Age modulates genetic risk for multiple common diseases

Xilin Jiang[1,2,3], Chris Holmes[1,2,4], Gil McVean[1]

Affiliations
[1]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford
[2]Department of Statistics, University of Oxford
[3]Wellcome Centre for Human Genetics, University of Oxford
[4]The Alan Turing Institute, London, UK

Abstract

Genetic variation contributes substantially to complex disease and can be used to predict individual risk. However, the extent to which genetic factors are equally relevant across age, or influence risk within particular age intervals, remains largely unknown. We used a proportional hazards model within an interval-based censoring approach to estimate age-varying individual variant contributions to genetic risk for 24 common diseases within the British ancestry subset of UK Biobank. We use a Bayesian clustering approach to group variants by their risk profile over age and permutation testing for age dependency and multiplicity of profiles. We find evidence for age-varying risk profiles in nine diseases, including hypertension, skin cancer, atherosclerotic heart disease, hypothyroidism and calculus of gallbladder, several of which show evidence for multiple distinct profiles of genetic risk. The predominant pattern shows genetic risk factors having greatest impact on risk of early disease, which can only partially be explained through the concept of frailty, in which unobserved covariates such as environmental factors and interactions involving genetic risk factors contribute to heterogeneity. Our findings have implications for the estimation and use of genetic risk scores in prediction.

Introduction

Many studies have demonstrated the potential utility of using genetic risk factors for predicting individual risk of common diseases, ranging from heart disease (Rossouw 2002; Ripatti et al. 2010) to breast cancer (Mavaddat et al. 2015) and auto-immune conditions (Cotsapas et al. 2011). Genetic risk coefficients can be estimated from cross-sectional genome-wide association studies, which estimate enrichment of common genetic variants among clinically-ascertained (or sometimes self-reported) cases. Genome-wide scores, typically referred to as polygenic risk scores (PRS), are usually constructed as linear combinations of individual variant effects, though there is considerable variation in how variants are selected for inclusion and how coefficients are estimated (Choi, Mak, and O'Reilly 2020). Nevertheless, validation on independent data sets has demonstrated odds-ratios for PRSs that are comparable to established risk factors, both lifestyle-related (Mosley et al. 2020) and monogenic (Khera et al. 2018), thus providing an impetus for their adoption within health management, both at individual and population levels; though see (Mosley et al. 2020).

One aspect of genetic risk estimation that has received relatively little attention is the role of age in modulating effects. Several studies have identified variants that influence age-at-onset for diseases including type 1 diabetes (Ide et al. 2002) , Alzheimer's disease (Wollmer et al. 2003) and multiple-sclerosis (Moutsianas et al. 2015). Often, variants identified are the same as those affecting lifetime risk. Similarly, individuals with high PRS risk tend to have earlier age-at-onset than those who have low genetic risk, but nevertheless get the disease (Nalls et al. 2015; Harbo et al. 2014) . Both results can be explained by a proportional-hazards model, in which genetic risk factors multiply a baseline (and potentially time-varying) rate of disease risk. Those entering the disease state earliest will tend, therefore, to be those with the highest burden of risk factors. Nevertheless, these results raise the possibility that genetic risk factors may play larger or smaller roles in influencing risk of disease during different age intervals. Similarly, genetic analyses of quantitative traits including blood pressure, lipid levels and BMI have identified genetic variants whose effect size changes with age (Lasky-Su et al. 2008; Shi et al. 2009; Simino et al. 2014; Dumitrescu et al. 2011).

Here, we use a proportional hazards model within an interval-censoring approach to investigate age-dependency of effect size of genetic risk factors. Because the information available for single variants is relatively weak, we use a Bayesian clustering approach to identify sets of variants that show similar profiles with age. We use permutation strategies to test deviations from uniformity with age and for the presence of multiple profiles, demonstrating through simulation that the method is accurate and robust in realistic settings. Finally, we apply the method to 24 diseases within the British ancestry component of UK Biobank, identifying multiple complex diseases where genetic risk profiles change with age.

Method overview

We model the impact of a genetic risk variant as influencing the instantaneous risk of disease incidence using the proportional hazards model on a set of variants that had previously been associated with common diseases within the UK Biobank (Cortes et al. 2020; Bycroft et al. 2018) (see Methods). To estimate age-specific effects of variants we divide age into a series of intervals and use an interval-censoring approach in which the hazard rate for the risk factor is estimated by comparing those whose first disease event occurs during the interval in question to those who have a non-disease censoring event during the interval (such as death from a different disease, or drop-out from the study for reasons unrelated to disease) and those who have neither a disease nor a censoring event during the interval (Figure 1). For a given variant, we estimate the effect size and its standard error for each interval using a proportional-hazards approach, matching additional covariates such as date of birth, sex, BMI and 40 genetic principal components (see Methods). Effect sizes for individual SNPs were estimated in both univariate and multivariate settings (see below). Because estimated variant-interval coefficients have high uncertainty, we used a Bayesian clustering approach to estimate latent profiles of age-specific genetic risk, encouraging smoothness of profiles through splines. Finally, to test for deviations from homogeneity of risk over age, and to test for the presence of multiple age-specific risk profiles, we use a permutation strategy. Full details of the method are given in the Methods and Supplementary Note.

To evaluate the methodology under the assumptions of the fitted model, we used stochastic simulation, varying the number of distinct profiles and their departure from uniformity. We first considered a likelihood ratio test (LRT) approach, fitting a linear model for risk profiles over age. Under realistic assumptions about the magnitude of effect sizes and number of associated variants we found that the multivariate approach is well-calibrated in its rejection of the null model of uniformity (i.e. when effect sizes are constant over time the LRT test has a false positive rate of 0.048 at $P \leq 0.05$) and when effect sizes are constant across variants and the absolute slope is greater than 0.003 (a change in 0.6% per year on average), we have over 90% power to reject uniformity (Figure 2A). When quadratic splines were used to capture a wider range of possible risk profiles, we found that the LRT was less well calibrated under the null (false positive rate of 0.0725 and $P \leq 0.05$; Figure 2A), hence we adopted a permutation strategy for analysing empirical data. When applying the quadratic model to data simulated under a linear profile, we find a good match between true and inferred profiles (Figure 2B).

To simulate multiple cluster profiles, we modelled 10% of the variants as having a shared linear slope and again used an LRT to assess the evidence for multiple risk profiles. This required an absolute slope of at least 0.02 (4% per year change on average) in order to achieve 90% power (at $P \leq 0.05$) to detect multiple clusters (Figure 2C). Under the null (all variants have a constant profile) the test has a false positive rate of 0.063 for the linear and 0.088 for the quadratic polynomial fitting at $P \leq 0.05$. When using the quadratic model to fit risk profiles we find a good match between true and inferred profiles (Figure 2D). We therefore conclude that the approach has sufficient power to detect deviations from constant profiles and provide unbiased estimates

of risk profiles in data sets of comparable size and complexity to the UK Biobank. When analysing multiple diseases we used a FDR approach to correct for multiple testing.

