Assessing thyroid cancer risk using polygenic risk scores

Sandya Liyanarachchi[a,b,1], Julius Gudmundsson[c,1], Egil Ferkingstad[c], Huiling He[a,b], Jon G. Jonasson[d,e,f], Vinicius Tragante[c,g], Folkert W. Asselbergs[g,h,i,j,k], Li Xu[l,m], Lambertus A. Kiemeney[n], Romana T. Netea-Maier[o], Jose I. Mayordomo[p], Theo S. Plantinga[q], Hannes Hjartarson[r], Jon Hrafnkelsson[s], Erich M. Sturgis[l,m], Pamela Brock[t], Fadi Nabhan[u], Gudmar Thorleifsson[c], Matthew D. Ringel[u], Kari Stefansson[c,e,2,3], and Albert de la Chapelle[a,b,2,3]

[a]Human Cancer Genetics Program, Comprehensive Cancer Center, The Ohio State University, Columbus, OH 43210; [b]Department of Cancer Biology and Genetics, Comprehensive Cancer Center, The Ohio State University, Columbus, OH 43210; [c]Department of Population Genomics, deCODE genetics/Amgen Inc., 101 Reykjavik, Iceland; [d]Department of Pathology, Landspitali-University Hospital, 101 Reykjavik, Iceland; [e]Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland; [f]Department of Epidemiology, The Icelandic Cancer Registry, 105 Reykjavik, Iceland; [g]Department of Cardiology, Heart and Lung Division, University Medical Center Utrecht, University of Utrecht, 3584 CX Utrecht, The Netherlands; [h]Durrer Center for Cardiovascular Research, Netherlands Heart Institute, 3511 EP Utrecht, The Netherlands; [i]Institute of Cardiovascular Science, Faculty of Population Health Sciences, University College London, WC1E 6BT London, United Kingdom; [j]Farr Institute of Health Informatics Research, University College London, NW1 2DA London, United Kingdom; [k]Institute of Health Informatics, University College London, NW1 2DA London, United Kingdom; [l]Department of Head and Neck Surgery, The University of Texas MD Anderson Cancer Center, Houston, TX 77030; [m]Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030; [n]Radboud Institute for Health Sciences, Radboud University Medical Centre, 6500HB Nijmegen, The Netherlands; [o]Division of Endocrinology, Department of Internal Medicine, Radboud Institute for Health Sciences, Radboud University Medical Centre, 6500HB Nijmegen, The Netherlands; [p]UCHealth Diane O'Connor Thompson Breast Center, University of Colorado Hospital, Aurora, CO 80045; [q]Department of Pathology, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, 6500HB Nijmegen, The Netherlands; [r]Head&Neck Surgery Department, Landspitali-University Hospital, 101 Reykjavik, Iceland; [s]Department of Oncology, Landspitali-University Hospital, 101 Reykjavik, Iceland; [t]Department of Internal Medicine, Comprehensive Cancer Center, The Ohio State University, Columbus, OH 43210; and [u]Division of Endocrinology, Diabetes, and Metabolism, Department of Internal Medicine, The Ohio State University, Columbus, OH 43210

Contributed by Albert de la Chapelle, January 6, 2020 (sent for review November 20, 2019; reviewed by Carmelo Nucera and Michael Xing)

Genome-wide association studies (GWASs) have identified at least 10 single-nucleotide polymorphisms (SNPs) associated with papillary thyroid cancer (PTC) risk. Most of these SNPs are common variants with small to moderate effect sizes. Here we assessed the combined genetic effects of these variants on PTC risk by using summarized GWAS results to build polygenic risk score (PRS) models in three PTC study groups from Ohio (1,544 patients and 1,593 controls), Iceland (723 patients and 129,556 controls), and the United Kingdom (534 patients and 407,945 controls). A PRS based on the 10 established PTC SNPs showed a stronger predictive power compared with the clinical factors model, with a minimum increase of area under the receiver-operating curve of 5.4 percentage points ($P \leq 1.0 \times 10^{-9}$). Adding an extended PRS based on 592,475 common variants did not significantly improve the prediction power compared with the 10-SNP model, suggesting that most of the remaining undiscovered genetic risk in thyroid cancer is due to rare, moderate- to high-penetrance variants rather than to common low-penetrance variants. Based on the 10-SNP PRS, individuals in the top decile group of PRSs have a close to sevenfold greater risk (95% CI, 5.4–8.8) compared with the bottom decile group. In conclusion, PRSs based on a small number of common germline variants emphasize the importance of heritable low-penetrance markers in PTC.

thyroid cancer | GWAS | polygenic risk score | risk prediction

Recent advances in genetic and genomic research have led to the development of efficient methods to detect and evaluate diagnostic and prognostic factors in individual patients. In numerous monogenic and congenital disorders, diagnoses can be made based on the occurrence of germline variants. In contrast, polygenic and acquired disorders can be assessed by the study of somatic mutations in the appropriate cell or tissue. Typically, this applies to many cancers for which the study of genetic mutations in the tumor itself can be diagnostic, prognostic, or informative for choice of therapy (1–3). The present study investigates methods for thyroid cancer risk assessment at the germline level.

