Novel Genome-Wide Association Study–Based Candidate Loci for Differentiated Thyroid Cancer Risk

Gisella Figlioli, Aleksandra Köhler, Bowang Chen, Rossella Elisei, Cristina Romei, Monica Cipollini, Alfonso Cristaudo, Franco Bambi, Elisa Paolicchi, Per Hoffmann, Stefan Herms, Michał Kalemba, Dorota Kula, Susana Pastor, Ricard Marcos, Antonia Velázquez, Barbara Jarząb, Stefano Landi, Kari Hemminki, Asta Försti,* and Federica Gemignani*

Department of Biology (G.F., M.C., E.P., S.L., F.G.), University of Pisa, 56123 Pisa, Italy; Molecular Genetic Epidemiology (A.K., B.C., K.H., A.F.), German Cancer Research Center, 69120 Heidelberg, Germany; Department of Endocrinology and Metabolism (R.E., C.R., A.C.), University of Pisa, 56124 Pisa, Italy; Blood Centre (F.B.), Azienda Ospedaliero Universitaria A. Meyer, 50139 Firenze, Italy; Department of Genomics (P.H., S.H.), Life and Brain Center and Institute of Human Genetics (P.H., S.H.), University of Bonn, D-53127 Bonn, Germany; Division of Medical Genetics (P.H., S.H.), University Hospital Basel and Department of Biomedicine, University of Basel, CH-4058 Basel, Switzerland; Department of Nuclear Medicine and Endocrine Oncology (M.K., D.K., B.J.), Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Gliwice Branch, 44–101 Gliwice, Poland; Grup de Mutagènesi (S.P., R.M., A.V.), Departament de Genètica i de Microbiologia, Facultat de Biociències, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallés, Barcelona, Spain; Centro de Investigación Biomédica en Red y Epidemiologia y Salud Pública (S.P., R.M., A.V.), Instituto de Salud Carlos III, 28029 Madrid, Spain; and Center for Primary Health Care Research (K.H., A.F.), Clinical Research Center, Lund University, 205 02 Malmö, Sweden

Context: Genome-wide association studies (GWASs) on differentiated thyroid cancer (DTC) have identified robust associations with single nucleotide polymorphisms (SNPs) at 9q22.33 (*FOXE1*), 14q13.3 (*NKX2–1*), and 2q35 (*DIRC3*). Our recently published GWAS suggested additional susceptibility loci specific for the high-incidence Italian population.

Objective: The purpose of this study was to identify novel Italian-specific DTC risk variants based on our GWAS and to test them further in low-incidence populations.

Design: We investigated 45 SNPs selected from our GWAS first in an Italian population. SNPs that showed suggestive evidence of association were investigated in the Polish and Spanish cohorts.

Results: The combined analysis of the GWAS and the Italian replication study (2260 case patients and 2218 control subjects) provided strong evidence of association with rs10136427 near *BATF* (odds ratio [OR] =1.40, $P = 4.35 \times 10^{-7}$) and rs7267944 near *DHX35* (OR = 1.39, $P = 2.13 \times 10^{-8}$). A possible role in DTC susceptibility in the Italian populations was also found for rs13184587 (*ARSB*) ($P = 8.54 \times 10^{-6}$) and rs1220597 (*SPATA13*) ($P = 3.25 \times 10^{-6}$). Only the associations between rs10136427 and rs7267944 and DTC risk were replicated in the Polish and the Spanish populations with little evidence of population heterogeneity (GWAS and all replications combined, OR = 1.30, $P = 9.30 \times 10^{-7}$ and OR = 1.32, $P = 1.34 \times 10^{-8}$, respectively). In silico analyses provided new insights into the possible functional consequences of the SNPs that showed the strongest association with DTC.

Conclusions: Our findings provide evidence for novel DTC susceptibility variants. Further studies are warranted to identify the specific genetic variants responsible for the observed associations and to functionally validate our in silico predictions. **(*J Clin Endocrinol Metab* 99: E2084–E2092, 2014)**

ISSN Print 0021-972X ISSN Online 1945-7197
Printed in U.S.A.

Received March 14, 2014. Accepted July 7, 2014.
First Published Online July 16, 2014

* A.F. and F.G. contributed equally to the study.

Abbreviations: ASR, age-standardized rate; CI, confidence interval; DTC, differentiated thyroid cancer; GWAS, genome-wide association study; LD, linkage disequilibrium; OR, odds ratio; PTC, papillary thyroid cancer; SNP, single nucleotide polymorphism; TC, thyroid cancer.

Thyroid cancer (TC) represents the most common malignancy of the human endocrine system (1). Although the annual incidence rate is decreasing worldwide, several reports have described a continuous increase in some regions of the world. It remains unclear whether this trend is true or whether it depends on increased detection rates of small tumors due to the technological advances in morphological and functional imaging techniques over the last few decades (2, 3).

Approximately 90% of diagnosed TCs arise from follicular cells or thyrocytes, the thyroid hormone–producing cells, and are classified as differentiated thyroid cancer (DTC). Among them, the most frequent subtype is papillary (75%), followed by follicular (10%), Hürthle cells (5%), and poorly differentiated carcinomas (1%–6%) (4).

There are few known DTC risk factors except for a previous benign thyroid disease, exposure to ionizing radiation, and female sex (5). In addition, a significantly higher risk in first-degree relatives of patients with DTC than in the general population suggests that genetic variation may contribute to the risk of the disease (6). To date, candidate-gene association studies remain the most prevalent type of investigation to identify common DTC susceptibility alleles. More than 100 candidate-gene studies have been published in the last decade, evaluating >300 single nucleotide polymorphisms (SNPs). Whereas some of these variants may represent true associations with DTC, many more are false-positive associations and fail to replicate among additional populations (7).