Results

Age-specific genetic risk profiles in common diseases

We applied the approach to data on 409,694 individuals within the UK Biobank who self-identify as being of British Isles ancestry. We analysed 24 diseases, identified by specific ICD-10 codes for which at least 20 variants had been identified previously as associated with the disease (Cortes et al. 2020) and that have a prevalence of at least 0.5% (Table 1; Supplementary Table 1). We used eight age intervals of five years each.

When effects for variants are estimated jointly and fitted to a linear latent profile, we identified, through permutation, nine diseases with evidence ($P < 0.05$) of a departure from uniform genetic risk over age (Table 1). These are: C44.3 “other and unspecified malignant neoplasm of skin of other and unspecified parts of face”; C44.5 “unspecified malignant neoplasm of skin of trunk”; E03.9 “hypothyroidism, unspecified”; E78.0 “pure hypercholesterolemia”; I10 “essential (primary) hypertension”; I20.9 “angina pectoris, unspecified”; I25.1 “atherosclerotic heart disease of native coronary artery”; I25.2 “Old myocardial infarction” and; K80.2 “calculus of gallbladder without cholecystitis”. All diseases have $Q < 0.1$ after FDR analysis. To model non-linearity we compared polynomial and cubic spline models with different degrees of freedom (Supplementary Figure 2) and selected the quadratic polynomial model using likelihood ratio tests. No additional diseases were identified as having non-constant risk profiles when fitting a quadratic polynomial and only four of the original nine (E78.0, I10, I25.1 and C44.5) remain significant (Table 1). However, we find one additional disease (I20.0 “unstable angina”) and three of the above diseases (C44.3, E78.0 and I25.1) show evidence for more than one age-related risk profile ($P < 0.05$; Table 1, though only I25.1 has $Q < 0.1$).

A common feature of the estimated risk profiles over age is a trend towards smaller effect sizes with increasing age (Figure 3A). For example, for I25.1, we find posterior of effect size drops by 50% from 45 years old to 75 years old and for C44.5 we find the posterior drops by 58% over the same interval. (Supplementary Table 2). Where diseases may have multiple risk profiles (Figure 3B), at least one of these is also typically decreasing with age. Profiles for all 24 diseases are shown in Supplementary Figure 3 and Supplementary Figure 4. We find no compelling examples of increasing risk over age. These results are consistent with the effects of genetic risk factors to have a larger impact on the risk of early disease (de Miguel-Yanes et al. 2011), rather than late disease, though it is important to note that the absolute rate of disease typically increases with age for all diseases studied here. Estimates of risk profiles are provided in Supplementary Table 3.

The impact of frailty

For any causal covariate of interest, the presence of unmeasured and causally-associated uncorrelated covariates has the effect of generating (at the population level) additional variability in hazard rates, centred on the effect size. Such heterogeneity, typically referred to as frailty (Govindarajulu et al. 2011), has the potential to induce bias in effect sizes over time, somewhat remarkably even if independent of the covariate of interest, due to the increased rate at which individuals with high unmeasured risk enter into a disease state. Over time, those individuals with a risk-increasing covariate, but who do not have the disease, will become enriched for a protective background. Frailty will thus tend to lead to an underestimate of true effect sizes in older populations and, consequently, can even lead to biased effect size estimates (typically underestimates) in regression analysis of the entire cohort (Lin, Psaty, and Kronmal 1998).

To investigate the extent to which unmeasured genetic factors might be responsible for the diminishing of risk over time we compared the results of univariate and multivariate analyses of the variants analysed here (Figure 4A). We found that results were essentially identical under the two approaches, suggesting that genetically-arising frailty cannot explain the pattern. We next attempted to estimate general parameters of frailty using incidence data from the UK Biobank by fitting a parametric model in which the underlying disease incidence (baseline hazard rate) increases in proportion to age as a power function of age, but where there is a distribution of rates within the population, parameterised as a gamma distribution with a mean of one and an unknown variance (Aalen 1988; Vaupel, Manton, and Stallard 1979); see Methods and Supplementary Note. Estimates of parameters are provided in Supplementary Table 4, along with the significance value for a goodness-of-fit test for the inferred model. We find substantial variation in the inferred parameters. For example, the baseline hazard rate of K80.2 “calculus of gallbladder without cholecystitis” is estimated to increase proportional to age to the power of 1.9, but with substantial frailty (scale parameter = 1.87, goodness-of-fit P = 0.93; Figure 4B). In contrast, the baseline hazard rate of C44.3 “other and unspecified malignant neoplasm of skin of other and unspecified parts of face” is estimated to increase more rapidly with age (power of 3.58), but with lower frailty (scale parameter = 0.94; P = 0.76). It should be noted that the simple parametric model can be rejected at $P < 0.01$ for one (J45.9, “other and unspecified asthma”) of the 24 disorders, with the main discrepancy being a reduction in incidence among the eldest UK Biobank participants compared to the fitted model, which may potentially be explained by selection bias in recruitment and competing risks of multi-morbidity.

Previous work has demonstrated that the magnitude of the diluting impact of frailty can be predicted using the incidence and frailty distribution parameters (Aalen 1988); notably the implied effect size at a given age is reduced by a factor proportional to the prevalence at that age multiplied by the variance of frailty distribution; see Methods. We therefore compared inferred (univariate) curves for genetic variants against that implied by the fitted model (Figure 4C). In 17 of the 24 diseases we find that while the estimated frailty predicts a decreasing genetic effect size with age, the observed decrease both starts earlier and is of a larger magnitude than expected (Supplementary Figure 7; Supplementary Figure 8). Importantly, the estimated effect size tends to decrease substantially even when the prevalence of the disease

is very low. We therefore conclude that, even after accounting for independent unmeasured factors that influence disease risk, genetic risk decreases with age.

