Scienze naturali
Biologia
Istituto Lombardo (Rend. Scienze) 150, 191-202 (2016)

IDENTIFICATION AND FUNCTIONAL CHARACTERIZATION OF GERMLINE GENETIC VARIANTS PREDISPOSING TO DIFFERENTIATED THYROID CANCER RISK

GISELLA FIGLIOLI (*), *et al.*[#]

Nota presentata dal s.c. Andrea Mattevi
(Adunanza del 3 novembre 2016)

SUNTO. – Il carcinoma tiroideo differenziato (DTC) è il tumore endocrino più frequente ed in Italia si registra una elevata incidenza. Il rischio d'insorgenza di questa malattia è significativamente maggiore nei parenti di primo grado degli individui affetti suggerendo che anche le varianti genetiche germinali possano contribuire alla sua insorgenza. Precedenti studi di associazione sull'intero genoma (GWASs) hanno permesso l'identificazione di polimorfismi a singolo nucleotide (SNPs) associati con il rischio di DTC localizzati nelle regioni 2q35, 9q22.33 e 14q13.3. Al fine di individuare nuovi *loci* di suscettibilità al DTC, è stato eseguito un GWAS sulla popolazione italiana. In seguito, gli SNPs più interessanti sono stati analizzati in una popolazione italiana più ampia e in altre casistiche europee. I risultati di questo studio hanno confermato il ruolo del *loci* 2q35 e 9q22.33 nella suscettibilità al DTC. Inoltre, i polimorfismi nelle regioni 14q24.3 e 20q11.22-q12 sono risultati associati ad un incremento del rischio di DTC nell'analisi combinata di tutte le popolazioni analizzate, e gli SNPs localizzati in 3q25.32, 5q14, 7q21, 9q34.3, 11p15, 13q12.12 e 20p11 sono risultati associati alla malattia soltanto negli Italiani. Secondo i dati del progetto ENCODE, molti di questi polimorfismi si trovano all'interno di regioni di regolazione della trascrizione e lo studio delle eQTL ha mostrato che cinque degli SNPs identificati in questo studio sono associati alla regolazione dei loro geni più vicini in diversi tessuti, compreso il tessuto tiroideo. In conclusione, tramite questo studio sono stati identi-

(*) IFOM, the FIRC Institute of Molecular Oncology, Milan, Italy.
E-mail: gisella.figlioli@ifom.eu

(#) La nota è il risultato di un lavoro di ricerca condotto in collaborazione con: Aleksandra Köhler, Bowang Chen, Rossella Elisei, Cristina Romei, Kari Hemminki, Stefano Landi, Asta Försti, Federica Gemignani.

ficati nuovi alleli di suscettibilità al DTC ed è stato proposto un loro possibile ruolo funzionale.

***

ABSTRACT. – Differentiated thyroid cancer (DTC) is the most common endocrine tumor, showing a high incidence in Italy. A significant higher risk of this cancer is described in first-degree relatives of DTC patients compared to the general population suggesting that germline genetic variants may contribute to its development. Previous genome-wide association studies (GWASs) on DTC have identified robust associations with single nucleotide polymorphisms (SNPs) at chromosomes 2q35, 9q22.33 and 14q13.3. In order to identify additional DTC susceptibility *loci*, a novel GWAS on the Italian population was conducted. The GWAS was followed by validation studies, where the most interesting SNPs were replicated in a larger Italian population and other European cohorts. Previously observed association for 2q35 and 9q22.33 was confirmed. Moreover, a strong relationship of DTC risk was found with SNPs on 14q24.3 and 20q11.22-q12 across all populations and SNPs on 3q25.32, 5q14, 7q21, 9q34.3, 11p15, 13q12.12 and 20p11 only among Italians. According to ENCODE Project data, many of these SNPs are located in transcription regulatory regions and eQTL analyses showed that five of the associated SNPs may affect the expression levels of their closest genes in different human tissues, including thyroid. In conclusion, novel DTC risk alleles were identified and new insights into their possible functional role were discovered.