During the past few years genome-wide association studies (GWASs) have rapidly emerged worldwide as an important strategy to reveal loci predisposing to common diseases, such as DTC. Validation of the most significant associations should include as a first step a replication in a population similar to that used in the GWAS with the purpose of ruling out false positivity. Replication in geographically distinct populations is also an important task because the frequency of the predisposing alleles and environmental factors may show variations and account for the observed large differences in the incidence rates worldwide (8, 9).

The first GWAS on DTC was published by Gudmundsson et al (10). The first stage of this study was based on 192 histopathologically confirmed Icelandic case patients with TC and 37 196 control subjects. The follow-up study confirmed the associations of 2 variants, rs965513 on 9q22.33 (near *FOXE1*) and rs944289 on 14q13.3 (near *NKX2–1*), with DTC in 2 case-control groups of European descent (10). These robust associations were also identified in the second GWAS focused on radiation-related papillary thyroid cancer (PTC) and further confirmed by independent studies performed in different populations (11–16). Additional TC risk variants were revealed in a GWAS based on circulating TSH levels. Four TSH level–associated SNPs, rs116909374 (*NKX2–1*) on 14q13.3, rs966423 (*DIRC3*) on 2q35, rs2439302 (*NRG1*) on 8p12, and rs334725 (*NFIA*) on 1p31.3, were further associated with TC both in the Icelandic population and in the replication studies, including individuals from the United States (Ohio), The Netherlands, and Spain (17).

Recently, we reported the results of a GWAS in a high-incidence Italian population (age-standardized rate [ASR] = 13.5/100 000; http://eco.iarc.fr/eucan), showing a robust association with the *FOXE1* locus. The other most significant variants were first replicated in 2 additional Italian populations and then in 3 low-incidence populations: Polish, United Kingdom, and Spanish (ASR = 4.1/100 000, ASR = 3.8/100 000, and ASR = 3.8/100 000, respectively; http://eco.iarc.fr/eucan). Although the combined analysis of all populations revealed a genome-wide significant association with DTC only for rs6759952 (*DIRC3*) on 2q35, the analysis of the Italian cohorts alone showed suggestive associations with rs7617304 (within *RARRES1*) on 3q25.32, rs10238549 and rs7800391 (*IMMP2L*) on 7q21, and rs10781500 (*SNAPC4*) on 9q34.3 (18).

Here, we tested our hypothesis of Italian-specific DTC susceptibility loci further and report the results of a study focusing on 45 SNPs selected from our GWAS. We tested each variant in an Italian population similar to the population studied in the GWAS. Then, SNPs that showed suggestive evidence of association were investigated further in the Polish and the Spanish cohorts. We used in silico tools to predict functional consequences for the most significantly associated variants.

Materials and Methods

Ethics statement

Study participants were recruited according to the protocols approved by the institutional review boards in accordance with the Declaration of Helsinki. All subjects provided written informed consent.

Study populations

This study was conducted on 4 sets of samples (Italian GWAS and replication set, Polish and Spanish), which were described elsewhere (18) and are reported in Supplemental Table 1. All case patients and control subjects were of Caucasian origin. In brief, the GWAS comprised 701 Italian patients with PTC ascertained through the University Hospital Cisanello in Pisa and 499 healthy individuals from the Meyer Hospital in Florence (390 blood donors and 109 workers from the Meyer Hospital). The

Italian replication cohort included 1539 patients with DTC attending the University Hospital Cisanello in Pisa. The control group (1719) was recruited from individuals without any thyroid disease and cancer history: 1079 were workers at the same hospital of Pisa and 640 were blood donors from the Meyer Hospital in Florence. The Polish group comprised 468 patients with DTC and 470 healthy control subjects from the Department of Nuclear Medicine and Endocrine Oncology, Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology in Gliwice. The Spanish cohort consisted of 446 patients with DTC, recruited by the Department of Genetics and Microbiology of Autonomous University of Barcelona and 420 healthy individuals.

DNA isolated from peripheral blood leukocytes (Italian, Polish, and Spanish cohorts) and oral mucosa cells (Spanish cohort) were used. DNA was extracted according to the protocols used in the respective institutions that provided the samples. The DNA concentration was evaluated with a NanoDrop spectrophotometer. For the Italian replication set and Polish samples, whole-genome amplification was performed using an Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare), according to the manufacturer's protocol.

SNP selection

Candidate SNPs were selected based on the results of the Italian GWAS reported by Köhler et al (18), in which the best 100 SNPs had already been investigated. Here, the following 150 SNPs were visually screened for the quality of their clustering pattern. The Manhattan plots (±100 kb from the SNP position) were also investigated, and the SNPs were screened for other SNPs in linkage disequilibrium (LD) with the variant of interest. Finally, we selected 45 SNPs for further evaluation. All of them represented a region with at least 2 SNPs associated with DTC.

SNP genotyping

Depending on the SNP, genotyping was performed using either the TaqMan SNP genotyping assay (Life Technologies) or the KASPar v4.0 allele-specific PCR (KBioscience), according to the manufacturer's guidelines. All populations were genotyped using the same assay for the same SNP. To assure the genotyping reliability, repeated analysis was performed in a randomly selected 5% of samples (the average concordance rate was 99%). After excluding samples with >50% missing genotypes, all markers had a call rate of >95%, with a mean call rate of 97%. The original GWAS samples were regenotyped with TaqMan (rs10136427, rs7267944, and rs1220597) or KASPar (rs13184587) assays, and the results confirmed the GWAS data (concordance of >99%).

