

Cancer Diagnosis, Polygenic Risk, and Longevity-Associated Variants

Laura H. Goetz¹, Janith Don¹, Andrew J. Schork^{1,2}, David Duggan¹, Nathan D. Price³, Daniel S. Evans⁴, Steve Cummings⁴, Thomas Perls⁵, Paola Sebastiani^{5,6}, Nicholas J. Schork^{1,7*}

Affiliations:

1. The Translational Genomics Research Institute (TGen), Phoenix, AZ; 2. Institute of Biological Psychiatry, Mental Health Center Sct. Hans, Roskilde, Denmark; 3. The Institute for Systems Biology, Seattle, WA; 4. California Pacific Medical Center Research Institute, San Francisco, CA; 5. Boston University, Boston, MA; 6. Tufts University, Boston, MA; 7. The City of Hope National Medical Center

Keywords: Polygenic Risk Scores, Association Analysis, Cancer, Variant Annotations, Genetic Epidemiology

***Address correspondence to:**

Nicholas J. Schork, Ph.D.

Department of Quantitative Medicine

The Translational Genomics Research Institute (TGen)

445 North Fifth Street

Phoenix, AZ 85004

nschork@tgen.org

ABSTRACT

Background: Polygenic risk scores (PRS) have been developed to predict individual cancer risk and their potential clinical utility is receiving a great deal of attention. However, the degree to which the predictive utility of individual cancer-specific PRS may be augmented or refined by the incorporation of other cancer PRS, non-cancer disease PRS, or the protective effects of health and longevity-associated variants, is largely unexplored.

Methods: We constructed PRS for different cancers from public domain data as well as genetic scores for longevity ('Polygenic Longevity Scores' or 'PLS') for individuals in the UK Biobank. We then explored the relationships of these multiple PRS and PLS among those with and without various cancers.

Results: We found statistically significant associations between some PLS and individual cancers, even after accounting for cancer-specific PRS. None of the PLS in their current form had an effect pronounced enough to motivate clinical cancer risk stratification based on its combined use with cancer PRS. A few variants at loci used in the PLS had known associations with Alzheimer's disease and other diseases.

Conclusion: Underlying heterogeneity behind cancer susceptibility in the population at large is not captured by PRS derived from analytical models that only consider marginal associations of individual variants with cancer diagnoses. Our results have implications for the derivation and calculation of PRS and their use in clinical and biomedical research settings.

Impact: Extensions of analyses like ours could result in a more refined understanding of cancer biology and how to construct PRS for cancer.

INTRODUCTION

While cancer treatments and outcomes have improved significantly over the past 40 years, overall cancer incidence continues to rise.(1) In 2019, an estimated 16.9 million cancer survivors lived in the United States, a seven-fold increase since 1971.(2) Cancers are primarily diseases of aging, with over 80% of cancers occurring in individuals over age 50.(1) Since this proportion of the US population is projected to increase steadily over the next several decades,(3) there is an urgent need to address this impending public health and medical crisis through the development of strategies to decrease rates of cancers in aging adults. Developing appropriate strategies will require identifying individuals at risk for cancer and providing them with specific risk mitigation interventions that go beyond current prevention and screening recommendations. In this light, cancer risk assessment can be greatly enhanced by using genetic profiling. Both monogenic and polygenic influences play an important role in susceptibility to and pathogenesis of cancers, however, cancers in the aged are less likely to be associated with monogenic germline variants(4-6) The degree to which polygenic variants contribute specifically to aging-related cancer risk is the subject of much current research.(7,8)

GWAS have identified thousands of common genetic variants associated with many different types of cancer.(9,10) Although each individual variant has a small effect on risk, especially relative to those contributing to monogenic forms of cancer, their cumulative effects can be pronounced. These *polygenic effects* can make individual contributions to risk equal to that of the individual rare variants contributing to monogenic forms of hereditary cancer (11) and also modify or compound the risk of these monogenic driver variants(12-14). These common

variant, polygenic effects are often summarized at the individual level as a polygenic risk score (PRS) that aggregates effects across multiple variants into a single estimate of the total cancer susceptibility burden in a given genome. Cancer PRS are derived in different ways, but often use simple weighted multiple regression and related prediction modeling techniques.(15-17).

Supplementary Table 1 provides a list of recent cancer PRS studies.

Interest in PRS grew from developing models that could discriminate specific cancer cases from controls, but has now expanded into multiple other areas. For example, PRS have been used to explore genetic, as opposed to phenotypic, correlations between cancers to better characterize shared genetic determinants of different cancers;(8,18) correlations between the PRS themselves for different cancer-related phenotypes;(19) the pleiotropic effects of variants across multiple different cancers;(20) and the causal effects of various measurable and potentially modifiable genetically-mediated factors on cancer via, e.g., Mendelian Randomization tests.(21-23) This work suggests that being susceptible to one type of cancer or phenotype can influence susceptibility - positively or negatively - to another type.(21,24) The degree to which this is the case is an open question.

Fundamental processes mediating cancer susceptibility that are shared across cancers are also likely to mediate susceptibility to other diseases and conditions beyond cancer. Such processes are quite likely to be associated with aging and longevity, given that the manifestation of many common chronic diseases is age-related. By bringing together genetic insights into multiple cancers, diseases, aging and longevity, one might be able to develop not only better prediction models through, e.g., risk stratification, but also more insight into the

genetic etiology of cancer. Identifying variants associated with health span and longevity is challenging. Different definitions of longevity are used, sample sizes in studies focusing on extreme longevity are small, and the actual heritability of human lifespan is debated, as is the need to account for environmental factors in relevant studies.(25-27) Large-scale analyses of long-lived cohorts have enabled pooled analysis of multiple GWAS, as has been the focus of the National Institute on Aging's Longevity Consortium (LC), the Integrated Longevity Omics (ILO), and Long Life Family Study (LLFS) projects.(25) The results of these studies have yet to be integrated with studies of various diseases. GWAS focused on genetic 'surrogates' for individual lifespan, such as parental lifespan, have also been pursued.(28,29) Although not all studies are consistent, many of these analyses suggest that long lived individuals possess different disease PRS profiles from short lived individuals.(30-33) Some individuals possessing a genetic background consistent with a long-life – i.e., having an elevated Polygenic Longevity Score (PLS) – may have inherited genetic variants that protect them from cancer and other diseases. However, these variants may not have individual protective effects on cancer pronounced enough to be detected as significantly associated with a cancer diagnosis in cancer-specific GWAS.(34)

We explored the relationships among PRS built from published GWAS results and PLS derived from different longevity-focused GWAS and their relationships to specific cancer diagnoses, general cancer risk, and age of cancer onset in the UK Biobank.(35,36) We also considered the development and performance of modified cancer risk prediction models that included longevity associated variants. We find that although many PLS are associated with cancers at some level, these effects are not pronounced enough to radically impact the use of

cancer PRS prediction models. We do, however, argue that the relationship between longevity, cancers and disease protection needs greater exploration, especially given that some of the variants used in the cancer-associated PLS are strongly negatively associated with other diseases, like Alzheimer's disease, raising important biological questions.

METHODS

Data Sources

We calculated polygenic risk scores (PRS) for 16 cancers using publicly available data,(7,37,38) as well as polygenic longevity scores (PLS) using 8 different longevity measures derived from different publicly available longevity GWAS summary statistics and different p-value thresholds (**Table 1**; SNPs and weights used for each PLS are provided in the supplementary material).

PLS were labeled according to the source reference for the summary statistics ('**dl**', '**dl90**', '**tim**', etc. See Table 1). We note that some PLS were based on a subset of SNP markers identified from the various longevity-focused GWAS because there were no reported effect sizes and some were not among those in version-3 of the UK Biobank total SNP list, including imputed SNPs. All PRS and PLS were computed on each of the individuals in the UK Biobank with genotype calls from the UK Biobank version-3 genotype data using the linear scoring function in PLINK v2 software and published PRS effect sizes.(36,39,40) We restricted our

analyses to White British individuals (based on UK Biobank Data-Field 21000) as the PRS and PLS we used were derived from individuals with Northern European ancestry and the simultaneous study of multiple ancestries can confound PRS inference.

Logistic Regression Modeling

We used logistic regression to fit a number of prediction models to the UK Biobank data, with a specific cancer diagnosis or cancer PRS as a dependent variable and age, sex, PLS and the PRS for other cancer diagnoses as independent variables. For these analyses we included the first 10 principal components (PC) from the genome wide genetic relationship matrix (GRM) of the participants, although there is debate about the need or meaningfulness of including PCs in relevant analyses when the focus is on polygenic effects.(41-44) We also carried out stepwise logistic regression, using both forward and backward steps in a ‘both-way’ approach to calculate the best model, keeping independent variables in models in each step based on the Akaike information criterion (AIC) until no improvements in the model occurred. More details about the various models are described in corresponding sections in the Results section. All analyses were carried out using R version 3.6.3.

Survival Analyses

We performed survival analyses on each of the 16 cancers and for ‘any-cancer’ diagnosis using age-of-onset information as the dependent variable and an individual’s age at assessment, if that individual was not diagnosed with cancer, as a censoring time variable. The UK Biobank recruited individuals between the ages of 40 and 70. We therefore did not include any cancer age-of-onsets that were reported after age 70 in follow-up studies with UK

Biobank participants. Survival analyses were pursued using both the Kaplan-Meier models coupled with log-rank tests of differences in survival curves as a function of PLS, and Cox proportional hazards models with the first 10 PCs from the GRM (see above), cancer PRS and PLS as independent variables as implemented in the ‘survival’ R package.(45)

Prediction Evaluation

We evaluated Receiver-Operator Characteristic (ROC) curves for each logistic regression model fit with cancer diagnosis as a dependent variable. We computed the Area-Under-the-Curve statistic (AUC) from the ROC analyses to gauge the performance of our logistic regression-based prediction models with age, sex, cancer PRS and PLS as independent variables. The R package pROC(46) was used for the relevant analyses.

RESULTS

Evidence for a shared genetic basis for different cancers.

We first considered evidence that there might be a shared genetic basis to susceptibility to all cancers. We note that even if there is no evidence for a shared genetic basis for susceptibility across all or even most cancers, this does not preclude the possibility that one or a few cancers may share genetically-mediated determinants with longevity. Since individuals who live a long life did not die from cancer at earlier ages by definition, this suggests some degree of overlap in genetic determinants between cancers – and all diseases – and longevity (i.e., it may be that individuals who live a long life likely do not have genetic variants contributing to cancer susceptibility to the same degree as individuals who die earlier from cancer). We

explored evidence for common genetic factors among cancers in three ways: 1. we considered studies in the literature whose results were consistent with a shared genetic basis for cancers; 2. we analyzed data in the UK Biobank to test the hypothesis of a shared genetic basis; and 3. we explored the consistency between genetic correlations and PRS correlations among cancers.

Supplementary Table 2 shows the number of individuals in the UK Biobank diagnosed with one cancer and individuals with at least two different cancers. There were only 45 individuals with 3 or more cancers and there was a maximum of 4 cancers observed for any individual. **Supplementary Table 3** depicts the AUC statistics from logistic regression models with specific cancer diagnosis as a dependent variable and age as well as other cancer diagnoses as independent variables, chosen based on stepwise logistic regression (see Methods). We note that some cancers are less frequent than others and as such some of these logistic regression analyses were likely underpowered, despite the large overall sample size of the UK Biobank. Both **Supplementary Tables 2 and 3** suggest that many cancer diagnoses are significantly associated with, and predictive of, other cancer diagnoses. Many of the relationships between cancers that we identified are well known: e.g., breast and ovarian cancer; endometrial and colorectal cancer; but some are not: e.g., bladder and breast cancer. We acknowledge and emphasize that the associations between cancers could be, at least in part, attributable to environmental factors that contribute to both cancers.

