prediction or as an intervention target for disease prevention, will have to consider a trade-off in risk between cancer and non-neoplastic diseases. For example, a number of Wellness companies have been established that offer telomere length measurement services to the public (via a requesting physician) claiming that shorter telomeres are a general indicator of poor health status, older biological age and that information on telomere length can be used to motivate healthy lifestyle choices in patients. The conflicting direction of association between telomere length and risk of cancer and non-neoplastic diseases suggests, however, that such services to the general public may be premature.

Conclusion

Genetically longer telomeres are associated with increased risk for several cancers, but the relative increase in risk is highly heterogeneous across cancer types, and with reduced risk for some non-neoplastic diseases, including cardiovascular diseases. Further research is required to resolve whether telomere length is a useful predictor of risk that can help guide lifestyle modification, to clarify the shape of any dose-response relationship, and to characterise the nature of the association in population subgroups.

Acknowledgements

This work was supported by CRUK grant number C18281/A19169 (the Integrative Cancer Epidemiology Programme). Dr Haycock is supported by CRUK Population Research Postdoctoral Fellowship C52724/A20138. The MRC Integrative Epidemiology Unit is supported by grants MC_UU_12013/1 and MC_UU_12013/2. Dr Martin is supported by the National Institute for Health Research (NIHR), the Bristol Nutritional Biomedical Research Unit and the University of Bristol.

We gratefully acknowledge the assistance and contributions of Dr Julia Gumy and Ms Lisa Wright.

374

Table 1. Single nucleotide polymorphisms used as genetic proxies for telomere length

SNPs	Chr	Pos	Gene	EA	OA	EAF*	Beta*	SE*	P-value*	Phet*	No. studies*	Sample size*	Discovery p-value	% variance explained	Discovery study
rs11125529	2	54248729	ACYP2	A	C	0.16	0.065	0.012	0.000606	0.313	6	9177	8.00E-10	0.080	Codd[19]
rs6772228	3	58390292	PXK	T	A	0.87	0.041	0.014	0.049721	0.77	6	8630	3.91E-10	0.200	Pooley[15]
rs12696304	3	169763483	TERC	C	G	0.74	0.090	0.011	5.41E-08	0.651	6	9012	4.00E-14	0.319	Codd[20]
rs10936599	3	169774313	TERC	C	T	0.76	0.100	0.011	1.76E-09	0.087	6	9190	3.00E-31	0.319	Codd[19]
rs1317082	3	169779797	TERC	A	G	0.71	0.097	0.011	4.57E-09	0.029	6	9176	1.00E-08	0.319	Mangino[16]
rs10936601	3	169810661	TERC	C	T	0.74	0.087	0.011	8.64E-08	0.433	6	9150	4.00E-15	0.319	Pooley[15]
rs7675998	4	163086668	NAF1	G	A	0.80	0.048	0.012	0.008912	0.077	6	9161	4.35E-16	0.190	Codd[19]
rs2736100	5	1286401	TERT	C	A	0.52	0.085	0.013	2.14E-05	0.54	4	5756	4.38E-19	0.310	Codd[19]
rs9419958	10	103916188	OBFC1	T	C	0.13	0.129	0.013	5.26E-11	0.028	6	9190	9.00E-11	0.171	Mangino[16]
rs9420907	10	103916707	OBFC1	C	A	0.14	0.142	0.014	1.14E-11	0.181	6	9190	7.00E-11	0.171	Codd[19]
rs4387287	10	103918139	OBFC1	A	C	0.14	0.120	0.013	1.40E-09	0.044	6	8541	2.00E-11	0.171	Levy[23]
rs3027234	17	8232774	CTC1	C	T	0.83	0.103	0.012	2.75E-08	0.266	6	9108	2.00E-08	0.292	Mangino[16]
rs8105767	19	22032639	ZNF208	G	A	0.25	0.064	0.011	0.000169	0.412	6	9096	1.11E-09	0.090	Codd[19]
rs412658	19	22176638	ZNF676	T	C	0.35	0.086	0.010	1.83E-08	0.568	6	9156	1.00E-08	0.484	Mangino[16]
rs6028466	20	39500359	DHX35	A	G	0.17	0.058	0.013	0.003972	0.533	6	9190	2.57E-08†	0.041	Mangino[16] & Gu[18]
rs755017	20	63790269	ZBTB46	G	A	0.17	0.019	0.0129	0.339611	0.757	5	8026	6.71E-09	0.090	Codd[19]

*Summary data from Mangino et al[16]; Chr, chromosome; pos, base-pair position (GRCh38.p3); EA, effect allele, OA, other allele, Beta, standard deviation change in telomere length per copy of the effect allele; SE, standard error; EAF - effect allele frequency; Phet - p value for between-study heterogeneity in association between SNP and telomere length; †from a meta-analysis of Mangino[16] and Gu[18] performed in the present study.