Statistical analysis

We tested the genotype distributions in controls for Hardy-Weinberg equilibrium by using the χ^2 test. For each SNP, logistic regression analysis was performed to determine allelic odds ratios (ORs) with 95% confidence intervals (CIs) and allelic *P* values. The calculations were done for each cohort separately for unadjusted models as well as with adjustment for sex and age/age at diagnosis. The results for adjusted models were similar to the unadjusted results and are not reported. For the Italian cohorts, further adjustments for the enrollment center (University Hospital Cisanello in Pisa and Meyer Hospital in Florence) or the place of birth (Southern Italy or Northern and Central Italy) did not substantially change the associations. For all replication studies combined, calculations were performed, correcting for age, sex, and cohort. The Cochran *Q* statistics were calculated to test for heterogeneity, and the I^2 statistics were calculated to quantify the proportion of the total variation due to heterogeneity. All analyses were performed using SAS version 9.2 (SAS Institute Inc).

Imputation

To evaluate the associated loci more thoroughly, we imputed genotypes of all SNPs that were not genotyped in the GWAS located 100 kb upstream or downstream of the most significant SNP. We used genotype information from the CEU panels of the publicly available HapMap3 (www.hapmap.org/) and 1000 Genomes Project (www.1000genomes.org/) databases. We used the software IMPUTE2 to perform imputation analysis on each associated locus. Regional plots were generated using LocusZoom (http://csg.sph.umich.edu/locuszoom/).

Computational analysis

We searched for SNPs in a strong LD ($r^2 > 0.8$) with SNPs that showed the strongest associations with DTC predisposition based on the CEU data of the 1000 Genomes Project pilot release (www.1000genomes.org/) (Supplemental Figure 1). To explore the epigenetic profile of these regions, we checked the chromatin state segmentation profile (ChromHMM) in the lymphoblastoid cell line (GM12878) data generated by the ENCODE project and available at the University of California, Santa Cruz (UCSC) Genome Browser (http://genome.ucsc.edu/). To assess the possible functional role of each SNP we used the freely available tool HaploReg v2 (www.broadinstitute.org/mammals/haploreg), an ENCODE-based tool for investigation of noncoding genome variants. In particular, it provides evidence of function associated with a SNP due to alterations in regulatory protein binding, chromatin structure, histone modifications, and putative transcription factor binding sites. In addition, we examined the expression quantitative trait loci data available for the closest gene to each SNP (*cis*-eQTL) by using SNPexp (http://tinyurl.com/snpexp). We tested each SNP for the 210 unrelated HapMap phase II individuals pooled together using the additive model, and we calculated the unadjusted *P* value and the *P* value after Bonferroni correction. For the intronic SNPs, the effect on the splicing mechanism was investigated by the online prediction software developed within the Berkeley Drosophila Genome Project, freely available at http://www.fruitfly.org/seq tools/splice.html. This software analyzes the structure of the splice donor and acceptor sites using a separate neural network recognizer for each site. Moreover, we used SpliceAid (http://www.introni.it/splicing.html) to explain the effect of the intronic alteration on the intronic splicing enhancers or silencers.

Results

Recently, we conducted a GWAS on the high-incidence Italian population, including 701 case patients and 499 control subjects. After application of stringent GWAS quality control criteria, 572 042 SNPs were tested for association with DTC in 690 case patients and 497 control subjects (18). Here, using 3 independent cohorts (Italian,

Spanish, and Polish), we tested 45 new candidate loci identified in this GWAS. Characteristics of the participants in the initial GWAS and in the present study were described elsewhere (18) and are reported in Supplemental Table 1.

Association study

First, we genotyped 1539 Italian patients with DTC and 1719 healthy control subjects to validate the 45 selected SNPs (Supplemental Table 2). Two SNPs (rs10464672 and rs1053005) demonstrated a deviation from Hardy-Weinberg equilibrium in control subjects ($P < .005$) and were excluded from the analyses.

A statistically significant association ($P < .05$) at the same direction as in the GWAS was found for SNPs rs13184587 in *ARSB*, rs1220597 in *SPATA13*, rs10136427 upstream of *BATF*, and rs7267944 downstream of *DHX35*. Moreover, rs2281016 upstream of *TIPRL*, rs1159444 upstream of *GPD1L*, and rs2245026 upstream of *DACH1* showed the same direction of association as in the GWAS, although not statistically significant (Table 1). Combining the GWAS results with the present Italian data (2260 case patients and 2218 control subjects), we found that the most significant associations were observed for rs10136427 (OR = 1.40; 95% CI, 1.23–1.60; $P = 4.35 \times 10^{-7}$) and for rs7267944 (OR = 1.39; 95% CI, 1.24–1.56; $P = 2.13 \times 10^{-8}$). Moreover, borderline associations were reported for rs13184587 (OR = 1.28; 95% CI, 1.15–1.43; $P = 8.54 \times 10^{-6}$) and rs1220597 (OR = 1.26; 95% CI, 1.14–1.38; $P = 3.25 \times 10^{-6}$) (Table 1).

We proceeded to genotype the 7 SNPs in the Polish and the Spanish cohorts (Table 1). Only the associations for the SNPs rs10136427 and rs7267944 showed little evidence of between-study heterogeneity and the joint analysis of all replication studies (Italian, Polish, and Spanish), including 2453 case patients and 2609 controls, revealed an association with DTC: rs10136427 (OR = 1.24; 95% CI, 1.10−1.39; $P = 3.45 \times 10^{-4}$) and rs7267944 (OR = 1.25; 95% CI, 1.12–1.40; $P = 4.56 \times 10^{-5}$). When we combined the statistical evidence from the previous GWAS with the present replication studies (2985 case patients and 3727 control subjects), these 2 SNPs surpassed the threshold of genome-wide significance (OR = 1.30; 95% CI, 1.17–1.44; $P = 9.30 \times 10^{-7}$ and OR = 1.32, 95% CI, 1.20–1.46; $P = 1.34 \times 10^{-8}$, respectively). The highly significant associations observed in both Italian cohorts for rs13184587 and rs1220597 decreased after addition of the Polish and the Spanish samples, also indicated by the considerable between-study heterogeneity (Table 1).

