Polygenic background modifies penetrance of monogenic variants conferring risk for coronary artery disease, breast cancer, or colorectal cancer

Akl C. Fahed, M.D., M.P.H.,[*,1,2] Minxian Wang, Ph.D.,[*,2] Julian R. Homburger, Ph.D.,[*,3] Aniruddh P. Patel, M.D.,[1,2] Alexander G. Bick, M.D., Ph.D.,[1,2] Cynthia L. Neben, Ph.D.,[3] Carmen Lai, B.S.,[3] Deanna Brockman, M.S.,[1,2] Anthony Philippakis, M.D.,[2] Ph.D., Patrick T. Ellinor, M.D., Ph.D.,[1,2] Christopher A. Cassa, Ph.D.,[4] Matthew Lebo, Ph.D.,[5] Kenney Ng, Ph.D.,[8] Eric S. Lander, Ph.D.,[2,6,7] Alicia Y. Zhou, Ph.D.,[3] Sekar Kathiresan, M.D.,[9] Amit V. Khera, M.D., M.Sc[1,2]

[1] Center of Genomic Medicine and Division of Cardiology, Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts
[2] Cardiovascular Disease Initiative, Broad Institute of MIT and Harvard, Cambridge, Massachusetts
[3] Color Genomics, Burlingame, California
[4] Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts
[5] Laboratory for Molecular Medicine, Partners HealthCare Personalized Medicine, Boston, Massachusetts
[6] Department of Biology, MIT, Cambridge, Massachusetts
[7] Department of Systems Biology, Harvard Medical School, Boston, Massachusetts
[8] Center for Computational Health, IBM Research, Cambridge, Massachusetts
[9] Verve Therapeutics, Cambridge, Massachusetts

[*] Drs. Fahed, Wang, and Homburger contributed equally

Corresponding author:
Amit V. Khera, MD MSc
Center for Genomic Medicine
Massachusetts General Hospital
185 Cambridge Street | CPZN 6.256
Boston, MA 02114
Tel: 617-643-3388
E-mail: avkhera@mgh.harvard.edu

ABSTRACT

Background: Genetic variation can predispose to disease both through (i) monogenic risk variants in specific genes that disrupt a specific physiologic pathway and have a large effect on disease risk and (ii) polygenic risk that involves large numbers of variants of small effect that affect many different pathways. Few studies have explored the interaction between monogenic risk variants and polygenic risk.

Methods: We identified monogenic risk variants and calculated polygenic scores for three diseases, coronary artery disease, breast cancer, and colorectal cancer, in three study populations — case-control cohorts for coronary artery disease (UK Biobank; N=12,879) and breast cancer (Color Genomics; N=19,264), and an independent cohort of 49,738 additional UK Biobank participants.

Results: In the coronary artery disease case-control cohort, increased risk for carriers of a monogenic variant ranged from 1.3-fold for those in the lowest polygenic score quintile to 12.6-fold for those in the highest. For breast cancer, increased risk ranged from 2.4 to 6.9-fold across polygenic score quintiles. Among the 49,738 UK Biobank participants who carried a monogenic risk variant, the probability of disease at age 75 years was strongly modified by polygenic risk. Across individuals in the lowest to highest percentiles of polygenic risk, the probability of disease ranged from 17% to 78% for coronary artery disease; 13% to 76% for breast cancer; and 11% to 80% for colon cancer.

Conclusions: For three important genomic conditions, polygenic risk powerfully modifies the risk conferred by monogenic risk variants.

Background

For a range of common heritable diseases, a small subset of the population inherits a rare monogenic variant that causes a large increase in disease risk by disrupting a specific physiological pathway. More recently, polygenic scores have been developed that integrate the effects of many common genetic variants on disease risk. While the common variants have small individual effects (in the range of a few percent) on disease risk, they can cumulatively have large effects—producing, in some individuals, risks equivalent to the strong monogenic variants.[1–4]

A key open question is how monogenic and polygenic risk interact: Can disease risk from a monogenic variant that causes major disruption to a specific pathway be meaningfully modified by polygenic risk factors that involve small perturbations to a wide range of cellular pathways? Taking familial hypercholesterolemia as an example, monogenic variants predispose to premature coronary artery disease through major dysregulation of clearance of LDL cholesterol from the circulation. By contrast, only a small minority (~20%) of common DNA variants that predispose to coronary artery disease operate via cholesterol-related pathways, with the remainder affecting non-cholesterol related pathways (such as inflammation, cellular proliferation, vascular tone) and many additional pathways yet to be discovered.[5] Similarly, pathway analyses indicate that less than 20% of common variants linked to breast or colorectal cancer affect genes involved in DNA repair pathways, which are implicated in monogenic hereditary breast and ovarian cancer or Lynch syndrome.[3,6]

We explored this question by performing genetic analysis for three diseases—familial hypercholesterolemia, hereditary breast and ovarian cancer, and Lynch syndrome predisposing to colorectal cancer—for which about 1% of the population inherits a monogenic variant having a strong effect on disease risk.[7,8]

Although such variants are associated with several-fold increased risk of disease, it has long been recognized that the variants have incomplete penetrance and variable expressivity. For example, in one U.S. healthcare system, more than 50% of familial hypercholesterolemia variant carriers and more than 70% of female hereditary breast and ovarian cancer variant carriers remained free of disease well into middle age.[9,10]

Polygenic risk might help explain variation in outcomes for carriers of monogenic variants. If so, it could have important fundamental implications both for understanding disease physiology and for genetic counseling.