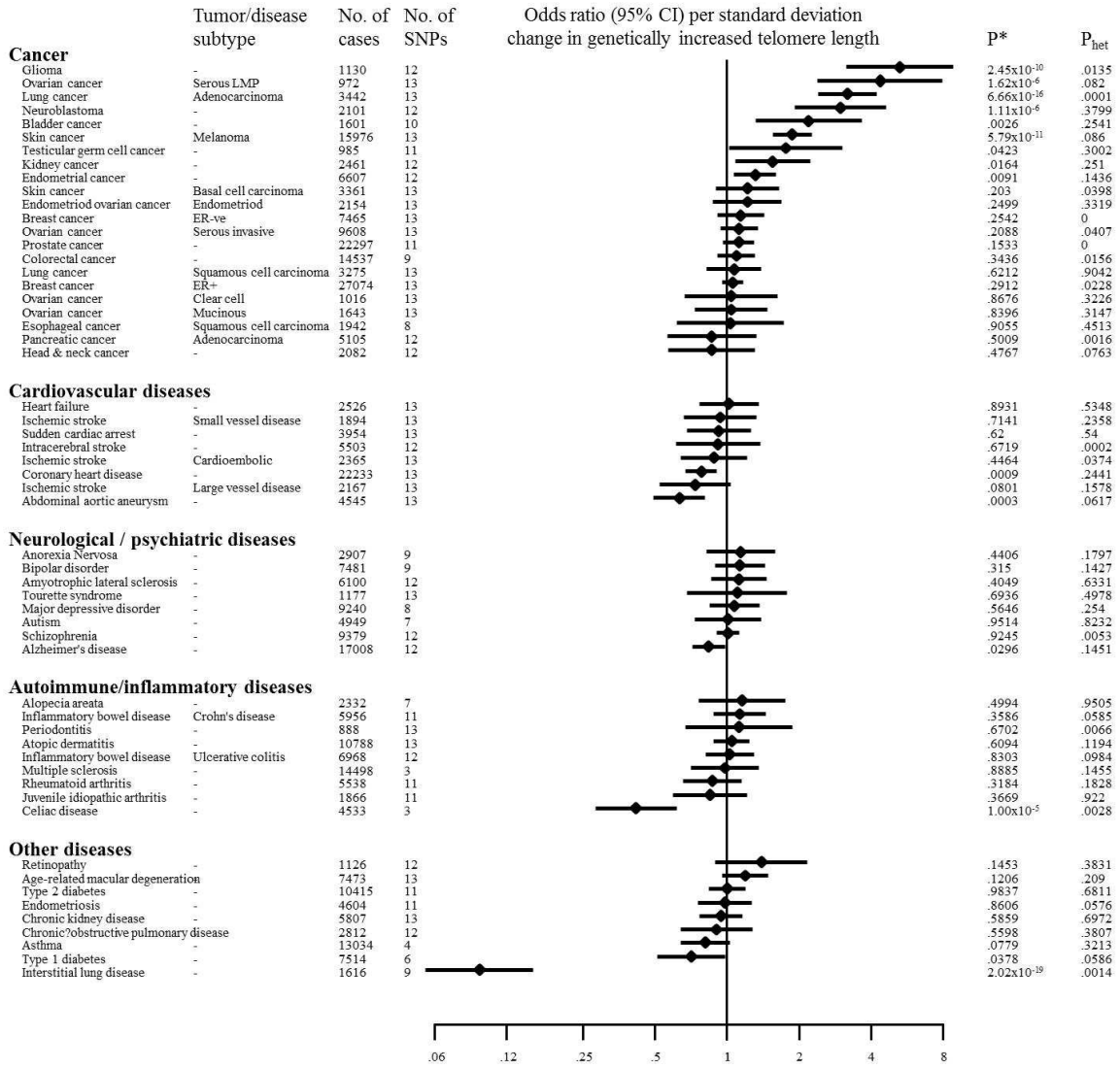
Table 2. Study characteristics for primary non-communicable diseases

	No. cases	No. controls	No. SNPs	Statistical power	Pop.	Study / First author
Cancer						
Bladder cancer	1601	1819	10	0.62	EUR	NBCS[75]
Breast cancer	48155	43612	13	1.00	EUR	BCAC[15,76]
Estrogen receptor –ve	7465	42175	13	1.00	EUR	BCAC[15,76]
Estrogen receptor +ve	27074	41749	13	1.00	EUR	BCAC[15,76]
Colorectal cancer	14537	16922	9	1.00	EUR	CORECT/GECC[61,77]
Endometrial cancer	6608	37925	12	1.00	EUR	ECAC[78,79]
Esophageal SCC	1942	2111	11	0.64	EA	Abnet[80]
Glioma	1130	6300	12	0.72	EUR	Wrensch[81] & Walsh[63]
Head & neck cancer	2082	3477	12	1.00	EUR	McKay et al[82]
Kidney cancer	2461	5081	12	0.99	EUR	KIDRISK[83]
Lung cancer	11348	15861	13	1.00	EUR	ILCCO[84]
Adenocarcinoma	3442	14894	13	1.00	EUR	ILCCO[84]
Squamous cell carcinoma	3275	15038	13	1.00	EUR	ILCCO[84]
Skin cancer						
Melanoma	15976	26451	13	1.00	EUR	MC[85]
Basal cell carcinoma	3361	11518	13	1.00	EUR	NHS/HPFS[86]
Neuroblastoma	2101	4202	12	0.87	EUR	Diskin[87]
Ovarian cancer	15397	30816	13	1.00	EUR	OCAC[15,88]
Clear cell	1016	30816	13	0.76	EUR	OCAC[15,88]
Endometriod	2154	30816	13	0.98	EUR	OCAC[15,88]
Mucinous	1643	30816	13	0.94	EUR	OCAC[15,88]
Serous invasive	9608	30816	13	1.00	EUR	OCAC[15,88]
Serous LMP	972	30816	13	0.73	EUR	OCAC[15,88]
Pancreatic cancer	5105	8739	12	1.00	EUR	PanScan (incl. EPIC)[89]
Prostate cancer	22297	22323	11	1.00	EUR	PRACTICAL[90,91]
Testicular germ cell cancer	986	4946	11	0.52	EUR	Turnbull[92] & Rapley[93]
Autoimmune/inflammatory diseases						
Alopecia areata	2332	5233	7	0.60	EUR	Betz[94]
Atopic dermatitis	10788	30047	13	1.00	EUR	EAGLE[95]
Celiac disease	4533	10750	3	0.82	EUR	Dubois[96]
Inflammatory bowel disease						
Crohn's disease	5956	14927	11	1.00	EUR	IIBDGC[97]
Ulcerative colitis	6968	20464	12	1.00	EUR	IIBDGC[97]
Juvenile idiopathic arthritis	1866	14786	11	0.87	EUR	Thompson[98]†
Multiple sclerosis	14498	24091	3	1.00	EUR	IMSGC[99]
Aggressive periodontitis	888	6789	13	0.63	EUR	Schaefer[100]
Rheumatoid arthritis	5538	20163	11	1.00	EUR	Stahl[101]
Cardiovascular diseases						
Abdominal aortic aneurysm	4972	99858	13	1.00	EUR	AC[102–107]
Coronary heart disease	22233	64762	13	1.00	EUR	CARDIoGRAM[108]
Heart failure	2526	20926	13	0.99	EUR	CHARGE-HF[109]
Hemorrhagic stroke	2963	5503	12	0.96	EUR	METASTROKE/ISGC[110]
Ischemic stroke	12389	62004	13	1.00	EUR	METASTROKE/ISGC[111,112]
large vessel disease	2167	62004	13	0.99	EUR	METASTROKE/ISGC[111,112]
small vessel disease	1894	62004	13	0.97	EUR	METASTROKE/ISGC[111]
cardioembolic	2365	62004	13	0.99	EUR	METASTROKE/ISGC[111]
Sudden cardiac arrest	3954	21200	13	1.00	EUR	Unpublished
Diabetes						
Type 1 diabetes	7514	9045	6	0.95	EUR	T1Dbase[113]
Type 2 diabetes	10415	53655	11	1.00	EUR	DIAGRAM[114]
Eye disease						
AMD	7473	51177	13	1.00	EUR	AMD Gene[115]

