IJC
International Journal of Cancer

Fine-mapping of two differentiated thyroid carcinoma susceptibility loci at 9q22.33 and 14q13.3 detects novel candidate functional SNPs in Europeans from metropolitan France and Melanesians from New Caledonia

Catherine Tcheandjieu[1], Fabienne Lesueur[2,3,4,5], Marie Sanchez[1], Dominique Baron-Dubourdieu[6], Anne-Valerie Guizard[7,8], Claire Mulot[9], Pierre Laurent-Puig[9], Claire Schvartz[10], Therese Truong[1] and Pascal Guenel[1]

[1] CESP, INSERM, Université Paris-Sud, Université Paris-Saclay, Villejuif, France
[2] Inserm, U900, Paris, France
[3] Institut Curie, Paris, France
[4] PSL Research University, Paris, France
[5] Mines ParisTech, Fontainebleau, France
[6] Laboratory of Anatomy and Cytopathology, Noumea, New Caledonia, France
[7] Registre Général des tumeurs du Calvados, Centre François Baclesse, Caen, France
[8] U1086 Inserm-UCNB, Cancers and Prevention, Caen, France
[9] Université Paris Descartes, Inserm UMR 5775 EPIGENETEC, Paris, France
[10] Centre de Lutte Contre le Cancer Jean GODINOT, Reims, France

Incidence of differentiated thyroid carcinoma varies considerably between countries and ethnic groups, with particularly high incidence rates in Melanesians of New Caledonia. Differentiated thyroid cancer (DTC) has a familial relative risk higher than other cancers, highlighting the contribution of inherited factors to the disease. Recently, genome-wide association studies (GWAS) identified several DTC susceptibility loci. The most robust associations were reported at loci 9q22 (rs965513 and rs1867277) and 14q13 (rs944289 and rs116909734). In this study, we performed a fine-mapping study of the two gene regions among Europeans and Melanesians from Metropolitan France and New Caledonia. We examined 81 single nucleotide polymorphisms (SNPs) at 9q22 and 561 SNPs at 14q13 in Europeans (625 cases/776 controls) and in Melanesians (244 cases/189 controls). The association with the four SNPs previously identified in GWAS was replicated in Europeans while only rs944289 was replicated in Melanesians. Among Europeans, we found that the two SNPs previously reported at 9q22 were not independently associated to DTC and that rs965513 was the predominant signal; at 14q13, we showed that the haplotype rs944289[C]-rs116909374[C]-rs999460[T] was significantly associated with DTC risk and that the association with rs116909374 differed by smoking status (p-interaction = 0.03). Among Melanesians, a new independent signal was observed at 14q13 for rs1755774 which is strongly correlated to rs2787423; this latter is potentially a functional variant. Significant interactions with parity ($p < 0.05$) and body mass index were observed for rs1755774 and rs2787423. This study contributed to a better characterization of the DTC loci 9q22 and 14q13 in Europeans and in Melanesians and has identified novel variants to be prioritized for further functional studies.

Key words: thyroid cancer, cancer genetics, case–control study, fine-mapping study

Abbreviations: BMI: body mass index; DTC: differentiated thyroid cancer; eQTL: expression quantitative traits; FTC: follicular thyroid cancer; GWAS: genome-wide association study; PTC: papillary thyroid cancer; PWM: position weight matrix; SNP: single nucleotide polymorphism; TC: thyroid cancer; TSH: thyroid stimulating hormone

Additional Supporting Information may be found in the online version of this article.

T.T. and P.G. are co-last authors

Conflict of interest: Nothing to report

Grant sponsors: Fondation de France (FDF), the Association pour la Recherche contre le Cancer (ARC), the Agence Française de Sécurité Sanitaire de l'Environnement et du Travail (AFSSET), Ligue Nationale Contre le Cancer, Institut National du Cancer, EDF

DOI: 10.1002/ijc.30088

History: Received 4 Dec 2015; Accepted 15 Feb 2016; Online 16 Mar 2016

Correspondence to: Therese Truong, PhD, Inserm, CESP, Epidemiology of Cancers, Genes and Environment, 16 Avenue, Paul Vaillant Couturier, Villejuif, France, Tel.: +33-1-45-59-50-31, E-mail: therese.truong@inserm.fr

What's new?

People of Melanesian descent in the South Pacific have an especially high risk of differentiated thyroid cancer (DTC). Certain genetic loci are thought to increase DTC risk. In this case-control, SNP fine-mapping study, the authors compared those candidate loci in Europeans and Melanesians in France and New Caledonia. They found that certain novel SNP variants may be worth investigating further for their role in DTC risk.

Thyroid cancer (TC) is the most common malignancy of the human endocrine system; its incidence is characterized by considerable ethnic and geographic variation.[1,2] One of the highest incidence rates was observed in Melanesian women of New Caledonia (71.4/100,000 person-years),[1–3] a French territory in South Pacific. In metropolitan France, the incidence rate was 12.6/100,000 in women and 4.1/100,000 in men[2] in 2012.

Papillary thyroid carcinoma (PTC) and follicular thyroid carcinoma (FTC) are the most frequent subtypes of differentiated thyroid carcinoma (DTC), representing about 90% of all thyroid cancers. Exposure to ionizing radiations during childhood is a well-established risk factor for DTC. Overweight, deficiencies in iodine intake, parity and late age at menarche are suspected DTC risk factors.[4–6] In addition, excess risk for DTC has consistently been observed among first-degree relatives of DTC patients[7] and it has been shown that TC is one of the cancers with the highest familial risk.[7–9] Linkage studies in multiple-case DTC families have reported several susceptibility loci but no high-penetrance gene could be identified.[10–12] Studies on candidate gene or pathways have reported associations of DTC with common susceptibility variants but only a few have been replicated in independent investigations.[13–15]

Recently, genome-wide association studies (GWAS) on DTC have identified several other susceptibility loci. Rs965513 at locus 9q22 and rs944289 at locus 14q13 were first identified in a GWAS on DTC conducted in Iceland in 192 DTC cases and 37,196 controls.[16] Rs965513 was also associated with PTC in a GWAS conducted in a population exposed to radiation in the Chernobyl area[17] and this association was independently confirmed in a population of Belarusian children exposed to fallout from the power plant accident.[18] Another GWAS on thyroid stimulating hormone (TSH) levels highlighted 22 single nucleotide polymorphisms (SNPs), among which rs116909374 on 14q13 was also associated with DTC risk in addition to rs965513 and rs944289.[19] These findings were replicated in a GWAS conducted in 690 DTC cases and 479 controls from an Italian population with high DTC incidence and in another GWAS conducted in 398 cases and 502 controls from Spain.[20,21] These two GWAS identified additional susceptibility SNPs at other loci.[20,21]

So far, the most robust associations from GWAS were reported for the SNPs rs965513 (9q22, near *FOXE1*) and rs944289 (14q13, between *PTCSC3* and *NKX2-1*). Both *FOXE1* and *NKX2-1* encode thyroid-specific transcription factors (TTFs) involved in thyroid development, differentiation and regulation of thyroid function, making them strong candidate susceptibility genes that may harbor likely pathogenic sequence variants.

Follow-up studies of the 9q22 locus identified a functional SNP rs1867277 located in the 5′ UTR of *FOXE1*, which affects the transcriptional regulation of *FOXE1* through the recruitment of USF1/USF2 (upstream stimulating factors).[22–24] Rs1867277 was suggested as the *FOXE1* causal variant at 9q22 for TC risk by functional studies.[22] Further studies also consistently associated DTC risk with the expansion of a trinucleotide repeat in the unique exon of *FOXE1* encoding polyalanine tract which could result in a gain-of-function mutation.[18,23,25] Recently, He *et al.* showed that rs965513 may also have a functional role in DTC as this SNP is associated with decreased expression of *PTCSC2* [a new long intergenic noncoding RNA (lincRNA) gene] expressed in PTC tissues.[26] Rs965513 is located in the unspliced transcript of *PTCSC2* and in an intron of the spliced *PTCSC2,* and is 60 kb upstream of rs1867277; these two variants are not correlated in Europeans [$r^2 = 0.39$ and $D' = 0.73$ in the European sample of the 1000 genomes (EUR)].