We explored the consistency between the strength of the genetic correlations and the strength of the PRS correlations, knowing that the different PRS used different sets of SNPs (see

Discussion section). **Figure 1** provides a scatter plot of the estimates of genetic correlation obtained from Lindstrom et al(47) and corresponding PRS correlations computed on the UK Biobank subjects using the PRS of several cancers generated from the publicly available summary statistics (see Methods).(7)

We note that genetic and PRS correlations can be computed under different assumptions, using different methods with different overall orientations. For example, genetic correlations consider all the variants in a GWAS, whereas PRS correlations only use variants in the individual PRS. **Figure 1** suggests there is more variation in the genetic correlations than there is in the PRS correlations, perhaps reflecting the fact that the PRS correlations involve a small number of SNPs, even though many PRS correlations are significant. **Figure 2** provides a heatmap of the cancer PRS correlations for the UK Biobank participants and suggests that many cancers exhibit positive and negative PRS correlations.

Note we included PRS correlations for cancers that are sex-specific for both sexes since such correlations may be meaningful biologically in terms of the pathways and underlying processes implicated. Since an obvious potential source for PRS correlations is overlap in the variants considered in the PRS being correlated, we tallied the number of overlapping variants used in the PRS we considered. The Supplementary Material provides a matrix with the number of SNPs in the summary statistics used to generate cancer PRS and number of intersecting SNPs across the cancers. More in-depth studies of the linkage-disequilibrium (LD) relationships between variants used in different PRS is called for to rule out more subtle sets of overlapping variant effects in different PRS.

Associations Between Polygenic Longevity Score and Cancer

We assessed the relationships between the 8 different PLS listed in **Table 1** and cancer diagnosis, as well as cancer PRS. We note that one of our PLS, the PLS derived from the Timmers et al study, 'tim,'(28) was derived from the UK Biobank data and studies involving actual UK Biobank participants, thus it should be kept in mind throughout our discussions that the study of this specific PLS could be confounded by an overfitting effect in our analyses. We took each cancer diagnosis as a dependent variable, and then considered the corresponding PRS for that cancer, PRS for the other cancers, age, and the different PLS as independent variables in stepwise logistic regression models. These analyses were pursued separately for each sex. The AUC statistics derived from these analyses consider: 1. age only; 2. age and cancer PRS; and 3. age, cancer PRS, and the PLS, and are summarized in **Table 2 and Supplementary Table 4**. (Specific models for each cancer with regression coefficients, etc. are available from the authors).

These tables suggest that PLS and other cancer PRS are associated with some cancers – Bladder, Colon, Pancreas, Prostate and Testicular cancer for men and Bladder and Colon for women – independently of the individual PRS for those cancers. However, the increase in the AUC score for these models when PLS are included are not large, and certainly not indicative of their potential for risk stratifying individuals in their current instantiation. We also computed correlations between the 16 cancer PRS and the 8 PLS. (**Supplementary Table 5**) We found small correlations, mostly negative as expected, although many were highly statistically significant. **Supplementary Table 6** provides a matrix of the correlations of each of the

cancers with each of the eight PLS. Excluding the correlations for the 'tim' PLS, the largest negative correlation was between the 'seb10' PLS and the PRS for lung cancer and the largest positive correlation was between 'sebp5' and oral cancer.

Age of Cancer Onset and Polygenic Longevity Score

Since the age-of-onset of cancers can be influenced by a number of factors, including genetic factors, we explored the association of the 8 PLS to cancer age-of-onset. We took information about the age-of-onset of each cancer in the UK Biobank and considered it in Cox proportional hazards models as well as a standard Kaplan-Meier (KM) survival analysis (see Methods). The age of individuals at the time they participated in the UK Biobank studies was used as a censoring age for individuals without a cancer diagnosis. **Supplementary Tables 7 and 8** provides the results of the Cox and KM analyses for each cancer and PLS, as well as analyses considering the age-of-onset of any cancer, and suggest that many cancers have an age-of-onset that is highly statistically significant in its association with one or another PLS, particularly the 'dl', 'dl90' and 'dl99' PLS. To depict the relationships between PLS and age-of-onset of cancers, we generated KM-based survival curves for individuals whose PLS were in different PLS distribution percentiles. **Figure 3** depicts the KM curves for analyses using age-of-onsets of any cancer diagnosis as well as for female breast cancer using the 'dl' PLS, contrasting individuals with 'dl' PLS in the upper and lower 10th percentiles of the PLS scores among the UK Biobank participants. (See **Supplementary Figure 1** for KM curves for cervical cancer and PLS).

Cancer Prediction Modeling Performance

To determine the clinical utility of cancer prediction modeling that includes PLS in addition to cancer PRS (and age and sex), we computed Receiver Operator Characteristic (ROC) curves and the area-under-the-curve (AUC) statistic from logistic regressions with all the 16 cancer diagnoses combined and treated as a dependent variable (i.e., a dummy dependent variable was constructed whereby 1 = any-cancer diagnosis, and 0 = no cancer diagnosis) and all cancer PRS, age and sex as independent variables. We then added PLS as independent variables to the models and tested to see if the AUC was improved significantly, but did not find any statistically significant improvement (**Table 3**). We also carried out a similar analysis for each of the cancers; first using only age as an independent variable and then adding a single PLS. We also did not find significant improvements in AUCs (**Supplementary Table 9**). Thus, although PLS are significantly associated with cancer diagnoses and age at onset, their inclusion in clinical prediction models is not justified at this time.

DISCUSSION

Many cancers share underlying etiological processes and factors (e.g., faulty DNA repair capacity, immune surveillance dysfunctions, etc.),⁽⁴⁸⁾ suggesting that the genetic determinants of these processes and factors mediate all or different subsets of cancers. The degree of overlap in the genetic determinants of cancers has been explored through the examination of the genetic correlations derived from individual GWAS results and data, as well as studies of the pleiotropic effects of individual genetic variants on multiple

cancers.(7,18,20,49,50) In addition, many cancers have also been shown to be genetically correlated with non-cancer-related diseases.(24,51) These genetic correlations, possibly exacerbated by relevant environmental factors, are also quite likely to lead to cancer patients developing a second primary cancer (as opposed to recurrent cancers) or have comorbid illness. In this light, many other less obvious diseases have been found to be genetically correlated to specific cancers, such as psychiatric illness and lung and breast cancer.(52) Also, and perhaps more germane to our study of PRS and PLS, a recent study by Witte et al.(7) found that many variants associated with one form of cancer were also associated with another (e.g., exhibited pleiotropic effects across different cancer types) and that variants used in the PRS of one cancer are often included in the PRS of other cancers. They also found that PRS values were correlated for many pairs of cancers, as have many others.(8,18-20,53)

Since cancers can compromise lifespan, it is possible that individuals who live a long time: 1. do not have the variants that predispose to cancer to the same degree as individuals who have not lived a long time; 2. possess variants that protect them from the deleterious effects of cancer predisposing variants; 3. have been exposed to environmental factors (e.g., diet, lifestyle, etc.) that protect them from cancer in some way; or 4. had or have a combination of the above. Since many published studies provide evidence that genetic variants exist that predispose to a long life,(30,31,34) a good question concerns the degree to which such variants are absent from cancer PRS, or whether the possession of longevity-enhancing variants protects individuals against the deleterious effects of having an elevated cancer PRS. This is of particular interest given that cancer is primarily a disease of aging, and thus cancer's genetically-mediated relationship to longevity may be pronounced. In addition, exploring and

characterizing the positive genetic correlations between cancers, between cancers and other diseases, and the likely negative genetic correlations between cancers and longevity, can shed light on the pathogenesis of aging-related diseases, and also potential risk stratification strategies for identifying individuals at risk for a specific cancer, multiple cancers, disease sequelae and correlated morbidities, and complex disease trajectories whose interventions may be complicated.

Our study of the relationship between PLS and cancer diagnosis, age-of-onset and PRS in the UK Biobank suggests that although PLS are primarily negatively associated with many cancers, these associations are not strong enough to be used to modify risk assessments in clinical contexts in their current state. Our study used the UK Biobank data to evaluate associations between PLS, cancer PRS and cancer diagnoses and age-of-onsets, but considered PLS and cancer PRS for these analyses that were derived independently of the UK Biobank, with the exception of the ‘**tim**’ PLS, to avoid overfitting. Despite this, there are several limitations to our study that should motivate further research. First, we did not consider an analysis involving the genetic basis of many modifiable risk factors that could impact susceptibility to, or protection from, cancer, such as diet, stress, exercise, smoking, etc. that themselves might be under genetic control.(54,55) Second, our study, like most GWAS-based studies, is only relevant to individuals of Northern European descent, and it is known that many cancers are more prevalent in individuals of non-European descent.(1) This is a well-documented criticism of PRS studies(56) and highlights the importance of conducting more diverse large-scale genomics research, inclusive of racially and ethnically diverse populations, such as the All-of-U.s project, (57) before PRS (and PLS) can be exploited in a clinical setting.

Third, our analyses focused on common variants of the type used in most PRS calculations based on GWAS data and as such ignores rare variants that have been shown to enhance longevity and protect against diseases.(58-60) Fourth, most PRS have been derived from cross-sectional data (i.e., cases and controls) and yet are thought to be useful for both classifying individuals with and without cancer diagnoses, and predicting disease. Future analyses should consider the association of PLS and cancer development with longitudinal data that considers age, aging, long term exposure to environmental factors, and overall health trajectory to get a clearer picture of their interplay. Fifth, for reasons of scope, we did not consider detailed functional assessments of the individual genes harboring the variants, nor the linkage disequilibrium patterns they exhibit, in the various longevity and cancer PRS, which, if pursued in future studies, could shed light on the more basic biology behind the relationship between the longevity and cancer. We do, however, point out that many of the variants used to construct the 8 PLS we studied have been shown to be negatively correlated with other diseases, suggesting there are important biological relationships between PLS and PRS variants. For example, the ‘dl’ PLS we studied is made up of only two SNPs, but those SNPs are known to be associated with Alzheimer’s disease (AD), suggesting a relationship between multilocus influences on AD and cancers.(61,62)[See the supplementary material for SNP information.]

An additional issue concerns why variants associated with longevity used in the PLS we studied are not included in cancer PRS if, in fact, the PLS are associated with cancer diagnosis and cancer PRS. Although there are a number of explanations for this which should motivate further study, we believe this is ultimately due to small marginal effects each variant used to construct a cancer PRS (or PLS) has on cancer susceptibility and the fact that the

nearly the same score for a PRS can be obtained with different combinations of genotypes. Small individual effects make it unlikely that all biologically relevant variants are identified in studies used to derive PRS (or PLS) from a single finite data set. Also, there is likely great locus heterogeneity underlying cancer susceptibility and most diseases, as well as lifespan. Finally, cancer PRS are often trained on cases that possess a number of susceptibility variants, against controls that do not, and, as such, the variants used in PRS that are predictive of cancer diagnosis (or longevity) act as surrogates for the presence of others. Despite these issues, our results suggest that genetically-mediated factors that contribute to cancer susceptibility are indeed complex, but can be teased apart to some degree by leveraging insights into the genetic basis of not only different cancers, but also health-positive traits such as longevity. More detailed and focused studies of specific genetically-mediated factors that might contribute to both cancer risk reduction and longevity, including studies involving longitudinal data, are needed.

ACKNOWLEDGEMENTS

This research was conducted with the UK Biobank Resource. The Project ID is 43036. Aspects of this work were funded in part by NIH grants UH2 AG064706, U19 AG023122, U24 AG051129, U24 AG051129-04S1; NSF grant (FAIN number) 2031819; Dell, Inc.; and the Ivy and Ottesen Foundations.

FINANCIAL DISCLOSURES

The authors declare no financial conflicts

Table 1: Key to PRS/PLS labels with references of summary statistics used for each score.