To increase the knowledge about the associated regions, we conducted imputation over a 200-kb interval spanning the rs10136427, rs7267944, rs13184587, and rs1220597 loci (Figure 1). In chromosome 14, the region near the *BATF* gene was associated with DTC predisposition, as demonstrated by the low *P* values of SNPs rs10136427, rs3759738, rs17103360, and rs10144160 analyzed in our GWAS and by several imputed SNPs. Imputation analysis in the region surrounding rs7267944 did not reveal any additional SNPs that were more significantly associated with DTC risk. In chromosome 5, many associations resulted from SNPs in strong LD with the marker SNP rs13184587 in the intronic region of *ARSB*. In chromosome 13, rs1220597 within *SPATA13* and less associated common variants, not in LD with the SNP of interest, defined the association.

Prediction of the biological effect of the associated regions

To discover the functional relevance of the associated loci, we investigated the epigenetic profiles of the association signals and we functionally annotated the LD block containing each SNP.

The region including rs10136427 at chromosome 14 was reported as a weak enhancer region (Figure 1A). Many SNPs in the LD block containing rs10136427 were predicted to alter transcription factor binding sites by HaploReg v2, and, in particular, the risk C allele of rs10136427 removes a FOXD3 binding site. Several regulatory consequences were reported for rs4903324, located 18 kb upstream of rs10136427 and in strong LD ($r^2 = 0.92$) with it. Analysis of the position weight matrix suggested that rs4903324 is located in the binding site of transcription factors Gm397 and Pax-6. Chromatin immunoprecipitation sequencing data revealed that this variant also alters the binding of several regulatory proteins, such as SRF, YY1, and FOS. Moreover, it seems that rs4903324 is located at a site of multiple histone markers (Supplemental Table 3). Nevertheless, according to SNPexp none of the SNPs in the LD block of rs10136427 are associated with the expression level of *BATF* (Supplemental Table 4).

For chromosome 20, the chromatin state segmentation profile did not reveal any functional consequences for rs7267944 (Figure 1B). HaploReg v2 analysis indicated that the minor allele of rs7267944 (downstream *DHX35*) may affect the binding affinities of CEBPB (CCAAT/enhancer-binding protein β) transcription factor (Supplemental Table 3). eQTL analysis indicated that rs7267944, rs6101462, and rs12329618 are associated with transcription level of *DHX35* gene ($P < .05$) and the association of rs7267944 with the expression was statistically significant also after the Bonferroni correction ($P = .03$) (Supplemental Table 4).

Table 1. Risk of DTC Associated With the 7 SNPs in All Cohorts

dbSNP Identification No.	Chr	Position[a]	Gene	Risk Allele	Cohort	No. of Case Patients/ Control Subjects	Risk Allele Frequency (Case Patients/ Control Subjects)	Allelic OR (95% CI)[b]	*P*[b]
rs2281016	1	166414516	*TIPRL*	A	GWAS	646/414	0.27/0.20	1.49 (1.21–1.83)	1.84×10^{-4}
					Italian[c]	1437/1534	0.25/0.23	1.11 (0.99–1.26)	0.07
					Both Italian cohorts	2083/1948		1.21 (1.08–1.35)	1.16×10^{-3}
					Polish[c]	448/424	0.18/0.20	0.87 (0.68–1.10)	0.24
					Spanish[c]	375/408	0.26/0.21	1.32 (1.05–1.67)	0.02
					All replications	2260/2366		1.09 (0.98–1.21)	0.10
					Joint	2906/2708		1.16 (1.06–1.27)	2.03×10^{-3}
								$P_{het} = 5.4 \times 10^{-3}$, $I^2 = 76.34\%$	
rs1159444	3	32121192	*GPD1L*	T	GWAS	646/414	0.14/0.09	1.67 (1.26–2.20)	2.83×10^{-4}
					Italian[d]	1486/1595	0.13/0.11	1.15 (0.98–1.34)	0.08
					Both Italian cohorts	2132/2009		1.37 (1.18–1.58)	3.35×10^{-5}
					Polish[d]	456/447	0.06/0.09	0.59 (0.41–0.85)	3.8×10^{-3}
					Spanish[d]	375/405	0.15/0.12	1.40 (1.04–1.87)	0.02
					All replications	2317/2447		1.15 (1.00–1.31)	0.05
					Joint	2963/2861		1.23 (1.09–1.39)	9.13×10^{-4}
								$P_{het} = 1 \times 10^{-4}$, $I^2 = 85.95\%$	
rs13184587	5	78242539	*ARSB*	G	GWAS	646/414	0.77/0.69	1.51 (1.24–1.83)	3.60×10^{-5}
					Italian[c]	1429/1541	0.76/0.72	1.19 (1.06–1.33)	4.15×10^{-3}
					Both Italian cohorts	2075/1955		1.28 (1.15–1.43)	8.54×10^{-6}
					Polish[c]	442/434	0.76/0.74	1.10 (0.88–1.36)	0.41
					Spanish[c]	382/403	0.72/0.74	0.91 (0.73–1.14)	0.42
					All replications	2253/2378		1.10 (0.99–1.21)	0.07
					Joint	2899/2792		1.17 (1.07–1.27)	7.16×10^{-4}
								$P_{het} = 6.4 \times 10^{-3}$, $I^2 = 75.61\%$	
rs1220597	13	23716013	*SPATA13*	C	GWAS	646/414	0.48/0.39	1.42 (1.20–1.70)	7.11×10^{-5}
					Italian[d]	1458/1575	0.47/0.43	1.20 (1.08–1.33)	5.02×10^{-4}
					Both Italian cohorts	2104/1989		1.26 (1.14–1.38)	3.25×10^{-6}
					Polish[d]	435/449	0.44/0.43	1.01 (0.84–1.22)	0.92
					Spanish[d]	332/407	0.47/0.47	1.00 (0.82–1.23)	0.99
					All replications	2225/2431		1.09 (1.00–1.19)	0.06
					Joint	2871/2845		1.16 (1.07–1.25)	2.64×10^{-4}
								$P_{het} = 0.02$, $I^2 = 70.16\%$	
rs2245026	13	71387788	*DACH1*	G	GWAS	644/414	0.21/0.14	1.56 (1.32–1.97)	1.62×10^{-4}
					Italian[d]	1446/1558	0.19/0.17	1.13 (0.99–1.29)	0.07
					Both Italian cohorts	2090/1972		1.26 (1.11–1.42)	3.84×10^{-4}
					Polish[d]	444/439	0.23/0.23	1.04 (0.83–1.29)	0.75
					Spanish[d]	249/404	0.24/0.23	1.07 (0.82–1.39)	0.61
					All replications	2139/2401		1.09 (0.97–1.22)	0.14
					Joint	2783/2815		1.17 (1.06–1.30)	2.09×10^{-3}
								$P_{het} = 0.04$, $I^2 = 65.14\%$	
rs10136427	14	75049642	*BATF*	C	GWAS	646/414	0.88/0.81	1.62 (1.28–2.06)	5.73×10^{-5}
					Italian[d]	1515/1598	0.87/0.84	1.26 (1.09–1.45)	1.32×10^{-3}
					Both Italian cohorts	2158/2012		1.40 (1.23–1.60)	4.35×10^{-7}
					Polish[d]	458/444	0.79/0.76	1.19 (0.95–1.48)	0.12
					Spanish[d]	380/406	0.84/0.83	1.05 (0.81–1.37)	0.72
					All replications	2350/2448		1.24 (1.10–1.39)	3.45×10^{-4}
					Joint	2999/2862		1.30 (1.17–1.44)	9.30×10^{-7}
								$P_{het} = 0.12$, $I^2 = 49.06\%$	
rs7267944	20	37380848	*DHX35*	C	GWAS	646/414	0.26/0.18	1.54 (1.24–1.90)	6.60×10^{-5}
					Italian[d]	1477/1601	0.23/0.19	1.29 (1.14–1.45)	5.83×10^{-5}
					Both Italian cohorts	2123/2015		1.39 (1.24–1.56)	2.13×10^{-8}
					Polish[d]	454/442	0.17/0.16	1.10 (0.85–1.41)	0.45
					Spanish[d]	374/406	0.21/0.18	1.23 (0.96–1.57)	0.11
					All replications	2305/2449		1.25 (1.12–1.40)	4.56×10^{-5}
					Joint	2951/2863		1.32 (1.20–1.46)	1.34×10^{-8}
								$P_{het} = 0.18$, $I^2 = 38.63\%$	