Discussion

Genetic factors influence lifetime risk for common and complex diseases through modulating a large number of molecular, cellular and tissue phenotypes, many of which are also likely to be affected by acute exposure and persistent environment (Corominas et al. 2014; Bønnelykke and Ober 2016; Stranger et al. 2017). Despite such complexity, remarkable progress has been made in identifying factors, both genetic and non-genetic, that influence risk, each of which may only have a small effect, but which, in aggregate, have substantial and clinically relevant predictive value (Jostins and Barrett 2011; Gandal et al. 2016; Manolio 2013). To date, relatively little attention has been paid to the extent to which risk prediction can be improved by allowing genetic risk to be modulated by context, such as age, sex and environment (though note (Mühlenbruch et al. 2013; Favé et al. 2018)). Here, we set out to ask whether one specific aspect of individual context, namely age, has a modulating effect on genetic risk. For example, whether there are windows during which genetic risks are particularly relevant to disease and, conversely, other windows in which genetics plays a lesser role. Our principal finding is that genetic risk factors are consistently most important in predicting risk of early disease. This does not mean that they are not relevant in predicting later disease, which is typically when most diseases occur. Rather, our results can be thought of as implying that the factor by which genetic risk factor increases risk above baseline for someone in their 40s may be exponentially higher than for an equivalent person in their late 70s (Supplementary Table 4). For example, the factor by which being in the highest decile of genetic risk for I25.1 “Atherosclerotic heart disease of native coronary artery” increases incidence over baseline between 45 to 50 years old is 6.6, compared to only 2.4 between 70 to 75 years old. Moreover, for a limited number of diseases, we find some, albeit relatively weak, evidence for multiple distinct profiles of changing risk over age.

The explanation for age-varying risk profiles is unclear, though a number of potential models could explain the observation. First, genetic risk factors, unlike environmental ones, are present from birth, while non-genetic risk factors tend to accumulate and evolve over time. Such a difference could lead to a reduced impact of genetics over time if, for example, genetic risk were mediated by developmental pathways (whose relevance will decrease over time) while non-genetic risk is mediated by separate pathways, such as those involved in adult homeostasis. Nevertheless, there are several contexts where genetic and non-genetic risk is, at least in part, mediated by the same factor, such as the impact of LDL cholesterol on cardiovascular disease. A second possible explanation is the presence of multiple gene-by-environment or gene-by-gene interactions. Such effects would exacerbate the diluting effect of frailty (Figure 5A). However, we note that while some GxE and GxG interactions have been described for complex diseases (Cordell 2009; Moutsianas et al. 2015), these are typically relatively small compared to the main genetic effects and thus unlikely to have a major impact on effect size. A third possibility is that modelling genetic risk as a multiplier of baseline risk is a poor model for the mechanistic basis of disease. Generalised risk processes such as threshold models (Duggirala

et al. 1997) provide a potentially richer framework in which to consider the impact and interactions among risk factors over time (Figure 5B), though which parameterisations (and implied mechanistic models) might be consistent with the observations here is unknown.

Whatever the cause of age-varying genetic risk, our results have several implications for the use of genetic risk factors in the genetic analysis and prediction of disease risk. First, and most obvious, is that genetic risk prediction for early disease is likely to be more effective than for later disease. For most of the diseases studied here, the inference of a single age-profile does mean that the rank order of genetic risk for an individual is stable over time. However, it implies that integrated prediction from genetic and non-genetic risk factors (Thomas 2010; Aschard et al. 2012; Kraft et al. 2007) will have to consider the evolving contribution of genetics over age. For diseases with multiple age profiles, even the rank order of genetic risk among individuals may change over time. Second, the biasing impact of unmeasured covariates, even when independent of a covariate of interest, introduces analytical complexity that cannot easily be overcome and will typically lead to an underestimate of genetic effect sizes in older age groups. Finally, because contexts beyond age, such as sex and environment, modulate genetic risk (Ober, Loisel, and Gilad 2008; Thomas 2010), each of these will induce its own age-specific profiles. As a consequence, effective genetic prediction will most likely be driven by empirical models that can benefit from access to large and well-measured populations, such as population-scale biobanks.

Acknowledgements

This research has been conducted using the UK Biobank Resource; application number 12788. This work uses data provided by patients and collected by the NHS as part of their care and Support. Funded by Wellcome (BST00080-H503.01 to XJ, 100956/Z/13/Z to GM); the Li Ka Shing Foundation (to GM); The Alan Turing Institute, Health Data Research UK, the Medical Research Council UK, the Engineering and Physical Sciences Research Council (EPSRC) through the Bayes4Health programme Grant EP/R018561/1, and AI for Science and Government UK Research and Innovation (UKRI) (to CH). Computation used the Oxford Biomedical Research Computing (BMRC) facility, a joint development between the Wellcome Centre for Human Genetics and the Big Data Institute supported by Health Data Research UK and the NIHR Oxford Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. We thank Brieuc Lehmann for discussion and comments on the manuscript.

Conflict of interest statement

G.M. is a director of and shareholder in Genomics PLC, and is a partner in Peptide Groove LLP. The other authors declare no competing financial interests.

Tables

ICD-10 code	Description	Prevalence in UK Biobank	Number of associated variants	Mean reported age of onset	P1	Q1	P2	Q2	P3	Q3
C44.3	Other and unspecified malignant neoplasm of skin of other and unspecified parts of face	1.41%	34	62.64	0.0159*	0.0633*	0.0739	0.2531	0.0092*	0.1093
C44.5	Other and unspecified malignant neoplasm of skin of trunk	0.48%	25	61.45	0.0016**	0.0091**	0.0143*	0.0853*	0.5644	0.7964
C78.7	Secondary malignant neoplasm of liver and intrahepatic bile duct	0.64%	30	64.20	0.3753	0.5124	0.6920	0.8368	0.5973	0.7964
C79.5	Secondary malignant neoplasm of bone and bone marrow	0.52%	29	64.56	0.4649	0.5579	0.8287	0.8647	0.3122	0.6242
E03.9	Hypothyroidism, unspecified	3.46%	32	60.61	0.0329*	0.0876*	0.0592	0.2501	0.8010	0.8738
E10.9	Type 1 diabetes mellitus without complications	0.63%	28	57.32	0.4975	0.5686	0.7828	0.8539	0.0566	0.2353
E11.9	Type 2 diabetes mellitus without complications	4.38%	75	61.95	0.4103	0.5182	0.5358	0.7610	0.2710	0.6242
E66.9	Obesity, unspecified	2.46%	21	60.71	0.0981	0.2067	0.2609	0.5350	0.5420	0.7964
E78.0	Pure hypercholesterolemia	8.12%	43	62.59	0.0001**	0.0001**	0.0001**	0.0001**	0.0449*	0.2353
I10	Essential (primary) hypertension	18.98%	80	61.87	0.0001**	0.0001**	0.0001**	0.0001**	0.3764	0.6604
I20.0	Unstable angina	1.26%	46	59.92	0.9021	0.9021	0.7323	0.8368	0.0356*	0.2353
I20.9	Angina pectoris, unspecified	3.84%	78	61.84	0.0244*	0.073*	0.1269	0.3382	0.1902	0.6242
I21.9	Acute myocardial infarction, unspecified	0.92%	24	62.08	0.7556	0.7884	0.3121	0.5350	0.6356	0.8028
I25.1	Atherosclerotic heart disease of native coronary artery	4.46%	116	61.57	0.0001**	0.0001**	0.0001**	0.0001**	0.0001**	0.0001**
I25.2	Old myocardial infarction	1.72%	83	64.32	0.0096*	0.0456*	0.1053	0.3156	0.2198	0.6242
I25.9	Chronic ischaemic heart disease, unspecified	2.66%	86	63.69	0.1285	0.2202	0.4428	0.7084	0.2452	0.6242
I50.1	Left ventricular failure, unspecified	0.83%	22	63.22	0.0696	0.1669	0.2971	0.5350	0.9950	0.9950
J44.9	Chronic obstructive pulmonary disease, unspecified	1.78%	24	64.40	0.3329	0.5124	0.5391	0.7610	0.8452	0.8819
J45.9	Other and unspecified asthma	6.34%	35	57.54	0.1056	0.2067	0.2027	0.4863	0.5087	0.7964
K29	Gastritis and duodenitis	7.02%	33	58.87	0.3843	0.5124	0.7056	0.8368	0.0589	0.2353
K80.2	Calculus of gallbladder without cholecystitis	2.11%	26	57.99	0.0236*	0.073*	0.0626	0.2501	0.7942	0.8738
M06.9	Rheumatoid arthritis, unspecified	0.99%	54	60.16	0.3548	0.5124	0.6271	0.8361	0.3853	0.6604
M19.9	Osteoarthritis, unspecified site	3.70%	22	62.80	0.1120	0.2067	0.2981	0.5350	0.2989	0.6242
M54.5	Low back pain	1.90%	29	56.75	0.6320	0.6894	0.8935	0.8935	0.7793	0.8738