1. INTRODUCTION

Thyroid cancer (TC) comprises approximately 1% of all human malignancies and it is the most common endocrine malignancy, representing up to 80% of all cancers originating from endocrine organs [1]. Worldwide, thyroid cancer incidence varies in different geographic regions and it is overall higher in more economically developed countries. In Europe particularly elevated age-standardised rates (ASRs) were observed in Lithuania (ASR=15.5/100,000), Italy (ASR=13.5/100,000) and Austria (ASR=12.4/100,000) (http://eco.iarc.fr/EUCAN/ and http://globocan.iarc.fr).

Approximately 90% of diagnosed TCs arise from the thyrocytes, the thyroid hormone–producing cells, and are classified as differentiated thyroid cancers (DTCs). Among them, the most frequent subtype is papillary (PTC, 75%), followed by follicular (FTC, 10%), Hurthle cells (5%), and poorly differentiated carcinomas (1%–6%). Only a small proportion of TCs, called medullary thyroid cancer (MTC), originates from thyroid parafollicular cells [2].

There are few known TC risk factors except for exposure to ionizing radiation, female gender and a previous benign thyroid disease

[3]. Interestingly, TC is characterized by having one of the highest heritability among all cancer sites, as demonstrated by a significantly higher risk in first-degree relatives of patients with DTC than in the general population and suggesting that also germline genetic variants (*e.g.* single nucleotide polymorphisms, SNPs) may contribute to the risk of the disease [4].

In order to identify SNPs associated with DTC genetic predisposition case-control association studies were conducted [5]. To date, candidate gene/pathway association studies remain the most prevalent type of investigation with more than 100 studies published and more than 300 SNPs examined. Of them, SNPs within genes involved in DNA repair pathways, cell-cycle control, xenobiotic metabolism and in the MAPK pathway were frequently investigated. While some of these variants could represent true associations with DTC, many failed to be replicated among additional populations and could be false-positive [6].

During the past decade genome-wide association studies (GWASs) emerged worldwide as an unbiased approach, independent of *a priori* knowledge on the disease, to reveal SNPs and genomic regions associated with human cancer risk. GWASs rely on the phenomenon of linkage disequilibrium (LD) wherein SNPs are not inherited individually but instead are in LD blocks, with many nearby SNPs being highly correlated. This allows obtaining information on ~50,000 variants by analyzing only one SNP (formally named "tag-SNP") and therefore it may be used to study the entire genome by evaluating ~500,000-700,000 tag-SNPs [7]. GWASs are commonly performed using a multistage approach. In the first stage (*e.g.* the discovery phase), all tag-SNPs are tested on a small subset of cases and control. During the second stage (*e.g.* the validation phase), significant tag-SNPs, are analyzed in a population similar to that used in the previous phase with the purpose of ruling out false positive associations. Moreover, validations in different populations represent also an important task to check if the observed effects are specific for the discovery population or if they may be extended to different geographic groups [8, 9]. Once the association of a SNP is confirmed, the next challenge lies in discovering its functional role and in the identification of its target gene. While few GWAS-identified SNPs are predicted to disrupt protein-coding regions, the great majority of them, approximately 88%, lie in intergenic or intronic regions and could be located within transcriptional regulatory elements (such as promoters, enhancers, transcription factor binding sites [TFBSs] and DNaseI hypersensitive sites [DHSs]).

All these functional elements of the human genome were annotated by the ENCODE (ENCylopedia Of DNA Elements) project and they could be mined to uncover the possible functional role of GWAS-identified SNPs [10, 11]. Moreover, it was clearly reported that inherited genetic variants within several genomic regions, called expression quantitative-trait *loci* (eQTLs), are associated with expression of many transcripts. Thus, studying the associations between SNPs and gene expression levels may represent a useful way to connect risk variants to their putative genes or transcripts. eQTLs can be located either near (1 Mb) the gene they regulate (*cis*-eQTL) or at a significant distance away from it (*trans*-eQTL) [12, 13]. At this regard, the Genotype-Tissue Expression (GTEx) program has recently provided expression levels of human genes in transformed fibroblasts and in several tissues from healthy individuals allowing the characterization of the eQTLs and the interpretation of GWAS-identified SNPs [14].