Retinopathy	1122	18289	12	0.75	EUR	Jensen[116]
Lung diseases						
Asthma	13034	20638	4	1.00	EUR	Ferreira/GABRIEL[117,118]
COPD	2812	2534	12	0.85	EUR	COPDGene[119]
Interstitial lung disease	1616	4683	9	0.60	EUR	Fingerlin[120]
Neurological / psychiatric diseases						
ALS	6100	7125	12	1.00	EUR	SLAGEN/ALSGEN[121]
Alzheimer's disease	17008	37154	12	1.00	EUR	IGAP[122]
Anorexia nervosa	2907	14860	9	0.93	EUR	GCAN[123]
Autism	4949	5314	7	0.82	EUR	PGC[124]
Bipolar disorder	7481	9250	9	1.00	EUR	PGC[125]
Major depressive disorder	9240	9519	8	0.99	EUR	PGC[126]
Schizophrenia	35476	46839	12	1.00	EUR	PGC[127]
Tourette syndrome	1177	4955	13	0.74	EUR	Scharf[128]
Other						
Chronic kidney disease	5807	56430	13	1.00	EUR	CKDGen[129]
Endometriosis	4604	9393	11	1.00	Mix	Nyholt[130]

Study acronyms: AC, the aneurysm consortium; **ALSGEN,** the International Consortium on Amyotrophic Lateral Sclerosis Genetics**; AMD Gene**, Age-related Macular Degeneration Gene Consortium; **BCAC**, Breast Cancer Association Consortium; **CARDIoGRAM**, Coronary ARtery DIsease Genome wide Replication and Meta-analysis; **CHARGE-HF**, Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium – Heart Failure Working Group; **COPDGene**, the genetic epidemiology of COPD; **CKDGen**, Chronic Kidney Disease; **CORECT**, ColoRectal Transdisciplinary Study; **DIAGRAM**, DIAbetes Genetics Replication And Meta-analysis; **EAGLE**, EArly Genetics & Lifecourse Epidemiology Eczema Consortium (excluding 23andMe); **ECAC**, Endometrial Cancer Association Consortium; **GCAN**, Genetic Consortium for Anorexia Nervosa; **GECCO**, Genetics and Epidemiology of Colorectal Cancer Consortium; **IGAP**, International Genomics of Alzheimer's Project; **HPFS**, Health Professionals Follow-Up Study; **ILCCO**, International Lung Cancer Consortium; **IMSGC**, International Multiple Sclerosis Genetic Consortium; **IIBDGC**, International Inflammatory Bowel Disease Genetics Consortium; **KIDRISK**, Kidney cancer consortium; **MC, the melanoma meta-analysis consortium; METASTROKE/ISGC**, METASTROKE project of the International Stroke Genetics Consortium; **NBCS**, Nijmegen Bladder Cancer Study; **NHS**, Nurses' Health Study; **OCAC**, Ovarian Cancer Association Consortium; **NCCC**, Dartmouth-Hitchcock Norris Cotton Cancer Center; **PANSCAN**, Pancreatic Cancer Cohort Consortium; **PGC**, Psychiatric Genomics Consortium; **PRACTICAL**, Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome; **SLAGEN**, Italian Consortium for the Genetics of Ayotrophic Lateral Sclerosis. **Abbreviations**: **ALS**, amyotrophic lateral sclerosis; **AMD**, age-related macular degeneration; **COPD**, chronic obstructive pulmonary disease; **EUR**, European; **EA**, East Asian; **LMP**, low malignant potential; **No.**, number; **Pop.**, population; **SCC**, squamous cell carcinoma; **SNP**, single nucleotide polymorphism; **-ve**, negative; **+ve**, positive; †plus previously unpublished data.