[a] According to National Center for Biotechnology Information (NCBI) build 36.

[b] For each analyzed cohort, unadjusted allelic ORs with corresponding 95% CIs and *P* values are shown. The joint analysis of all cohorts was adjusted for age, sex, and cohort. Threshold *P* values using the Bonferroni correction: GWAS, $P = 5 \times 10^{-8}$; Italian replication, $P = .001$; Polish and Spanish replications, $P = .01$.

[c] Genotyped with the KASPar assay.

[d] Genotyped with the TaqMan assay.

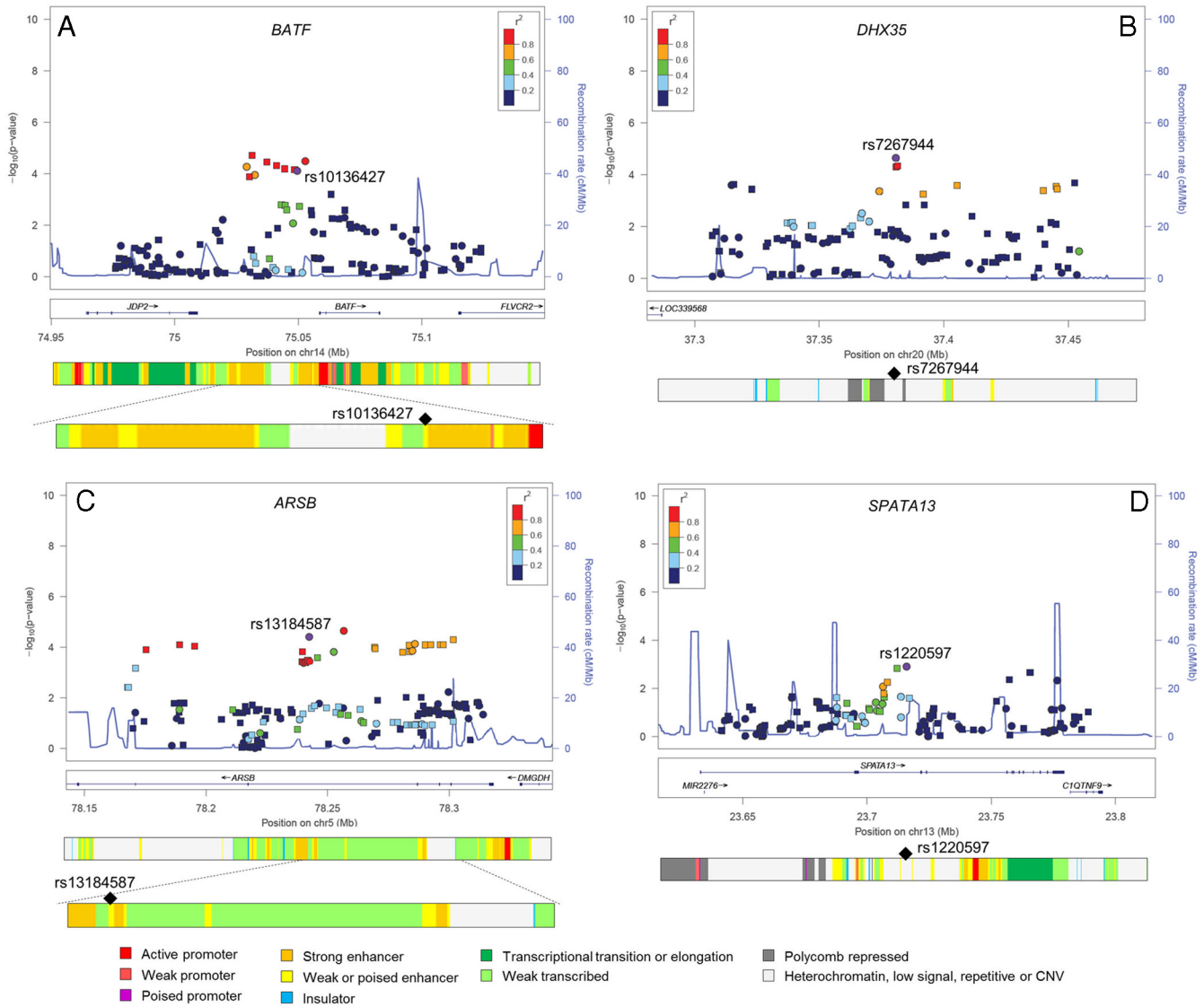

Figure 1. Regional association plots and chromatin state segmentation profile of the strongest associated variants after imputation. A–D, rs10136427 is located upstream of *BATF* (A), rs7267944 is located downstream of *DHX35* (B), rs13184587 is located in an intron of *ARSB* (C), and rs1220597 is located in an intron of *SPATA13* (D). In each plot, −log (10)*P* values (y-axis) of SNPs are shown according to their chromosomal position (x-axis). SNPs of interest are indicated by a violet circle. SNPs that were genotyped in the GWAS are marked by circles; imputed SNPs are marked as squares. The color of the SNPs represents the strength of the LD with the SNP of interest. The blue line indicates the local recombination rate (cM/Mb). On the bottom, the chromatin state segmentation profile (ChromHMM) in a lymphoblastoid cell line (GM12878) is shown.

Our in silico prediction revealed many functional roles for the SNPs in the LD block of rs13184587, located in a weak or poised enhancer region (Figure 1C). For example, rs13154740 was predicted to affect binding sites for several transcription factors (eg, SOX) and other regulatory proteins (EGR1 [early growth response protein 1], TCF12 [transcription factor 12], NFKB [nuclear factor-κB], and OCT 2 [POU class 2 homeobox 2, POU2F2]). Moreover, these genetic variants affect chromatin structure and chromatin state (Supplemental Table 3), suggesting that they may have an impact on *ARSB* regulation. According to NNSplice, none of these intronic variants alter the acceptor or donor splice sites (data not shown), but most of them may have an effect on the intronic splicing enhancer and silencer motifs (Supplemental Table 3).

The SNP rs1220597 is located in an intron of *SPATA13* and according to the 1000 Genomes Project, there are no SNPs in LD with it. The chromatin state segmentation profile and HaploReg v2 analysis did not reveal any functional consequence for rs1220597 (Figure 1D and Supplemental Table 3). The variant C allele did not alter the acceptor or the donor splice sites (data not shown). However, the substitution T to C was predicted to introduce an additional RNA sequence (GCCCAG) target for the hnRNP-K factor, a component of the intronic enhancer complex (Supplemental Table 3).