Follow-up study of the 14q13 locus described a lincRNA named *PTCSC3* located 64 kb upstream of the GWAS SNP rs944289.[27] The authors demonstrated that the expression of *PTCSC3* was strongly downregulated in PTC and that the risk allele T of rs944289 was associated with low expression of *PTCSC3* in PTC. The second GWAS SNP at 14q13, rs116909374, is located in a gene desert region, 89 kb downstream of rs944289; these two SNPs are not correlated ($r^2 = 0.006$, $D' = 0.4$ in the EUR population) suggesting the presence of two distinct signals at this locus.

Following the mapping of the two DTC loci, association studies on familial DTC, sporadic DTC and radio-induced DTC have consistently replicated the association between DTC and 9q22 polymorphisms but were less consistent for 14q13 polymorphisms.[18,25,28,29] However, most of the variants identified in these loci are located outside of gene regions. Also it remains unclear whether the polymorphisms identified within each of these two loci are independent and whether they are causally related to DTC.

In this study, using data from studies conducted in Europeans and Melanesians, we aimed to conduct a fine-mapping analysis at loci 9q22 and 14q13 in these two populations and to functionally annotate the most associated variants in order to propose novel candidate risk variants of DTC. Another

objective was to examine whether differences in frequency of risk alleles between Europeans and Melanesians might account for the difference in DTC incidence rates between the two groups. We also explored the interaction between susceptibility alleles and nongenetic risk factors for DTC such as body mass index (BMI), parity, age at menarche, smoking and alcohol drinking.

Material and Methods

Study populations

We used data from two case–control studies on thyroid cancer conducted in metropolitan France (CATHY study) and in New Caledonia, a French territory in South Pacific. All participants signed written consent.

The CATHY study is a population-based case–control study conducted in Marne, Ardennes and Calvados, three French administrative areas (départements) covered by a cancer registry. Eligible cases were all patients diagnosed with DTC between 2002 and 2007 residing in these areas. Controls were frequency-matched to cases by sex and 5-year age groups and were randomly selected among individuals aged 25 years and over, residing in the same areas at the time of cases' diagnosis. From the 621 cases and 706 controls initially included in the study, saliva (Oragene DNA®) or cells from buccal swabs were available for 583 cases and 643 controls; among those, 556 cases and 630 controls declared to be of European ancestry.

The New Caledonia (NC) study is a country-wide population-based case–control study.[6] The case group included patients with DTC diagnosed between 1993 and 1999, who had been living for at least 5 years in NC at the time of diagnosis. The cases were identified from the two pathology laboratories of NC and active searches in medical records of the main hospitals. Age and sex frequency-matched controls were randomly selected from recently updated electoral rolls. A total of 332 cases and 412 controls were included. Of these, 244 cases and 189 controls declared themselves as Melanesians and 42 cases and 133 controls declared themselves as Europeans. DNA was obtained from saliva samples available for 284 Melanesians (164 cases and 120 controls) and 110 Europeans (28 cases and 82 controls).

In both studies, information on ethnicity, personal and familial history of thyroid disease, reproductive factors, exogenous hormone use, weight, height, dietary habits, alcohol intake, tobacco smoking, residential and occupational histories was collected during in-person interviews.

Markers selection and genotyping

For the fine-mapping analysis, we defined at locus 9q22, an interval that encompassed rs965513 and the *FOXE1* gene [62 kb, positions 100,556,108–100,671,478 (NCBI build 37/hg19 assembly)] and at locus 14q13, an interval that encompassed *PTCSC3* and *NKX2-1* genes [375 kb; positions 36,601,061–36,999,637 (NCIB build 37/hg19 assembly)]. We used tagger software[30] (pairwise approach with $r^2 \geq 0.8$) to capture the genetic variation within the selected regions. We selected tagging SNPs with a minimum minor allele frequency (MAF) of 0.05 in the Caucasian population (CEU) as reported in the HapMap Project (Data Release 21/Phase II).

At locus 9q22, 11 tagging SNPs (including the GWAS SNP rs965513 and the functional SNP rs1867277) were selected. At locus 14q13, 47 tagging SNPs (including the GWAS SNPS rs944589 and rs116909374) were selected. A total of 58 SNPs were genotyped using the Fluidigm Dynamic array chip.

Quality control

Oligonucleotide primers and probes were validated on 90 CEPH DNA samples. All genotyped SNPs had a call rate- $> 95\%$. From The CATHY study, 64 cases/65 controls with DNA prepared from oral brush sample were not genotyped due to their lower DNA concentration (<5 µl). We excluded 19 cases/11 controls because of genotyping completion rate $<90\%$. Among those, nine cases/nine controls were from NC study. The deviation of the genotype proportions from Hardy–Weinberg equilibrium (HWE) was assessed in both European controls and Melanesian controls using chi-squared test with one degree of freedom. We applied a Bonferroni correction for multiple testing using a p values threshold of 8×10^{-4}; no SNP deviated from HWE. Because the genotype frequencies of controls in Europeans from the New Caledonia study and in Europeans from the CATHY study were similar, we combined these two populations in the analyses. In total, 1,134 Europeans (508 cases/626 controls) and 268 Melanesians (156 cases/112 controls) were considered for analysis (Table 1).

Regional imputation

Ungenotyped SNPs with MAF > 0.02 in the selected regions were imputed using IMPUTE2.2 software[31] and the cosmopolitan reference panel from the 1000 genomes dataset (September 2013 release). The use of this panel as reference haplotype is particularly recommended for admixed population such as the Oceanian population.[32] We imputed separately genotypes of Europeans and genotypes of Melanesians for each locus. We evaluated the accuracy of the imputation by masking a random subset of the available genotypes and checking whether these genotypes could be recovered by imputation. The concordance between imputed and true genotypes was 95% in the two populations. For the analyses, we excluded the imputed SNPs that deviated from HWE in controls and the imputed SNPs with imputation performance (INFO-score < 0.70).

Among Europeans, respectively, 56 imputed SNPs at loci 9q22 and 478 imputed SNPs 14q13 were considered for analysis; in Melanesians 69 imputed SNPs at locus 9q22 and 464 imputed SNPs at locus 14q13 were considered for analysis.

Statistical analyses

Single marker association tests were performed using unconditional logistic regression assuming a log-additive model. Imputed SNPs were analyzed as expected genotype count (gene dosage) with SNPtest V2.4.1 software. We conducted

Table 1. Characteristics of the study sample stratified by ethnicity

	Europeans		Melanesians	
Characteristics	Cases, *n* (%)	Controls, *n* (%)	Cases, *n* (%)	Controls, *n* (%)
Mean age (years)	50.9	51.4	45.5	45.7
Sex (%)				
Females	399 (75.2)	459 (64.4)	148 (94.9)	104 (92.8)
Males	109 (24.8)	167 (35.6)	8 (5.1)	8 (7.1)
Study population				
CATHY	481 (94.7)	544 (87.0)	–	–
NC1	29 (5.3)	82 (13.0)	156 (100)	112 (100)
Area of residence in CATHY (%)				
Ardennes	111 (21.8)	99 (15.8)	–	–
Calvados	152 (29.9)	189 (30.2)	–	–
Marne	218 (42.9)	256 (40.9)	–	–
Province of residence in NC[1] (%)				
North	3 (0.11)	6 (0.07)	32 (20.5)	26 (23.2)
South	23 (0.85)	75 (0.91)	54 (34.6)	47 (41.9)
Island	1 (0.04)	1 (0.04)	70 (44.8)	39 (34.8)
Type of cancer (%)				
Papillary	447 (88.0)	–	131 (84.0)	–
Vesicular	61 (12.0)	–	25 (16.0)	–
Total (*N*)	508	626	156	112

[1]NC: New Caledonia.

conditional analyses to test for residual association after accounting for a key SNP. We applied Bonferroni correction based on the effective number of independent tests estimated by the method of Li *et al.*[33] We considered that SNPs with p values $< 4.4 \times 10^{-4}$ (based on our estimation of 115 independent tests) were significantly associated with DTC. Haplotype analysis was carried out using the R package Haplot.Stat. All the analyses were conducted separately in Europeans and Melanesians and were adjusted on age, sex and area of residence. Europeans of NC were grouped into one single area of residence (Table 1).