Thyroid cancer is the ninth most common type of cancer in the world, with an annual incidence of >500,000 cases (>50,000 occurring in the United States) (4, 5). Although surgery and other therapies solve most cases, morbidity among patients is high, and in some cases the tumors exhibit more aggressive behavior. Thyroid cancer can be categorized by histology. The medullary type, which accounts for ∼5% of all cases, arises from parafollicular C cells of the thyroid. The remaining 95% of all thyroid cancer cases are of the nonmedullary type and arise in cells of follicular origin. There are three major histological forms of nonmedullary thyroid

Significance

Thyroid cancer shows a high degree of heritability compared with other cancers. Genome-wide association studies (GWASs) have identified at least 10 single-nucleotide polymorphisms (SNPs) associated with papillary thyroid cancer risk. How these risk factors might help individualize the assessment of thyroid cancer risk clinically has been unexplored. We present a polygenic risk score (PRS) analysis with consistent results in three large cohorts (United States, Iceland, and United Kingdom). The 10 GWAS SNPs have additive effects on cancer predisposition, and the 10-SNP PRS has equally strong risk predictive power as a PRS with >500,000 common variants. Our work demonstrates that the 10 low-penetrance variants have the potential to be applied in medicine to improve individualized cancer risk assessment.

Author contributions: S.L., J.G., E.F., H. He, P.B., F.N., G.T., M.D.R., K.S., and A.d.l.C. designed research; S.L., J.G., and G.T. performed research; J.G.J., V.T., F.W.A., L.X., L.A.K., R.T.N.-M., J.I.M., T.S.P., H. Hjartarson, J.H., E.M.S., and G.T. contributed new reagents/analytic tools; J.G., H. He, J.G.J., V.T., F.W.A., L.X., L.A.K., R.T.N.-M., J.I.M., T.S.P., H. Hjartarson, J.H., E.M.S., P.B., F.N., M.D.R., K.S., and A.d.l.C. recruited and/or contributed samples/research materials; S.L., J.G., E.F., and G.T. analyzed data; and S.L., J.G., E.F., H. He, V.T., L.A.K., G.T., K.S., and A.d.l.C. wrote the paper.

Reviewers: C.N., Beth Israel Deaconess Medical Center, Harvard Medical School; and M.X., Johns Hopkins University School of Medicine.

Competing interest statement: J.G., E.F., V.T., G.T., and K.S. are employees of deCODE/Amgen. The other authors have no conflicts of interest to declare.

Data deposition: The data (weights) used to calculate the polygenic risk scores, for all three datasets discussed in the paper, are available at deCODE genetics, https://www.decode.com/summarydata (Liyanarachchi, et al. Assessing thyroid cancer risk using polygenic risk scores).

[1]S.L. and J.G. contributed equally to this work.

[2]K.S. and A.d.l.C. contributed equally to this work.

[3]To whom correspondence may be addressed. Email: kstefans@decode.is or Albert.delaChapelle@osumc.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1919976117/-/DCSupplemental.

First published March 4, 2020.

Table 1. Demographic characteristics of the Ohio, Iceland, and UKB study groups

Characteristic	Patients	Controls
Ohio		
Total	1,544	1,593
Gender, *n* (%)		
Male	395 (26)	423 (27)
Female	1,149 (74)	1,170 (73)
Age, y, mean ± SD	42.9 ± 15.1	45.2 ± 14.0
Year of birth, median (range)	1962 (1913 to 2001)	1963 (1918 to 1991)
First- or second-degree relative diagnosed with thyroid cancer, *n* (%)		
Yes	135 (8.7)	9 (0.6)
No	1,409 (91.3)	1,584 (99.4)
Iceland		
Total	723	129,556
Gender, *n* (%)		
Male	183 (25.3)	60,282 (46.5)
Female	540 (74.7)	69,274 (53.5)
Age, y, mean ± SD	49.3 ± 17.3	60.4 ± 18.0
Year of birth, median (range)	1946 (1911 to 1989)	1956 (1890 to 1990)
First- or second-degree relative diagnosed with thyroid cancer, *n* (%)		
Yes	84 (11.6)	6,455 (5.0)
No	639 (88.4)	123,101 (95.0)
UKB		
Total	534	407,945
Gender, *n* (%)		
Male	131 (24.5)	187,661 (46.0)
Female	403 (75.5)	220,818 (54.0)
Age, y, mean SD	51.8 ± 12.18	64.1 ± 8.00
Year of birth, median (range)	1949 (1938 to 1969)	1950 (1934 to 1969)
First- or second-degree relative diagnosed with thyroid cancer, *n* (%)*		
Yes	NA	NA
No	NA	NA

NA, not applicable.
*No information about family history of thyroid cancer is available for the UKB samples.

cancer: papillary (PTC), follicular (FTC), and anaplastic (ATC). PTC alone accounts for ~85% of all thyroid cancers.