Here, we studied in 81,881 individuals to examine whether polygenic risk can account for variation in outcomes for carriers of monogenic variants—a topic that has both scientific implications about disease physiology and clinical implications for genetic counseling.

Methods:

Study populations

We studied three diseases: coronary artery disease, breast cancer, and colorectal cancer. For coronary artery disease, we designed and studied a case-control cohort (6,449 cases and 6,430 controls) derived from the UK Biobank, a prospective cohort study that enrolled middle-aged adult participants between 2006 and 2010.[11] We followed up these results in an independent cohort of 49,738 participants of the UK Biobank who previously underwent exome sequencing.[12] For breast cancer, we studied an all female case-control cohort (1,920 cases and 17,344 controls) from a commercial testing laboratory (Color Genomics; Burlingame, CA), and then followed up the results in 27,144 female UK Biobank participants. For colorectal cancer, we studied 49,738 UK Biobank participants.

For participants in the UK Biobank, cases of coronary artery disease, breast cancer, and colorectal cancer were identified based on a combination of self-reported data confirmed by trained healthcare professionals, hospitalization records, and national procedural, cancer, and death registries. For participants in the commercial testing cohort, breast cancer status was defined based on self-report at the time of enrollment.[13] Additional details on study cohorts, genetic testing, and disease ascertainment are provided in the Supplementary Appendix.

Monogenic variant classification

For each of three conditions, observed variants were classified as pathogenic or likely pathogenic by a laboratory geneticist according to current clinical standards blinded to any phenotype data.[14] We analyzed three genes causal for familial hypercholesterolemia (*LDLR*, *APOB,* and *PCSK9*), two for hereditary breast and ovarian cancer syndrome (*BRCA1* and *BRCA2*), and four for Lynch syndrome (*MLH1*, *MSH2*, *MSH6,* and *PMS2)*.

Polygenic background

To quantify the contribution of polygenic background to disease, we calculated previously validated polygenic scores for each of the three diseases.[1–3] Ancestry-corrected reference distributions for each score were generated, as described previously (Fig. S1 in the Supplementary Appendix).[15]

Statistical Analysis

In the coronary artery disease and breast cancer case-control studies, logistic regression was used to determine risk among participants stratified by monogenic variant status and polygenic score. Within the cohort of 49,738 UK Biobank participants, we modeled the odds ratio for disease according to monogenic variant status and polygenic score percentiles. A second analysis estimated probability of disease by age 75 years using Cox regression, standardized to the mean of all covariates in each population. All regression models were adjusted for age, sex, and genetic ancestry — as quantified by the first four principal components — except for breast cancer analyses, which were restricted to females.[16]

Results

Polygenic background modifies risk of coronary artery disease conferred by familial hypercholesterolemia variants

We first studied the interaction of monogenic risk variants and polygenic scores in coronary artery disease. To identify individuals with monogenic variants causal for familial hypercholesterolemia, we sequenced the three genes related to the condition — *LDLR, APOB,* and *PCSK9* — in 6,449 coronary artery disease cases and 6,430 controls derived from the UK Biobank (Table 1). Each of the observed genetic variants was reviewed by a laboratory geneticist blinded to any phenotype data. A total of 28 distinct genetic variants were classified as pathogenic or likely pathogenic; they were present in 43 (0.67%) cases and 13 (0.20%) controls. The presence of a familial hypercholesterolemia variant conferred a 3.17-fold increased risk of coronary artery disease (95% confidence interval [CI] 1.70 to 5.92).

We next examined the effect of participants' polygenic background on risk of coronary artery disease, by computing a previously validated polygenic score in all cases and controls. Even among carriers of a familial hypercholesterolemia variant, the observed risk varied substantially according to the polygenic score. We classified individuals as having low polygenic score (bottom quintile), intermediate polygenic score (quintiles 2-4), or high polygenic score (top quintile). Compared to individuals who did not carry a pathogenic mutation, the risk among mutation carriers ranged from 1.30-fold (95% CI 0.39 to 4.31) for those in the lowest quintile of the polygenic score distribution to 12.59 (95% CI 2.96 to 53.53) in the highest quintile (Figure 1A).

We next examined an independent cohort of 49,738 UK Biobank participants, after confirming that monogenic variants and the polygenic score were both associated with coronary artery disease as expected (Tables S1-2 in the Supplementary Appendix). A laboratory geneticist identified a familial hypercholesterolemia variant in 131 (0.26%) of these participants, which conferred a 5.02-fold (95%CI 2.97 to 8.46) increased risk of coronary artery disease at

the time of enrollment. The polygenic score for coronary artery disease was normally distributed in the population and strongly associated with disease — odds ratio per standard deviation increment of 1.64 (95%CI 1.57 to 1.72). Within a logistic regression model, the relationship between the polygenic score and prevalent disease conformed to a linear model (Fig. S2 and Table S3 in the Supplementary Appendix).