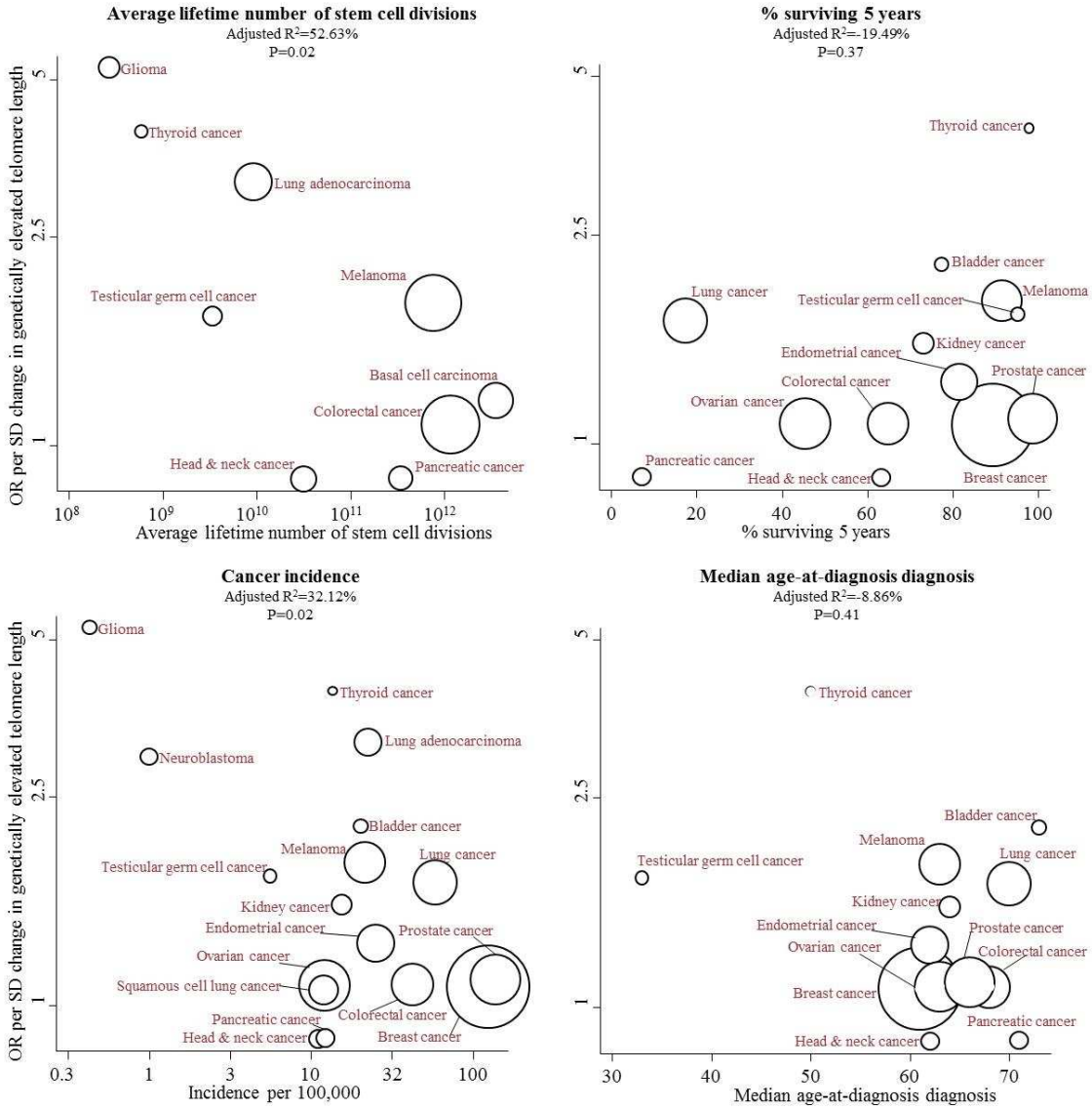
Figure 1. The association between genetically longer telomeres and odds of primary non-communicable diseases



Legend to Figure 1

*P value for association between genetically longer telomeres and disease from maximum likelihood; †the effect estimate for heart failure is a hazard ratio (all others are odds ratios); P_{het}, p value for heterogeneity amongst SNPs in the genetic risk score; SNP, single nucleotide polymorphism; CI, confidence interval; LMP, low malignancy potential; ER, estrogen receptor; -VE, negative; +VE, positive.

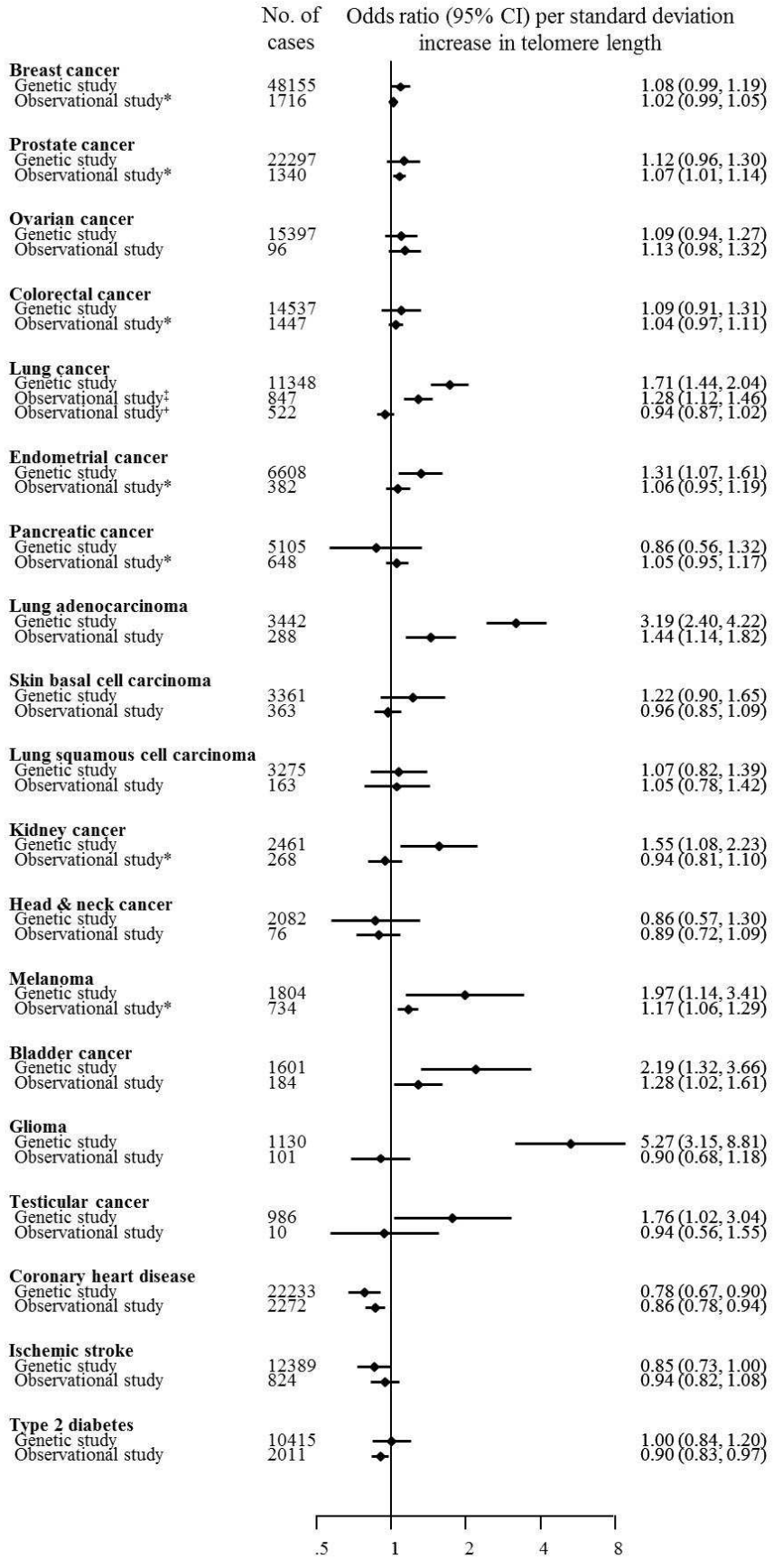
Figure 2. The association between genetically longer telomeres and odds of cancer as a function of selected characteristics