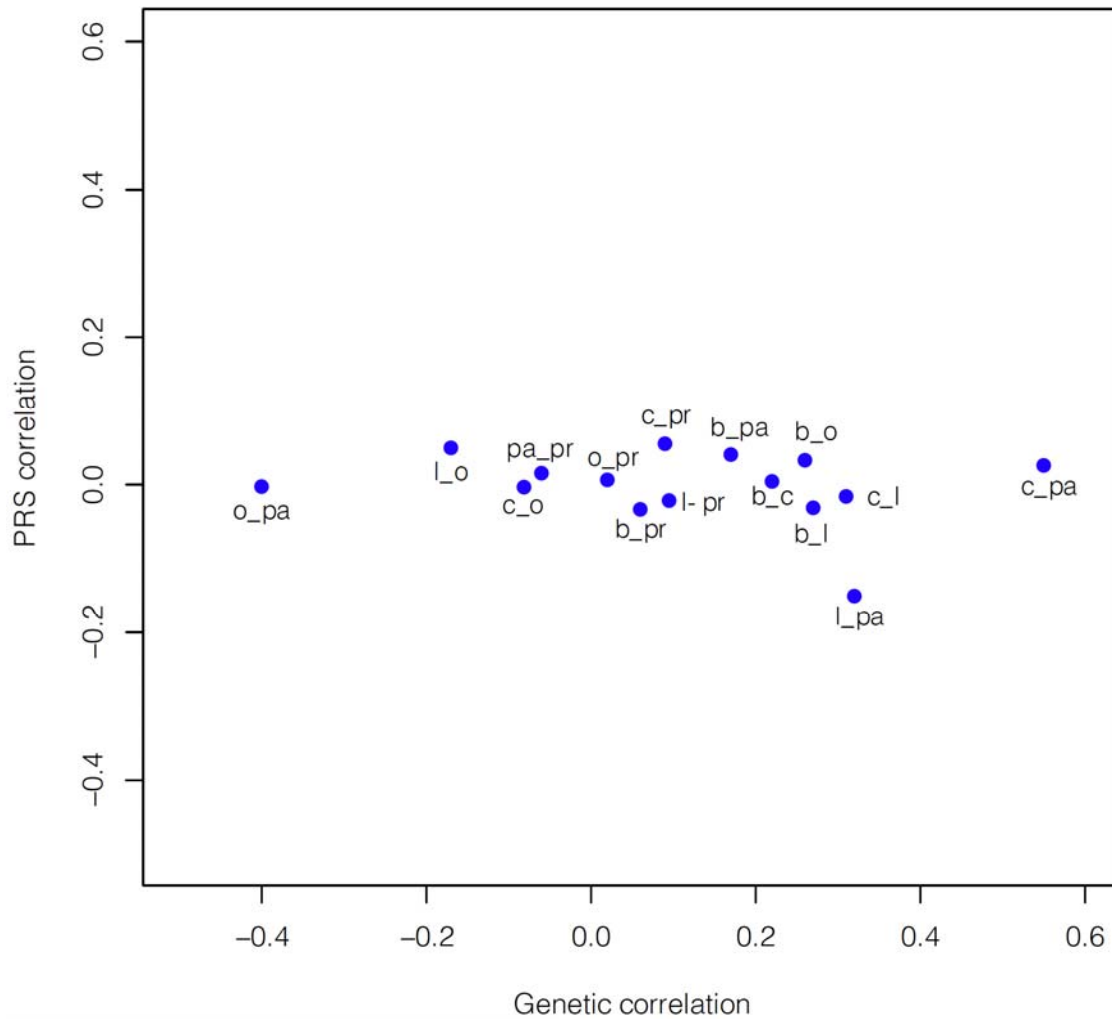
Labels	Key	Notes	Reference	# SNPS*
16 cancers**	Labeled as cancer PRS	Variants obtained from the supplementary tables of the referenced paper.	Graff (7)	See Graff et al.
dl	Abbreviation of 'Deelen'	Two snps only (subset of dl90)	Deelen (30) GWAS Catalog(37)	2/2
dl90		Longevity (age >90th survival percentile)	GWAS Central(38)	8/6
dl99		Longevity (age >99th survival percentile)	GWAS Central(38)	7/3
tim	Abbreviation of 'Timmers'	19 variants included	Timmers (28)	19/19
seb10	Abbreviation of 'Sebastiani'	seb10 includes the 10 most significant SNPs reported in the paper	Sebastiani (25)	10/10
sebp2		Includes SNPs filtered by p-value < 10^{-2}		85715/83359
sebp5		Includes SNPs filtered by p-value < 10^{-5}		272/263
sebp8		Includes SNPs filtered by p-value < 10^{-8}		30/29

* # SNPs lists the number of SNPs from the longevity GWAS / the number found in UK Biobank and used in the calculation of PLS.

** The 16 cancers included melanoma, oral cancer, pancreas, colorectal, Non-Hodgkin's Lymphoma, leukemia, kidney, testicular, endometrial, prostate, thyroid, and ovarian.

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

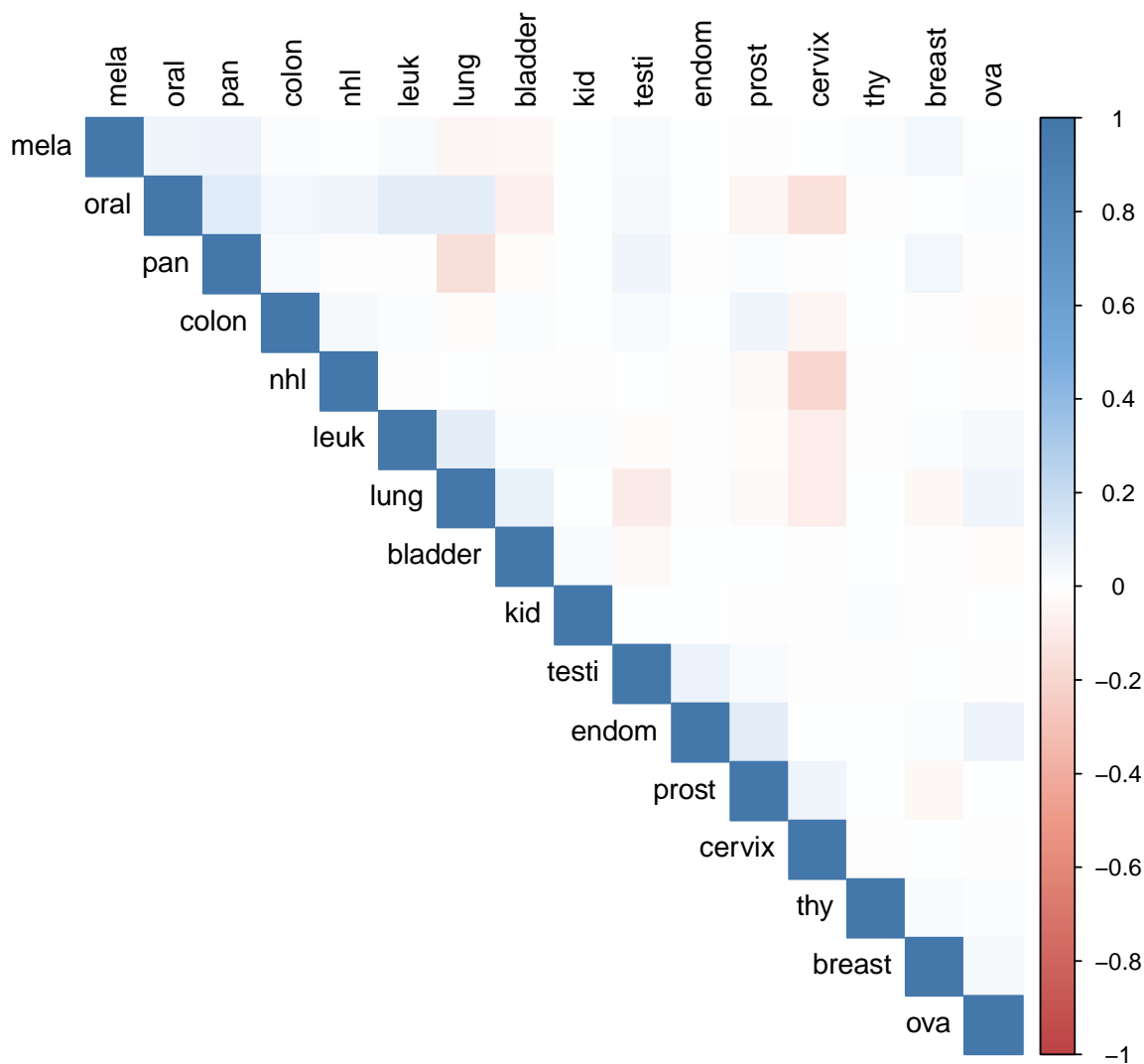
Figure 1: Scatter plot of the correlation strengths between 15 pairs of cancer PRS against the genetic correlations of those cancers based on the results of Lindstrom et al,(47) reflecting greater variation amongst genetic correlations and minimal variation amongst PRS correlations. (b=breast; c=colorectal; l=lung; o=ovarian; pa=pancreatic; pr=prostate)



It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/) .

Figure 2: Colored correlation table reflecting the strength of the PRS correlations among 16 cancers.

Red indicates a negative correlation and blue indicates a positive correlation. The stronger the correlation, the darker the color. The strongest correlations seen are a negative correlation between cervical cancer and Non-Hodgkin's Lymphoma, cervical cancer and oral cancer, and lung cancer and pancreas cancer. (mela=melanoma, oral=oral cancer, pan=pancreas, colon=colorectal, nhl=Non-Hodgkin's Lymphoma, leuk=leukemia, kid=kidney, testi=testicular, endom=endometrial, prost=prostate, thy=thyroid, ova=ovarian)



It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/) .

Table 2. AUC values derived for ROC analyses of the results of logistic regression modeling with specific cancer diagnosis as a dependent variable and the corresponding PRS for that cancer, PRS of the other cancers, age, and the different PLS as independent variables, including the first 10 Principal Components(PC) from the whole genome genetic relationship matrix (see Methods). We fit three models to determine the significance of the contribution of PLS: a ‘trivial model’ with just age as an independent variable; a ‘basic’ model with age and the cancer-specific PRS as independent variables; and a ‘final’ model in which age, the cancer-specific PRS, the PRS of the other cancers, and the eight PLS were independent variables subjected to forward and backward stepwise logistic regression with age fixed in all models. P value column corresponds to the most significant p values for the ‘final’ AUC model. The significant ‘final’ AUC values are shown in bold.