Joint modeling of monogenic variant carrier status and polygenic score indicate substantial gradients in risk of coronary artery disease according to inherited DNA variation that can be assessed from the time of birth. Odds ratio for coronary artery disease in monogenic variant carriers — as compared to noncarriers with average polygenic score — ranged from 1.59 to 21.24 across percentiles of the polygenic score (Figure 1B). Modeling the probability of disease by age 75 years indicated striking gradients in risk, ranging from 4.8% for individuals who were both noncarriers and in the lowest percentile of the polygenic score to 77.7% for individuals with a monogenic risk variant who were also in the highest polygenic score percentile (Figure 1C).

Polygenic background modifies risk of breast cancer conferred by familial hereditary breast and ovarian cancer variants

We next set out to apply the same analysis to breast cancer. We first identified monogenic risk variants by sequencing the *BRCA1* and *BRCA2* genes in 1920 breast cancer cases and 17,344 controls, all female, from the Color Genomics commercial testing laboratory (Table 2). A pathogenic or likely pathogenic variant was identified in 176 (9.1%) cases and 674 (3.9%) controls, corresponding to a 3.48-fold (95% CI 2.81 to 4.21) increased risk of breast cancer in variant carriers. We then calculated polygenic risk for breast cancer, using a previously validated polygenic score. As we saw for coronary artery disease, breast cancer risk was strongly affected by polygenic background. Compared to noncarriers with intermediate polygenic score, increased risk among carriers ranged from 2.40-fold (95% CI 1.58 to 3.65) for

those in the lowest quintile of the polygenic score distribution to 6.85-fold (95% CI 4.71 to 9.96) in the highest quintile (Figure 2B).

We extended these results into the UK Biobank participants, this time focusing only on the 27,144 female participants. A laboratory geneticist reviewed all observed genetic variants in the *BRCA1* and *BRCA2* genes, identifying 116 carriers of pathogenic or likely pathogenic variants. These variants conferred a 4.45-fold increased risk of breast cancer (95%CI 2.75 to 7.20). The polygenic score was associated with an odds ratio per standard deviation increment of 1.61 (95%CI 1.52 to 1.70), with the relationship to breast cancer again conforming to a linear model (Fig. S2 and Table S3 in the Supplementary Appendix).

Joint modeling of both monogenic variant status and polygenic score indicated that the risk for breast cancer among carriers of a *BRCA1* or *BRCA2* variant ranges from 1.43 to 16.41 increased risk across percentile of the polygenic score. When modeled as probability of disease by age 75 years, risk among monogenic variant carriers ranged from 12.8 to 75.5% and risk among noncarriers ranged from 3.4 to 29.7%.

Risk of colorectal cancer varies according to Lynch syndrome monogenic variant carrier status and polygenic score

We explored a third disease, colorectal cancer, in the same set of 49,738 UK Biobank participants used above. A pathogenic or likely pathogenic Lynch syndrome variant was identified in 76 (0.15%) individuals, conferring an odds ratio for colorectal cancer of 27.69 (95%CI 14.27 to 53.72). The odds ratio per standard deviation increment in the colorectal cancer polygenic score was 1.66 (95%CI 1.48 to 1.85). Joint modeling of monogenic variants and the polygenic score — using noncarriers with average polygenic score as the reference group — noted odds ratios ranging from 8.34 to 117.53 for carriers of monogenic variants and 0.27 to 3.78 for noncarriers (Figure 3A). Absolute risk of colorectal cancer by age 75 years ranged from 11.6 to 79.5% for carriers and 0.7 to 8.7% for noncarriers (Figure 3B).

Discussion:

Our analysis of the interaction between monogenic risk variants and polygenic background — for three important genomic conditions: familial hypercholesterolemia, hereditary breast and ovarian cancer syndrome, and Lynch syndrome — has two important implications.

First, we were surprised that risk conferred by monogenic risk variants, which act by perturbing a specific molecular pathway, can be so substantially modified by polygenic background, which appears to act by affecting a diverse set of physiological processes. From a physiological standpoint, it is not clear why the major disruptions caused by monogenic variants can be offset by other factors. Yet, the risk for monogenic variant carriers with the lowest polygenic risk scores approached the population average. Understanding the physiological basis for these interactions may suggest therapeutic strategies for monogenic variant carriers in general.

Second, our findings indicate that accounting for polygenic background is likely to increase the accuracy of risk estimation for individuals who inherit a monogenic risk variant. An important example is the decision about whether to undergo prophylactic mastectomy (as opposed to serial imaging) faced by carriers of a *BRCA1* variant.[17] At present, up to 50% of women opt for prophylactic mastectomy, but rates are highly variable.[18,19] Here, we find a broad spectrum of risk across percentiles of the polygenic score that may better inform shared decision making — odds ratios compared to noncarriers ranging from 1.4 to 16.4 and an absolute risk at age 75 years ranging from 13 to 76%. Similarly, refined risk estimates may improve decisions about the timing and intensity of lipid-lowering therapy for individuals with familial hypercholesterolemia and whether to undergo serial colonoscopies or prophylactic colectomy for individuals with Lynch syndrome variants.

In a clinical setting, assessing both monogenic risk variants and polygenic risk requires high-coverage sequencing of genes associated with monogenic risk and an approach to assay common variants across the genome. At present, this can be accomplished by various

approaches, including (i) high-coverage whole genome sequencing,[15] (ii) whole exome sequencing and genotyping array, or (iii) targeted high-coverage sequencing of individuals genes and with low-coverage sequencing across the genome.[20] Ongoing efforts to improve the cost and accessibility of these technologies will improve the feasibility of incorporating this information into routine clinical practice.