We investigated the interaction between SNPs showing the strongest association with DTC risk (including the GWAS SNPs) at each locus and suspected lifestyle risk factors of DTC such as BMI, alcohol intake and cigarette smoking; we also investigated interaction with parity, age at menarche and irregular periods in women. We used the likelihood ratio test comparing models with and without the interaction term. These models were adjusted for age, sex, area of residence and all the suspected lifestyle risk factors.

We used the R package ggplot2 and locuszoom1.3 software[34] for plotting our result of single markers association and conditional regression at each locus.

Functional annotation

To assess the possible functional role of the most associated SNPs we used the following freely available tools that investigate noncoding genome variants: RegulomeDB (http://regulome.stanford.edu) and Haploreg V2 (www.broadinstitute.org/mammals/haploreg). These tools predict whether a given SNP may alter the regulatory protein binding sites, chromatin structure, histone modifications and putative transcription factors (TFs) binding sites. The prediction is based on information from Chip-Seq data or position weight matrix analysis (PWMs). In addition, we used is-rSNP (http://bioinformatics.research.nicta.com.au/software/is-rsnp) to determine the statistical significance of the nucleotide variation for altering binding of TFs using a PWM score. To examine the association between associated SNPs and expression level of a nearby gene, we used the freely available expression quantitative trait loci (eQTL) data of SNPexp tools (http://app3.titan.uio.no/biotools). With this tool, we calculated an adjusted p values (with Bonferroni correction) applying additive model, the correlation between genotype of each SNP and expression levels for probes representing the gene of interest using the 210 HapMap phase II dataset. We also used UCSC (http://genome-euro.ucsc.edu) and Ensembl (http://www.ensembl.org) genomes databases to identify possible uncharacterized lincRNA, predicted transcripts and ESTs (Express Sequence Tags) at each locus.

Results

The characteristics of the study sample are presented in Table 1. Men represented 25% of the cases in Europeans and 5% in Melanesians.

Locus 9q22

In Europeans and in Melanesians, rs965513 and rs1867277, previously associated with DTC risk in GWAS or functional studies, were in separate LD blocks. The LD block containing rs1867277 encompasses *FOXE1* and part of a gene desert region; this LD block was slightly larger in Melanesians than in Europeans (Supporting Information Fig. S1).

In Europeans, we analyzed 69 genotyped or imputed SNPs. The associations with alleles rs965513[A] (OR = 1.52, $p = 6 \times 10^{-6}$) and rs1867277[A] (OR = 1.38, $p = 3 \times 10^{-4}$) were replicated (Table 2; Fig. 1*a*). The most significant association was observed for the imputed SNP rs13295081[T] (OR = 1.44, $p = 5 \times 10^{-6}$). This marker was strongly correlated with rs965513 ($r^2 = 1$) that displayed the second most significant association (Fig. 2*a*). From the 50 other SNPs that reached statistical significance ($p < 4.4 \times 10^{-4}$) (Fig. 1*a* and Supporting Information Table S1), 12 SNPs were strongly to moderately correlated with rs13295081 ($0.6 < r^2 \leq 1.0$) (Fig. 2*a*) and 34 SNPs were strongly correlated with rs1867277 ($r^2 > 0.8$).

When conditioning the regression test on rs13295081 or rs965513 (Fig. 2*b*), no SNP (including rs1867277) remained significantly associated with DTC. When conditioning on rs1867277, the *p*-values for rs13295081 and rs965513 remained suggestive of an association ($p = 6 \times 10^{-3}$) (Fig. 2*c*), indicating predominant signal from the LD block containing rs965513 and a nonindependence of the signals from the LD blocks containing rs965513 and rs1867277. Haplotype analysis did not reveal any association beyond the single SNP analysis (data not shown).

In Melanesians, we analyzed 75 genotyped or imputed SNPs. No SNP was statistically significantly associated with DTC (Fig. 1*b*), but suggestive associations were observed for rs965513[A] (OR = 1.43; $p = 0.03$) and rs1867277[A] (OR = 1.34; $p = 0.06$), with effects of the same order of magnitude as in Europeans (Table 2).

Locus 14q13

The size of the LD blocks at 14q13 was slightly smaller in Melanesians than in Europeans. Moreover, the allele frequencies of several SNPs differed considerably between the two ethnic groups (Supporting Information Fig. S2; Supporting Information Table S3).

In Europeans, 522 genotyped or imputed SNPs were considered. Association between DTC and GWAS SNPs rs944289[C] (OR = 0.74, $p = 4.7 \times 10^{-4}$) and rs116909374[T] (OR = 1.97; $p = 3.3 \times 10^{-3}$) was at the limit of statistical significance after correction for multiple testing (Table 2), whereas the association with the imputed SNP rs7494749[A] (OR = 0.73, $p = 4.3 \times 10^{-4}$) was statistically significant (Table 2, Fig. 1*c*). This marker is located in the *PTCSC3* third intron and is strongly correlated with rs944289 ($r^2 = 0.98$; Fig. 2*d*), which displayed the second strongest association with DTC in Europeans.

Table 2. Single markers association results for DTC, for a selection of SNPs (GWAS SNP and possible functional SNP that are highly correlated to a top SNP) for loci 9q22.33 and 14q13.3 in Europeans and Melanesians

			Allele		Europeans				Melanesians			
					MAF[1]				MAF[1]			
Locus	SNPs	Position	Ref[2]	Risk	Case	Control	OR (95% CI)	*p*-values	Case	Controls	OR (95% CI)	*p* values
9q22.33	rs13295081[3]	100559011	C	T	0.49	0.40	1.44 (1.22–1.70)	5.46×10^{-6}	–	–	–	–
	rs965513[4]	100556109	G	A	0.43	0.33	1.52 (1.28–1.81)	5.87×10^{-6}	0.38	0.30	1.43 (1.00–2.07)	0.03
	rs10759944	100556972	G	A	0.43	0.34	1.48 (1.25–1.76)	2.25×10^{-5}	0.39	0.31	1.42 (0.95–2.05)	0.03
	rs1867277[4]	100615914	G	A	0.46	0.38	1.38 (1.17–1.63)	2.59×10^{-4}	0.36	0.29	1.34 (0.93–1.94)	0.06
14q13.3	rs1755774	36688516	A	G	0.29	0.34	0.76 (0.63–0.92)	7.29×10^{-3}	0.53	0.70	0.50 (0.35–0.71)	1.62×10^{-5}
	rs2787423	36687054	G	A	0.29	0.34	0.78 (0.65–0.94)	8.78×10^{-3}	0.54	0.69	0.51 (0.35–0.73)	3.69×10^{-5}
	rs7494749	36615427	G	A	0.38	0.45	0.73 (0.62–0.87)	4.34×10^{-4}	0.64	0.76	0.57 (0.39–0.83)	7.94×10^{-4}
	rs944289[4]	36649246	T	C	0.34	0.41	0.74 (0.61–0.89)	4.78×10^{-4}	0.62	0.74	0.57 (0.39–0.83)	1.13×10^{-3}
	rs116909374[4]	36738362	C	T	0.05	0.03	1.97 (1.26–3.10)	3.25×10^{-3}	0.01	0.004	1.44 (0.13–15.96)	0.76
	rs999460	36984511	C	T	0.34	0.39	0.8 (0.68–0.96)	5.18×10^{-3}	0.21	0.19	1.14 (0.74–1.76)	0.59

[1]MAF, minor allele frequency and frequency of risk allele. The odds ratios are adjusted for age, sex and area of residence.
[2]Ref, reference allele.
[3]rs13295081 did not pass quality control in Melanesians.
[4]GWAS SNPs.