Legend to Figure 2

The plotted data show how the strength of the relationship between genetically longer telomeres and cancer varies by the selected characteristic. The R^2 statistic indicates how much of the variation between cancers can be explained by the selected characteristic. P values are from meta-regression models. Circle sizes are proportional to the inverse of the variance of the log odds ratio. The hashed line indicates the null of no association between telomere length and cancer (i.e. an odds ratio of 1). Data for percentage survival 5 years after diagnosis, cancer incidence and median age-at-diagnosis was downloaded from the Surveillance, Epidemiology, and End Results Program.[30] Data for average lifetime number of stem cell divisions was downloaded from Tomasetti and Vogelstein.[31] SD, standard deviation; OR, Odds ratio. Not all cancers had information available for the selected characteristics (hence the number of cancers varies across the subplots). Information was available for 12 cancers for tissue-specific rates of stem cell division, 18 cancers for percentage surviving 5 years post-diagnosis, 23 cancers for cancer incidence and 18 cancers for median age-at-diagnosis.

Figure 3. Comparison of genetic and prospective observational studies[†] of the association between telomere length and disease



Legend to Figure 3

*from fixed-effects meta-analysis of independent observational studies described in Table S3; †search strategy and characteristics for observational studies are described in Tables S3 and S4; ‡CCHS and CGPS; +PLCO, ATBC & SWHS (acronyms explained in Table S3); **CI**, confidence interval

REFERENCES

1. Blackburn EH, Epel ES, Lin J. Human telomere biology: A contributory and interactive factor in aging, disease risks, and protection. Science (80-) 2015;350(6265):1193–8.

2. Weischer M, Bojesen SE, Nordestgaard BG. Telomere shortening unrelated to smoking, body weight, physical activity, and alcohol intake: 4,576 general population individuals with repeat measurements 10 years apart. PLoS Genet 2014;10(3):e1004191.

3. Houben JMJ, Moonen HJJ, van Schooten FJ, Hageman GJ. Telomere length assessment: biomarker of chronic oxidative stress? Free Radic Biol Med 2008;44(3):235–46.

4. Marchesi V. Risk factors: Short telomeres: association with cancer survival and risk. Nat Rev Clin Oncol 2013;10(5):247.

5. Haycock PC, Heydon EE, Kaptoge S, Butterworth a. S, Thompson A, Willeit P. Leucocyte telomere length and risk of cardiovascular disease: systematic review and meta-analysis. BMJ 2014;349(jul08 3):g4227–g4227.

6. 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.

7. Zhao J, Miao K, Wang H, Ding H, Wang DW. Association between telomere length and type 2 diabetes mellitus: a meta-analysis. PLoS One 2013;8(11):e79993.

8. Weischer M, Nordestgaard BG, Cawthon RM, Freiberg JJ, Tybjærg-Hansen A, Bojesen SE. Short telomere length, cancer survival, and cancer risk in 47102 individuals. J Natl Cancer Inst 2013;105(7):459–68.

9. Ma H, Zhou Z, Wei S, et al. Shortened telomere length is associated with increased risk of cancer: a meta-analysis. PLoS One 2011;6(6):e20466.

10. Wentzensen IM, Mirabello L, Pfeiffer RM, Savage SA. The association of telomere length and cancer: a meta-analysis. Cancer Epidemiol Biomarkers Prev 2011;20(6):1238–50.

11. Pooley KA, Sandhu MS, Tyrer J, et al. Telomere length in prospective and retrospective cancer case-control studies. Cancer Res 2010;70(8):3170–6.

12. Hou L, Joyce BT, Gao T, et al. Blood Telomere Length Attrition and Cancer Development in the Normative Aging Study Cohort. EBioMedicine 2015;2(6):591–6.

13. Welter D, MacArthur J, Morales J, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res 2014;42(Database issue):D1001-6.

14. Hindorff LA LA, MacArthur J, Morales J, et al. A catalog of published genome-wide association studies [Internet]. [cited 2015 Jan 15];Available from: www.genome.gov/gwastudies

15. Pooley KA, Bojesen SE, Weischer M, et al. A genome-wide association scan (GWAS) for mean telomere length within the COGS project: identified loci show little association with hormone-related cancer risk. Hum Mol Genet 2013;22(24):5056–64.

16. Mangino M, Hwang S-J, Spector TD, et al. Genome-wide meta-analysis points to CTC1 and ZNF676 as genes regulating telomere homeostasis in humans. Hum Mol Genet 2012;21(24):5385–94.

17. Prescott J, Kraft P, Chasman DI, et al. Genome-wide association study of relative telomere length. PLoS One 2011;6(5):e19635.

18. Gu J, Chen M, Shete S, et al. A genome-wide association study identifies a locus on chromosome 14q21 as a predictor of leukocyte telomere length and as a marker of susceptibility for bladder cancer. Cancer Prev Res (Phila) 2011;4(4):514–21.

19. Codd V, Nelson CP, Albrecht E, et al. Identification of seven loci affecting mean telomere length and their association with disease. Nat Genet 2013;45(4):422–7.

20. Codd V, Mangino M, van der Harst P, et al. Common variants near TERC are associated with mean telomere length. Nat Genet 2010;42(3):197–9.

21. Liu Y, Cao L, Li Z, et al. A genome-wide association study identifies a locus on TERT for mean telomere length in Han Chinese. PLoS One 2014;9(1):e85043.

22. Saxena R, Bjonnes A, Prescott J, et al. Genome-wide association study identifies variants in casein kinase II (CSNK2A2) to be associated with leukocyte telomere length in a Punjabi Sikh diabetic cohort. Circ Cardiovasc Genet 2014;7(3):287–95.

23. 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 U S A 2010;107(20):9293–8.