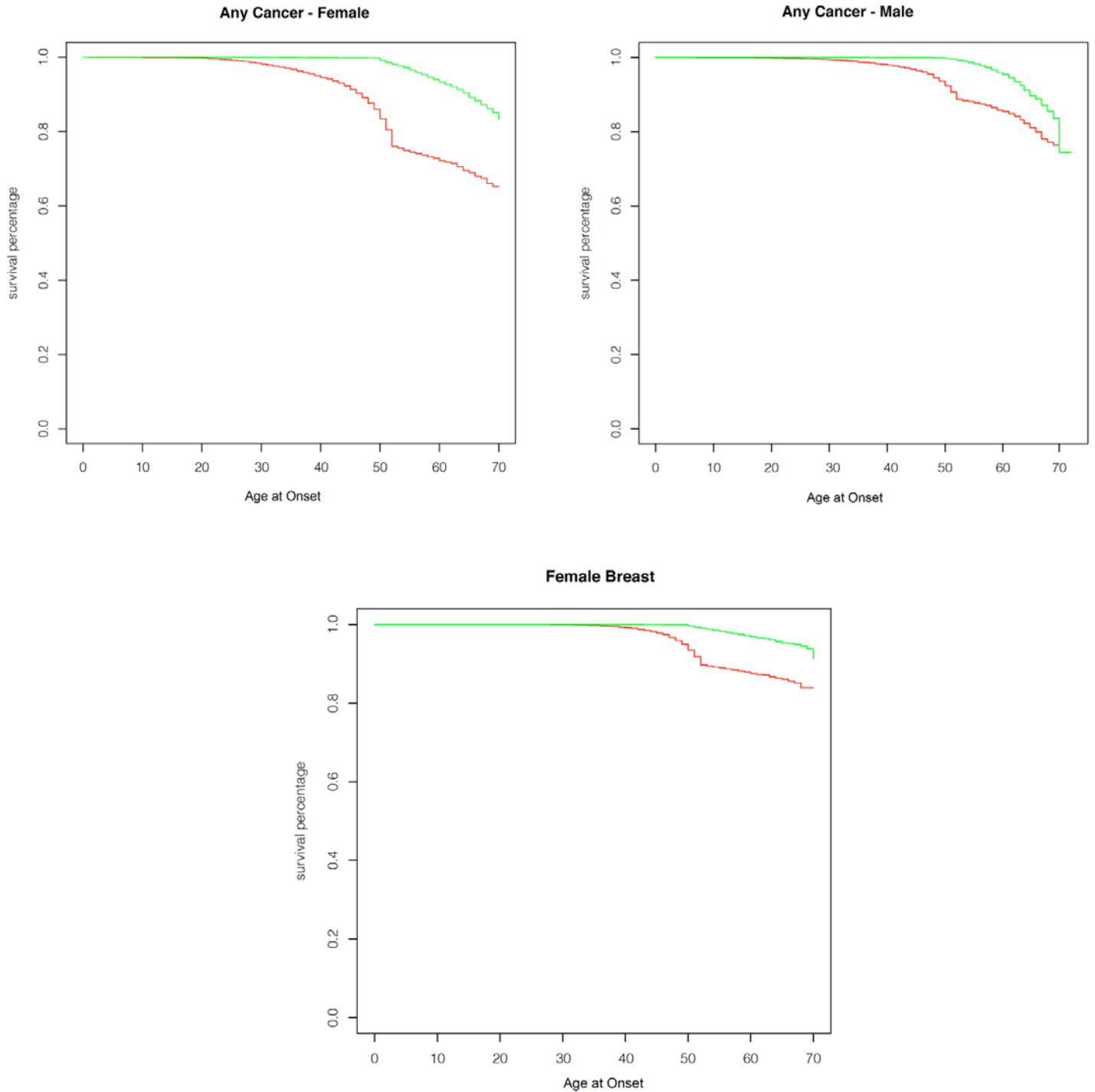
Cancer	Number of Cases	AUC (trivial)	AUC (basic)	AUC (final)	p value	Additional PRS/PLS in the Final Model
MALES						
Bladder	835	0.7046	0.7202	0.7253	ns	lung, dl99, kidney, thyroid
Colon	835	0.7047	0.7340	0.7369	0.006	dl90, oral, prostate
Thyroid	90	0.5366	0.6598	0.6742	ns	kidney, tim
Melanoma	1492	0.6031	0.6312	0.6312		-
Kidney	377	0.6366	0.6488	0.6500	ns	sebp2
Leukemia	397	0.6199	0.6652	0.6710	ns	bladder, colon, melanoma
Lung	227	0.6886	0.6889	0.6961	ns	prostate, colon, kidney
NHL	513	0.5875	0.6134	0.6165	ns	kidney
Oral	293	0.6312	0.6336	0.6476	ns	dl90, lung, melanoma, tim
Pancreas	49	0.6900	0.6903	0.7416	0.0027	lung, dl99, tim, thyroid
Prostate	3700	0.7480	0.8032	0.8036	0.05	kidney, melanoma, sebp8+ testicular, pancreas, NHL
Testicular	774	0.5803	0.7284	0.7296	ns	dl99, bladder
FEMALES						

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/) .

Bladder	302	0.7019	0.7379	0.7462	0.003	colon, thyroid, dl99, tim, NHL, leukemia
Colon	638	0.6937	0.7277	0.7293	0.036	dl, pancreas
Thyroid	330	0.5325	0.6356	0.6406	ns	endometrium, dl99
Melanoma	2101	0.5730	0.6189	0.6201	ns	kidney, endometrium, oral
Kidney	219	0.6360	0.6404	0.6603	ns	NHL, lung, endometrium, melanoma, pancreas
Leukemia	276	0.5955	0.6306	0.6377	ns	thyroid, p2, cervix
Lung	174	0.6854	0.6906	0.7009	ns	colon, ovarian, kidney
NHL	383	0.5897	0.6229	0.6329	ns	Leukemia, bladder, lung
Oral	184	0.6039	0.6050	0.6307	0.0038	sebp8, sebp2, cervix, sebp5
Pancreas	34	0.6044	0.6138	0.6559	ns	thyroid, p2
Cervix	3569	0.5483	0.5674	0.5695	0.04	lung, p2, oral
Ovarian	755	0.6019	0.6099	0.6130	ns	thyroid, endometrium, breast
Breast	10091	0.6232	0.6724	0.6732	0.0143	colon, oral, p5, dl99, tim, endometrium, melanoma
Endometrium	1137	0.6693	0.6741	0.6774	0.0047	Leukemia, kidney, dl90, dl99, cervix

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Figure 3: Kaplan-Meier curves showing effect of the ‘dl’ PLS on age-of-onset of select cancers. Red indicates individuals in the lower 10th percentile of PLS and green indicates individuals in the upper 10th percentile of PLS scores. Upper left panel: females with any cancer; Upper right panel: males with any cancer. Lower panel: Females with breast cancer. All p-values for tests of the relationship of the percentile categories and cancer age-of-onset are $p < 10^{-16}$ and thus highly significant even after multiple comparisons corrections (see Methods).



It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/) .

Table 3. AUC values derived from ROC curve analysis of logistic regression models in which all 16 cancer diagnoses were combined and treated as a dependent variable and all cancer PRS, age, sex, and the PLS were considered independent variables. Independent variables for each case are shown inside the brackets. (Analysis results involving each individual cancer are provided in **Supplementary Table 7.**)

Population	AUC (age)	AUC (age, cancer PRS)	AUC (age, cancer and PLS)
male	0.66610	0.68355	0.68360
female	0.59181	0.61491	0.61518
general	0.64256	0.65164	0.65174

REFERENCES

1. American Cancer Society Cancer Facts and Figures 2020 <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2020/cancer-facts-and-figures-2020.pdf> Accessed June 16, 2020.
2. American Association for Cancer Research Cancer Progress Report 2019 <https://cancerprogressreport.org/Pages/default.aspx> Accessed June 16, 2020.
3. United States Census Bureau [Internet] Washington, D.C. 'Older People Projected to Outnumber Children for First Time in U.S. History' Release Number CB18-41. Available from: <https://www.census.gov/newsroom/press-releases/2018/cb18-41-population-projections.html> Accessed June 22, 2020.
4. Tomasetti C, Vogelstein B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* **2015**;347:78-81
5. Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. *Nat Med* **2004**;10:789-99
6. Bozic I, Antal T, Ohtsuki H, Carter H, Kim D, Chen S, *et al.* Accumulation of driver and passenger mutations during tumor progression. *Proc Natl Acad Sci U S A* **2010**;107:18545-50
7. Graff RE, Cavazos TB, Thai KK, Kachuri L, Rashkin SR, Hoffman JD, *et al.* 2020 June 26, 2020. Cross-Cancer Evaluation of Polygenic Risk Scores for 17 Cancer Types in Two Large Cohorts. bioRxiv Web Site <<https://www.biorxiv.org/content/10.1101/2020.01.18.911578v3>>. June 26, 2020.
8. Zhang YD, Hurson AN, Zhang H, Choudhury PP, Easton DF, Milne RL, *et al.* Assessment of polygenic architecture and risk prediction based on common variants across fourteen cancers. *Nat Commun* **2020**;11:3353
9. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet* **2017**;101:5-22
10. Sud A, Kinnersley B, Houlston RS. Genome-wide association studies of cancer: current insights and future perspectives. *Nat Rev Cancer* **2017**;17:692-704
11. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* **2018**;50:1219-24
12. Fahed AC, Wang M, Homburger JR, Patel AP, Bick AG, Neben CL, *et al.* Polygenic background modifies penetrance of monogenic variants conferring risk for coronary artery disease, breast cancer, or colorectal cancer. medRxiv **2019**:19013086
13. Mars N, Widén E, Kerminen S, Meretoja T, Pirinen M, Palta P, *et al.* 2020 June 26, 2020. Polygenic risk, susceptibility genes, and breast cancer over the life course. <<https://www.medrxiv.org/content/10.1101/2020.04.17.20069229v1.full.pdf>>. June 26, 2020.
14. Gallagher S, Hughes E, Wagner S, Tshiaba P, Rosenthal E, Roa BB, *et al.* Association of a Polygenic Risk Score With Breast Cancer Among Women Carriers of High- and Moderate-Risk Breast Cancer Genes. *JAMA Netw Open* **2020**;3:e208501
15. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. *Nat Rev Genet* **2018**;19:581-90
16. Martens F, Janssens A. How the Intended Use of Polygenic Risk Scores Guides the Design and Evaluation of Prediction Studies. *Current Epidemiology Reports* **2019**;6:1-7
17. Newcombe PJ, Nelson CP, Samani NJ, Dudbridge F. A flexible and parallelizable approach to genome-wide polygenic risk scores. *Genet Epidemiol* **2019**;43:730-41
18. Sampson JN, Wheeler WA, Yeager M, Panagiotou O, Wang Z, Berndt SI, *et al.* Analysis of Heritability and Shared Heritability Based on Genome-Wide Association Studies for Thirteen Cancer Types. *J Natl Cancer Inst* **2015**;107:djv279

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/) .

19. Fritsche LG, Gruber SB, Wu Z, Schmidt EM, Zawistowski M, Moser SE, *et al.* Association of Polygenic Risk Scores for Multiple Cancers in a Phenome-wide Study: Results from The Michigan Genomics Initiative. *Am J Hum Genet* **2018**;102:1048-61
20. Jiang X, Finucane HK, Schumacher FR, Schmit SL, Tyrer JP, Han Y, *et al.* Shared heritability and functional enrichment across six solid cancers. *Nat Commun* **2019**;10:431
21. Qian DC, Byun J, Han Y, Greene CS, Field JK, Hung RJ, *et al.* Identification of shared and unique susceptibility pathways among cancers of the lung, breast, and prostate from genome-wide association studies and tissue-specific protein interactions. *Hum Mol Genet* **2015**;24:7406-20
22. Richardson TG, Harrison S, Hemani G, Davey Smith G. An atlas of polygenic risk score associations to highlight putative causal relationships across the human phenome. *Elife* **2019**;8
23. Wainschtein P, Jain DP, Yengo L, Zheng Z, Cupples LA, Shadyab AH, *et al.* 2019 June 26, 2020. Recovery of trait heritability from whole genome sequence data. *bioRxiv Web Site*. <https://www.biorxiv.org/content/biorxiv/early/2019/03/25/588020.full.pdf>. June 26, 2020.
24. Pierce BL, Ahsan H. Genetic susceptibility to type 2 diabetes is associated with reduced prostate cancer risk. *Hum Hered* **2010**;69:193-201
25. Sebastiani P, Gurinovich A, Bae H, Andersen S, Malovini A, Atzmon G, *et al.* Four Genome-Wide Association Studies Identify New Extreme Longevity Variants. *J Gerontol A Biol Sci Med Sci* **2017**;72:1453-64
26. Broer L, Buchman AS, Deelen J, Evans DS, Faul JD, Lunetta KL, *et al.* GWAS of longevity in CHARGE consortium confirms APOE and FOXO3 candidacy. *J Gerontol A Biol Sci Med Sci* **2015**;70:110-8
27. van den Berg N, Beekman M, Smith KR, Janssens A, Slagboom PE. Historical demography and longevity genetics: Back to the future. *Ageing Res Rev* **2017**;38:28-39
28. Timmers PR, Mounier N, Lall K, Fischer K, Ning Z, Feng X, *et al.* Genomics of 1 million parent lifespans implicates novel pathways and common diseases and distinguishes survival chances. *Elife* **2019**;8
29. Pilling LC, Atkins JL, Bowman K, Jones SE, Tyrrell J, Beaumont RN, *et al.* Human longevity is influenced by many genetic variants: evidence from 75,000 UK Biobank participants. *Aging (Albany NY)* **2016**;8:547-60
30. Deelen J, Evans DS, Arking DE, Tesi N, Nygaard M, Liu X, *et al.* A meta-analysis of genome-wide association studies identifies multiple longevity genes. *Nat Commun* **2019**;10:3669
31. Revelas M, Thalamuthu A, Oldmeadow C, Evans TJ, Armstrong NJ, Riveros C, *et al.* Exceptional Longevity and Polygenic Risk for Cardiovascular Health. *Genes (Basel)* **2019**;10
32. Beekman M, Nederstigt C, Suchiman HE, Kremer D, van der Breggen R, Lakenberg N, *et al.* Genome-wide association study (GWAS)-identified disease risk alleles do not compromise human longevity. *Proc Natl Acad Sci U S A* **2010**;107:18046-9
33. Carlson P, Wojczynski MK, Druley T, Lee JH, Zmuda JM, Thyagarajan B. Prevalence of clinically actionable disease variants in exceptionally long-lived families. *BMC Med Genomics* **2020**;13:61
34. Timmers P, Wilson JF, Joshi PK, Deelen J. Multivariate genomic scan implicates novel loci and haem metabolism in human ageing. *Nat Commun* **2020**;11:3570
35. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* **2015**;12:e1001779
36. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **2018**;562:203-9
37. 2020 June 10, 2020. GWAS Catalog: The NHGRI-EBI Catalog of human genome-wide association studies. <https://www.ebi.ac.uk/gwas/studies/GCST008598>. June 10, 2020.
38. GWAS Central <https://www.gwascentral.org/study/HGVST3991/results> Accessed June 10, 2020.
39. 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**;4:7
40. Chang C. Software: PLINK 2.0 www.cog-genomics.org/plink/2.0/ Accessed August 4, 2020.