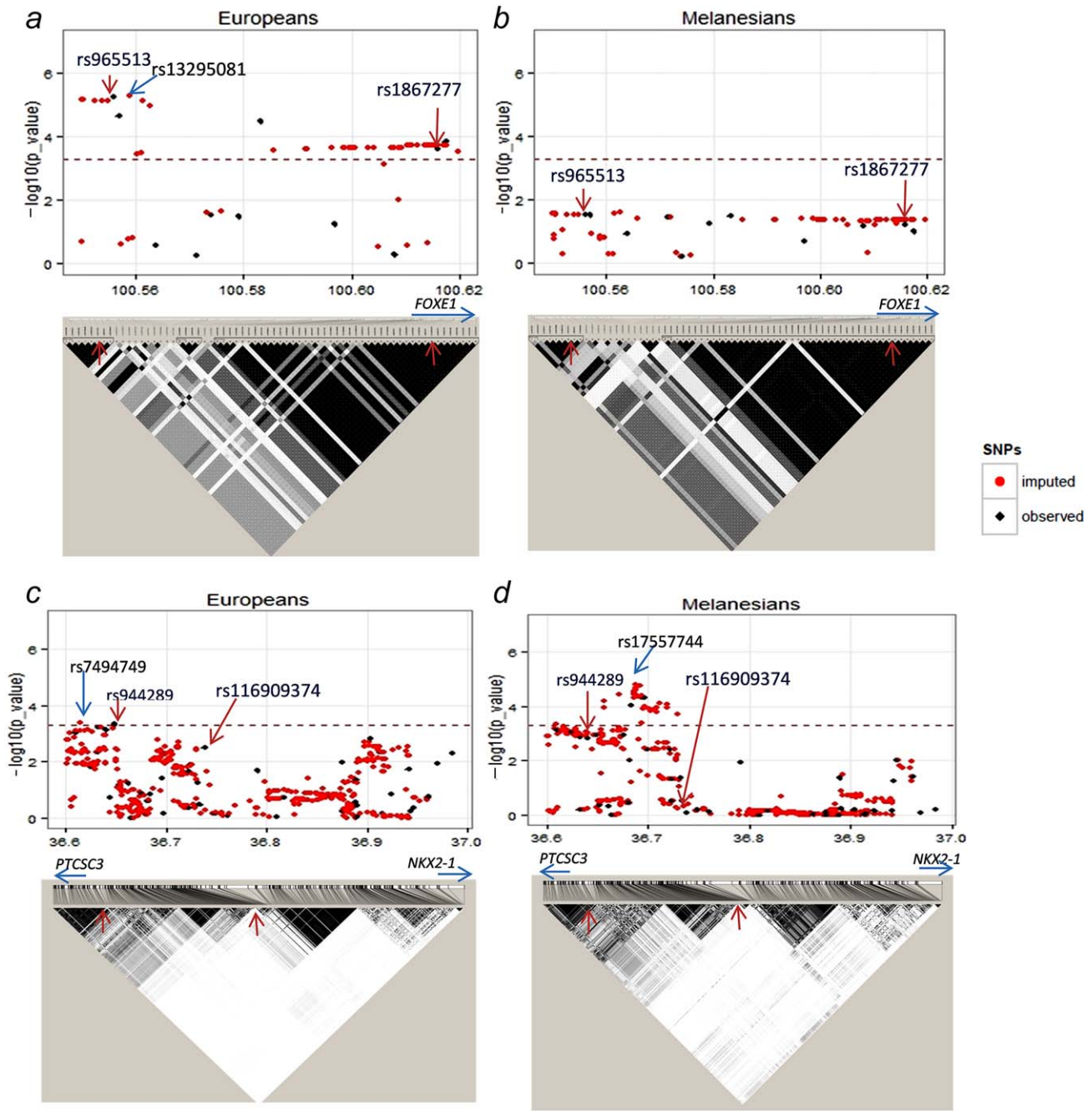

Figure 1. Manhattan plot of the DTC risk loci 9q22 and 14q13 and heatmap of correlation between SNPs at each locus for Europeans (*a*) and for Melanesians (*b*). Genotyped and imputed SNPs are plotted according to their chromosomal positions on the *x* axis and their overall *p* values (−log10 *p* values) on the *y* axis. The black rhombuses indicate observed SNPs and the red circle, the imputed SNP. The red line indicates the threshold of significance, the red arrow indicates the GWAS SNPs and the blue arrow indicates the most robust associated SNPs in our study sample. On the heat map, the blue arrow indicates the gene and its position on the chromosome. (*a*) Manhattan plot of locus and heatmap of correlation between SNPs at locus 9q22 in Europeans. (*b*) Manhattan plot of locus and heatmap of correlation between SNPs at locus 9q22 in Melanesians. (*c*) Manhattan plot of locus and heatmap of correlation between SNPs at locus 14q13 in Europeans. (*d*) Manhattan plot of locus and heatmap of correlation between SNPs at locus 14q13 in Melanesians.

Conditional analysis on rs7494749 or rs944289 attenuated the association *p*-values of DTC with other SNPs, except for the GWAS SNP rs116909374 and SNPs near *NKX2-1* for which, the *p*-values remained suggestive of an association ($p < 4 \times 10^{-3}$; Fig. 2*e*). When conditioning the analysis on the GWAS SNPs rs944289 and rs116909374, rs999460 remained the most associated SNP with a *p* values at the limit of significance (Fig. 2*f*). These results suggest that several SNPs in the locus are independently associated with DTC risk. We performed an haplotype analysis based on rs944289, rs116909374