24. Chene G, Thompson SG. Methods for Summarizing the Risk Associations of Quantitative Variables in Epidemiologic Studies in a Consistent Form. Am J Epidemiol 1996;144(6):610–21.

25. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG, EPIC-InterAct Consortium. Using published data in Mendelian randomization : a blueprint for efficient identification of causal risk factors. Eur J Epidemiol 2015;30(7):543–52.

26. Thomas DC, Lawlor DA, Thompson JR. Re: Estimation of bias in nongenetic observational studies using "Mendelian triangulation" by Bautista et al. Ann Epidemiol 2007;17(7):511–3.

27. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol 2015;44(2):512–25.

28. VanderWeele TJ, Tchetgen Tchetgen EJ, Cornelis M, Kraft P. Methodological challenges in mendelian randomization. Epidemiology 2014;25(3):427–35.

29. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. Genet Epidemiol

30. National Cancer Institute. Surveillance, Epidemiology, and End Results Program

[Internet]. [cited 2015 Aug 1];Available from: www.seer.cancer.gov

31. Tomasetti C, Vogelstein B. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. Science 2015;347(6217):78–81.

32. U.S. Census Bureau [Internet]. [cited 2016 Jul 11];Available from: U.S. Census Bureau

33. R Core Team. A language and environment for statistical computing. 2013;

34. Burgess S, Thompson SG. Bias in causal estimates from Mendelian randomization studies with weak instruments. Stat Med 2011;30(11):1312–23.

35. Davey Smith G, Ebrahim S. What can mendelian randomisation tell us about modifiable behavioural and environmental exposures? BMJ Br Med J 2005;330(7499):1076–9.

36. Anic GM, Sondak VK, Messina JL, et al. Telomere length and risk of melanoma, squamous cell carcinoma, and basal cell carcinoma. Cancer Epidemiol 2013;37(4):434–9.

37. Pellatt AJ, Wolff RK, Torres-Mejia G, et al. Telomere length, telomere-related genes, and breast cancer risk: the breast cancer health disparities study. Genes Chromosomes Cancer 2013;52(7):595–609.

38. Caini S, Raimondi S, Johansson H, et al. Telomere length and the risk of cutaneous melanoma and non-melanoma skin cancer: a review of the literature and meta-analysis. J Dermatol Sci 2015;80(3):168–74.

39. Sanchez-Espiridion B, Chen M, Chang JY, et al. Telomere length in peripheral blood leukocytes and lung cancer risk: a large case-control study in Caucasians. Cancer Res 2014;74(9):2476–86.

40. Campa D, Mergarten B, De Vivo I, et al. Leukocyte telomere length in relation to pancreatic cancer risk: a prospective study. Cancer Epidemiol Biomarkers Prev 2014;23(11):2447–54.

41. Cui Y, Cai Q, Qu S, et al. Association of leukocyte telomere length with colorectal cancer risk: nested case-control findings from the Shanghai Women's Health Study. Cancer Epidemiol Biomarkers Prev 2012;21(10):1807–13.

42. De Vivo I, Prescott J, Wong JYY, Kraft P, Hankinson SE, Hunter DJ. A prospective study of relative telomere length and postmenopausal breast cancer risk. Cancer Epidemiol Biomarkers Prev 2009;18(4):1152–6.

43. Han J, Qureshi AA, Prescott J, et al. A prospective study of telomere length and the risk of skin cancer. J Invest Dermatol 2009;129(2):415–21.

44. Hofmann JN, Lan Q, Cawthon R, et al. A prospective study of leukocyte telomere length and risk of renal cell carcinoma. Cancer Epidemiol Biomarkers Prev 2013;22(5):997–1000.

45. Julin B, Shui I, Heaphy CM, et al. Circulating leukocyte telomere length and risk of overall and aggressive prostate cancer. Br J Cancer 2015;112(4):769–76.

46. Kim S, Sandler DP, Carswell G, et al. Telomere length in peripheral blood and breast cancer risk in a prospective case-cohort analysis: results from the Sister Study. Cancer

Causes Control 2011;22(7):1061–6.

47. Lan Q, Cawthon R, Shen M, et al. A prospective study of telomere length measured by monochrome multiplex quantitative PCR and risk of non-Hodgkin lymphoma. Clin Cancer Res 2009;15(23):7429–33.

48. Lee I-M, Lin J, Castonguay AJ, Barton NS, Buring JE, Zee RYL. Mean leukocyte telomere length and risk of incident colorectal carcinoma in women: a prospective, nested case-control study. Clin Chem Lab Med 2010;48(2):259–62.

49. Liang G, Qureshi AA, Guo Q, De Vivo I, Han J. No association between telomere length in peripheral blood leukocytes and the risk of nonmelanoma skin cancer. Cancer Epidemiol Biomarkers Prev 2011;20(5):1043–5.

50. Lynch SM, Major JM, Cawthon R, et al. A prospective analysis of telomere length and pancreatic cancer in the alpha-tocopherol beta-carotene cancer (ATBC) prevention study. Int J Cancer 2013;133(11):2672–80.

51. McGrath M, Wong JYY, Michaud D, Hunter DJ, De Vivo I. Telomere length, cigarette smoking, and bladder cancer risk in men and women. Cancer Epidemiol Biomarkers Prev 2007;16(4):815–9.

52. Nan H, Du M, De Vivo I, et al. Shorter telomeres associate with a reduced risk of melanoma development. Cancer Res 2011;71(21):6758–63.

53. Prescott J, McGrath M, Lee I-M, Buring JE, De Vivo I. Telomere length and genetic analyses in population-based studies of endometrial cancer risk. Cancer 2010;116(18):4275–82.

54. Qu S, Wen W, Shu X-O, et al. Association of leukocyte telomere length with breast cancer risk: nested case-control findings from the Shanghai Women's Health Study. Am J Epidemiol 2013;177(7):617–24.

55. Risques RA, Vaughan TL, Li X, et al. Leukocyte telomere length predicts cancer risk in Barrett's esophagus. Cancer Epidemiol Biomarkers Prev 2007;16(12):2649–55.

56. 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):4090–8.

57. 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):133–7.