P1: permutation test for fitting a linear profile over age; P2: permutation test for fitting a quadratic polynomial profile over age; P3: permutation test for multiple profiles over age **P < 0.005 *P < 0.05
Q1: FDR adjusted P1 value; Q2: FDR adjusted P2 value; Q3: FDR adjusted P3 value **Q < 0.01 *Q < 0.1

Table 1. Summary of ICD-10 disease codes analysed and evidence for age-varying effect sizes and number of age-profile classes.

Figures

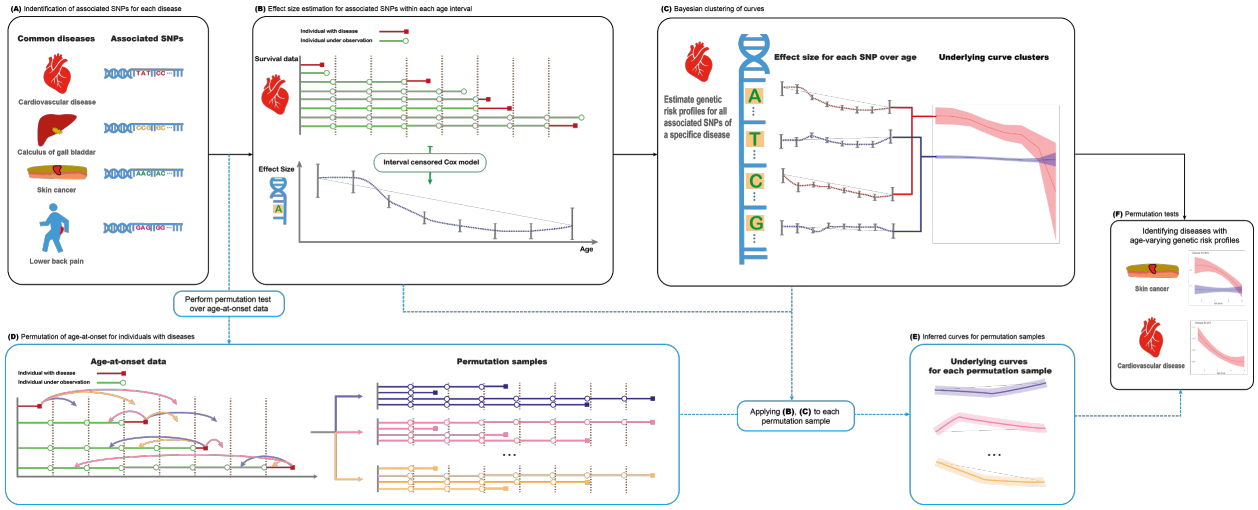

Figure 1. Schematic representation of methodology. A) Independent variants associated with a trait of interest are identified by analysis of the entire UK Biobank cohort using the TreeWAS methodology (Cortes et al. 2017). B) An interval-censored proportional hazards model (Finkelstein 1986) is used to estimate the effect (and associated standard error) of each variant on the trait of interest within each of eight age intervals. C) Bayesian clustering is used to estimate age-profiles of risk, using either linear models or quadratic polynomials to encourage smoothness. D-F) Permutation is used to test for age-homogeneity of effect size as well as to assess the evidence for multiple age profiles.