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/) .

41. Lawson DJ, Davies NM, Haworth S, Ashraf B, Howe L, Crawford A, *et al.* Is population structure in the genetic biobank era irrelevant, a challenge, or an opportunity? *Hum Genet* **2020**;139:23-41
42. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics C, *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **2015**;47:291-5
43. Morris TT, Davies NM, Hemani G, Smith GD. Population phenomena inflate genetic associations of complex social traits. *Sci Adv* **2020**;6:eaay0328
44. Haworth S, Mitchell R, Corbin L, Wade KH, Dudding T, Budu-Aggrey A, *et al.* Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. *Nat Commun* **2019**;10:333
45. Therneau TM. 2020 A Package for Survival Analysis in R. R package version 3.2-3} <https://CRAN.R-project.org/package=survival> Accessed August 4, 2020.
46. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **2011**;12:77
47. Lindstrom S, Finucane H, Bulik-Sullivan B, Schumacher FR, Amos CI, Hung RJ, *et al.* Quantifying the Genetic Correlation between Multiple Cancer Types. *Cancer Epidemiol Biomarkers Prev* **2017**;26:1427-35
48. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* **2011**;144:646-74
49. Shi Z, Yu H, Wu Y, Lin X, Bao Q, Jia H, *et al.* Systematic evaluation of cancer-specific genetic risk score for 11 types of cancer in The Cancer Genome Atlas and Electronic Medical Records and Genomics cohorts. *Cancer Med* **2019**;8:3196-205
50. Kachuri L, Graff RE, Smith-Byrne K, Meyers TJ, Rashkin SR, Ziv E, *et al.* June 30, 2020. Integration of polygenic risk scores with modifiable risk factors improves risk prediction: results from a pan-cancer analysis. bioRxiv Web Site <https://www.biorxiv.org/content/10.1101/2020.01.28.922088v1>. Accessed 2020 June 30, 2020.
51. Ryan BM, Wolff RK, Valeri N, Khan M, Robinson D, Paone A, *et al.* An analysis of genetic factors related to risk of inflammatory bowel disease and colon cancer. *Cancer Epidemiol* **2014**;38:583-90
52. Zuber V, Jonsson EG, Frei O, Witoelar A, Thompson WK, Schork AJ, *et al.* Identification of shared genetic variants between schizophrenia and lung cancer. *Sci Rep* **2018**;8:674
53. Zhang Y, Wilcox AN, Zhang H, Choudhury PP, Easton DF, Milne RL, *et al.* 2019 June 26, 2020. Assessment of Polygenic Architecture and Risk Prediction based on Common Variants Across Fourteen Cancers. In bioRxiv. <https://www.biorxiv.org/content/biorxiv/early/2019/08/09/723825.full.pdf>. June 26, 2020.
54. Matoba N, Akiyama M, Ishigaki K, Kanai M, Takahashi A, Momozawa Y, *et al.* GWAS of smoking behaviour in 165,436 Japanese people reveals seven new loci and shared genetic architecture. *Nat Hum Behav* **2019**;3:471-7
55. Bohmer AC, Hecker J, Schroder J, Gharahkhani P, May A, Gerges C, *et al.* Shared Genetic Etiology of Obesity-Related Traits and Barrett's Esophagus/Adenocarcinoma: Insights from Genome-Wide Association Studies. *Cancer Epidemiol Biomarkers Prev* **2019**
56. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet* **2019**;51:584-91
57. Sankar PL, Parker LS. The Precision Medicine Initiative's All of Us Research Program: an agenda for research on its ethical, legal, and social issues. *Genet Med* **2017**;19:743-50
58. McCarrison J, Girke T, Goetz LH, Miller RA, Schork NJ. Genetic Support for Longevity-Enhancing Drug Targets: Issues, Preliminary Data, and Future Directions. *J Gerontol A Biol Sci Med Sci* **2019**;74:S61-S71
59. Harper AR, Nayee S, Topol EJ. Protective alleles and modifier variants in human health and disease. *Nat Rev Genet* **2015**;16:689-701
60. Hornstrup LS, Frikke-Schmidt R, Nordestgaard BG, Tybjaerg-Hansen A. Genetic stabilization of transthyretin, cerebrovascular disease, and life expectancy. *Arterioscler Thromb Vasc Biol* **2013**;33:1441-7

It is made available under a [CC-BY-NC-ND 4.0 International license](#) .

61. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* **2013**;45:1452-8
62. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, *et al.* Discovery and refinement of loci associated with lipid levels. *Nat Genet* **2013**;45:1274-83
63. Shieh Y, Fejerman L, Lott PC, Marker K, Sawyer SD, Hu D, *et al.* A polygenic risk score for breast cancer in U.S. Latinas and Latin-American women. *J Natl Cancer Inst* **2019**
64. Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, *et al.* Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. **2019**;104:21-34
65. Seibert TM, Fan CC, Wang Y, Zuber V, Karunamuni R, Parsons JK, *et al.* Polygenic hazard score to guide screening for aggressive prostate cancer: development and validation in large scale cohorts. *BMJ* **2018**;360:j5757
66. Yang X, Leslie G, Gentry-Maharaj A, Ryan A, Intermaggio M, Lee A, *et al.* Evaluation of polygenic risk scores for ovarian cancer risk prediction in a prospective cohort study. *J Med Genet* **2018**;55:546-54
67. Loveday C, Law P, Litchfield K, Levy M, Holroyd A, Broderick P, *et al.* Large-scale Analysis Demonstrates Familial Testicular Cancer to have Polygenic Aetiology. *Eur Urol* **2018**;74:248-52
68. Lecarpentier J, Silvestri V, Kuchenbaecker KB, Barrowdale D, Dennis J, McGuffog L, *et al.* Prediction of Breast and Prostate Cancer Risks in Male BRCA1 and BRCA2 Mutation Carriers Using Polygenic Risk Scores. *J Clin Oncol* **2017**;35:2240-50
69. Kuchenbaecker KB, McGuffog L, Barrowdale D, Lee A, Soucy P, Dennis J, *et al.* Evaluation of Polygenic Risk Scores for Breast and Ovarian Cancer Risk Prediction in BRCA1 and BRCA2 Mutation Carriers. *J Natl Cancer Inst* **2017**;109

Supplementary Material

Supplementary Table 1. Recent papers on polygenic risk scores and cancer*

Study	Year	Cancer Type	Number of SNPs	p value cutoff	Cases	Controls	AUC
Shieh(63)**	2020	Breast	180	$p < 5 \times 10^{-8}$	4658	7622	0.63
Mavaddat(64)	2019	Breast	313	$p < 10^{-5}$	94,075	75,017	0.63
Seibert(65)	2018	Prostate	54	$p < 10^{-7}$	18,868	12,879	n/a
Yang(66)	2018	Ovarian	96	$p < 10^{-5}$	750	1428	0.60
Loveday(67)	2018	Testicular	37	$p < 10^{-8}$	4167	12,368	n/a
Lecarpentier(68)	2017	Male Breast	88	$p < 10^{-5}$	308	1469	0.59
Lecarpentier(68)	2017	Prostate	103	$p < 10^{-5}$	255	1469	0.62
Kuchenbaecker(69)	2017	Breast	88	$p < 10^{-8}$	12,127	11,336	0.56-0.58
Kuchenbaecker(69)	2017	Ovarian	17	$p < 10^{-8}$	3093	20,370	0.58-0.63

*Ancestry represented in the studies is predominantly European unless indicated.

**Latinas and Latina-American Women

Supplementary Tables 2a-b. Matrices with cancer diagnosis by sex. Central diagonal contains the number of individuals diagnosed with an individual cancer. Corresponding non-diagonal positions contain the number of individuals diagnosed with a combination of cancers.

2a) Male individuals with cancers

	blad	colon	thy	mel	pros	leuk	kid	pan	testi	oral	nhl	lung
<i>blad</i>	835	6	1	9	58	1	9	0	8	3	3	4
<i>colon</i>	6	835	0	7	22	0	2	2	2	2	3	21
<i>thy</i>	1	0	90	2	2	0	1	0	1	0	1	0
<i>mel</i>	9	7	2	1492	39	5	6	1	11	5	9	3
<i>pros</i>	58	22	2	39	3700	11	22	0	9	5	11	11
<i>leuk</i>	1	0	0	5	11	397	2	0	2	0	5	1
<i>kid</i>	9	2	1	6	22	2	377	1	5	0	1	12
<i>pan</i>	0	2	0	1	0	0	1	49	0	0	0	0
<i>testi</i>	8	2	1	11	9	2	5	0	774	1	6	8
<i>oral</i>	3	2	0	5	5	0	0	0	1	293	2	2
<i>nhl</i>	3	3	1	9	11	5	1	0	6	2	513	1
<i>lung</i>	4	21	0	3	11	1	12	0	8	2	1	227

2b) Female individuals with cancers

	blad	colon	brst	ova	thy	cer	mel	endom	leuk	kid	pan	oral	nhl	lung
<i>blad</i>	302	4	21	2	1	8	1	4	0	5	0	0	1	2
<i>colon</i>	4	638	46	10	2	14	8	13	1	3	1	1	1	7
<i>brst</i>	21	46	10091	55	20	116	94	80	12	10	1	8	12	26
<i>ova</i>	2	10	55	755	0	21	8	28	0	1	0	0	2	8
<i>thy</i>	1	2	20	0	330	7	5	1	0	0	0	2	0	1
<i>cer</i>	8	14	116	21	7	3569	35	41	6	4	0	3	5	5
<i>mel</i>	1	8	94	8	5	35	2101	16	4	6	2	3	6	3
<i>endom</i>	4	13	80	28	1	41	16	1137	3	2	0	1	1	4
<i>leuk</i>	0	1	12	0	0	6	4	3	276	0	0	1	0	1
<i>kid</i>	5	3	10	1	0	4	6	2	0	219	0	1	0	0
<i>pan</i>	0	1	1	0	0	0	2	0	0	0	34	0	0	0
<i>oral</i>	0	1	8	0	2	3	3	1	1	1	0	184	1	1
<i>nhl</i>	1	1	12	2	0	5	6	1	0	0	0	1	383	0
<i>lung</i>	2	7	26	8	1	5	3	4	1	0	0	1	0	174

Supplementary Table 3. AUCs of ROC curves based on a logistic regression model with specific cancer diagnosis as a dependent variable and age as well as other cancer diagnoses as independent variables. No cancer PRS were included in these models. Other cancers are chosen after both-way stepwise logistic regression keeping age fixed. We call this model the final model.