58. Walcott F, Rajaraman P, Gadalla SM, et al. Telomere length and risk of glioma. Cancer Epidemiol 2013;37(6):935–8.

59. Zee RYL, Castonguay AJ, Barton NS, Buring JE. Mean telomere length and risk of incident colorectal carcinoma: a prospective, nested case-control approach. Cancer Epidemiol Biomarkers Prev 2009;18(8):2280–2.

60. Willeit P, Willeit J, Mayr A, et al. Telomere length and risk of incident cancer and cancer mortality. Jama 2010;304(1538–3598 (Electronic)):69–75.

61. Zhang C, Doherty J a., Burgess S, et al. Genetic determinants of telomere length and risk of common cancers: a Mendelian randomization study. Hum Mol Genet

2015;24(18):5356–66.

62. Iles MM, Bishop DT, Taylor JC, et al. The effect on melanoma risk of genes previously associated with telomere length. J Natl Cancer Inst 2014;106(10).

63. Walsh KM, Codd V, Rice T, et al. Longer genotypically-estimated leukocyte telomere length is associated with increased adult glioma risk. Oncotarget 2015;6(40):42468–77.

64. Ojha J, Codd V, Nelson CP, et al. Genetic Variation Associated with Longer Telomere Length Increases Risk of Chronic Lymphocytic Leukemia. Cancer Epidemiol Biomarkers Prev 2016;25(7):1043–9.

65. Armanios M, Blackburn EH. The telomere syndromes. Nat Rev Genet 2012;13(10):693–704.

66. Armanios M. Syndromes of telomere shortening. Annu Rev Genomics Hum Genet 2009;10(46):45–61.

67. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144(5):646–74.

68. Blasco MA. Telomere length, stem cells and aging. Nat Chem Biol 2007;3(10):640–9.

69. Stone RC, Horvath K, Kark JD, Susser E, Tishkoff SA, Aviv A. Telomere Length and the Cancer-Atherosclerosis Trade-Off. PLoS Genet 2016;12(7):e1006144.

70. Pierce BL, Burgess S. Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators. Am J Epidemiol 2013;178(7):1177–84.

71. Burgess S, Butterworth AS, Thompson JR. Beyond Mendelian randomization: how to interpret evidence of shared genetic predictors. J Clin Epidemiol 2015;1–9.

72. Centers for Disease Control and Prevention [Internet]. [cited 2016 Jul 14];Available from: http://www.cdc.gov/nchs/fastats/deaths.htm

73. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. Cell 2013;153(6):1194–217.

74. Media THE, McCartney M. Would you like your telomeres tested? BMJ 2012;344(February):1–2.

75. Rafnar T, Sulem P, Thorleifsson G, et al. Genome-wide association study yields variants at 20p12.2 that associate with urinary bladder cancer. Hum Mol Genet 2014;23(20):5545–57.

76. Michailidou K, Hall P, Gonzalez-Neira A, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet 2013;45(4):353–61, 361-2.

77. Schumacher FR, Schmit SL, Jiao S, et al. Genome-wide association study of colorectal cancer identifies six new susceptibility loci. Nat Commun 2015;6:7138.

78. Spurdle AB, Thompson DJ, Ahmed S, et al. Genome-wide association study identifies a common variant associated with risk of endometrial cancer. Nat Genet 2011;43(5):451–4.

79. Painter JN, O'Mara TA, Batra J, et al. Fine-mapping of the HNF1B multicancer locus

identifies candidate variants that mediate endometrial cancer risk. Hum Mol Genet 2015;24(5):1478–92.

80. Abnet CC, Freedman ND, Hu N, et al. A shared susceptibility locus in PLCE1 at 10q23 for gastric adenocarcinoma and esophageal squamous cell carcinoma. Nat Genet 2010;42(9):764–7.

81. Wrensch M, Jenkins RB, Chang JS, et al. Variants in the CDKN2B and RTEL1 regions are associated with high-grade glioma susceptibility. Nat Genet 2009;41(8):905–8.

82. McKay JD, Truong T, Gaborieau V, et al. A genome-wide association study of upper aerodigestive tract cancers conducted within the INHANCE consortium. PLoS Genet 2011;7(3):e1001333.

83. Purdue MP, Johansson M, Zelenika D, et al. Genome-wide association study of renal cell carcinoma identifies two susceptibility loci on 2p21 and 11q13.3. Nat Genet 2010;43(1):60–5.

84. Wang Y, McKay JD, Rafnar T, et al. Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. Nat Genet 2014;46(7).

85. Law MH, Bishop DT, Lee JE, et al. Genome-wide meta-analysis identifies five new susceptibility loci for cutaneous malignant melanoma. Nat Genet 2015;47(9):987–95.

86. Zhang M, Song F, Liang L, et al. Genome-wide association studies identify several new loci associated with pigmentation traits and skin cancer risk in European Americans. Hum Mol Genet 2013;22(14):2948–59.

87. Diskin SJ, Capasso M, Schnepp RW, et al. Common variation at 6q16 within HACE1 and LIN28B influences susceptibility to neuroblastoma. Nat Genet 2012;44(10):1126–30.

88. Pharoah PDP, Tsai Y-Y, Ramus SJ, et al. GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. Nat Genet 2013;45(4):362–70, 370-2.

89. Wolpin BM, Rizzato C, Kraft P, et al. Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer. Nat Genet 2014;46(9):994–1000.

90. Eeles RA, Olama AA Al, Benlloch S, et al. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. Nat Genet 2013;45(4):385–91, 391-2.

91. Al Olama AA, Kote-Jarai Z, Berndt SI, et al. A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. Nat Genet 2014;46(10):1103–9.

92. Turnbull C, Rapley E a, Seal S, et al. Variants near DMRT1, TERT and ATF7IP are associated with testicular germ cell cancer. Nat Genet 2010;42(7):604–7.

93. Rapley EA, Turnbull C, Al Olama AA, et al. A genome-wide association study of testicular germ cell tumor. Nat Genet 2009;41(7):807–10.

94. Betz RC, Petukhova L, Ripke S, et al. Genome-wide meta-analysis in alopecia areata resolves HLA associations and reveals two new susceptibility loci. Nat Commun

2015;6:5966.