The concept of this study has already been used in numerous investigations dealing with risk assessment in cancers (4–8). Here we address an aspect of this field that has not yet been fully examined: the phenotypic effect of low-penetrance mutations and their use in individualized thyroid cancer risk assessment. PTC mutations with high penetrance have been found but account for a very minor part of all PTCs. We and others have suggested that common alleles, each conferring a slightly increased risk, may account for many PTCs and provide an explanation for the high heritability of PTC (9–11). Indeed, many common, low-penetrance

Table 2. Effect estimates used in the PRS model in each study group

					Ohio		Iceland		UKB	
Marker	Locus	Position (bp)*	OA	EA	EAF	OR[†]	EAF	OR[‡]	EAF	OR[§]
rs12129938	1q42.2	233,276,815	G	A	0.81	1.2	0.783	1.16	0.775	1.32
rs11693806	2q35	217,427,435	G	C	0.318	1.43	0.285	1.37	0.279	1.43
rs6793295	3q26.2	169,800,667	C	T	0.755	1.2	0.78	1.16	0.729	1.23
rs73227498	5q22.1	112,150,207	T	A	0.891	1.28	0.855	1.33	0.863	1.37
rs2466076	8p12	32,575,278	T	G	0.528	1.32	0.467	1.3	0.452	1.32
rs1588635	9q22.33	97,775,520	C	A	0.476	1.64	0.356	1.72	0.326	1.69
rs7902587	10q24.33	103,934,543	C	T	0.12	1.25	0.095	1.22	0.098	1.41
rs368187	14q13.3	36,063,370	C	G	0.626	1.33	0.523	1.28	0.55	1.39
rs116909374	14q13.3	36,269,155	C	T	0.044	1.71	0.049	1.73	0.038	1.71
rs2289261	15q22.33	67,165,147	G	C	0.7	1.14	0.669	1.22	0.647	1.23

*Chromosomal position with reference to build 38, the effect allele (EA) and the other allele (OA), the effect allele frequency (EAF) in controls for each study group, and the OR used to calculate the polygenic risk score for each study group.
[†]OR from the meta analysis after excluding results for the Ohio study group.
[‡]OR from the meta analysis after excluding results for the Icelandic study group.
[§]OR from the meta analysis after excluding results for the UKB study group.

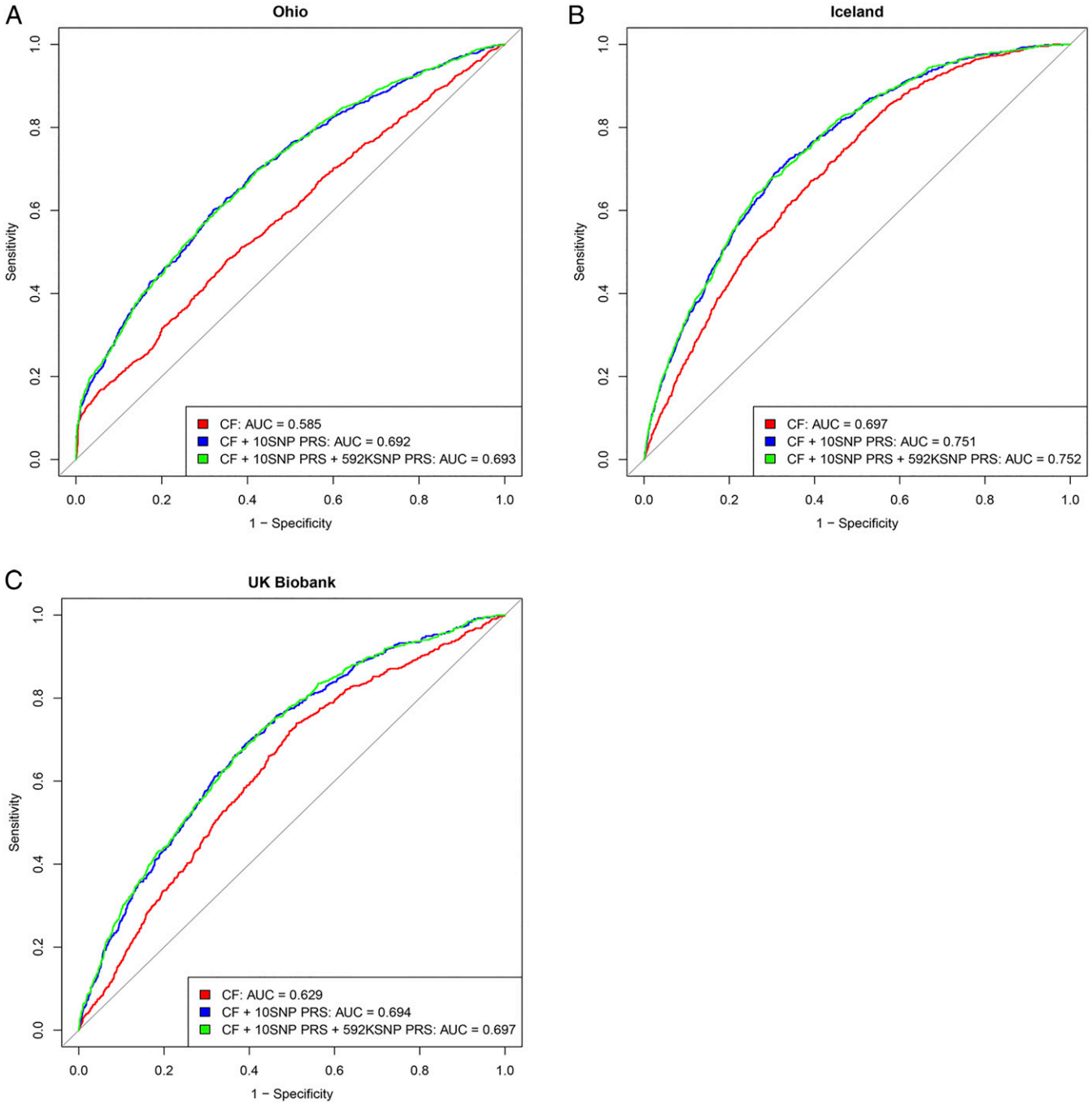

Fig. 1. ROC curves assessing the discriminative power of the PRS models for the Ohio (*A*), Iceland (*B*), and UKB (*C*) study groups. The CF model includes year of birth, gender, ancestry, and familiality for all groups except the UKB group, for which no information was available about family history of thyroid cancer.