From a statistical standpoint, our data indicate that polygenic risk contributes to the incomplete penetrance of the monogenic risk variants. Additionally, our data suggest a roughly additive interaction between monogenic risk variants and polygenic risk, although the statistical power to detect non-additive interaction was limited by the number of individuals carrying monogenic risk variants. Within the cohort of 49,738 UK Biobank participants, the p-values for interaction were 0.07, 0.53, and 0.49 for coronary artery disease, breast cancer, and colorectal cancer, respectively.

Our results should be interpreted in the context of potential limitations. First, UK Biobank participants tend to be healthier than the general population. Disease risk models should be calibrated for a given target population prior to clinical use.[21] Second, our analysis focused only on the role of the monogenic variants for the three conditions studied; for example, it did not consider the risk of the breast and ovarian cancer variants on ovarian cancer.[22] Third, we aggregated together all pathogenic and likely pathogenic variants for each monogenic condition; however; there may be heterogeneity among these variants.[22–24] Fourth, further efforts are needed to understand how best to disclose integrated genomic risk assessments to patients and treating clinicians and how to integrate them with existing predictors based on nongenetic factors.[25–27]

Finally, we highlight an important equity issue. The fact that our knowledge concerning the monogenic risk variants and the development of the polygenic scores has been based primarily on patients of European ancestry affects the utility for patients of other ancestries.[15,28–30] In particular, the polygenic scores are known to be less precise for other ancestry groups.[30] It is

important for the biomedical community to invest in the development of more diverse population allele frequency databases,[31] disease association studies in other ancestral backgrounds, new computational algorithms that better account for ancestral background,[32] and new technology or machine learning algorithms to enable unbiased high-throughput functional assessments of variants.[33,34]

References:

1. Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet 2018;
2. Mavaddat N, Michailidou K, Dennis J, et al. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. Am J Hum Genet 2019;104(1):21–34.
3. Huyghe JR, Bien SA, Harrison TA, et al. Discovery of common and rare genetic risk variants for colorectal cancer. Nat Genet 2019;51(1):76–87.
4. Chatterjee N, Shi J, García-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. Nat Rev Genet 2016;17(7):392–406.
5. Khera AV, Kathiresan S. Genetics of coronary artery disease: discovery, biology and clinical translation. Nat Rev Genet 2017;18(6):331–44.
6. Michailidou K, Lindström S, Dennis J, et al. Association analysis identifies 65 new breast cancer risk loci. Nature 2017;551(7678):92–4.
7. Bowen MS, Kolor K, Dotson WD, Ned RM, Khoury MJ. Public health action in genomics Is now needed beyond newborn screening. Public Health Genomics 2012;15(6):327–34.
8. Dewey FE, Murray MF, Overton JD, et al. Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. Science 2016;354(6319):aaf6814.
9. Manickam K, Buchanan AH, Schwartz MLB, et al. Exome Sequencing–Based Screening for BRCA1/2 Expected Pathogenic Variants Among Adult Biobank Participants. JAMA Netw Open 2018;1(5):e182140–e182140.
10. Abul-Husn NS, Manickam K, Jones LK, et al. Genetic identification of familial hypercholesterolemia within a single U.S. health care system. Science 2016;354(6319):aaf7000.
11. Bycroft C, Freeman C, Petkova D, et al. Genome-wide genetic data on ~500,000 UK Biobank participants. bioRxiv 2017;

12. Hout CVV, Tachmazidou I, Backman JD, et al. Whole exome sequencing and characterization of coding variation in 49,960 individuals in the UK Biobank. bioRxiv 2019;572347.
13. Neben CL, Zimmer AD, Stedden W, et al. Multi-Gene panel testing of 23,179 individuals for hereditary cancer risk identifies pathogenic variant carriers missed by current genetic testing guidelines. J Mol Diagn 2019;21(4):646–57.
14. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med Off J Am Coll Med Genet 2015;17(5):405–24.
15. Khera AV, Chaffin M, Zekavat SM, et al. Whole-genome sequencing to characterize monogenic and polygenic contributions in patients hospitalized with early-onset myocardial infarction. Circulation 2019;139(13):1593–602.
16. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38(8):904–9.
17. Jolie A. Opinion | My Medical Choice by Angelina Jolie [Internet]. N. Y. Times. 2013 [cited 2019 Sep 30];Available from: https://www.nytimes.com/2013/05/14/opinion/my-medical-choice.html
18. the Hereditary Breast Cancer Clinical Study Group, Metcalfe K, Eisen A, et al. International trends in the uptake of cancer risk reduction strategies in women with a BRCA1 or BRCA2 mutation. Br J Cancer 2019;121(1):15–21.
19. Liede A, Mansfield CA, Metcalfe KA, et al. Preferences for breast cancer risk reduction among BRCA1/BRCA2 mutation carriers: a discrete-choice experiment. Breast Cancer Res Treat 2017;165(2):433–44.
20. Homburger JR, Neben CL, Mishne G, Zhou AY, Kathiresan S, Khera AV. Low coverage

whole genome sequencing enables accurate assessment of common variants and calculation of genome-wide polygenic scores. bioRxiv 2019;716977.

21. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of sociodemographic and health-related characteristics of UK biobank participants with those of the general population. Am J Epidemiol 2017;186(9):1026–34.
22. Antoniou A, Pharoah PDP, Narod S, et al. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case series unselected for family history: a combined analysis of 22 studies. Am J Hum Genet 2003;72(5):1117–30.
23. Khera AV, Won HH, Peloso GM, et al. Diagnostic yield and clinical utility of sequencing familial hypercholesterolemia genes in patients with severe hypercholesterolemia. J Am Coll Cardiol 2016;67(22):2578–89.
24. Bonadona V, Bonaïti B, Olschwang S, et al. Cancer risks associated with germline mutations in MLH1, MSH2, and MSH6 genes in Lynch syndrome. JAMA 2011;305(22):2304–10.
25. Kurian AW, Munoz DF, Rust P, et al. Online tool to guide decisions for BRCA1/2 mutation carriers. J Clin Oncol 2012;30(5):497–506.
26. Pérez de Isla Leopoldo, Alonso Rodrigo, Mata Nelva, et al. Predicting cardiovascular Events in familial hypercholesterolemia. Circulation 2017;135(22):2133–44.
27. Møller P, Seppälä T, Bernstein I, et al. Cancer incidence and survival in Lynch syndrome patients receiving colonoscopic and gynaecological surveillance: first report from the prospective Lynch syndrome database. Gut 2017;66(3):464–72.
28. Landry LG, Rehm HL. Association of racial/ethnic categories with the ability of genetic Tests to detect a cause of cardiomyopathy. JAMA Cardiol 2018;3(4):341–5.
29. Manrai AK, Funke BH, Rehm HL, et al. Genetic misdiagnoses and the potential for health disparities. N Engl J Med 2016;375(7):655–65.
30. 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(4):584–91.

31. Karczewski KJ, Francioli LC, Tiao G, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of- function intolerance across human protein-coding genes. bioRxiv 2019;
32. Marquez-Luna C, Loh PR, South Asian Type 2 Diabetes C, Sigma Type 2 Diabetes Consortium, Price AL. Multiethnic polygenic risk scores improve risk prediction in diverse populations. Genet Epidemiol 2017;41(8):811–23.
33. Findlay GM, Daza RM, Martin B, et al. Accurate classification of BRCA1 variants with saturation genome editing. Nature 2018;562(7726):217–22.
34. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, et al. Predicting splicing from primary sequence with deep learning. Cell 2019;176(3):535-548.e24.

Figure Legends:

Figure 1: Polygenic background and risk of coronary artery disease in familial hypercholesterolemia

In panel A, case-control participants for coronary artery disease in the UK Biobank (n=12,879) were stratified into three groups according to their polygenic score — low, intermediate, or high defined as the lowest quintile, the middle three quintiles, and the highest quintile of the polygenic score distribution respectively. For carriers and noncarriers of familial hypercholesterolemia in each polygenic score group, the odds ratio for coronary artery disease was calculated in a logistic regression model with age, sex, and the first four principal components of ancestry as covariates. Non-carriers with intermediate polygenic score served as the reference group. In panels B and C, for UK Biobank participants (n=49,738), shown are the predicted odds ratios of coronary artery disease (panel B) and predicted probability of coronary artery disease by age 75 years (panel C) in each percentile (dots) of the polygenic score distribution for carriers (blue) and noncarriers (brown) of familial hypercholesterolemia variants. For calculating the odds ratios, a logistic regression model with age, sex, and the first four principal components of ancestry was used, and the predicted odds ratios were calculated by referencing to the risk of average polygenic risk score and conditioning on the mean value of each covariate. The predicted probability of disease as a function of monogenic variant carrier status and polygenic score was modeled using Cox regression models with age, sex, and the first four principal components of ancestry as covariates. Prevalent cases at baseline enrollment in the UK Biobank and incident cases during follow-up were included (Table S2 in the Supplementary Appendix). The shaded area around the dots represents the 95% confidence interval. The horizontal dashed lines show the odds ratio or probability of disease for people with average polygenic risk score. FH — familial hypercholesterolemia

Figure 2: Polygenic background and risk of breast cancer in hereditary breast and ovarian cancer

In panel A, case-control participants for breast cancer in Color Genomics (n=19,264) were stratified into three groups according to their polygenic score — low, intermediate, or high defined as the lowest quintile, the middle three quintiles, and the highest quintile of the polygenic score distribution respectively. For carriers and noncarriers of hereditary breast and ovarian cancer variants in each polygenic score group, the odds ratio for breast cancer was calculated in a logistic regression model with age and the first four principal components of ancestry as covariates. Non-carriers with intermediate polygenic score served as the reference group. In panels B and C, for female UK Biobank participants (n=27,144), shown are the predicted odds ratios of breast cancer (panel B) and predicted probability of breast cancer by age 75 years (panel C) in each percentile (dots) of the polygenic score distribution for carriers (blue) and noncarriers (brown) of hereditary breast and ovarian cancer variants. For calculating the odds ratios, a logistic regression model with age and the first four principal components of ancestry was used, and the predicted odds ratios were calculated by referencing to the risk of average polygenic risk score and conditioning on the mean value of each covariate. The predicted probability of disease as a function of monogenic variant carrier status and polygenic score was modeled using Cox regression models with age and the first four principal components of ancestry as covariates. Prevalent cases at baseline enrollment in the UK Biobank and incident cases during follow-up were included (Table S2 in the Supplementary Appendix). The shaded area around the dots represents the 95% confidence interval. The horizontal dashed lines show the odds ratio or probability of disease for people with average polygenic risk score. HBOC — hereditary breast and ovarian cancer