Discussion

To uncover new DTC risk variants, we investigated 45 SNPs from the loci that showed evidence of association with DTC in a high-incidence Italian population in our recent GWAS, which comprised 572 042 SNPs (18). The associations between the SNPs rs10136427 and rs7267944 and DTC reached close to a genome-wide significance in the combined Italian populations. These associations were slightly strengthened by the Polish and Spanish study populations. In addition, SNPs rs13184587 and rs1220597 showed a borderline association with DTC in the Italian populations only.

rs10136427 at chromosome 14 is located in an intergenic region upstream from *BATF* (basic leucine zipper transcription factor, ATF-like). BATF proteins form heterodimers with the Jun proteins and although the BATF/Jun dimers bind as well as the Fos/Jun dimers to the target DNA sequences, they possess a reduced ability to transactivate gene expression. For this reason, BATF proteins are considered "AP-1 inhibitors" and experiments on mouse myeloid leukemia cells have suggested that they can act as tumor suppressors by promoting growth arrest and differentiation (19, 20). Thus, the possibility that BATF could play a similar role in other cell lineages, such as the follicular cells of the thyroid, cannot be ruled out, but these kinds of studies are still lacking.

In silico analysis demonstrated that the risk allele of rs10136427 removes the consensus sequence 5′-A[AT]T-[AG]TTTGTTT-3′ recognized by the FOXD3 transcription factor. FOXD3 belongs to the forkhead family of transcription factors and acts as a tumor suppressor in melanoma cells, colorectal cancers, and gastric cancers (21–23). A heterozygous mutation in the promoter region of *FOXD3* was identified in a large family with autosomal dominant vitiligo (OMIM: 607836), an autoimmune disorder frequently associated with other autoimmune diseases, comprising Hashimoto thyroiditis (OMIM: 140300) (24, 25). Moreover, many genetic variants in the LD block containing rs10136427 may alter the binding of proteins and the function of regulatory elements (eg, enhancers). Nevertheless, they are not associated with the expression level of the nearest gene (*BATF*). Thus, it is also possible that this locus acts as a *trans*-regulatory region controlling the expression of genes that reside further away in the same or even in a different chromosome (*trans*-eQTL).