Cancer	Cases	Controls	AUC (age)	AUC (final)	Cancers in the Final Model
MALES					
Bladder	835	196731	0.7020	0.7123	prostate, kidney, testicular, lung
Colon	835	196731	0.6991	0.7074	lung, pancreas, leukemia
Thyroid	90	197476	0.5090	0.5090	-
Melanoma	1492	196074	0.6015	0.6037	testicular, NHL
Kidney	377	197189	0.6317	0.6565	lung, bladder, prostate, testicular, pancreas
Leukemia	397	197169	0.6161	0.6217	NHL, colon
Lung	227	197339	0.6758	0.7409	colon, kidney, testicular, oral, bladder, prostate
NHL	513	197053	0.5836	0.5891	leukemia, testicular, melanoma
Oral	293	197273	0.5959	0.5969	lung
Pancreas	49	197517	0.6413	0.6705	colon, prostate, kidney
Prostate	3700	193866	0.7476	0.7494	bladder, kidney, pancreas
Testicular	774	196792	0.5764	0.5872	lung, bladder, NHL, melanoma, kidney
FEMALES					
Bladder	302	232761	0.6962	0.7068	kidney, lung, colon, cervix, melanoma
Colon	638	232425	0.6929	0.7041	lung, ovarian, endometrium, kidney, bladder, pancreas, breast, cervix
Thyroid	330	232733	0.5212	0.5248	oral, ovarian
Melanoma	2101	230962	0.5689	0.5702	kidney, pancreas, bladder
Kidney	219	232844	0.6299	0.6430	bladder, melanoma, colon
Leukemia	276	232787	0.5921	0.5921	-
Lung	174	232889	0.6631	0.7229	ovarian, breast, colon, bladder, endometrium, oral
NHL	383	232680	0.5814	0.5859	breast
Oral	184	232879	0.5879	0.5907	thyroid
Pancreas	34	233029	0.5390	0.5824	melanoma, colon
Cervix	3569	229494	0.5448	0.5521	endometrium, breast, ovarian, bladder, colon
Ovarian	755	232308	0.5983	0.6204	endometrium, lung, colon, breast, cervix, thyroid, leukemia
Breast	10091	222972	0.6226	0.6244	lung, ovarian, cervix, endometrium, NHL, colon
Endometrium	1137	231926	0.6671	0.6821	ovarian, cervix, colon, breast, lung

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Supplementary Table 4. AUC values derived for ROC analyses of the results of logistic regression modeling with specific cancer diagnosis as a dependent variable and the corresponding PRS for that cancer, PRS of the other cancers, age, and the different PLSs as independent variables. We fit three models to determine the significance of the contribution of PLS: a ‘trivial model’ with just age as an independent variable; a ‘basic’ model with age and the cancer-specific PRS as independent variables; and a ‘final’ model in which age, the cancer-specific PRS, the PRSs of the other cancers, and PLSs were independent variables subjected to forward and backward stepwise logistic regression with age fixed in all models. The bolded entries included a significant PLS effect.

Cancer	Number of Cases	AUC (trivial)	AUC (basic)	AUC (final)	Additional PRS/PLS in the Final Model	P-values of PLS in Final Model
MALES						
Bladder	835	0.7020	0.7185	0.7235	lung, dl99, kidney, thyroid	dl99=0.06
Colon	835	0.6991	0.7309	0.7340	dl90, oral, prostate	dl90=0.0057
Thyroid	90	0.5090	0.6440	0.6604	kidney, tim	tim=0.1
Melanoma	1492	0.6015	0.6306	0.6306	-	-
Kidney	377	0.6317	0.6448	0.6454	pancreas	-
Leukemia	397	0.6161	0.6612	0.6673	bladder, colon, melanoma	-
Lung	227	0.6758	0.6761	0.6834	prostate, colon, kidney	-
NHL	513	0.5836	0.6076	0.6107	kidney	-
Oral	293	0.5959	0.5994	0.6285	dl90, lung, melanoma, tim, pancreas	dl90=0.15 tim=0.09
Pancreas	49	0.6413	0.6416	0.7193	lung, dl99, tim, thyroid	dl99=0.002 tim=0.023
Prostate	3700	0.7476	0.8030	0.8035	kidney, melanoma, sebp8, testicular, pancreas, NHL	sebp8=0.0549
Testicular	774	0.5764	0.7275	0.7287	dl99, bladder	dl99=0.0919

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/) .

FEMALES						
Bladder	302	0.6962	0.7356	0.7440	colon, thyroid, dl99, tim, NHL, leukemia	dl99=0.003 tim=0.018
Colon	638	0.6929	0.7275	0.7291	dl, pancreas	dl=0.036
Thyroid	330	0.5212	0.6310	0.6356	endometrium, dl99	dl99=0.147
Melanoma	2101	0.5689	0.6159	0.6173	kidney, endometrium, oral	-
Kidney	219	0.6299	0.6344	0.6496	NHL, lung, endometrium	-
Leukemia	276	0.5921	0.6278	0.6344	thyroid, sebp2	sebp2=0.0959
Lung	174	0.6631	0.6705	0.6806	colon, ovarian, kidney	-
NHL	383	0.5814	0.6206	0.6289	Leukemia, bladder	-
Oral	184	0.5879	0.5918	0.6252	sebp8, sebp2, cervix, sebp5	sebp2=0.051, sebp5=0.137 sebp8=0.0004
Pancreas	34	0.5390	0.5730	0.6277	thyroid, sebp2	sebp2=0.126
Cervix	3569	0.5448	0.5647	0.5667	lung, sebp2, oral	sebp2=0.04
Ovarian	755	0.5983	0.6065	0.6094	thyroid, endometrium, breast	-
Breast	10091	0.6226	0.6723	0.6731	colon, oral, sebp5, dl99, endometrium, tim, melanoma	sebp5=0.02 dl99=0.014 tim=0.11
Endometrium	1137	0.6671	0.6723	0.6754	Leukemia, kidney, dl90, dl99	dl90=0.012 dl99=0.0055

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Supplementary Table 5. Correlations of Cancer-PRS vs Longevity Scores.

The table shows p-values for significant correlations (p-value < 0.05). Bold type-face indicates a statistically significant negative correlation or positive correlation between the specific longevity score and cancer type.

Cancer	dl (p-val)	dl90 (p-val)	dl99 (p-val)	tim (p-val)	seb10 (p-val)	sebp2 (p-val)	sebp5 (p-val)	sebp8 (p-val)
Bladder	0.0023	0.0015	0.0024	-0.0004	0.0011	0.0017	-0.0017	0.0002
Colon	0.0009	0.0047 (0.002)	0.0008	0.0349 ($<10^{-100}$)	-0.0004	0.0002	0.0300 ($<10^{-50}$)	-0.0003
Thyroid	-0.0033 (0.028)	-0.002	-0.0033 (0.032)	-0.0029	-0.0025	0.0015	-0.0003	-0.0021
Melanoma	-0.0032 (0.033)	-0.0014	-0.0027	-0.0011	-0.0041 (0.0068)	-0.0067 (10^{-5})	-0.0013	-0.0026
Kidney	-0.0010	-0.0023	-0.0014	-0.0007	0.0001	0.0017	-0.0018	-0.0009
Leukemia	0.0002	-0.0003	-0.0001	-0.0061 (0.00006)	0.0005	-0.0074 ($<10^{-5}$)	-0.0042 (0.0055)	0.0015
Lung	0.0004	-0.0001	0.0005	-0.1925 ($<10^{-100}$)	-0.0134 ($<10^{-17}$)	-0.0081 ($<10^{-6}$)	0.0166 ($<10^{-27}$)	0.0007
NHL	0.0002	0.0002	0.0001	-0.0243 ($<10^{-50}$)	-0.0005	-0.0004	0.0006	-0.0005
Oral	-0.0002	0.0005	-0.0002	0.0054 (0.0004)	-0.0041 (0.0077)	0.0103 ($<10^{-10}$)	0.0551 ($<10^{-100}$)	0.0001
Pancreas	0.0005	0.0012	0.0003	-0.0311 ($<10^{-50}$)	0.0008	-0.0009	0.0012	0.0003
Prostate	0.0008	-0.0007	0.0008	-0.0124 ($<10^{-15}$)	-0.0021	-0.0054 (0.00035)	-0.0001	-0.0006
Testicular	-0.0004	-0.0006	-0.0004	0.0018	0.0002	0.0306 ($<10^{-50}$)	0.0030	0.0004
Cervix	-0.0043 (0.005)	-0.0035 (0.02)	-0.0039 (0.01)	-0.0535 ($<10^{-100}$)	-0.0028	-0.0057 (0.0002)	-0.0001	-0.0023
Ovarian	-0.0013	-0.0025	-0.0013	-0.0035 (0.023)	-0.0014	-0.0016	-0.0012	-0.0014
Breast	0.0002	0.0001	0.0003	-0.0029	-0.0006	0.0001	-0.0002	-0.0005
Endometrium	0.0009	0.0008	0.0012	-0.0005	0.0015	0.0110 ($<10^{-12}$)	0.0023	0.0002

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Supplementary Table 6. Matrix with number of SNPs in the summary statistics used to generate cancer PRSs (in the diagonal) and number of intersecting SNPs across the cancers (in corresponding non-diagonal positions).

	bladder	breast	cervix	colon	endom	kid	leuk	lung	mela	nhl	oral	ova	pan	prost	testi	thy
<i>bladder</i>	15	0	0	1	0	0	1	0	1	0	0	0	1	0	0	0
<i>breast</i>	0	177	0	0	0	1	0	0	0	0	0	0	0	2	0	0
<i>cervix</i>	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>colon</i>	1	0	0	103	0	0	1	0	0	0	0	0	0	0	0	0
<i>endom</i>	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0
<i>kid</i>	0	1	0	0	0	19	0	0	0	0	0	0	0	0	0	0
<i>leuk</i>	1	0	0	1	0	0	75	1	0	0	0	2	0	0	0	0
<i>lung</i>	0	0	0	0	0	0	1	71	0	0	0	1	0	0	1	0
<i>mela</i>	1	0	0	0	0	0	0	0	23	0	0	0	1	1	0	0
<i>nhl</i>	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	0
<i>oral</i>	0	0	0	0	0	0	0	0	0	0	13	0	0	0	0	0
<i>ova</i>	0	0	0	0	0	0	2	1	0	0	0	31	0	0	0	1
<i>pan</i>	1	0	0	0	0	0	0	0	1	0	0	0	21	0	0	0
<i>prost</i>	0	2	0	0	0	0	0	0	1	0	0	0	0	158	0	0
<i>testi</i>	0	0	0	0	0	0	0	1	0	0	0	0	0	0	51	0
<i>thy</i>	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	10

Supplementary Table 7. Cox proportional hazards models testing the difference in the upper and lower 50th percentiles of the 8 different PLS against specific cancer age-of-onsets. The first 10 PCs of the genome-wide GRM for individuals considered in the study as well as the cancer PRS for the specific cancer whose onset is of focus were included as covariates. The upper entries provide the estimated hazard ratios for those in the upper 50th percentile of the PLS distribution and the lower entries provide the p-value for the difference from those in the lower 50th percentiles. Bold entries have p<0.05.