SNPs have been found to convey PTC risk (12–21). In this study, we analyzed the combined genetic effects of 10 well-established thyroid cancer risk SNPs by constructing and evaluating their polygenic risk scores (PRSs). We also checked for any residual predictive power in the remaining genome by generating a genome-wide PRS using 592,495 tagging SNPs and adding it to the 10-SNP PRS, along with considering conventional clinical factors (CFs).

Results

Study Participants and Their Demographic Characteristics. The results reported here are based on previously published genome-wide association studies (GWASs) of thyroid cancer in populations from Columbus, Ohio and Houston, Texas; Iceland; The Netherlands; and Spain (16). In addition, we incorporated thyroid cancer GWAS data generated using genotypic information from the UK Biobank (UKB) (22, 23). The addition of UKB data to the meta-analysis of thyroid cancer did not reveal any new significantly associated genome-wide risk variants.

PRSs were calculated for the three largest sample sets from Ohio, Iceland, and the United Kingdom with chip genotyped data. The Ohio study group comprised 1,544 thyroid cancer samples and 1,593 controls; the Iceland group, 723 thyroid cancer samples and 129,556

Table 3. Classification results for different models in each study group

Model	AUC	95% CI	*P* value*
Ohio			
CF	0.585	0.565 to 0.605	Reference
CF+10SNPs_PRS	0.692	0.673 to 0.710	3.10E-21
CF+10SNPs_PRS+592KSNPs_PRS	0.693	0.675 to 0.712	0.34
Iceland			
CF	0.697	0.680 to 0.714	Reference
CF+10SNPs_PRS	0.751	0.736 to 0.768	3.00E-14
CF+10SNPs_PRS+592KSNPs_PRS	0.752	0.680 to 0.714	0.3
UKB			
CF	0.629	0.606 to 0.651	Reference
CF+10SNPs_PRS	0.694	0.673 to 0.716	1.00E-09
CF+10SNPs_PRS+592KSNPs_PRS	0.697	0.676 to 0.719	0.27

**P* values for the stepwise addition of factors shown in the Model column.

controls; and the UKB group, 534 thyroid cancer cases and 407,945 controls (Table 1).

PRS Analysis in Study Groups from Ohio, Iceland, and the United Kingdom and Association with Cancer Risk. The effect estimates included in the PRS analysis are based on the meta-analysis of thyroid cancer including all three aforementioned study groups. In short, we generated PRSs for the Ohio study group by using effect estimates after excluding all samples from the United States from our thyroid cancer GWAS meta-analysis, thereby omitting any potential confounding effects. Similarly, when generating PRS for Icelandic and British individuals, corresponding samples from those study groups were excluded from the meta-analysis (Table 2).

For each individual belonging to the Ohio, Icelandic, and UKB study group, a PRS was generated using the published 10 GWAS thyroid cancer risk SNPs (10-SNP PRS; Table 2) as well as 592,475 common SNPs with minor allele frequency >1% (592K-SNP PRS). The PRS for the 592,475 common SNPs was estimated based on the LDpred method, adjusting for GWAS summary statistics for the effects of linkage disequilibrium (24). The risk loci, risk allele frequencies, and effect estimates of the 10 GWAS SNPs included in the 10-SNP PRS are provided in Table 2. The PRSs of the 10-SNPs and the 592K-SNPs were approximately normally distributed among thyroid cases and controls (*SI Appendix*, Fig. S1) and were significantly different between thyroid cancer cases and controls in all three study groups ($P = 2.9 \times 10^{-58}$, $P = 3.3 \times 10^{-48}$ for Ohio; $P = 1.3 \times 10^{-48}$, $P = 4.7 \times 10^{-33}$ for Iceland; and $P = 1.3 \times 10^{-25}$, $P = 2.2 \times 10^{-23}$ for UKB for the 10-SNP PRS and the 592K-SNP PRS, respectively).

PRSs in Prediction Models. To investigate the predictive ability of PRSs, we evaluated prediction models using receiver operating characteristic (ROC) curves. With CFs, including year of birth, gender, the 10 first principal components, and familiality (not available for the UKB samples), we obtained an area under the ROC curve (AUC) of 0.585 (95% CI, 0.565 to 0.605) for the Ohio study group, 0.697 (95% CI, 0.680 to 0.714) for the Iceland group, and 0.629 (95% CI, 0.606 to 0.651) for the UKB group (Fig. 1 and Table 3). By adding the10-SNP PRS to the model with CF, we obtained a significantly increased AUC of 0.692 ($P = 3.1 \times 10^{-21}$) for the Ohio group, 0.751 ($P = 3.0 \times 10^{-14}$) for the Iceland group, and 0.694 ($P = 1.0 \times 10^{-09}$) for the UKB group (Fig. 1 and Table 3).