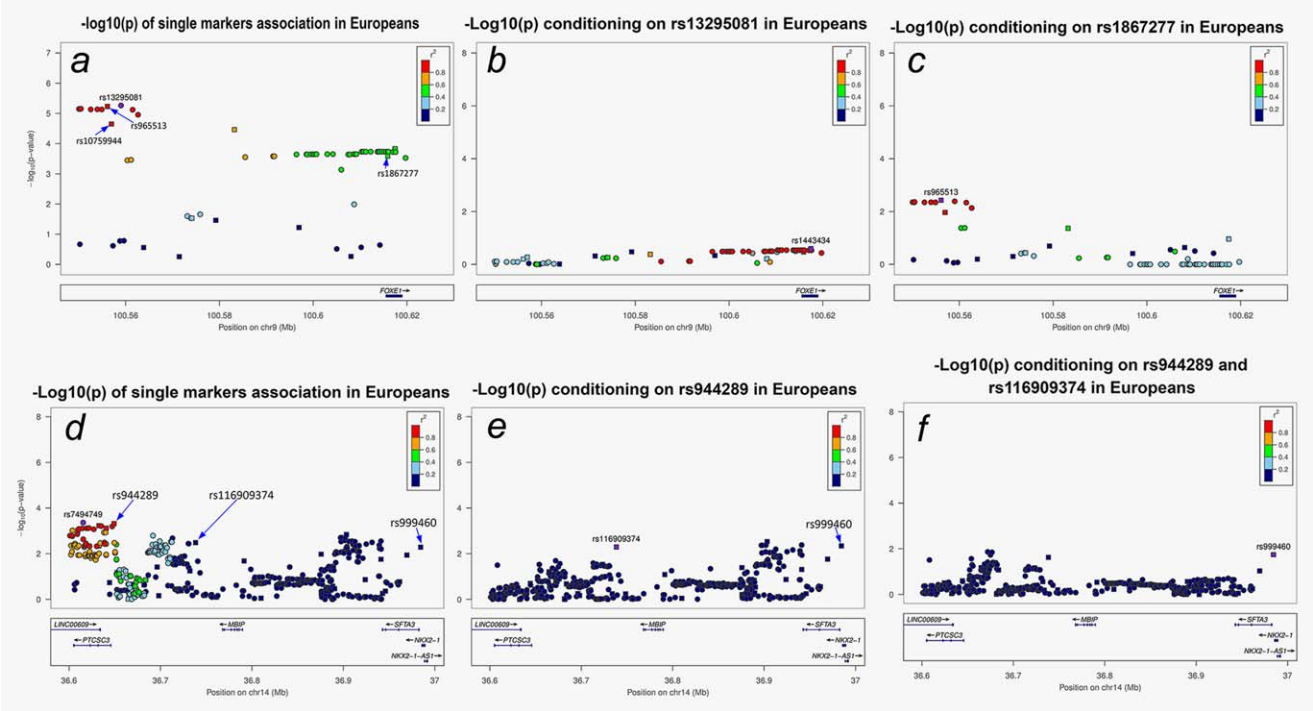

Figure 2. Regional plot of results of single-marker association and results of conditional regression, at loci 9q22 and 14q13, in Europeans. (*a* and *d*) Plot of –log10 (*p* values) of a single marker association results at loci 9q22 (*a*) and 14q13 (*d*) in Europeans. (*b* and *c*) The results of –log10 (*p* values) of SNPs after conditioning the regression test on the most robust signal rs13295081 at locus 9q22 (*b*) or on the previously describe functional SNP rs1867277 (*c*) in Europeans. (*e* and *f*) The results of –log10 (*p* values) of SNPs after conditioning the regression test on the most robust signal rs7494749 at locus 14q13 (*e*) or on the GWAS SNP rs116909374 (*f*) in Europeans. On *y* axis, are the –log10 (*p* values); on *x* axis, the position of SNPs and gene on the chromosome. The color of each SNP spot reflects its r^2 with the most associated SNP (in purple). The correlation between SNP is ranged from high (in red) to low (in blue); the arrows show the GWAS SNP in the region. The square indicated observed SNPs and the circle the imputed SNPs.

Table 3. OR estimations for common haplotypes based on rs944289, rs116909374, rs999460 at locus 14q13 in Europeans

				Frequencies					
Haplotype	rs944289	rs116909374	rs999460	Cases	Controls	Pools	OR	CI	*p* values
5	T	C	C	0.39	0.36	0.38	ref	–	–
1	C	C	C	0.23	0.23	0.23	0.93	(0.74–1.15)	0.48
6	T	C	T	0.22	0.21	0.21	0.93	(0.74–1.16)	0.52
2	C	C	T	0.1	0.17	0.13	0.51	(0.38–0.68)	3.11×10^{-6}
8	T	T	C	0.03	0.01	0.02	2.47	(1.25–4.86)	8.9×10^{-3}
7	C	T	T	0.02	0.006	0.01	3.16	(1.27–7.86)	0.01

The odds ratios are adjusted for age, sex and area of residence.

and rs999460 and we observed a significantly decreased risk of DTC in individuals carrying the haplotype rs944289[C]-rs116909374[C]-rs999460[T] (OR = 0.51, $p = 3.11 \times 10^{-6}$) when compared to individuals with the most frequent haplotype rs944289[T]-rs116909374[C]-rs999460[C] (Table 3).

In Melanesians, 511 imputed and genotyped SNPs were analyzed. The per-allele OR reported for the GWAS SNP rs944289[C] (OR = 0.57, $p = 1.1 \times 10^{-3}$) was at the limit of statistical significance whereas the other GWAS SNP rs116909374 was not associated with DTC (OR = 1.44, $p = 0.76$; Table 2). The OR for rs944289 in Melanesians was close to that reported in Europeans, but a much higher frequency of the rs944289[T] allele was observed. In this population, the imputed SNP rs1755774[G] was the most strongly associated SNP with DTC (OR = 0.50, $p = 1.6 \times 10^{-5}$; Table 2, Fig. 1*d*). Rs1755774 was weakly correlated with rs944289 ($r^2 = 0.27$) and rs116909374 ($r^2 = 0.14$) (Fig. 3*a*) in Melanesians. Of note, rs1755774 was also associated with DTC in Europeans but the association was at the limit of significance (OR = 0.76, $p = 7.3 \times 10^{-3}$; Fig. 1*c*). The additional 37 SNPs that reached statistical significance in Melanesians ($p < 4.4 \times 10^{-4}$; Supporting Information Table S3) were all located in the intergenic region between *PTCSC3* and *MBIP* and most of them were strongly correlated with rs1755774

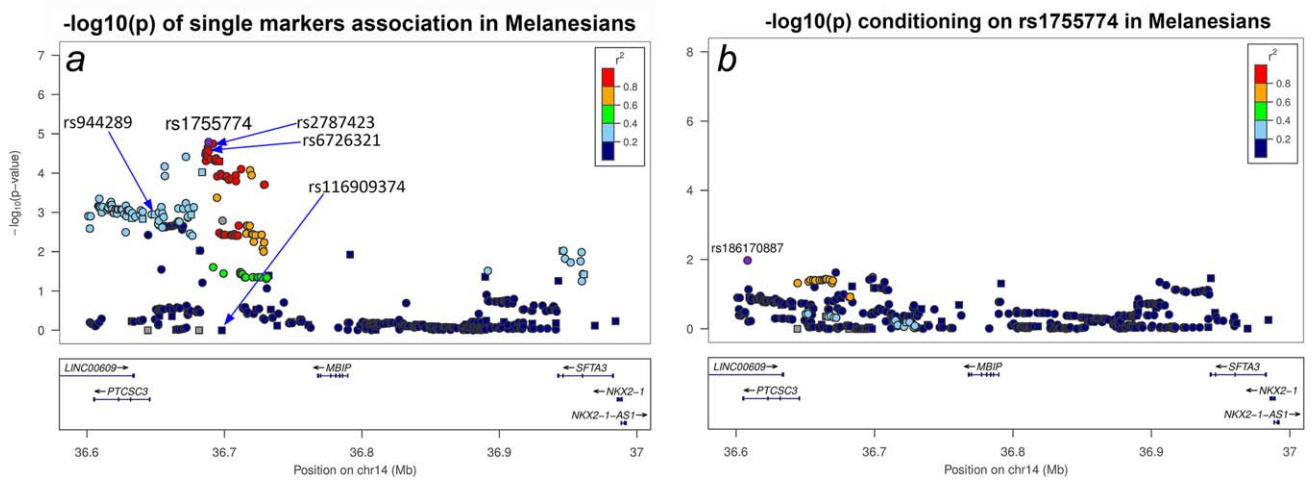

Figure 3. Regional plot of results of single-marker association and results of conditional regression, at locus 14q13, in Melanesians. (*a*) Plot of –log10 (*p* values) of a single marker association results at loci 14q13 in Melanesians. (*b*) Plot of –log10 (*p* values) of SNPs after conditioning the regression test on the most robust signal rs1755774 at locus 14q13 in Melanesians. On *y* axis, the –log10 (*p* values); on *x* axis, the position of SNPs and gene on the chromosome. The color of each SNP spot reflects its r^2 with the most associated SNP in purple. The correlation is ranged from high (in red) to low (in blue); the arrows show the GWAS SNP in the region. The squares represent observed SNPs and the circles the imputed SNPs.

($r^2 > 0.8$; Fig. 3*a*). No SNP remained statistically significant after conditioning the association tests on rs1755774 (Fig. 3*b*), suggesting one independent signal only at 14q13 in Melanesians.