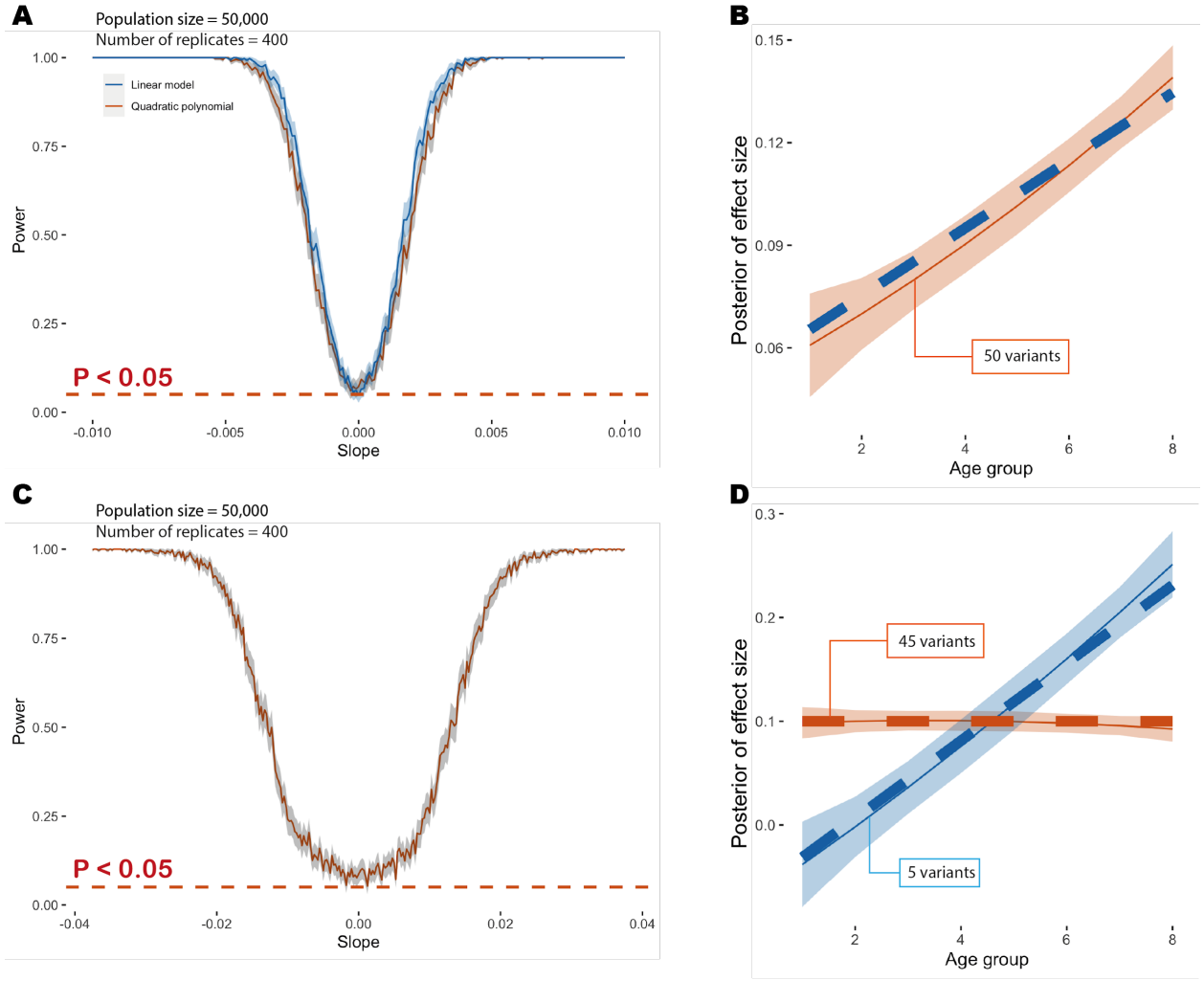

Figure 2. Overview of simulation results. A) Power at P ≤ 0.05 to detect deviation from age-homogeneity as a function of slope in a model where effect sizes change linearly with age. The blue line indicates the point estimate when using a linear model to fit, the red line indicates the point estimate with a quadratic polynomial model and the grey shading indicates the 95% confidence interval. B) Example showing the age-profile under which data are simulated (dashed blue line) and the inferred age profile (dashed red line) and 95% credible interval (red shading). C) Power at P ≤ 0.05 to detect multiple age profiles in a simulation where 90% of variants have a time-invariant profile and 10% have an effect size that increases with age. Note the change in x-axis scale compared to Fig. 2A. D) Example showing inferred age-profiles for the two components (mean posterior and 95% credible interval). Additional simulation details are provided in the Methods and Supplementary Figure 1.

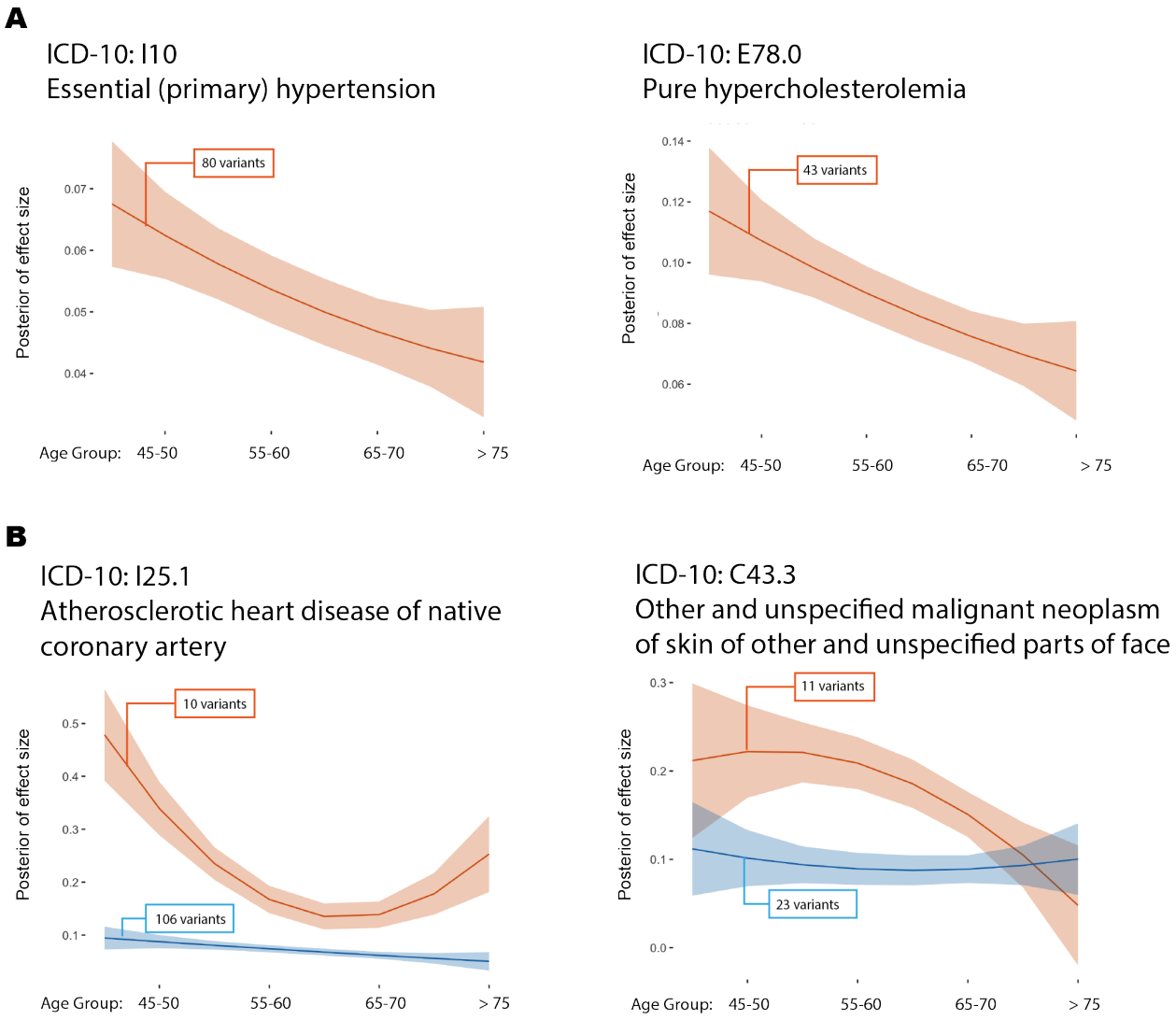

Figure 3. Age-varying disease risk profiles. A) Inferred cluster profiles for two disorders where there is evidence for single non-constant profile; "Primary (essential) hypertension" (ICD-10 code I10; P = 0.0001) and "pure hypercholesterolaemia" (E78.0; P = 0.0001). B) Inferred cluster profiles for two disorders where there is evidence for multiple non-constant profiles; "atherosclerotic heart disease of native coronary artery" (I25.1; P = 0.0001) and "other and unspecified malignant neoplasm of skin and unspecified parts of face" (C44.3; P = 0.0092). Curves for all diseases are shown in Supplementary Figure 3 and Supplementary Figure 4. The solid line indicates the posterior mean and the shaded area the 95% credible interval. Numbers in parentheses indicate the number of variants in each cluster. All estimates are made with quadratic models for age-varying risk profiles.