We further evaluated the prediction ability after adding the 592K-SNP PRS to the model with 10-SNP PRS and CF (*SI Appendix*, Table S2 for results for individual covariates). For the Ohio and the Icelandic samples, AUCs of 0.693 and 0.752, respectively, were obtained, showing only a 0.1-percentage point increase over the 10-SNP PRS model ($P = 0.34$ and 0.31, respectively) (Fig. 1 *A* and *B* and Table 3). In the UKB samples, an AUC of 0.697 was obtained, for a nonsignificant 0.3-percentage point increase ($P = 0.27$) (Fig. 1*C* and Table 3). Together, these results demonstrate that no significant improvement was achieved by adding the genome-wide PRS (592K-SNP PRS) to the model.

In a multicollinearity analysis, we found variance inflation factors (VIFs) of 2.32 and 2.32 for the Ohio group and 2.04 and 2.04 for the UKB group for the 10-SNP and 592K-SNP PRS scores, respectively, included in the all-predictive factors combined model. The VIF in the Icelandic study group was somewhat smaller 1.64 and 1.63 for the 10-SNP and 592K-SNP PRS scores, respectively. The VIF indicates how much larger the variance is compared to what it would be if the respective variables included in the model were not correlated with each other. Moreover, we estimated that the 10 SNPs under study explained ~8% of the familial risk of thyroid cancer in the Ohio study group.

Assessing Thyroid Cancer Risk by 10-SNP PRS Percentile Group Based on the Meta-Analysis Results from Ohio, Iceland, and the United Kingdom. We meta-analyzed results from Ohio, Iceland, and the UK after ranking 10-SNP PRS scores for each individual and correlating them with cancer status. Individuals in the top decile group of the PRSs had a 6.9-fold greater risk compared with the bottom decile group ($P = 5.1 \times 10^{-54}$) (Fig. 2). This difference is substantial and might be useful in clinical counseling.

Discussion

Our findings largely confirm a previous estimate that 11% of the genetic predisposition to PTC could be accounted for by the interaction of five common SNPs (4). In our present study involving 10 SNPs and larger numbers of cases, the proportion was slightly smaller (8%); this can probably be explained by a somewhat different sample set and higher genetic resolution applied in the present study. Nevertheless, the fraction of predisposition accounted for is quite low even when ~592,000 common SNPs were investigated.

Interestingly, our data indicate that the 10-SNP PRS in the prediction model of thyroid cancer performs equally well as the combined model with 10-SNP PRS and common 592K-SNP PRS

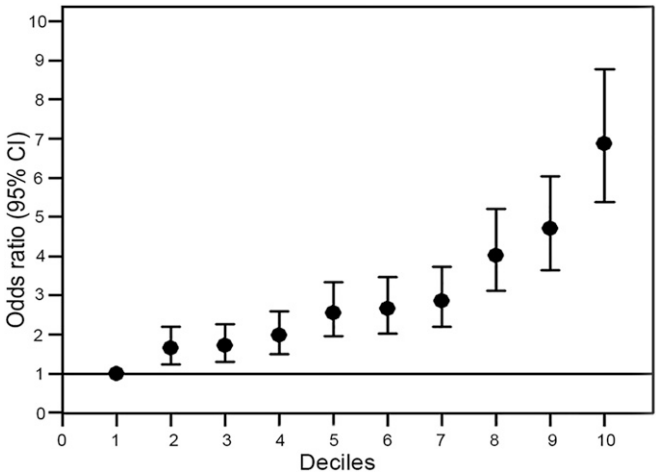

Fig. 2. OR estimates for 10-SNP PRS deciles of thyroid cancer status obtained from the meta-analysis results from the Ohio, Iceland, and the UKB study groups, using the bottom 10-SNP PRS decile (0 to 10%) as the reference group (shown as a horizontal solid line).

in all three study populations. Our observations further support the notion that the 10 variants previously detected by GWAS are important genetic factors conferring thyroid cancer risk (16). The majority of the variants (9 out of 10) are located either intronic or intergenic, while only one SNP rs6793295 is a missense variant in the *LRRC34* gene in 3q26.2 (16). This coding variant is also significantly associated with the risk of multiple myeloma, monoclonal gammopathy, and interstitial lung disease (25, 26). Interestingly, five noncoding variants—rs11693806, rs2466076, rs1588635, rs368187, and rs116909374—are also associated with serum levels of thyroid function-related hormones (thyroid-stimulating hormone, T3, and T4), and rs116909374 is associated with hypothyroidism (12, 13, 16, 27). The intergenic noncoding variant rs7902587 in 10q24.33 is significantly associated with lung cancer and ovarian cancer, and rs11693806 in 2q35 is associated with breast cancer (28–30).