Similarly, rs7267944 is located in a gene desert downstream of *DHX35* [DEAH (Asp-Glu-Ala-His) box polypeptide 35] on chromosome 20. The gene product belongs to the DEAD/H box protein family. These proteins are classified as RNA helicases because they are involved in every aspect of RNA synthesis and function (26). Deregulation of the enzymes implicated in such RNA processes could affect cellular homeostasis and contribute to malignant transformation. In fact, expression of the *DEAD/H* box genes has been found to be deregulated in many types of tumors (27). Their role in DTC has not yet been investigated, but a similar role could be supposed according to our computational prediction. DEAD/H box proteins may also be involved in the transcriptional regulation of genes implicated in cell cycle or growth regulation (28). For example, DP103 modulates the apoptotic activity of FOXL2 (29), belonging to the same FOX family as *FOXE1*, the strongest gene associated with DTC risk identified so far (10, 11, 14).

The rs7267944 risk allele destroys the CEBPB binding site in silico. CEBPB belongs to the CACAAT/enhancer-binding protein family involved in cell proliferation, differentiation, and inflammation (30). Moreover, functional studies performed on the strong DTC-associated variant rs944289 on 14q13.3 have shown that the minor allele of this SNP affects the CEBPB consensus sequence (31), suggesting that CEBPB could have a key function in DTC development and progression.

A possible role in DTC susceptibility was found for rs13184587 on chromosome 5 (intron of *ARSB*) and rs1220597 on chromosome 13 (intron of *SPATA13*) only in the Italian cohorts. By using the SpliceAid web tool, we found that both regions may be involved in deregulation of splicing events and thus in protein synthesis. However, the role of *ARSB* and *SPATA13* in thyroid tissue has never been investigated.

The strengths of the present study include the validation of the GWAS data in a large set of case patients and control subjects from the same high-incidence Italian population as in the original GWAS. A replication in the Polish and Spanish cohorts gave us an opportunity to further evaluate the associations in 2 low-incidence populations. Both genetic and environmental factors can contribute to the differences in the incidence rates of DTC in Europe (32). It is possible that the identified risk variants are in high LD with rare functional variants (eg, related to founder effects and eliciting intermediate/high risks) among Italians but not in the Polish or the Spanish populations used in the replication phase. Similar population differences have previously been shown, eg, for breast, prostate, and pancreatic cancer (33–35). Population-specific environmental factors or gene-environmental interactions may also contribute to the SNP associations. Unfortunately, lack of information on the demographic, dietary, occupational, or lifestyle factors of the participants did not allow us to analyze the interactions between these factors and the SNPs. Finally, the relatively small

sample size of the Polish and the Spanish cohorts limited the statistical power to detect small effect sizes and may partly explain the lack of replication.

In conclusion, we performed a case-control association study of GWAS candidate loci for DTC in three European populations. Although a strong significant association for SNPs rs10136427 and rs7267944 was found in the combined analyses of all cohorts, a possible role in DTC susceptibility was found for rs13184587 and rs1220597 only in the Italian cohorts. In silico analyses provided new insights into possible functional influence for these loci. Further studies are warranted to identify the specific genetic variants responsible for the observed associations as well as functional assays to validate our bioinformatic predictions and to reveal underlying biological processes.

Acknowledgments

We thank Alaa Abdel-Ghani Madi for technical assistance. We are grateful to the study participants and the study investigators and coordinators for work in recruitment of subjects.

Address all correspondence and requests for reprints to: Dr Asta Försti, Division of Molecular Genetic Epidemiology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 580, 69120 Heidelberg, Germany. E-mail: a.foersti@dkfz-heidelberg.de; or Federica Gemignani, Genetics Department of Biology, University of Pisa, Via Derna 1, 56126 Pisa, Italy. E-mail: fgemignani@biologia.unipi.it.

The Italian part of the study has received financial support from the Istituto Toscano Tumori. A.K. had a guest scientist fellowship from the German Cancer Research Center (DKFZ). The Polish part of the study was supported by the Polish National Science Center (Grant N N402 193740). The Spanish contribution to the study was supported by the Generalitat de Catalunya (Grant 2009SGR-725).

Present address for A.K.: II Medizinische Klinik, Gastrologie, Onkologie und Palliativmedizin, St.Agnes-Hospital Bocholt, Bocholt, Germany.

Disclosure Summary: The authors have nothing to disclose.