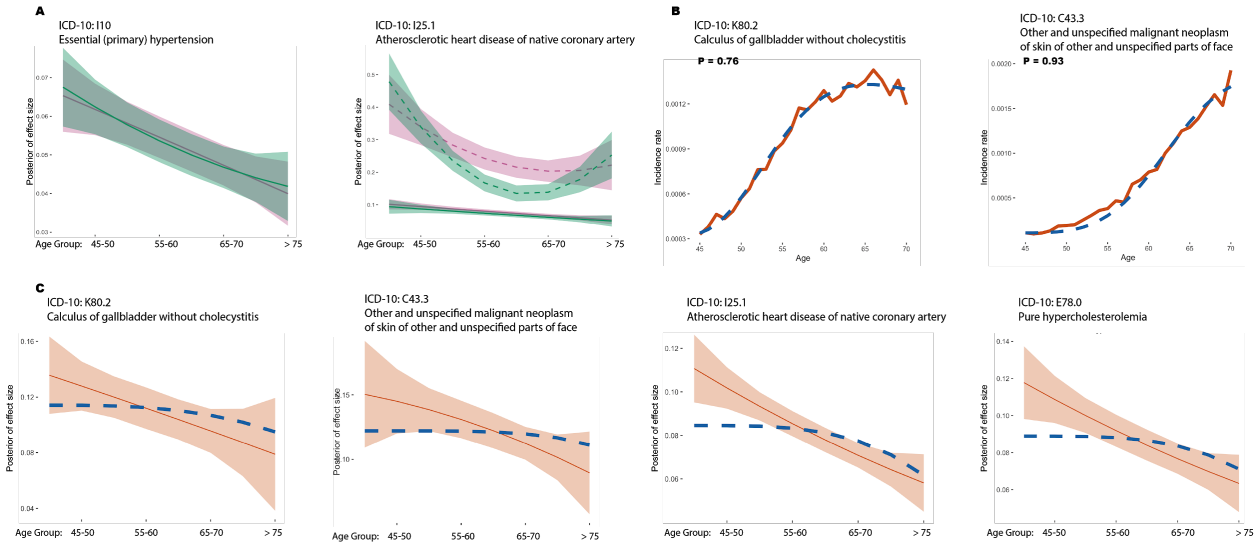

Figure 4. The impact of frailty on genetic risk profiles. A) Estimated age-profiles for genetic risk for I10 "essential (primary) hypertension" (left) and I25.1 "atherosclerotic heart disease of native coronary artery" (right) fitted under the univariate (purple) and multivariate (green) approaches. The solid line indicates the posterior mean and the shaded area the 95% credible interval. Comparisons for all diseases are shown in Supplementary Figure 5. B) Estimated incidence by age for K80.2 "Calculus of gallbladder without cholecystitis" (left) and C44.3 "Other and unspecified malignant neoplasm of skin and unspecified parts of face" (right). The red solid line indicates the rate estimated from the UK Biobank (see Methods) and the dotted blue line indicates the fitted incidence curve from the parametric model. The P value indicates the Goodness-of-Fit test. Curves for all diseases are shown in Supplementary Figure 6. C) Comparison of inferred genetic effect sizes (red curve) and those implied by the frailty parameters estimated from incidence rate within the UK Biobank (blue dashed curve).