In Europeans, 88% of cases were PTC while 12% were FTC; in Melanesians, the numbers were 84% and 16%, respectively (Table 1). When restricting the analyses to PTC, the ORs in both ethnic groups were similar to those observed for PTC and FTC combined (Supporting Information Tables S1 and S2).

Difference in allelic frequency between Europeans and Melanesians

We compared frequencies of risk alleles of the most significant SNPs at 9q22 and 14q13 between Melanesians and Europeans to see whether they may account for the differences in DTC incidence rates between ethnic groups. We found that the risk alleles were more frequent in Europeans than in Melanesians, and hence could not explain the high DTC incidence rates among Melanesians relative to Europeans.

Effect modification by nongenetic risk factors of DTC

Supporting Information Figures 4 and 5 show forest plots of the associations between DTC risk and candidate SNPs at loci 9q22 and 14q13, stratified on categories of BMI, alcohol consumption, smoking status, age at first menarche, irregular menstrual cycles and number of pregnancies among European and Melanesian women separately. Analysis of interaction in European men was not possible due to limited sample size in each strata.

At locus 9q22, there was no evidence of interaction between the previously highlighted SNPs and the nongenetic risk factors (Supporting Information Fig. S4).

At locus 14q13, we observed a statistically significant interaction between the GWAS SNP rs116909374 and smoking status among European women (p-interaction = 0.03). The association between rs116909374[T] allele and DTC risk was stronger in current smoker [OR = 11.06 (95% CI: 1.26–96.82)] than in former and never smokers [OR = 1.63 (95% CI: 0.50–5.38); OR = 0.77 (95% CI: 0.36–1.64), respectively] (Supporting Information Fig. S5). Interaction between smoking status and rs116909374 could not be examined in Melanesian women due to a very low risk allele frequency (MAF $\leq$ 0.01).

In Melanesians, we found that the association between DTC risk and rs1755774 differed by parity (p-interaction = 0.04) and BMI (p-interaction = 0.05). Indeed, the association between DTC and rs1755774 was weaker in women who had two or more pregnancies [OR = 0.26 (95% CI: 0.13–0.49)] than and in women who had <2 pregnancies [OR = 0.66 (95% CI: 0.22–1.97)] (Supporting Information Fig. S5). For the same SNP, women with BMI > 25 kg/m^2 had an OR of 0.26 (95% CI: 0.14–0.49) while women with BMI $\leq$ 25 kg/m^2 had an OR of 0.56 (95% CI: 0.18–1.70). No such effect was apparent in European women.

Discussion

At 9q22, the previously described associations between DTC risk and SNPs rs965513 and rs1867277[16,19,20,22,29] were replicated in our European population [OR = 1.52 for rs965513 ($p = 6 \times 10^{-6}$) and OR = 1.38 for of rs1867277 ($p = 3 \times 10^{-4}$)]. In Melanesians, effects of same magnitude were observed for rs965513 and rs1867277 although the p-values did not reach the level of statistical significance, possibly due to small sample size (OR = 1.43, $p = 0.03$ and OR = 1.34, $p = 0.06$, respectively, for rs965513 and rs1867277).

Using conditional analyses we showed that the associations of DTC with rs965513 and rs1867277 were not independent. This finding is consistent with the candidate gene study by Jones *et al.*[35] that reported reduced signals of the two SNPs after reciprocal adjustment. Because the predominant association signal at locus 9q22 lied in the LD block containing rs965513, we annotated this block *in silico* looking for the presence of a potential functional SNP.

Among SNPs predicted to alter TFs or regulatory protein-binding sites, we identified rs10759944 ($r^2 = 0.99$ with rs965513) as a possible functional SNP as it is located in the binding site of the regulatory proteins ERALPHA (ESR1) according to chip-Seq data and its risk allele A was suggested by PWM analysis to disrupt the binding site of some TFs such as EWSR1-FLI1, E2F1 and AP-2α (TFAP2A) (data not shown). Indeed, ESR1 is a TF activated by estrogen binding. Steroid hormones and their receptors are involved in the regulation of gene expression and affect cellular proliferation and differentiation in target tissues. Overexpression of ESR1 receptors (ERα) was reported in PTC[36] or thyroid cells.[37] E2F1 is a TF controlling cell cycle and acting as tumor suppressor protein.[38] AP-2α (TFAP2A) is a protein that interacts with cellular enhancer elements to regulate transcription of selected genes. These proteins are overexpressed in PTC[36,38,39] and seem to play an important role in PTC pathogenesis.

The biological role of rs10759944 in thyroid tissue was investigated in a recent study that performed *in vivo* experiments on variants prioritized from a larger region of locus 9q22 than in our study.[40] This study reported that at least three regulatory elements functioning as enhancers lie in a LD block containing rs965513, and that these enhancers harbor at least four functional SNPs: rs7864322, rs12352658, rs7847449 and rs10759944.[40] They also showed that the three enhancer elements interact with the shared promoter region of *FOXE1* and *PTCSC2* in a human PTC cell line and this interaction was also retrieved in unaffected thyroid tissue for the enhancer containing rs1079944. Among these four SNPs, we analyzed only rs10759944; rs7864322 was not imputed due to its location over 1,000 kb upstream of our fine-mapping boundary; imputation of rs12352658 and rs7847449 did not pass the quality controls and was excluded from the analysis (info-score < 0.7).

We reported that rs10759944 was associated with DTC risk in Europeans (OR = 1.48, $p = 2 \times 10^{-5}$) and in Melanesians (OR = 1.42, $p = 0.03$). This association was also reported by the GWAS conducted in Iceland, the GWAS on radiation-related PTC in Belarussian patients, and a study on familial DTC cases.[16,17,25]

At locus 14q13, the associations previously reported for the GWAS SNPs rs944289 (OR = 0.74, $p = 5 \times 10^{-4}$) and rs116909374 (OR = 1.97, $p = 3 \times 10^{-3}$) were replicated in Europeans. In Melanesians, rs944289 was replicated (OR = 0.57, $p = 1 \times 10^{-3}$), but not rs116909374 (OR = 1.44, $p = 0.76$), which is rare in this population (MAF = 0.004). Moreover, in Europeans we found one haplotype rs944289[C]-rs116909374[C]-rs999460[T] that was significantly associated with DTC risk (OR = 0.51, $p = 3 \times 10^{-6}$). This result suggests a real combined effect of rs944289, rs116909374 and rs999460 on DTC risk or a yet to be characterized causal SNP tagged by this haplotype.

Rs944289[T] is positioned in the binding site of the TFs CEBP/α and CEBP/β and had been associated with variation in expression levels of *PTCSC3* in PTC tissues.[27] To further investigate whether the risk alleles rs116909374 and rs999460 could also be biologically relevant to DTC pathogenesis, we used range of annotation tools.

Interestingly, chip-Seq data revealed that rs116909374 was located in the binding site of the regulatory protein ESR1. Rs116909374 is also located in an enhancer region in H1 cell line and in a promoter of an uncharacterized lincRNA named TCONS_0022711, which is expressed in thyroid tissues.

Rs999460 is positioned in the poised promoter in H1 cell line and in a repressed polycomb of numerous cell lines. Chip-Seq data revealed that rs999460 falls also in the binding site of EZH2, and analysis using PWMs indicates that its risk allele may alter the binding of the TFs ZNF143 and NKX2-1 (data not shown). EZH2 is the functional enzymatic component of the polycomb repressive complex 2 (PRC2) responsible for healthy embryonic development through the epigenetic maintenance of genes regulating development and differentiation.[41] EZH2 is also involved in upregulation of NKX2-1[41]; however, using SNPexp tools, we did not find significant association between the expression level of *NXK2-1* gene and rs999460. Therefore, further *in vivo* investigation regarding the potential role of rs116909374 and rs999460 in DTC pathogenesis is warranted.