References

1. **Sipos JA, Mazzaferri EL.** Thyroid cancer epidemiology and prognostic variables. *Clin. Oncol (R Coll Radiol)*. 2010;22:395–404.
2. **Morris LG, Myssiorek D.** Improved detection does not fully explain the rising incidence of well-differentiated thyroid cancer: a population-based analysis. *Am J Surg*. 2010;200:454–461.
3. **Yu GP, Li JC, Branovan D, McCormick S, Schantz SP.** Thyroid cancer incidence and survival in the national cancer institute surveillance, epidemiology, and end results race/ethnicity groups. *Thyroid*. 2010;20:465–473.
4. **Grande E, Diez JJ, Zafon C, Capdevila J.** Thyroid cancer: molecular aspects and new therapeutic strategies. *J Thyroid Res*. 2012;2012:847108.
5. **Dal Maso L, Bosetti C, La Vecchia C, Franceschi S.** Risk factors for thyroid cancer: an epidemiological review focused on nutritional factors. *Cancer Causes Control*. 2009;20:75–86.
6. **Fallah M, Pukkala E, Tryggvadottir L, et al.** Risk of thyroid cancer in first-degree relatives of patients with non-medullary thyroid cancer by histology type and age at diagnosis: a joint study from five Nordic countries. *J Med Genet*. 2013;50:373–382.
7. **Landa I, Robledo M.** Association studies in thyroid cancer susceptibility: are we on the right track? *J Mol Endocrinol*. 2011;47:R43–R58.
8. **Huo D, Olopade OI.** Interpretation of genome-wide association study results. *Oncology (Williston Park)*. 2010;24:643, 646.
9. **Stadler ZK, Gallagher DJ, Thom P, Offit K.** Genome-wide association studies of cancer: principles and potential utility. *Oncology (Williston Park)*. 2010;24:629–637.
10. **Gudmundsson J, Sulem P, Gudbjartsson DF, et al.** Common variants on 9q22.33 and 14q13.3 predispose to thyroid cancer in European populations. *Nat Genet*. 2009;41:460–464.
11. **Jones AM, Howarth KM, Martin L, et al.** Thyroid cancer susceptibility polymorphisms: confirmation of loci on chromosomes 9q22 and 14q13, validation of a recessive 8q24 locus and failure to replicate a locus on 5q24. *J Med Genet*. 2012;49:158–163.
12. **Liyanarachchi S, Wojcicka A, Li W, et al.** Cumulative risk impact of five genetic variants associated with papillary thyroid carcinoma. *Thyroid*. 2013;23:1532–1540.
13. **Matsuse M, Takahashi M, Mitsutake N, et al.** The FOXE1 and NKX2–1 loci are associated with susceptibility to papillary thyroid carcinoma in the Japanese population. *J Med Genet*. 2011;48:645–648.
14. **Takahashi M, Saenko VA, Rogounovitch TI, et al.** The FOXE1 locus is a major genetic determinant for radiation-related thyroid carcinoma in Chernobyl. *Hum Mol Genet*. 2010;19:2516–2523.
15. **Tomaz RA, Sousa I, Silva JG, et al.** FOXE1 polymorphisms are associated with familial and sporadic nonmedullary thyroid cancer susceptibility. *Clin Endocrinol (Oxf)*. 2012;77:926–933.
16. **Wang YL, Feng SH, Guo SC, et al.** Confirmation of papillary thyroid cancer susceptibility loci identified by genome-wide association studies of chromosomes 14q13, 9q22, 2q35 and 8p12 in a Chinese population. *J Med Genet*. 2013;50:689–695.
17. **Gudmundsson J, Sulem P, Gudbjartsson DF, et al.** Discovery of common variants associated with low TSH levels and thyroid cancer risk. *Nat Genet*. 2012;44:319–322.
18. **Köhler A, Chen B, Gemignani F, et al.** Genome-wide association study on differentiated thyroid cancer. *J Clin Endocrinol Metab*. 2013;98:E1674–E1681.
19. **Echlin DR, Tae HJ, Mitin N, Taparowsky EJ.** B-ATF functions as a negative regulator of AP-1 mediated transcription and blocks cellular transformation by Ras and Fos. *Oncogene*. 2000;19:1752–1763.
20. **Liao J, Humphrey SE, Poston S, Taparowsky EJ.** Batf promotes growth arrest and terminal differentiation of mouse myeloid leukemia cells. *Mol Cancer Res*. 2011;9:350–363.
21. **Abel EV, Aplin AE.** FOXD3 is a mutant B-RAF-regulated inhibitor of G_1-S progression in melanoma cells. *Cancer Res*. 2010;70:2891–2900.
22. **Schmid CA, Müller A.** FoxD3 is a novel, epigenetically regulated tumor suppressor in gastric carcinogenesis. *Gastroenterology*. 2013;144:22–25.
23. **van Roon EH, Boot A, Dihal AA, et al.** BRAF mutation-specific promoter methylation of FOX genes in colorectal cancer. *Clin Epigenet*. 2013;5:2.
24. **Alkhateeb A, Fain PR, Spritz RA.** Candidate functional promoter variant in the FOXD3 melanoblast developmental regulator gene in autosomal dominant vitiligo. *J Invest Dermatol*. 2005;125:388–391.
25. **Alkhateeb A, Stetler GL, Old W, et al.** Mapping of an autoimmunity susceptibility locus (AIS1) to chromosome 1p31.3-p32.2. *Hum Mol Genet*. 2002;11:661–667.

26. **Cordin O, Banroques J, Tanner NK, Linder P.** The DEAD-box protein family of RNA helicases. *Gene*. 2006;367:17–37.
27. **Yang L, Lin C, Liu ZR.** Phosphorylations of DEAD box p68 RNA helicase are associated with cancer development and cell proliferation. *Mol Cancer Res*. 2005;3:355–363.
28. **Abdelhaleem M.** Do human RNA helicases have a role in cancer? *Biochim Biophys Acta*. 2004;1704:37–46.
29. **Lee K, Pisarska MD, Ko JJ, et al.** Transcriptional factor FOXL2 interacts with DP103 and induces apoptosis. *Biochem Biophys Res Commun*. 2005;336:876–881.
30. **Zahnow CA.** CCAAT/enhancer-binding protein beta: its role in breast cancer and associations with receptor tyrosine kinases. *Expert Rev Mol Med*. 2009;11:e12.
31. **Jendrzejewski J, He H, Radomska HS, et al.** The polymorphism rs944289 predisposes to papillary thyroid carcinoma through a large intergenic noncoding RNA gene of tumor suppressor type. *Proc Natl Acad Sci USA*. 2012;109:8646–8651.
32. **Novembre J, Johnson T, Bryc K, et al.** Genes mirror geography within Europe. *Nature*. 2008;456:98–101.
33. **Campa D, Rizzato C, Bauer AS, et al.** Lack of replication of seven pancreatic cancer susceptibility loci identified in two Asian populations. *Cancer Epidemiol Biomarkers Prev*. 2013;22:320–323.
34. **Hutter CM, Young AM, Ochs-Balcom HM, Carty CL, et al.** Replication of breast cancer GWAS susceptibility loci in the Women's Health Initiative African American SHARe Study. *Cancer Epidemiol Biomarkers Prev*. 2011;20:1950–1959.
35. **Na R, Liu F, Zhang P, et al.** Evaluation of reported prostate cancer risk-associated SNPs from genome-wide association studies of various racial populations in Chinese men. *Prostate*. 2013;73:1623–1635.