Our data suggest that the current PRS models with either 10-SNP or 592K-SNP PRSs could still have the problem of missing heritability (31, 32). Most likely, only a few other common variants may remain to be discovered, as the 592K-SNP PRS was designed to assess and estimate the contribution of such variants (16). Therefore, we hypothesize that hitherto undetected low-frequency or rare DNA variants—particularly those located in regions of low linkage disequilibrium—may play a role in PTC risk prediction (33). Indeed, we and others have demonstrated a high degree of genetic heterogeneity in thyroid cancer (34–36). We have identified multiple rare or very-low-frequency DNA variants that may contribute to the predisposition of familial and sporadic PTC (35–38). Identification of additional low-frequency and/or rare germline DNA variants may benefit the assessment of additive genetic effects in PRS models and personalized medical diagnosis and treatment (32). We note with great interest that the recent studies by Vogelstein et al. (39, 40) reached the same conclusion, that the great majority of the driver mutations in PTCs are somatic events occurring randomly in stem cells of the target organ. These findings appear to predict that only few or very few additional high-penetrance germline variants will be found in the future. Nevertheless, the data presented here further support the idea that individual PTC-associated variants confer only a small or modest disease risk, but the combined effects of the known associated SNPs can be substantial in predicting cancer risk (4, 16, 20). Our data provide evidence that PRS could be used for profiling individuals in the highest and lowest relative risk groups for thyroid cancer, which has potential for the development of population-based risk screening and stratification programs, as has been demonstrated in other cancers (7, 41–44).

The strength of our present study is the availability of large numbers of case-control samples from three study groups representing different areas of Caucasians (Ohio, Iceland, and the United Kingdom). We used effect sizes obtained from meta-analyses of the Iceland/UKB study groups, Ohio/UKB study groups, and Ohio/Iceland study groups to estimate PRSs in the excluded population: Ohio, Iceland, and the United Kingdom, respectively. The PRSs constructed in this study omit any confounding effect from the study populations being used to evaluate the correlation between PRS and disease status. Overall, the data presented here provide further evidence that PRS exhibits strong association with thyroid cancer. Fritsche et al. (7) reported associations of PRS in multiple cancers, including thyroid cancer, using a PRS with 8 SNPs. Interestingly, they found an attenuated association between increasing thyroid cancer PRS and reduced risk for hypothyroidism (7).

In our study, thyroid cancer patients belonging to the top decile of the 10-SNP PRS have a close to sevenfold greater risk relative to the bottom decile, based on meta-analysis results including data from Ohio, Iceland, and the United Kingdom. We conclude that hereditary germline variants should be taken into account alongside the traditional high-penetrance variants/somatic mutations that have already become the standard of care in the clinical handling of PTC (45–47). Identifying individuals with high genetic risk may prove useful to optimize screening for thyroid cancer.

Materials and Methods

Study Populations. The thyroid cancer meta-analysis has been described previously (16). In short, the meta-analysis included a total of 3,001 nonmedullary thyroid cancer patients and 287,550 controls from Iceland, Ohio, Texas, The Netherlands, and Spain (*SI Appendix*, Table S1). In the present study, we added thyroid cancer GWAS data from the UKB (accessed under application no. 24711) comprising samples from 534 patients with International Classification of Diseases for Oncology code C73 (PTC, FTC, cancer/carcinoma, and rare nonmedullary) and 407,945 controls not known to have thyroid cancer (*SI Appendix*, Table S1) (22).

Genotyping. The genotyping and imputation for the study groups from Iceland, Ohio, Texas, The Netherlands, and Spain has been described previously (16). Genotyping of the UKB samples was performed using a custom-made Affymetrix chip, UK BiLEVE Axiom (48), and with the Affymetrix UKB Axiom array (49). Imputation was performed by the Wellcome Trust Centre for Human Genetics using the Haplotype Reference Consortium (HRC) and the UK10K haplotype resources (49). This yielded a total of 96 million imputed variants; however, only 40 million variants imputed using the HRC reference set were used in this study, owing to quality issues with the remaining variants.

Variants in the UKB imputation dataset were mapped to National Center for Biotechnology Information Build38 positions and matched to the variants in the Icelandic dataset based on allele variation. The results from all study groups were combined using a fixed-effects model in which the study groups were allowed to have different population frequencies for alleles and genotypes but were assumed to have a common odds ratio (OR) and weighted with the inverse of the variance. Heterogeneity (P_{het}) was tested by comparing the null hypothesis of the effect being the same in all populations with the alternative hypothesis of each population having a different effect using a likelihood ratio test. I^2 lies between 0 and 100% and describes the proportion of total variation in study estimates that is due to heterogeneity.

Polygenic Risk Score. To evaluate the additive genetic effect of variants, we created PRSs as the sums of effects of each allele representing selected sets of variants as described by Vilhjalmsson et al. (24). For the PRS analysis, we regenerated three meta-analysis datasets, each time excluding data from the study group in which we intended to assess the correlation between the PRSs and affection status; for example, when generating PRSs for Icelanders, we used effect estimates from a meta-analysis after excluding the Iceland data.

PRSs were generated using two different sets of variants: the 10 published GWAS thyroid cancer risk variants and 592,475 common variants based on the previously published meta-analysis (16), including the addition of the UKB data described above. All PRS scores were standardized to have a unit SD. The OR per unit SD increase is reported.