Cancer	dl	dl90	dl99	tim	seb10	sebp2	sebp5	sebp8
MALES								
Bladder	5.435e-01 < 2e-16	8.991e-01 0.1276	7.620e-01 0.00011	1.016e+00 0.82375	1.040e+00 0.57079	9.855e-01 0.83543	1.076e+00 0.29245	1.059e+00 0.41390
Colon	4.369e-01 <2e-16	8.033e-01 0.00167	5.880e-01 2.64e-14	8.786e-01 0.063	9.301e-01 0.298	1.104e+00 0.154	8.727e-01 0.0506	9.459e-01 0.424
Thyroid	3.413e-01 1.7e-06	1.064e+00 0.770	4.935e-01 0.00115	1.415e+00 0.107	1.266e+00 0.269	1.061e+00 0.783	8.919e-01 0.590	1.337e+00 0.175
Melanoma	3.881e-01 <2e-16	8.695e-01 0.00734	5.029e-01 < 2e-16	1.067e+00 0.21506	9.905e-01 0.85551	1.051e+00 0.33856	9.909e-01 0.86052	9.568e-01 0.39683
Kidney	4.461e-01 1.49e-14	9.644e-01 0.726774	5.846e-01 2.46e-07	1.068e+00 0.527939	1.020e+00 0.848966	9.478e-01 0.605769	1.120e+00 0.27491	9.442e-01 0.579380
Leukemia	3.949e-01 <2e-16	9.397e-01 0.541	6.008e-01 6.03e-07	1.110e+00 0.3043	9.766e-01 0.8155	1.038e+00 0.7116	1.036e+00 0.7264	9.023e-01 0.312
Lung	4.882e-01 1.15e-07	8.862e-01 0.3661	6.429e-01 0.000984	1.023e+00 0.8672	1.118e+00 0.4032	1.043e+00 0.7513	1.182e+00 0.2131	9.961e-01 0.9769
NHL	4.035e-01 < 2e-16	1.020e+00 0.8249	5.337e-01 2.71e-12	9.621e-01 0.6633	1.029e+00 0.7470	1.053e+00 0.5593	9.933e-01 0.9394	1.059e+00 0.5164
Oral	4.296e-01 1.59e-12	1.039e+00 0.7416	6.966e-01 0.00211	1.108e+00 0.3799	1.313e+00 0.0211	1.147e+00 0.2454	1.143e+00 0.2556	1.112e+00 0.3673
Pancreas	3.064e-01 6.45e-05	6.330e-01 0.116	5.097e-01 0.0193	1.037e+00 0.898	9.027e-01 0.721	1.326e+00 0.334	1.214e+00 0.499	8.000e-01 0.437
Prostate	8.402e-01 1.04e-06	9.856e-01 0.6656	1.030e+00 0.3939	1.044e+00 0.2018	1.037e+00 0.2772	9.729e-01 0.4118	1.032e+00 0.3492	1.062e+00 0.0743
Testicular	1.411e-01 <2e-16	8.024e-01 0.0023	1.820e-01 <2e-16	9.258e-01 0.2841	9.013e-01 0.1495	1.055e+00 0.4606	9.142e-01 0.2126	8.666e-01 0.0470
FEMALES								
Bladder	5.681e-01 1.60e-06	1.137e+00 0.2701	7.054e-01 0.00272	8.963e-01 0.3456	1.058e+00 0.6251	1.130e+00 0.2921	9.736e-01 0.8174	1.028e+00 0.8129
Colon	3.674e-01 <2e-16	7.812e-01 0.00192	5.434e-01 2.18e-14	9.854e-01 0.853	8.993e-01 0.182	1.004e+00 0.961	9.547e-01 0.559	8.664e-01 0.0711
Thyroid	2.711e-01 <2e-16	9.931e-01 0.950	4.105e-01 2e-14	1.067e+00 0.555	1.073e+00 0.521	8.544e-01 0.157	1.106e+00 0.363	1.008e+00 0.940
Melanoma	2.836e-01 <2e-16	8.804e-01 0.003662	3.905e-01 < 2e-16	9.733e-01 0.536857	9.741e-01 0.549792	1.037e+00 0.409485	1.020e+00 0.649891	1.043e+00 0.336585

It is made available under a [CC-BY-NC-ND 4.0 International license](#) .

Kidney	3.926e-01 1.46e-11	8.298e-01 0.169	5.381e-01 5.66e-06	9.636e-01 0.784	9.650e-01 0.793	1.065e+00 0.643	1.027e+00 0.842	1.007e+00 0.960
Leukemia	4.418e-01 3.75e-11	8.574e-01 0.2020	5.334e-01 2.59e-07	1.079e+00 0.5277	1.078e+00 0.533	1.188e+00 0.1553	9.885e-01 0.9233	1.338e+00 0.0170
Lung	4.131e-01 1.08e-08	7.580e-01 0.0706	7.654e-01 0.0821	9.999e-01 0.9994	8.252e-01 0.2103	9.259e-01 0.6151	8.107e-01 0.1706	9.743e-01 0.8647
NHL	3.334e-01 < 2e-16	8.360e-01 0.0818	5.694e-01 5.55e-08	1.028e+00 0.7882	8.910e-01 0.2624	1.130e+00 0.2381	1.188e+00 0.0958	9.995e-01 0.9964
Oral	2.569e-01 <2e-16	6.738e-01 0.00828	3.487e-01 1.81e-11	9.127e-01 0.536	8.955e-01 0.455	8.148e-01 0.168	9.145e-01 0.545	8.424e-01 0.246
Pancreas	3.980e-01 0.00944	1.338e+00 0.403	4.249e-01 0.0158	1.280e+00 0.475	1.143e+00 0.697	1.372e+00 0.365	1.413e+00 0.321	8.704e-01 0.686
Cervix	1.792e-01 < 2e-16	8.927e-01 0.000739	2.003e-01 < 2e-16	1.019e+00 0.573878	1.008e+00 0.818883	9.589e-01 0.213458	1.016e+00 0.627755	1.058e+00 0.092896
Ovarian	2.912e-01 < 2e-16	8.655e-01 0.048299	4.456e-01 < 2e-16	8.760e-01 0.070408	1.024e+00 0.742515	1.009e+00 0.907721	9.500e-01 0.482307	9.469e-01 0.455129
Breast	2.989e-01 < 2e-16	8.909e-01 7.19e-09	4.649e-01 < 2e-16	1.017e+00 0.411124	1.016e+00 0.437377	1.016e+00 0.433947	1.048e+00 0.017952	1.013e+00 0.506141
Endometrium	3.751e-01 < 2e-16	8.845e-01 0.039065	5.355e-01 < 2e-16	9.855e-01 0.80527	1.075e+00 0.22649	1.004e+00 0.94651	1.037e+00 0.53849	1.156e+00 0.01496

It is made available under a [CC-BY-NC-ND 4.0 International license](#) .

Supplementary Table 8. Kaplan-Meier analyses for each cancer and each PLS. Values represent the rounded p-values from comparison of the two survival curves based on the upper and lower halves (50th percentiles) of PLS. The values are tabulated for each of the PLS and for each of the 16 cancers as well as 'any cancer'. PLSs 'dl', 'dl90' and 'dl99' show consistently significant differences between the upper and lower 50th percentile curves.

Cancer	tim	dl	dl90	dl99	seb10	sebp2	sebp5	sebp8
MALES								
Any	0.1	<2e-16	7e-05	<2e-16	0.4	0.6	0.06	0.8
Bladder	0.9	<2e-16	0.2	4e-04	0.8	0.7	0.3	0.4
Colon	0.1	<2e-16	0.003	1e-13	0.2	0.2	0.1	0.5
Thyroid	0.1	7e-07	0.8	0.001	0.3	0.9	0.6	0.2
Melanoma	0.2	<2e-16	0.01	<2e-16	0.6	0.5	0.9	0.5
Kidney	0.6	4e-15	0.8	5e-07	1	0.6	0.3	0.6
Leukemia	0.3	<2e-16	0.6	1e-06	0.7	0.8	0.7	0.3
Lung	1	6e-08	0.4	0.002	0.5	0.8	0.2	0.9
NHL	0.5	<2e-16	0.8	9e-12	0.9	0.6	1	0.5
Oral	0.4	4e-13	0.7	0.004	0.03	0.3	0.2	0.3
Pancreas	0.9	2e-05	0.1	0.02	0.7	0.5	0.5	0.5
Prostate	0.5	1e-07	0.7	0.3	0.6	0.1	0.3	0.2
Testicular	0.2	<2e-16	0.004	<2e-16	0.07	0.3	0.2	0.05
FEMALES								
Any	0.9	<2e-16	<2e-16	<2e-16	0.1	0.2	0.02	0.04
Bladder	0.3	2e-06	0.2	0.005	0.8	0.4	0.8	0.8
Colon	0.9	<2e-16	0.003	8e-14	0.1	0.8	0.8	0.09
Thyroid	0.6	<2e-16	1	2e-14	0.7	0.1	0.3	0.9
Melanoma	0.5	<2e-16	0.006	<2e-16	0.2	0.9	0.6	0.3
Kidney	0.8	5e-12	0.2	7e-06	0.7	0.7	0.8	0.9
Leukemia	0.6	2e-11	0.2	5e-07	0.6	0.2	0.9	0.01
Lung	0.7	4e-09	0.08	0.1	0.2	0.5	0.2	0.9
NHL	0.9	<2e-16	0.1	1e-07	0.2	0.4	0.09	0.9
Oral	0.5	<2e-16	0.009	9e-12	0.4	0.1	0.6	0.3
Pancreas	0.5	0.008	0.4	0.02	0.7	0.5	0.3	0.7
Cervix	1	<2e-16	0.002	<2e-16	0.6	0.02	0.5	0.05
Ovarian	0.06	<2e-16	0.06	<2e-16	1	0.8	0.5	0.5
Breast	0.6	<2e-16	5e-08	<2e-16	0.8	0.6	0.006	0.2
Endometrium	0.8	<2e-16	0.06	<2e-16	0.4	0.8	0.5	0.01

Supplementary Table 9. AUCs of ROC of each cancer prediction with age (triv) and AUCs with age plus each of the PLSs.