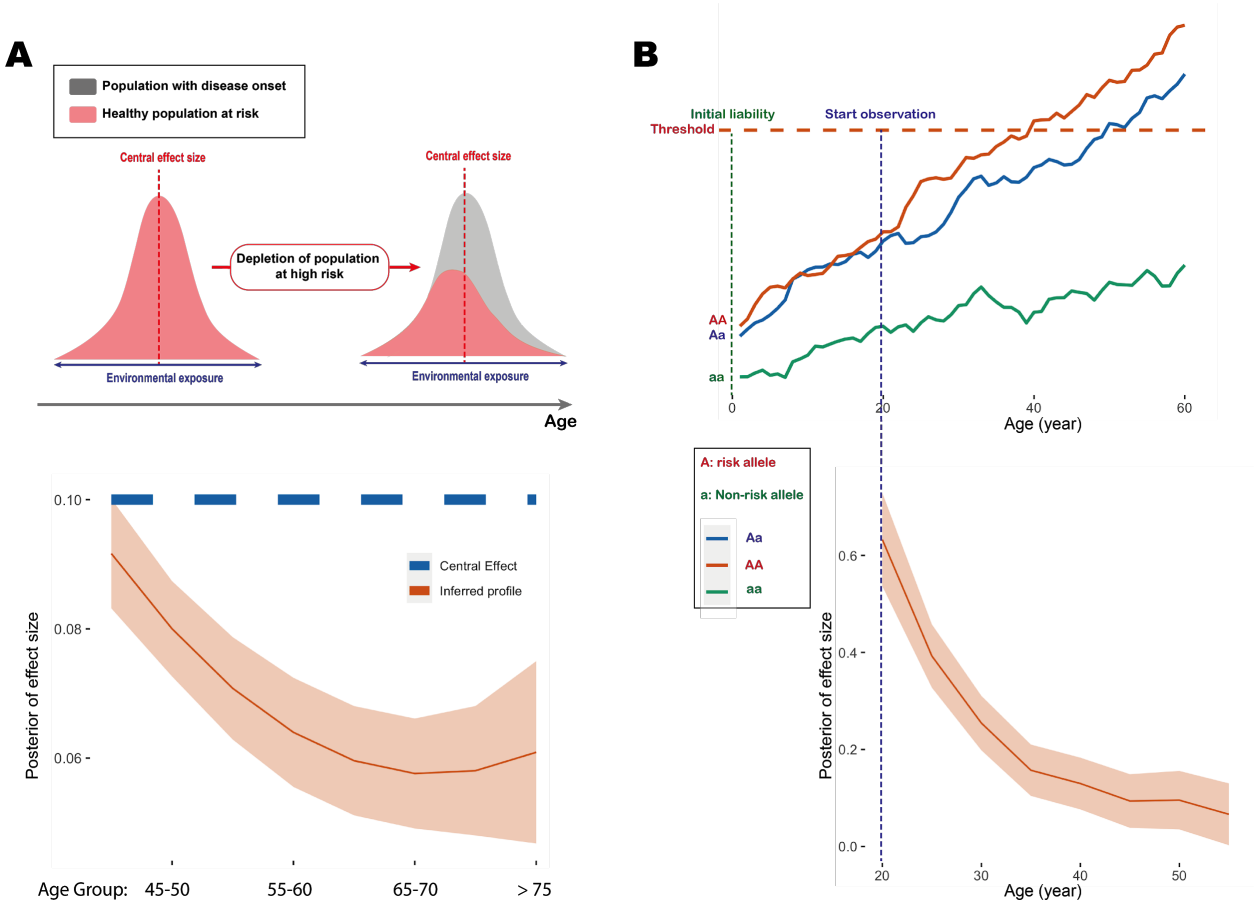

Figure 5. Possible models for decreasing genetic risk with age. A) Different individual contexts including environmental exposure could interact with genetic risk factors to create a distribution of effect size. Individuals at higher risk enter disease earlier, diluting the effect size estimation at a later age. The lower panel shows an simulation results using realistic parameters from UK Biobank. B) A threshold model when each individual has a disease "liability" which evolves over age. Disease onset occurs when liability crosses a threshold. Upper panel shows example trajectories where genetic risk alters only the liability baseline. Lower panel shows an estimation of the effect size from a simulated dataset of UK Biobank sample size. Details of simulations are provided in the Methods.

References

Aalen, O. O. 1988. “Heterogeneity in Survival Analysis.” *Statistics in Medicine* 7 (11): 1121–37.

Aschard, H., J. Chen, M. C. Cornelis, and L. B. Chibnik. 2012. “Inclusion of Gene-Gene and Gene-Environment Interactions Unlikely to Dramatically Improve Risk Prediction for Complex Diseases.” *The American Journal of*. https://www.sciencedirect.com/science/article/pii/S0002929712002509.

Bønnelykke, Klaus, and Carole Ober. 2016. “Leveraging Gene-Environment Interactions and Endotypes for Asthma Gene Discovery.” *The Journal of Allergy and Clinical Immunology* 137 (3): 667–79.

Bycroft, Clare, Colin Freeman, Desislava Petkova, Gavin Band, Lloyd T. Elliott, Kevin Sharp, Allan Motyer, et al. 2018. “The UK Biobank Resource with Deep Phenotyping and Genomic Data.” *Nature* 562 (7726): 203–9.

Choi, Shing Wan, Timothy Shin Heng Mak, and Paul F. O’Reilly. 2020. “A Guide to Performing Polygenic Risk Score Analyses.” https://doi.org/10.1101/416545.

Cordell, Heather J. 2009. “Detecting Gene–gene Interactions That Underlie Human Diseases.” *Nature Reviews. Genetics* 10 (6): 392–404.

Corominas, Roser, Xinping Yang, Guan Ning Lin, Shuli Kang, Yun Shen, Lila Ghamsari, Martin Broly, et al. 2014. “Protein Interaction Network of Alternatively Spliced Isoforms from Brain Links Genetic Risk Factors for Autism.” *Nature Communications* 5 (April): 3650.

Cortes, Adrian, Patrick K. Albers, Calliope A. Dendrou, Lars Fugger, and Gil McVean. 2020. “Identifying Cross-Disease Components of Genetic Risk across Hospital Data in the UK Biobank.” *Nature Genetics* 52 (1): 126–34.

Cortes, Adrian, Calliope A. Dendrou, Allan Motyer, Luke Jostins, Damjan Vukcevic, Alexander Dilthey, Peter Donnelly, Stephen Leslie, Lars Fugger, and Gil McVean. 2017. “Bayesian Analysis of Genetic Association across Tree-Structured Routine Healthcare Data in the UK Biobank.” *Nature Genetics* 49 (9): 1311–18.

Cotsapas, Chris, Benjamin F. Voight, Elizabeth Rossin, Kasper Lage, Benjamin M. Neale, Chris Wallace, Gonçalo R. Abecasis, et al. 2011. “Pervasive Sharing of Genetic Effects in Autoimmune Disease.” *PLoS Genetics* 7 (8): e1002254.

Duggirala, R., J. T. Williams, S. Williams-Blangero, and J. Blangero. 1997. “A Variance Component Approach to Dichotomous Trait Linkage Analysis Using a Threshold Model.” *Genetic Epidemiology* 14 (6): 987–92.

Dumitrescu, Logan, Kristin Brown-Gentry, Robert Goodloe, Kimberly Glenn, Wenjian Yang, Nancy Kornegay, Ching-Hon Pui, Mary V. Relling, and Dana C. Crawford. 2011. “Evidence for Age as a Modifier of Genetic Associations for Lipid Levels.” *Annals of Human Genetics* 75 (5): 589–97.

Favé, Marie-Julie, Fabien C. Lamaze, David Soave, Alan Hodgkinson, Héloïse Gauvin, Vanessa Bruat, Jean-Christophe Grenier, et al. 2018. “Gene-by-Environment Interactions in Urban Populations Modulate Risk Phenotypes.” *Nature Communications* 9 (1): 827.

Finkelstein, D. M. 1986. “A Proportional Hazards Model for Interval-Censored Failure Time Data.” *Biometrics* 42 (4): 845–54.

Gandal, Michael J., Virpi Leppa, Hyejung Won, Neelroop N. Parikshak, and Daniel H. Geschwind. 2016. “The Road to Precision Psychiatry: Translating Genetics into Disease Mechanisms.” *Nature Neuroscience* 19 (11): 1397–1407.

Govindarajulu, Usha S., Haiqun Lin, Kathryn L. Lunetta, and R. B. D’Agostino Sr. 2011. “Frailty Models: Applications to Biomedical and Genetic Studies.” *Statistics in Medicine* 30 (22): 2754–64.