Following the identification of the 14q13 locus in the first GWAS on DTC,[16] few studies replicated the association with rs944289 in other populations. Bonora *et al.*[25] investigated the candidate SNP and four other SNPs tagging the 14q33 region in families highly predisposed to DTC and showed that in addition to rs944289, rs999460 was significantly associated with the disease.[25] Also, consistently with the study of Gudmundsson *et al.*,[19] we showed that the GWAS SNPs rs944289 and rs116909374 are independently associated with DTC in Europeans.

In Melanesians, at 14q13, the pattern is different. The strongest association was reported for rs1755774 (OR = 0.50, $p = 1.6 \times 10^{-5}$), which is not correlated with rs944289 or rs116909374. No haplotype outperforming the single SNP analysis was shown to be associated with DTC, and conditional analysis suggested that rs1755774 is the only signal of association at this locus in Melanesians.

Fourteen SNPs located in the LD block containing rs1755774 and significantly associated to DTC in Melanesians showed putative functional prediction. Among those, rs2787423 and rs67263215 had predictive function that can be related to thyroid cancer. These SNPs were located in the binding site of the regulatory proteins STAT3 and FOS (Supporting Information Fig. S3). Rs2787423[G] also seems to alter the binding site of p300, CEBP/α and CEBP/β TFs, which are all expressed in thyroid and suppressed in tumor

tissue[27] while rs67263215[T] alters the binding site of BDP1, TATA box and YY1, which form together multiprotein complexes involved in the initiation of the transcription of some gene.[42,43] Moreover, rs2787423 and rs67263215 also fall in the enhancer region in human mammary epithelial cells (HMEC) upstream of an uncharacterized lincRNA or uncertain coding transcript named TCONS_0002269, which is abundantly expressed in thyroid tissues (Supporting Information Fig. S3). However, functional studies are necessary to confirm the importance of this variant and to better characterize the region.

So far, genetics studies on DTC have been conducted in different populations such as Chinese, Japanese, Cuban or Polynesians[28,29,44,45] but to our knowledge, this is the first study that investigated the role of common SNPs among Melanesians of New Caledonia, a population with one of the highest incidence rates of DTC in the world.[3] The incidence of DTC in Melanesian women was tenfold higher than in European women.[3] No clear explanation is currently available for this observed high incidence rates. Overdiagnosis of occult thyroid microcarcinomas due to intensive medical screening practices of thyroid nodules cannot be a unique explanation, as the incidence of large size carcinomas >10 mm remains elevated in Melanesians. In contrast to French Polynesia, no nuclear experiment was conducted in New Caledonia, making radiation exposure an unlikely explanation for the elevated incidence. In addition, striking difference in incidence rates between the ethnic groups in this territory does not support this hypothesis. We have shown previously that high parity and obesity may partly account for the elevated incidence of DTC in Melanesian women of New Caledonia.[46,47] In this study, we showed that the difference in allelic distribution of susceptibility SNPs at loci 9q22 and 14q13 may not contribute to the difference in DTC incidence between Europeans and Melanesians.

In addition to be the first study on the GWAS loci 9q22 and 14q13 in Melanesians, this is also to our knowledge, the first fine-mapping study of 14q13 locus in Europeans. Other strength of our study is the availability of data on lifestyle-related risk factors, which allowed studying gene–environment interactions. Therefore, we examined whether risk factors for DTC, including BMI, parity, tobacco smoking and alcohol consumption, modified the associations between SNPs and DTC. Gene–environment interactions in DTC have been investigated in only two previous studies where the authors investigated interaction between parity, BSA (body surface area), BMI, iodine intake or thyroid radiation dose and the GWAS SNPs at loci 9q22 and 14q13.[47,48] Consistently with these studies, we did not observe interactions between the SNPs and parity or BMI in European women. However, we reported that the association of rs116909374 with DTC in European women was stronger in current smokers (OR = 11.06, $p = 2.9 \times 10^{-2}$) than in never or former smokers (p-interaction = 0.03) (Supporting Information Figs. S5–SE). This interaction may be explained by the independent effects of cigarette smoking and rs116909374[T] on decreased TSH levels,[19,48] which in turn are associated with increased risk of DTC.[49] Gudmundsson *et al.*[19] hypothesized that low TSH levels could lead to less differentiation of the thyroid epithelium, causing a higher predisposition to malignant cell transformation. In Melanesians, we found that parity modified the association of DTC with rs2787423 or rs1755774 at 14q13. However, these results should be interpreted with caution given the small number of subjects in each stratum and need to be confirmed in other studies.

Our study has some limitation: first, the lack of power to detect associations in Melanesians; second, the fine-mapping study was based on imputed SNPs and we used *in silico* analysis to predict a potential biological function. This implied the need for replication of our results in other studies and the need for functional studies to confirm the biological prediction. The ethnicity was self-declared by study subjects possibly leading to misclassification particularly among Europeans. Also, DTC constitutes a heterogeneous set of diseases with various molecular signatures in the tumor genotype that have been associated with tumor aggressiveness.[50] It is possible that molecular signature of the tumor also differs according to the etiology of DTC, but this could not be examined in our study as no information on tumor molecular profile was available.

In summary, the association of DTC with SNPs at loci 9q22 and 14q13 was confirmed in Europeans. We showed that the two GWAS SNPs at locus 9q22 were not independently associated to DTC and that rs965513 is the predominant signal in this region. We also showed that rs10759944, strongly correlated to rs965513 and recently reported as a functional SNP, is associated with PTC in Europeans and Melanesians. At locus 14q13, we found a haplotype associated with DTC risk in Europeans and identified among Melanesians of New Caledonia, new susceptibility SNPs that are also associated to DTC risk in Europeans. We also performed a comprehensive bioinformatics assessment of the associated SNPs at the two loci and proposed some novel candidate SNPs that may be related to DTC pathogenesis although further functional studies are needed to confirm their role in DTC. Further studies on other suspected DTC susceptibility loci should also be conducted to explain the elevated incidence rates in Melanesians.

References

1. Moore MA, Baumann F, Foliaki S, et al. Cancer epidemiology in the pacific islands—past, present and future. *Asian Pac J Cancer Prev* 2010;11:99–106.
2. Ferlay J, Soerjomataram I, Ervik M, et al. GLOBOCAN 2012 v1.0, cancer incidence and mortality worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer. Available at: http://globocan.iarc.fr/Pages/references.aspx (accessed October 30, 2014).
3. Truong T, Rougier Y, Dubourdieu D, et al. Time trends and geographic variations for thyroid cancer in New Caledonia, a very high incidence area (1985-1999). *Eur J Cancer Prev* 2007;16:62–70.
4. Truong T, Orsi L, Dubourdieu D, et al. Role of goiter and of menstrual and reproductive factors in thyroid cancer: a population-based case-control study in New Caledonia (South Pacific), a very high incidence area. *Am J Epidemiol* 2005;161:1056–65.
5. Leux C, Guénel P. Risk factors of thyroid tumors: role of environmental and occupational exposures