Figure 3: Polygenic background and risk of colorectal cancer in Lynch syndrome

For UK Biobank participants (n=49,738), shown are the predicted odds ratios of colorectal cancer (panel A) and predicted probability of colorectal cancer by age 75 years (panel B) in each percentile (dots) of the polygenic score distribution for carriers (blue) and noncarriers (brown) of Lynch syndrome variants. For calculating the odds ratios, a logistic regression model with age, sex, and the first four principal components of ancestry was used, and the predicted odds ratios were calculated by referencing to the risk of average polygenic risk score and conditioning on the mean value of each covariate. The predicted probability of disease as a function of monogenic variant carrier status and polygenic score was modeled using Cox regression models with age, sex, and the first four principal components of ancestry as covariates. Prevalent cases at baseline enrollment in the UK Biobank and incident cases during follow-up were included (Table S2 in the Supplementary Appendix). The shaded area around the dots represents the 95% confidence interval. The horizontal dashed lines show the odds ratio or probability of disease for people with average polygenic risk score.

Tables:

Table 1: Baseline characteristics of coronary artery disease case-control study participants derived from the UK Biobank

	Cases with coronary artery disease (n=6449)	Controls (n=6430)
Age, mean (SD), y	58.8 (7.2)	58.7 (7.2)
Female sex, n (%)	2259 (35.0)	2239 (34.8)
Race, n (%)		
White	5980 (92.7)	6198 (96.4)
Black	72 (1.1)	58 (0.9)
Asian	257 (4.0)	90 (1.4)
Other	140 (2.2)	84 (1.3)
Hypertension, n (%)	4576 (71.0)	2013 (31.3)
Diabetes, n (%)	1559 (24.2)	356 (5.5)
Chronic kidney disease, n (%)	415 (6.4)	47 (0.7)
Current or former smoking, n (%)	4269 (66.7)	2868 (44.8)
Body mass index, mean (SD), kg/m^2	29.61 (5.4)	27.26 (4.5)
Family history of heart disease, n (%)	1996 (39.4)	1349 (26.6)

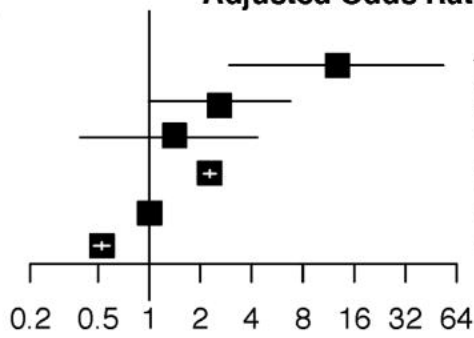
Table 2: Baseline characteristics of breast cancer case-control study participants derived from the Color Genomics commercial testing laboratory

	Cases with breast cancer (n=1920)	Controls (n=17344)
Age, mean (SD), y	57.4 (12.5)	45.9 (13.5)
Female sex, n (%)	1920 (100)	17344 (100)
Race, n (%)		
White	1375 (71.6)	12365 (71.3)
Black	30 (1.5)	410 (2.4)
Asian	83 (4.3)	695 (4.0)
Other	432 (22.6)	3874 (22.3)
Body mass index, mean (SD), kg/m^2	26.5 (6.7)	27.2 (6.8)
Family history of breast cancer, n (%)	855 (44.5)	7497 (43.2)

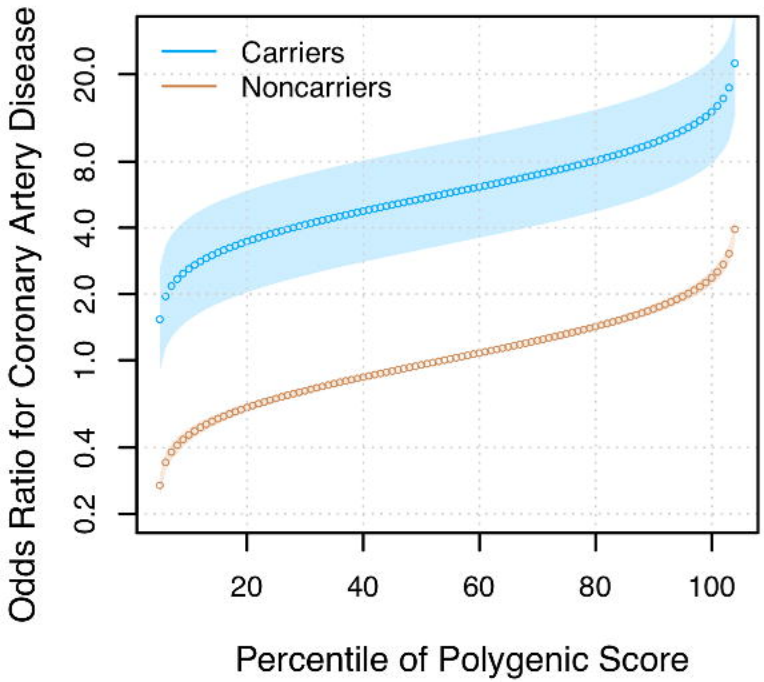
Figure 1. Polygenic background and risk of coronary artery disease in familial hypercholesterolemia