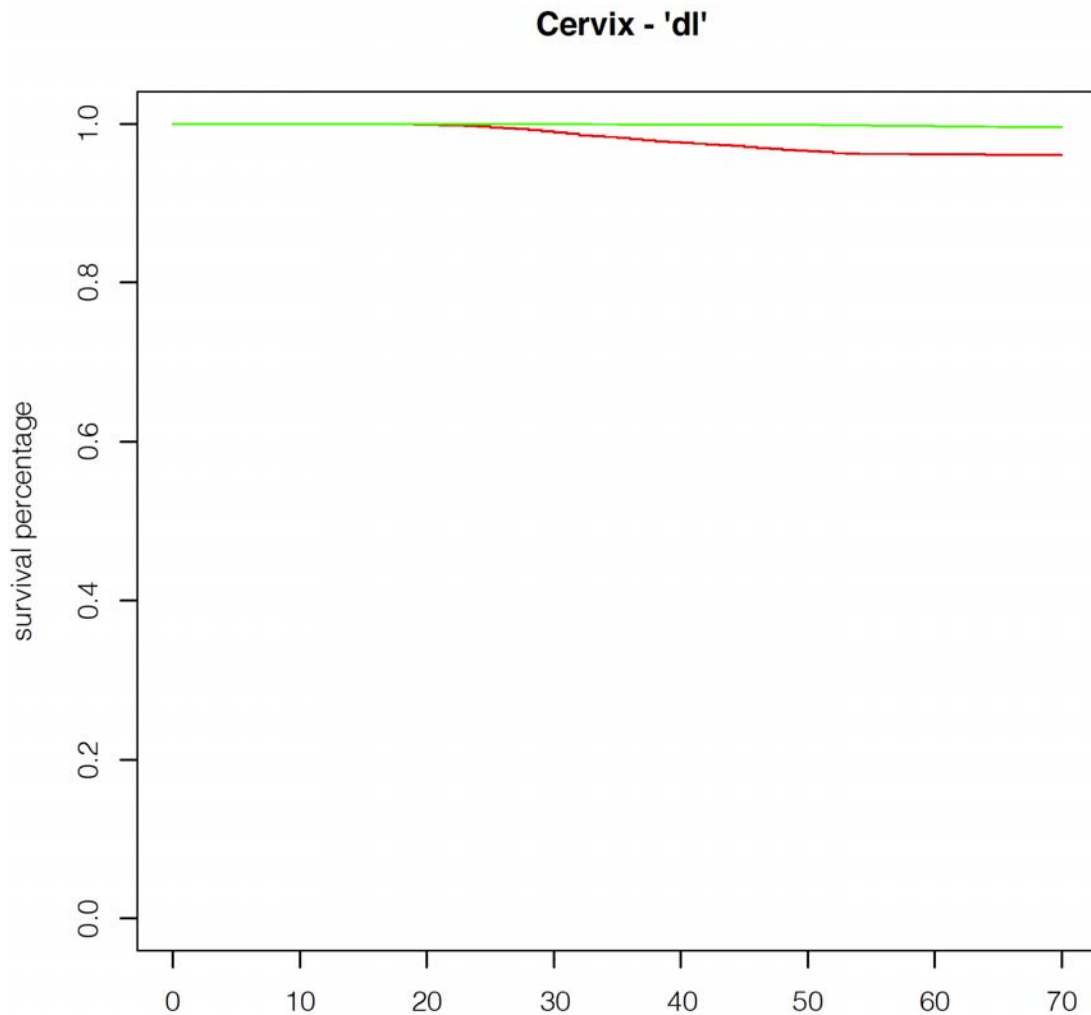
a) Males

	triv	p2	p5	p8	dl	tim	dl90	dl99	seb10
<i>blad</i>	0.702	0.702	0.703	0.703	0.703	0.702	0.703	0.703	0.703
<i>colon</i>	0.699	0.699	0.7	0.699	0.7	0.7	0.701	0.7	0.7
<i>thy</i>	0.509	0.511	0.515	0.538	0.539	0.555	0.537	0.536	0.536
<i>mel</i>	0.602	0.602	0.602	0.601	0.602	0.602	0.602	0.602	0.602
<i>pros</i>	0.748	0.748	0.748	0.748	0.748	0.748	0.748	0.748	0.748
<i>leuk</i>	0.616	0.616	0.616	0.618	0.616	0.618	0.616	0.616	0.616
<i>kid</i>	0.632	0.633	0.632	0.632	0.632	0.634	0.633	0.632	0.632
<i>pan</i>	0.641	0.642	0.641	0.666	0.666	0.641	0.66	0.664	0.663
<i>testi</i>	0.576	0.576	0.577	0.577	0.578	0.577	0.577	0.578	0.577
<i>oral</i>	0.596	0.596	0.597	0.601	0.606	0.605	0.607	0.605	0.605
<i>nhl</i>	0.584	0.584	0.584	0.584	0.584	0.584	0.584	0.584	0.584
<i>lung</i>	0.676	0.676	0.677	0.676	0.676	0.676	0.676	0.676	0.676

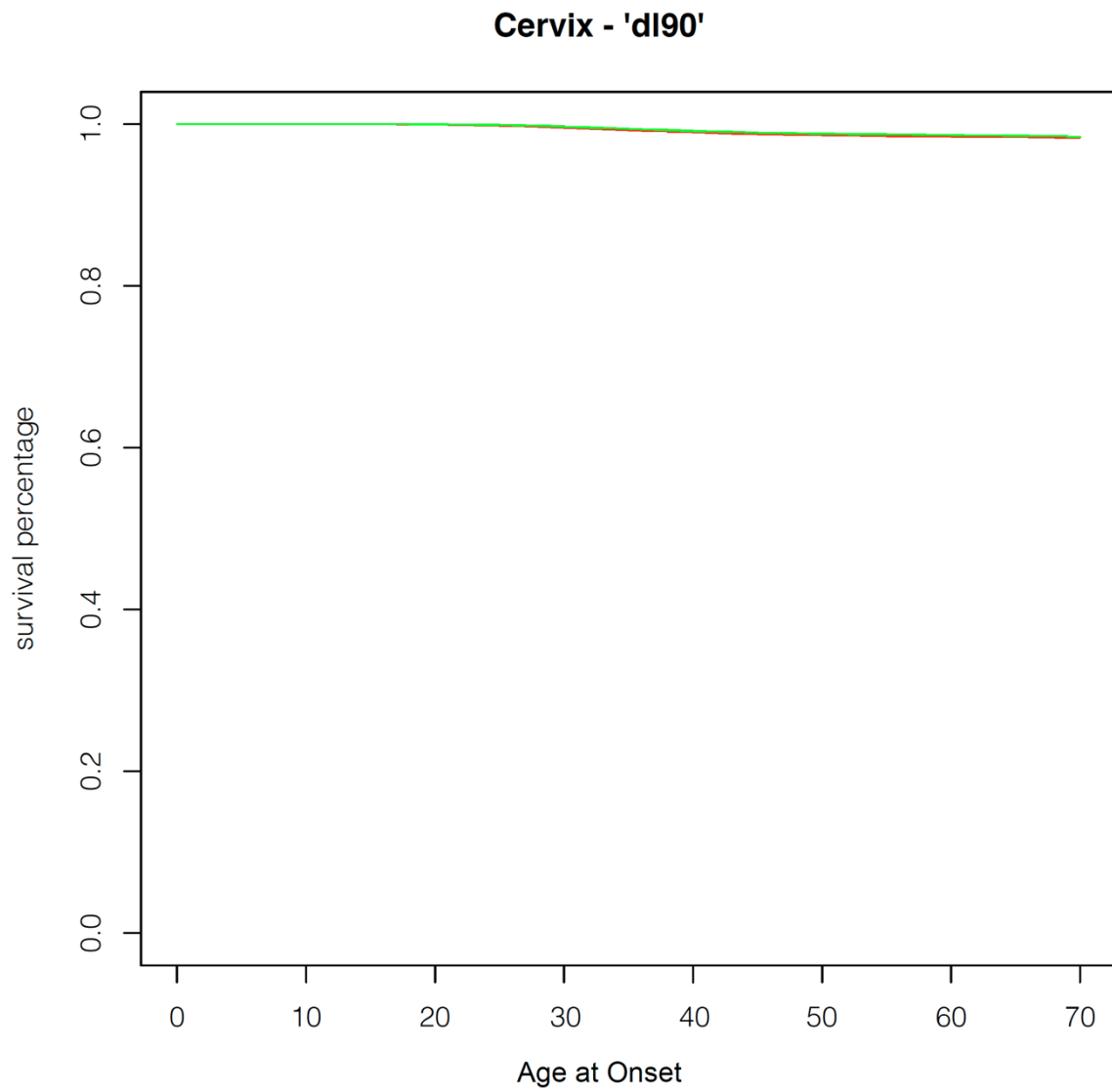
b) Females

	triv	p2	p5	p8	dl	tim	dl90	dl99	seb10
<i>blad</i>	0.696	0.696	0.696	0.698	0.699	0.697	0.698	0.699	0.697
<i>colon</i>	0.693	0.693	0.693	0.694	0.694	0.693	0.694	0.694	0.694
<i>brst</i>	0.623	0.623	0.623	0.623	0.623	0.623	0.623	0.623	0.623
<i>ova</i>	0.598	0.599	0.599	0.598	0.599	0.599	0.598	0.599	0.599
<i>thy</i>	0.521	0.521	0.521	0.529	0.531	0.53	0.525	0.532	0.523
<i>cer</i>	0.545	0.545	0.545	0.545	0.545	0.545	0.545	0.545	0.545
<i>mel</i>	0.569	0.569	0.569	0.569	0.569	0.569	0.57	0.569	0.569
<i>endom</i>	0.667	0.667	0.668	0.668	0.668	0.667	0.667	0.668	0.668
<i>leuk</i>	0.592	0.596	0.593	0.594	0.594	0.596	0.592	0.594	0.593
<i>kid</i>	0.63	0.63	0.632	0.63	0.63	0.63	0.63	0.63	0.631
<i>pan</i>	0.539	0.591	0.555	0.568	0.557	0.551	0.572	0.551	0.551
<i>oral</i>	0.588	0.598	0.592	0.615	0.612	0.6	0.609	0.612	0.604
<i>nhl</i>	0.581	0.582	0.582	0.583	0.581	0.581	0.582	0.582	0.581
<i>lung</i>	0.663	0.664	0.664	0.663	0.663	0.663	0.664	0.663	0.664

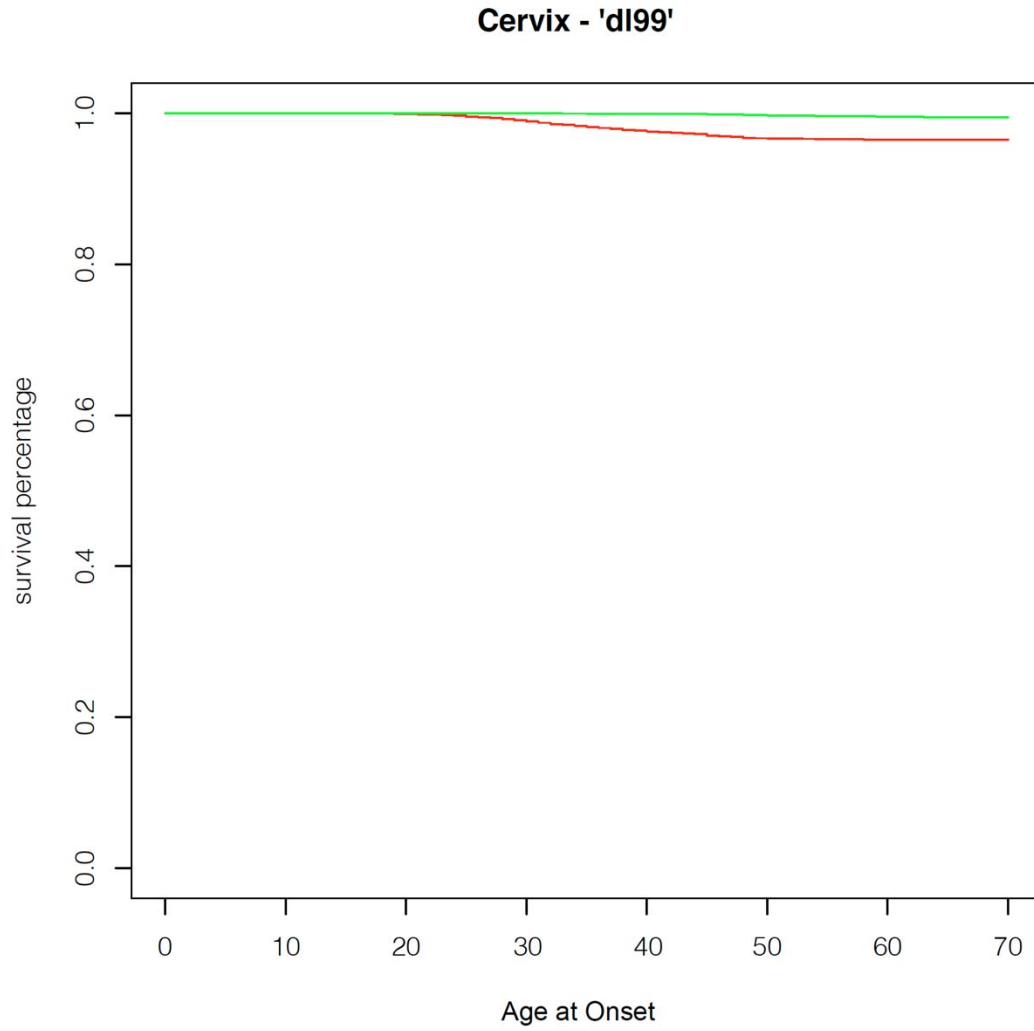
Supplementary Figure 1. Kaplan Meier Curves showing age at onset of cervical cancer as a function of different PLSs. Green indicates individuals in the lower 10th percentile of PLS and red indicates individuals in the upper 10th percentile of PLS. (a) Results for PLS 'dl' ($p < 0.0001$) (b) PLS 'dl90' ($p = 0.7$) (c) PLS 'dl99' ($p < 0.001$).



(a)



(b)



(c)