to chemical pollutants. *Rev Epidemiol Sante Publique* 2010;58:359–67.
6. Cléro E, Leux C, Brindel P, et al. Pooled analysis of two case-control studies in New Caledonia and French Polynesia of body mass index and differentiated thyroid cancer: the importance of body surface area. *Thyroid* 2010;20:1285–93.
7. Goldgar DE, Easton DF, Cannon-Albright LA, et al. Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. *J Natl Cancer Inst* 1994;86:1600–8.
8. Czene K, Lichtenstein P, Hemminki K. Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. *Int J Cancer* 2002;99:260–6.
9. Hemminki K, Eng C, Chen B. Familial risks for nonmedullary thyroid cancer. *J Clin Endocrinol Metab* 2005;90:5747–53.
10. Robledo M. Association studies in thyroid cancer susceptibility: are we on the right track? *J Mol Endocrinol* 2011;47:R43–58.
11. Lesueur F, Stark M, Tocco T, et al. Genetic heterogeneity in familial nonmedullary thyroid carcinoma: exclusion of linkage to RET, MNG1, and TCO in 56 families. *NMTC Consortium. J Clin Endocrinol Metab* 1999;84:2157–62.
12. Malchoff CD, Sarfarazi M, Tendler B, et al. Papillary thyroid carcinoma associated with papillary renal neoplasia: genetic linkage analysis of a distinct heritable tumor syndrome. *J Clin Endocrinol Metab* 2000;85:1758–64.
13. Granja F, Morari J, Morari EC, et al. Proline homozygosity in codon 72 of p53 is a factor of susceptibility for thyroid cancer. *Cancer Lett* 2004;210:151–7.
14. Akulevich NM, Saenko VA, Rogounovitch TI, et al. Polymorphisms of DNA damage response genes in radiation-related and sporadic papillary thyroid carcinoma. *Endocr Relat Cancer* 2009;16:491–503.
15. Sigurdson AJ, Land CE, Bhatti P, et al. Thyroid nodules, polymorphic variants in DNA repair and RET-related genes, and interaction with ionizing radiation exposure from nuclear tests in Kazakhstan. *Radiat Res* 2009;171:77–88.
16. 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;41:460–4.
17. 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;19:2516–23.
18. Damiola F, Byrnes G, Moissonnier M, et al. Contribution of ATM and FOXE1 (TTF2) to risk of papillary thyroid carcinoma in Belarusian children exposed to radiation. *Int J Cancer* 2014;134:1659–68.
19. Gudmundsson J, Sulem P, Gudbjartsson DF, et al. Discovery of common variants associated with low TSH levels and thyroid cancer risk. *Nat Genet* 2012;44:319–22.
20. Köhler A, Chen B, Gemignani F, et al. Genome-wide association study on differentiated thyroid cancer. *J Clin Endocrinol Metab* 2013;98:E1674–E1681.
21. Mancikova V, Cruz R, Inglada-Pérez L, et al. Thyroid cancer GWAS identifies 10q26.12 and 6q14.1 as novel susceptibility loci and reveals genetic heterogeneity among populations. *Int J Cancer* 2015;137:1870–8.
22. Landa I, Ruiz-Llorente S, Montero-Conde C, et al. The variant rs1867277 in FOXE1 gene confers thyroid cancer susceptibility through the recruitment of USF1/USF2 transcription factors. *PLoS Genet* 2009;5:e1000637.
23. Bullock M, Duncan EL, O'Neill C, et al. Association of FOXE1 polyalanine repeat region with papillary thyroid cancer. *J Clin Endocrinol Metab* 2012;97:E1814–E1819.
24. Fan Y, Ding Z, Yang Z, et al. Expression and clinical significance of FOXE1 in papillary thyroid carcinoma. *Mol Med Rep* 2013;8:123–7.
25. Bonora E, Rizzato C, Diquigiovanni C, et al. The FOXE1 locus is a major genetic determinant for familial nonmedullary thyroid carcinoma. *Int J Cancer* 2014;134:2098–107.
26. He H, Li W, Liyanarachchi S, et al. Genetic predisposition to papillary thyroid carcinoma: involvement of FOXE1, TSHR and a novel lincRNA gene, PTCSC2. *J Clin Endocrinol Metab* 2015;100:E164–72.
27. 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;109:8646–51.
28. Zheng J, Li C, Wang C, et al. Common genetic variant on 14q13.3 contributes to thyroid cancer susceptibility: evidence based on 12 studies. *Mol Genet Genomics* 2015;290:1125–33.
29. Kang J, Deng X-Z, Fan Y-B, et al. Relationships of FOXE1 and ATM genetic polymorphisms with papillary thyroid carcinoma risk: a meta-analysis. *Tumour Biol* 2014;35:7085–96.
30. de Bakker PIW, Yelensky R, Pe'er I, et al. Efficiency and power in genetic association studies. *Nat Genet* 2005;37:1217–23.
31. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. *G3 (Bethesda)* 2011;1:457–70. [22384356]
32. Li Y, Willer C, Sanna S, et al. Genotype imputation. *Annu Rev Genomics Hum Genet* 2009;10:387–406.
33. Li M-X, Yeung JMY, Cherny SS, et al. Evaluating the effective numbers of independent tests and significant p values thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum Genet* 2012;131:747–56.
34. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 2010;26:2336–7.
35. 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;49:158–63.
36. Di Vito M, De Santis E, Perrone GA, et al. Overexpression of estrogen receptor-α in human papillary thyroid carcinomas studied by laser-capture microdissection and molecular biology. *Cancer Sci* 2011;102:1921–7.
37. Xu S, Chen G, Peng W, et al. Oestrogen action on thyroid progenitor cells: relevant for the pathogenesis of thyroid nodules? *J Endocrinol* 2013;218:125–33.
38. Onda M, Nagai H, Yoshida A, et al. Up-regulation of transcriptional factor E2F1 in papillary and anaplastic thyroid cancers. *J Hum Genet* 2004;49:312–18.
39. Chiefari E, Brunetti A, Arturi F, et al. Increased expression of AP2 and Sp1 transcription factors in human thyroid tumors: a role in NIS expression regulation? *BMC Cancer* 2002;2:35.
40. He H, Li W, Liyanarachchi S, et al. Multiple functional variants in long-range enhancer elements contribute to the risk of SNP rs965513 in thyroid cancer. *Proc Natl Acad Sci USA* 2015;112:6128–33.
41. Morey L, Helin K. Polycomb group protein-mediated repression of transcription. *Trends Biochem Sci* 2010;35:323–32.
42. Klug J, Beato M. Binding of YY1 to a site overlapping a weak TATA box is essential for transcription from the uteroglobin promoter in endometrial cells. *Mol Cell Biol* 1996;16:6398–407.
43. Austen M, Lüscher B, Lüscher-Firzlaff JM. Characterization of the transcriptional regulator YY1. The bipartite transactivation domain is independent of interaction with the TATA box-binding protein, transcription factor IIB, TAFII55, or cAMP-responsive element-binding protein (CPB)-binding protei. *J Biol Chem* 1997;272:1709–17.
44. Pereda CM, Lesueur F, Pertesi M, et al. Common variants at the 9q22.33, 14q13.3 and ATM loci, and risk of differentiated thyroid cancer in the Cuban population. *BMC Genet* 2015;16:22.
45. Maillard S, Damiola F, Clero E, et al. Common variants at 9q22.33, 14q13.3, and ATM loci, and risk of differentiated thyroid cancer in the French Polynesian population. *PLoS One* 2015;10:e0123700.
46. Guignard R, Truong T, Rougier Y, et al. Alcohol drinking, tobacco smoking, and anthropometric characteristics as risk factors for thyroid cancer: a countrywide case-control study in New Caledonia. *Am J Epidemiol* 2007;166:1140–9.
47. Truong T, Baron-Dubourdieu D, Rougier Y, et al. Role of dietary iodine and cruciferous vegetables in thyroid cancer: a countrywide case-control study in New Caledonia. *Cancer Causes Control* 2010;21:1183–92.
48. Wiersinga WM. Smoking and thyroid. *Clin Endocrinol (Oxf)* 2013;79:145–51.
49. Rinaldi S, Plummer M, Biessy C, et al. Thyroid-stimulating hormone, thyroglobulin, and thyroid hormones and risk of differentiated thyroid carcinoma: the EPIC study. *J Natl Cancer Inst* 2014;106:dju097.
50. Yip L, Nikiforova MN, Yoo JY, et al. Tumor genotype determines phenotype and disease-related outcomes in thyroid cancer: a study of 1510 patients. *Ann Surg* 2015;262:519–25.