A. Risk of coronary artery disease by monogenic and polygenic risk strata (case-control cohort; N=12,879)

FH Variant	Polygenic Score	N of 6449 Cases	N of 6430 Controls	Adjusted Odds Ratio	95% CI	p–value
Carrier	High	23	2	12.59	[2.96; 53.53]	0.001
Carrier	Intermediate	14	6	2.59	[0.99; 6.77]	0.053
Carrier	Low	6	5	1.30	[0.39; 4.31]	0.67
Noncarrier	High	2352	1176	2.28	[2.09; 2.48]	<0.0001
Noncarrier	Intermediate	3412	3875	ref		
Noncarrier	Low	642	1366	0.53	[0.47; 0.59]	<0.0001



B. Odds ratio for coronary artery disease (UK biobank; N=49,738)



C. Probability of coronary artery disease (UK biobank; N=49,738)

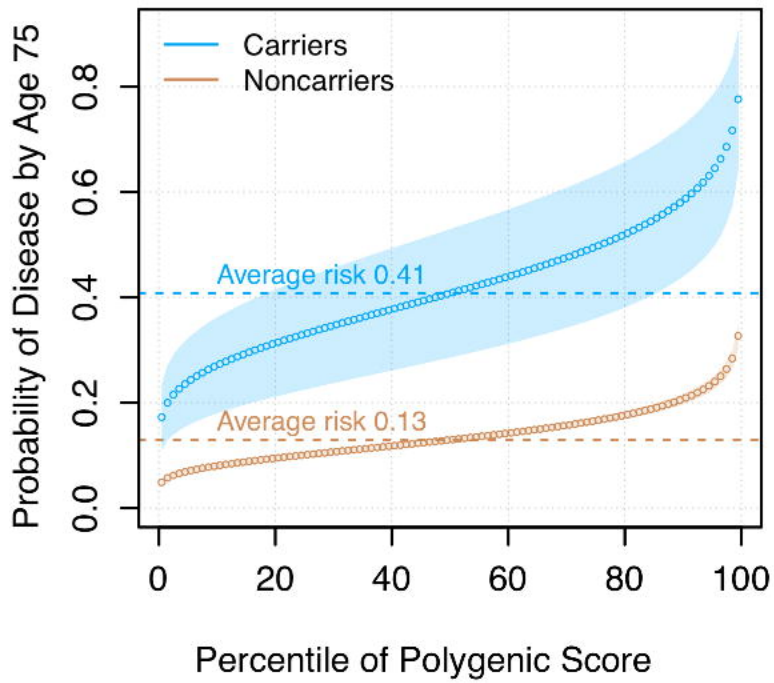
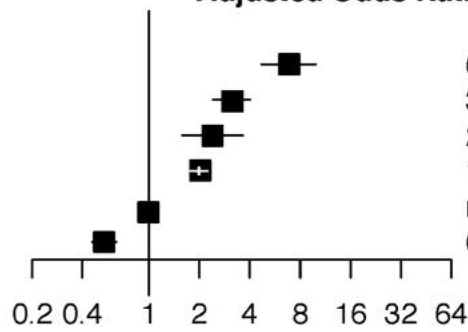


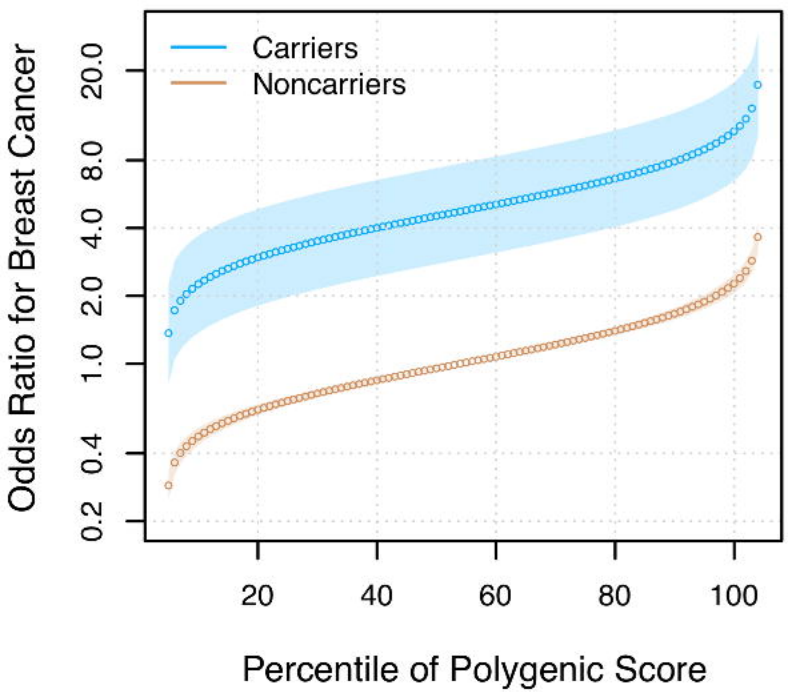
Figure 2. Polygenic background and risk of breast cancer in hereditary breast and ovarian cancer