Statistical Analysis. Logistic regression analysis was used to assess the association of PRSs with thyroid cancer status, adjusting for year of birth, gender, ancestry with 10 principal components, and familiality based on self-reported first- or second-degree relative information. Familiality information was not available for the UKB samples. The strength of prediction models to predict thyroid cancer against controls was assessed by comparing the AUC of the respective ROC curves that plots the true-positive rate against the false-positive rate. ROC curves were compared by applying DeLong's test (50). A higher AUC indicates better model performance. We examined all pairwise correlations and calculated VIFs associated with the PRS models (51). PRS percentile groups were used to create categorical predictors, and the risk of thyroid cancer between percentile groups was assessed by applying logistic regression. ORs per 1 SD increase to estimate the associations and AUCs to assess the discriminatory accuracy are presented with 95% CIs. Familial relative risk assessment is estimated with 10-SNPs, assuming an overall familial relative risk of 8.48 for thyroid cancer (16, 52, 53).

Data Availability. The data supporting the findings of this study are available in the paper, *SI Appendix*, and at https://www.decode.com/summarydata. The UKB data can be obtained on application (https://www.ukbiobank.ac.uk/).

ACKNOWLEDGMENTS. We thank Jan Lockman and Barbara Fersch for administrative help. This work was supported by National Cancer Institute Grants P30CA16058 and P01CA124570.