95. EArly Genetics and Lifecourse Epidemiology (EAGLE) Eczema Consortium, Australian Asthma Genetics Consortium(AAGC), Australian Asthma Genetics Consortium AAGC. Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. Nat Genet 2015;47(12):1449–56.

96. Dubois PC a, Trynka G, Franke L, et al. Multiple common variants for celiac disease influencing immune gene expression. Nat Genet 2010;42(4):295–302.

97. Liu JZ, van Sommeren S, Huang H, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat Genet 2015;47(9):979–86.

98. Thompson SD, Marion MC, Sudman M, et al. Genome-wide association analysis of juvenile idiopathic arthritis identifies a new susceptibility locus at chromosomal region 3q13. Arthritis Rheum 2012;64(8):2781–91.

99. Beecham AH, Patsopoulos NA, Xifara DK, et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. Nat Genet 2013;45(11):1353–60.

100. Schaefer AS, Richter GM, Nothnagel M, et al. A genome-wide association study identifies GLT6D1 as a susceptibility locus for periodontitis. Hum Mol Genet 2010;19(3):553–62.

101. Stahl EA, Raychaudhuri S, Remmers EF, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nat Genet 2010;42(6):508–14.

102. Bown MJ, Jones GT, Harrison SC, et al. Abdominal aortic aneurysm is associated with a variant in low-density lipoprotein receptor-related protein 1. Am J Hum Genet 2011;89(5):619–27.

103. Gretarsdottir S, Baas AF, Thorleifsson G, et al. Genome-wide association study identifies a sequence variant within the DAB2IP gene conferring susceptibility to abdominal aortic aneurysm. Nat Genet 2010;42(8):692–7.

104. Jones GT, Bown MJ, Gretarsdottir S, et al. A sequence variant associated with sortilin-1 (SORT1) on 1p13.3 is independently associated with abdominal aortic aneurysm. Hum Mol Genet 2013;22(14):2941–7.

105. Harrison SC, Smith AJP, Jones GT, et al. Interleukin-6 receptor pathways in abdominal aortic aneurysm. Eur Heart J 2013;34(48):3707–16.

106. Elmore JR, Obmann MA, Kuivaniemi H, et al. Identification of a genetic variant associated with abdominal aortic aneurysms on chromosome 3p12.3 by genome wide association. J Vasc Surg 2009;49(6):1525–31.

107. Borthwick K, Smelser D, Bock J, et al. Ephenotyping for Abdominal Aortic Aneurysm in the Electronic Medical Records and Genomics (emerge) Network: Algorithm Development and Konstanz Information Miner Workflow. Int J Biomed Data Min 2015;4(1).

108. Schunkert H, König IR, Kathiresan S, et al. Large-scale association analysis identifies

13 new susceptibility loci for coronary artery disease. Nat Genet 2011;43(4):333–8.

109. Smith NL, Felix JF, Morrison AC, et al. Association of genome-wide variation with the risk of incident heart failure in adults of European and African ancestry: a prospective meta-analysis from the cohorts for heart and aging research in genomic epidemiology (CHARGE) consortium. Circ Cardiovasc Genet 2010;3(3):256–66.

110. Woo D, Falcone GJ, Devan WJ, et al. Meta-analysis of genome-wide association studies identifies 1q22 as a susceptibility locus for intracerebral hemorrhage. Am J Hum Genet 2014;94(4):511–21.

111. Malik R, Freilinger T, Winsvold BS, et al. Shared genetic basis for migraine and ischemic stroke: A genome-wide analysis of common variants. Neurology 2015;84(21):2132–45.

112. Traylor M, Farrall M, Holliday EG, et al. Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE collaboration): a meta-analysis of genome-wide association studies. Lancet Neurol 2012;11(11):951–62.

113. Barrett JC, Clayton DG, Concannon P, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet 2009;41(6):703–7.

114. Morris ADPDPA, Voight BFB, Teslovich TMT, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet 2012;44(9):981–90.

115. Fritsche LG, Chen W, Schu M, et al. Seven new loci associated with age-related macular degeneration. Nat Genet 2013;45(4):433–9, 439-2.

116. Jensen RA, Sim X, Li X, et al. Genome-wide association study of retinopathy in individuals without diabetes. PLoS One 2013;8(2):e54232.

117. Moffatt MF, Kabesch M, Liang L, et al. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. Nature 2007;448(7152):470–3.

118. Ferreira MAR, Matheson MC, Duffy DL, et al. Identification of IL6R and chromosome 11q13.5 as risk loci for asthma. Lancet (London, England) 2011;378(9795):1006–14.

119. Cho MH, McDonald M-LN, Zhou X, et al. Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. Lancet Respir Med 2014;2(3):214–25.

120. Fingerlin TE, Murphy E, Zhang W, et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. Nat Genet 2013;45(6):613–20.

121. Fogh I, Ratti A, Gellera C, et al. A genome-wide association meta-analysis identifies a novel locus at 17q11.2 associated with sporadic amyotrophic lateral sclerosis. Hum Mol Genet 2014;23(8):2220–31.

122. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet 2013;45(12):1452–8.

123. Boraska V, Franklin CS, Floyd JAB, et al. A genome-wide association study of

anorexia nervosa. Mol Psychiatry 2014;19(10):1085–94.

124. Smoller JW, Hospital MG. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet 2013;381(9875):1371–9.

125. Sklar P, Ripke S, Scott LJ, et al. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nat Genet 2011;43(10):977–83.

126. Ripke S, Wray NR, Lewis CM, et al. A mega-analysis of genome-wide association studies for major depressive disorder. Mol Psychiatry 2013;18(4):497–511.

127. Ripke S, Neale BM, Corvin A, et al. Biological insights from 108 schizophrenia-associated genetic loci. Nature 2014;511(7510):421–7.

128. Scharf JM, Yu D, Mathews CA, et al. Genome-wide association study of Tourette's syndrome. Mol Psychiatry 2013;18(6):721–8.

129. Köttgen A, Pattaro C, Böger C a, et al. New loci associated with kidney function and chronic kidney disease. Nat Genet 2010;42(5):376–84.

130. Nyholt DR, Low S-K, Anderson CA, et al. Genome-wide association meta-analysis identifies new endometriosis risk loci. Nat Genet 2012;44(12):1355–9.