A. Risk of breast cancer by monogenic and polygenic strata (case-control cohort; N=19,264)

HBOC Variant	Polygenic Score	N of 1920 Cases	N of 17344 Controls	Adjusted Odds Ratio	95% CI	p–Value
Carrier	High	48	125	6.85	[4.71; 9.96]	<0.0001
Carrier	Intermediate	94	393	3.12	[2.42; 4.03]	<0.0001
Carrier	Low	32	153	2.40	[1.58; 3.65]	<0.0001
Noncarrier	High	585	3095	1.99	[1.77; 2.24]	<0.0001
Noncarrier	Intermediate	975	10096	ref		
Noncarrier	Low	186	3482	0.54	[0.46; 0.64]	<0.0001



B. Odds ratio for breast cancer (UK biobank; N=49,738)



C. Probability of breast cancer (UK biobank; N=49,738)

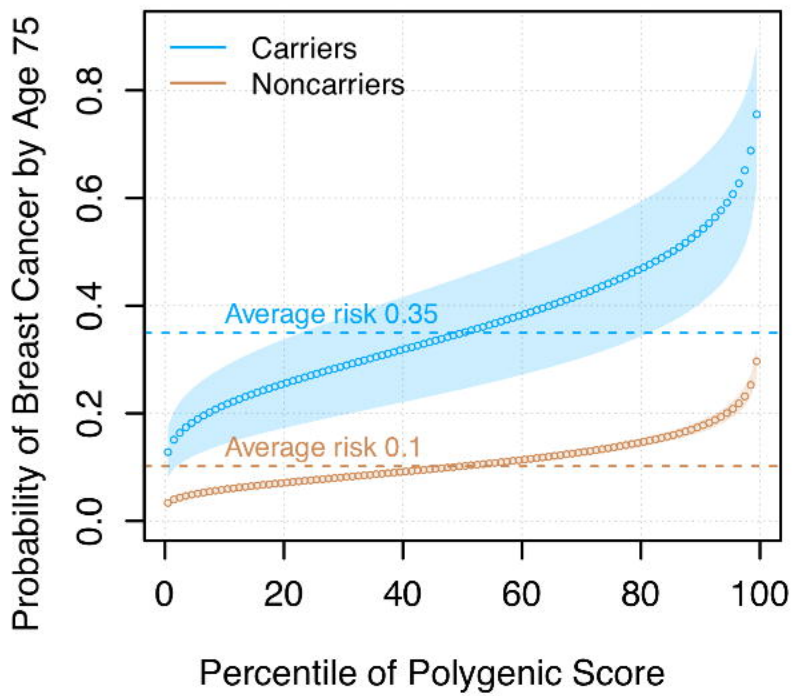
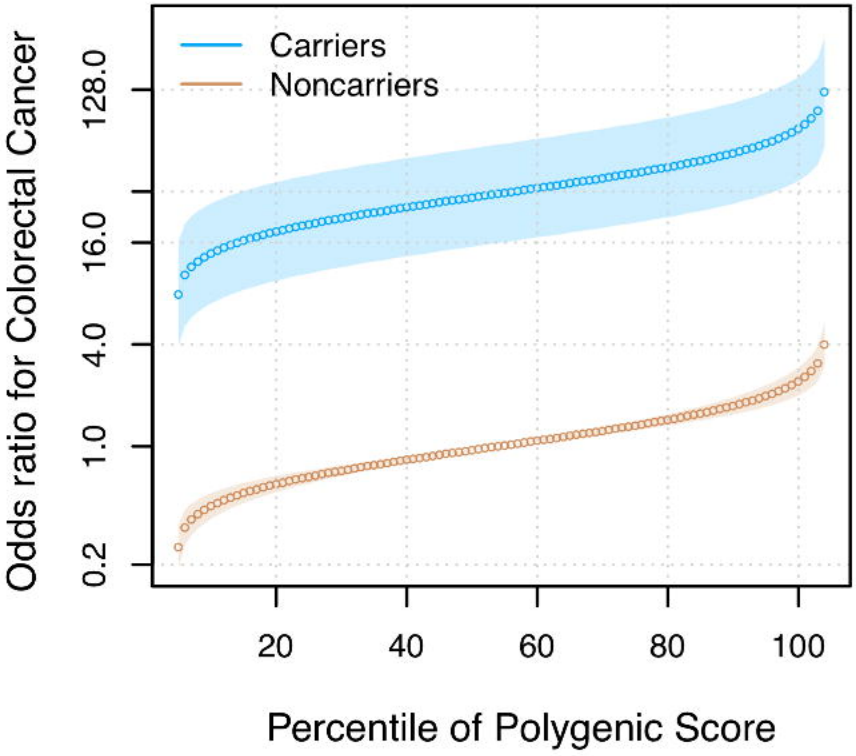


Figure 3. Polygenic background and risk of colorectal cancer in Lynch syndrome

A. Odds ratio for colorectal cancer (UK biobank; N=49,738)



B. Probability of colorectal cancer (UK biobank; N=49,738)