1. M. Xing, B. R. Haugen, M. Schlumberger, Progress in molecular-based management of differentiated thyroid cancer. *Lancet* **381**, 1058–1069 (2013).
2. P. D. Stenson *et al*., The Human Gene Mutation Database: Towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum. Genet.* **136**, 665–677 (2017).
3. A. P. Sokolenko, E. N. Imyanitov, Molecular diagnostics in clinical oncology. *Front. Mol. Biosci.* **5**, 76 (2018).
4. S. Liyanarachchi *et al*., Cumulative risk impact of five genetic variants associated with papillary thyroid carcinoma. *Thyroid* **23**, 1532–1540 (2013).
5. R. Szulkin *et al*.; Australian Prostate Cancer BioResource; Practical Consortium, Prediction of individual genetic risk to prostate cancer using a polygenic score. *Prostate* **75**, 1467–1474 (2015).
6. P. Maas *et al*., Breast cancer risk from modifiable and nonmodifiable risk factors among white women in the United States. *JAMA Oncol.* **2**, 1295–1302 (2016).
7. L. G. Fritsche *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.* **102**, 1048–1061 (2018).
8. N. Mavaddat *et al*.; ABCTB Investigators; kConFab/AOCS Investigators; NBCS Collaborators, Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *Am. J. Hum. Genet.* **104**, 21–34 (2019).
9. D. E. Goldgar, D. F. Easton, L. A. Cannon-Albright, M. H. Skolnick, Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. *J. Natl. Cancer Inst.* **86**, 1600–1608 (1994).
10. C. Dong, K. Hemminki, Modification of cancer risks in offspring by sibling and parental cancers from 2,112,616 nuclear families. *Int. J. Cancer* **92**, 144–150 (2001).
11. N. Risch, The genetic epidemiology of cancer: Interpreting family and twin studies and their implications for molecular genetic approaches. *Cancer Epidemiol. Biomarkers Prev.* **10**, 733–741 (2001).
12. J. Gudmundsson *et al*., Common variants on 9q22.33 and 14q13.3 predispose to thyroid cancer in European populations. *Nat. Genet.* **41**, 460–464 (2009).
13. J. Gudmundsson *et al*., Discovery of common variants associated with low TSH levels and thyroid cancer risk. *Nat. Genet.* **44**, 319–322 (2012).
14. A. Köhler *et al*., Genome-wide association study on differentiated thyroid cancer. *J. Clin. Endocrinol. Metab.* **98**, E1674–E1681 (2013).
15. H.-Y. Son *et al*., Genome-wide association and expression quantitative trait loci studies identify multiple susceptibility loci for thyroid cancer. *Nat. Commun.* **8**, 15966 (2017).
16. J. Gudmundsson *et al*., A genome-wide association study yields five novel thyroid cancer risk loci. *Nat. Commun.* **8**, 14517 (2017).
17. G. Figlioli *et al*., Novel genome-wide association study-based candidate loci for differentiated thyroid cancer risk. *J. Clin. Endocrinol. Metab.* **99**, E2084–E2092 (2014).
18. G. Figlioli *et al*., A comprehensive meta-analysis of case-control association studies to evaluate polymorphisms associated with the risk of differentiated thyroid carcinoma. *Cancer Epidemiol. Biomarkers Prev.* **25**, 700–713 (2016).
19. V. Mancikova *et al*., Thyroid cancer GWAS identifies 10q26.12 and 6q14.1 as novel susceptibility loci and reveals genetic heterogeneity among populations. *Int. J. Cancer* **137**, 1870–1878 (2015).
20. G. Figlioli *et al*., Novel genetic variants in differentiated thyroid cancer and assessment of the cumulative risk. *Sci. Rep.* **5**, 8922 (2015).
21. Y. Hwangbo *et al*., Genome-wide association study reveals distinct genetic susceptibility of thyroid nodules from thyroid cancer. *J. Clin. Endocrinol. Metab.* **103**, 4384–4394 (2018).
22. C. Bycroft *et al*., Genome-wide genetic data on ~500,000 UK Biobank participants. bioRxiv:10.1101/166298 (20 July 2017).
23. C. Bycroft *et al*., The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
24. B. J. Vilhjálmsson *et al*.; Schizophrenia Working Group of the Psychiatric Genomics Consortium, Discovery, Biology, and Risk of Inherited Variants in Breast Cancer (DRIVE) study, Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
25. B. Swaminathan *et al*., Variants in ELL2 influencing immunoglobulin levels associate with multiple myeloma. *Nat. Commun.* **6**, 7213 (2015).
26. T. E. Fingerlin *et al*., Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat. Genet.* **45**, 613–620 (2013).
27. G. Kichaev *et al*., Leveraging polygenic functional enrichment to improve GWAS power. *Am. J. Hum. Genet.* **104**, 65–75 (2019).
28. J. D. McKay *et al*.; SpiroMeta Consortium, Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. *Nat. Genet.* **49**, 1126–1132 (2017).
29. C. M. Phelan *et al*.; AOCS study group; EMBRACE Study; GEMO Study Collaborators; HEBON Study; KConFab Investigators; OPAL study group, Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. *Nat. Genet.* **49**, 680–691 (2017).
30. K. Michailidou *et al*.; NBCS Collaborators; ABCTB Investigators; ConFab/AOCS Investigators, Association analysis identifies 65 new breast cancer risk loci. *Nature* **551**, 92–94 (2017).
31. O. Zuk, E. Hechter, S. R. Sunyaev, E. S. Lander, The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 1193–1198 (2012).
32. A. I. Young, Solving the missing heritability problem. *PLoS Genet.* **15**, e1008222 (2019).
33. P. Wainschtein *et al*., Recovery of trait heritability from whole genome sequence data. bioRxiv:10.1101/588020 (25 March 2019).
34. F. Lesueur *et al*.; NMTC Consortium, Genetic heterogeneity in familial nonmedullary thyroid carcinoma: Exclusion of linkage to RET, MNG1, and TCO in 56 families. *J. Clin. Endocrinol. Metab.* **84**, 2157–2162 (1999).
35. H. He *et al*., SRGAP1 is a candidate gene for papillary thyroid carcinoma susceptibility. *J. Clin. Endocrinol. Metab.* **98**, E973–E980 (2013).
36. H. He *et al*., Ultra-rare mutation in long-range enhancer predisposes to thyroid carcinoma with high penetrance. *PLoS One* **8**, e61920 (2013).
37. J. Tomsic *et al*., A germline mutation in SRRM2, a splicing factor gene, is implicated in papillary thyroid carcinoma predisposition. *Sci. Rep.* **5**, 10566 (2015).
38. Y. Wang *et al*., Identification of rare variants predisposing to thyroid cancer. *Thyroid* **29**, 946–955 (2019).
39. C. Tomasetti, B. Vogelstein, Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* **347**, 78–81 (2015).
40. C. Tomasetti, L. Li, B. Vogelstein, Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science* **355**, 1330–1334 (2017).
41. N. Mavaddat *et al*., Prediction of breast cancer risk based on profiling with common genetic variants. *J. Natl. Cancer Inst.* **107**, djv036 (2015).
42. M. Frampton, R. S. Houlston, Modeling the prevention of colorectal cancer from the combined impact of host and behavioral risk factors. *Genet. Med.* **19**, 314–321 (2017).
43. P. Radice, P. D. P. Pharoah, P. Peterlongo, Personalized testing based on polygenic risk score is promising for more efficient population-based screening programs for common oncological diseases. *Ann. Oncol.* **27**, 369–370 (2016).
44. X. Yang *et al*., Evaluation of polygenic risk scores for ovarian cancer risk prediction in a prospective cohort study. *J. Med. Genet.* **55**, 546–554 (2018).
45. Y. E. Nikiforov, Role of molecular markers in thyroid nodule management: Then and now. *Endocr. Pract.* **23**, 979–988 (2017).
46. M. N. Nikiforova *et al*., Analytical performance of the ThyroSeq v3 genomic classifier for cancer diagnosis in thyroid nodules. *Cancer* **124**, 1682–1690 (2018).
47. M. Endo *et al*., Afirma gene sequencing classifier compared with gene expression classifier in indeterminate thyroid nodules. *Thyroid* **29**, 1115–1124 (2019).
48. L. V. Wain *et al*.; UK Brain Expression Consortium (UKBEC); OxGSK Consortium, Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): A genetic association study in UK Biobank. *Lancet Respir. Med.* **3**, 769–781 (2015).
49. S. Welsh, T. Peakman, S. Sheard, R. Almond, Comparison of DNA quantification methodology used in the DNA extraction protocol for the UK Biobank cohort. *BMC Genomics* **18**, 26 (2017).
50. E. R. DeLong, D. M. DeLong, D. L. Clarke-Pearson, Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* **44**, 837–845 (1988).
51. J. Fox, G. Monette, Generalized collinearity diagnostics. *J. Am. Stat. Assoc.* **87**, 178–183 (1992).
52. R. S. Houlston, D. Ford, Genetics of coeliac disease. *QJM* **89**, 737–743 (1996).
53. P. Broderick *et al*., Common variation at 3p22.1 and 7p15.3 influences multiple myeloma risk. *Nat. Genet.* **44**, 58–61 (2011).