Harbo, Hanne F., Noriko Isobe, Pål Berg-Hansen, Steffan D. Bos, Stacy J. Caillier, Marte W. Gustavsen, Inger-Lise Mero, et al. 2014. “Oligoclonal Bands and Age at Onset Correlate

with Genetic Risk Score in Multiple Sclerosis." *Multiple Sclerosis* 20 (6): 660–68.
Ide, Akane, Eiji Kawasaki, Norio Abiru, Fuyan Sun, Ryoko Takahashi, Hironaga Kuwahara, Naruhiro Fujita, et al. 2002. "Genetic Association between Interleukin-10 Gene Promoter Region Polymorphisms and Type 1 Diabetes Age-at-Onset." *Human Immunology* 63 (8): 690–95.
Jostins, Luke, and Jeffrey C. Barrett. 2011. "Genetic Risk Prediction in Complex Disease." *Human Molecular Genetics* 20 (R2): R182–88.
Khera, Amit V., Mark Chaffin, Krishna G. Aragam, Mary E. Haas, Carolina Roselli, Seung Hoan Choi, Pradeep Natarajan, et al. 2018. "Genome-Wide Polygenic Scores for Common Diseases Identify Individuals with Risk Equivalent to Monogenic Mutations." *Nature Genetics* 50 (9): 1219–24.
Kraft, Peter, Yu-Chun Yen, Daniel O. Stram, John Morrison, and W. James Gauderman. 2007. "Exploiting Gene-Environment Interaction to Detect Genetic Associations." *Human Heredity* 63 (2): 111–19.
Lasky-Su, Jessica, Helen N. Lyon, Valur Emilsson, Iris M. Heid, Cliona Molony, Benjamin A. Raby, Ross Lazarus, et al. 2008. "On the Replication of Genetic Associations: Timing Can Be Everything!" *American Journal of Human Genetics* 82 (4): 849–58.
Lin, D. Y., B. M. Psaty, and R. A. Kronmal. 1998. "Assessing the Sensitivity of Regression Results to Unmeasured Confounders in Observational Studies." *Biometrics*. https://www.jstor.org/stable/2533848?casa_token=9TDZ28lTPWsAAAAA:-wKtbGHa4A7Oy8faOleODACmOWZUh-yWVzr7Eozy5-7d-iWJYbSPY2LcxiIcn6DqXRm_0DA2QtFnAMPjpaJd8SMjNoGuH7AixnWmyPAPfR_ub3arL_ft.
Manolio, Teri A. 2013. "Bringing Genome-Wide Association Findings into Clinical Use." *Nature Reviews. Genetics* 14 (8): 549–58.
Mavaddat, Nasim, Paul D. P. Pharoah, Kyriaki Michailidou, Jonathan Tyrer, Mark N. Brook, Manjeet K. Bolla, Qin Wang, et al. 2015. "Prediction of Breast Cancer Risk Based on Profiling with Common Genetic Variants." *Journal of the National Cancer Institute* 107 (5). https://doi.org/10.1093/jnci/djv036.
Miguel-Yanes, Jose M. de, Peter Shrader, Michael J. Pencina, Caroline S. Fox, Alisa K. Manning, Richard W. Grant, Josèe Dupuis, et al. 2011. "Genetic Risk Reclassification for Type 2 Diabetes by Age below or above 50 Years Using 40 Type 2 Diabetes Risk Single Nucleotide Polymorphisms." *Diabetes Care* 34 (1): 121–25.
Mosley, Jonathan D., Deepak K. Gupta, Jingyi Tan, Jie Yao, Quinn S. Wells, Christian M. Shaffer, Suman Kundu, et al. 2020. "Predictive Accuracy of a Polygenic Risk Score Compared With a Clinical Risk Score for Incident Coronary Heart Disease." *JAMA: The Journal of the American Medical Association* 323 (7): 627–35.
Moutsianas, Loukas, Luke Jostins, Ashley H. Beecham, Alexander T. Dilthey, Dionysia K. Xifara, Maria Ban, Tejas S. Shah, et al. 2015. "Class II HLA Interactions Modulate Genetic Risk for Multiple Sclerosis." *Nature Genetics* 47 (10): 1107–13.
Mühlenbruch, Kristin, Charlotte Jeppesen, Hans-Georg Joost, Heiner Boeing, and Matthias B. Schulze. 2013. "The Value of Genetic Information for Diabetes Risk Prediction--Differences according to Sex, Age, Family History and Obesity." *PloS One* 8 (5). https://journals.plos.org/plosone/article/file?type=printable&id=10.1371/journal.pone.0064307.
Nalls, Mike A., Valentina Escott-Price, Nigel M. Williams, Steven Lubbe, Margaux F. Keller, Huw R. Morris, Andrew B. Singleton, and International Parkinson's Disease Genomics Consortium (IPDGC). 2015. "Genetic Risk and Age in Parkinson's Disease: Continuum Not Stratum." *Movement Disorders: Official Journal of the Movement Disorder Society* 30 (6): 850–54.
Ober, Carole, Dagan A. Loisel, and Yoav Gilad. 2008. "Sex-Specific Genetic Architecture of

Human Disease." *Nature Reviews. Genetics* 9 (12): 911–22.
Ripatti, Samuli, Emmi Tikkanen, Marju Orho-Melander, Aki S. Havulinna, Kaisa Silander, Amitabh Sharma, Candace Guiducci, et al. 2010. "A Multilocus Genetic Risk Score for Coronary Heart Disease: Case-Control and Prospective Cohort Analyses." *The Lancet* 376 (9750): 1393–1400.
Rossouw, Jacques E. 2002. "Hormones, Genetic Factors, and Gender Differences in Cardiovascular Disease." *Cardiovascular Research* 53 (3): 550–57.
Shi, Gang, Chi C. Gu, Aldi T. Kraja, Donna K. Arnett, Richard H. Myers, James S. Pankow, Steven C. Hunt, and Dabeeru C. Rao. 2009. "Genetic Effect on Blood Pressure Is Modulated by Age: The Hypertension Genetic Epidemiology Network Study." *Hypertension* 53 (1): 35–41.
Simino, Jeannette, Gang Shi, Joshua C. Bis, Daniel I. Chasman, Georg B. Ehret, Xiangjun Gu, Xiuqing Guo, et al. 2014. "Gene-Age Interactions in Blood Pressure Regulation: A Large-Scale Investigation with the CHARGE, Global BPgen, and ICBP Consortia." *American Journal of Human Genetics* 95 (1): 24–38.
Stranger, Barbara E., Lori E. Brigham, Richard Hasz, Marcus Hunter, Christopher Johns, Mark Johnson, Gene Kopen, et al. 2017. "Enhancing GTEx by Bridging the Gaps between Genotype, Gene Expression, and Disease." *Nature Genetics* 49 (12): 1664.
Thomas, Duncan. 2010. "Gene–environment-Wide Association Studies: Emerging Approaches." *Nature Reviews. Genetics* 11 (4): 259–72.
Vaupel, J. W., K. G. Manton, and E. Stallard. 1979. "The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality." *Demography* 16 (3): 439–54.
Wollmer, M. Axel, Johannes R. Streffer, Dieter Lütjohann, Magdalini Tsolaki, Vassiliki Iakovidou, Thomas Hegi, Thomas Pasch, et al. 2003. "ABCA1 Modulates CSF Cholesterol Levels and Influences the Age at Onset of Alzheimer's Disease." *Neurobiology of Aging* 24 (3): 421–26.