In 2009 Gudmusson and collaborators published the first GWAS on DTC. The strongest signals were found for the SNP rs965513 on 9q22.33, 57 kb upstream *FOXE1* with (OR=1.75, 95% CI 1.59-1.94) and rs944289 on 14q13.3, near *NKX2-1* (OR=1.37, 95% CI 1.24- 1.52) [15]. The importance of *FOXE1* in the disease genetic predisposition was confirmed in a second GWAS performed on radiation-related PTCs. In particular, the association for rs965513 was confirmed (OR=1.65, 95% CI 1.43-1.91) and a statistically significant association was found for rs1867277 (OR=1.48, 95% CI 1.27-1.71) [16]. Additional DTC risk variants were found through a GWAS on circulating TSH levels. In this study the strongest association was observed for rs116909374 on 14q13.3 with an OR of 2.09 (95% CI 1.68-2.60). Similarly, the variants rs966423 on 2q35 (within *DIRC3*) and rs2439302 on 8p12 (within *NRG1*) were significantly associated with DTC with OR of 1.34 (95% 1.22-1.47) and 1.36 (95% CI 1.23-1.50), respectively [17].

In order to search for additional variants predisposing to DTC a novel GWAS on the high-incidence Italian population was performed thanks to a successful collaboration among the Department of Endocrinology and Metabolism of the University Hospital Cisanello of Pisa (Italy), the Department of Biology of the University of Pisa (Italy) and the Division of Molecular Genetic Epidemiology of the German Cancer Research Center (DKFZ, Heidelberg, Germany). The GWAS was followed by the replication of SNPs that showed suggestive evidence of association in other European cohorts and by the assessment

of the cumulative effect of the identified SNPs. The functional role of the best-associated SNPs was also investigated by using ENCODE project experimental data and by eQTL analyses.

2. Risk variants associated with differentiated thyroid cancer risk

The local Ethical Committee approved this study and all participants gave their written informed consent to participate according to Helsinki declaration.

The discovery phase of the GWAS was performed on Italian histologically confirmed DTC patients and healthy controls. A total of 701 cases were recruited at the Department of Endocrinology and Metabolism of the University Hospital Cisanello in Pisa. The control group comprised 499 healthy individuals without any thyroid disease and cancer history and included workers who underwent a routine check-up at the same hospital of Pisa and blood donors from the Meyer Hospital in Florence. After the application of strict quality controls 572,042 SNPs were analyzed for association with DTC in 690 cases and 497 controls. The results of this phase confirmed the role of the SNPs identified in previous GWASs on DTC predisposition. In particular, a robust evidence for a relationship between *FOXE1* and the risk of the disease with multiple SNPs annotating in the region having *p-value*<5.0×10^{-8}. Because the association between *FOXE1* and DTC was already well-established, SNPs in this region were excluded from the following phases of analysis (*Fig. 1*) [18].

Validation studies were conducted on four European cohorts. The Italian cohort (Italian1+Italian2) included 1,539 patients and 1,719 controls collected in the same hospitals of the samples of the discovery phase. The Polish group comprised 468 patients with DTC and 470 healthy controls 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, recruited at the Department of Genetics and Microbiology of Autonomous University of Barcelona, and 420 healthy individuals. The United Kingdom series consisted of 509 cases, ascertained through the Institute of Cancer Research/Royal Marsden Hospital National Health Service Trust, and 1,118 controls recruited through the National Study of

Colorectal Cancer. A total of 109 SNPs, selected within regions showing stronger association signals (*e.g.* lower *p-value*) with DTC risk, were validated in the larger Italian population. Of them, 24 SNPs positively replicate the GWAS associations and were also analyzed in Polish, UK and Spanish cohorts. The joint analysis of all populations (2,985 cases and 3,727 controls) revealed a genome-wide significant association with DTC for rs6759952 near *DIRC3* on 2q35 (OR=1.25, 95% CI 1.16-1.34, *p-value*=6.4×10^{-10}), rs10136427 near *BATF* on 14q24.3 (OR=1.30, 95% CI 1.17-1.44, *p-value*=9.30×10^{-7}) and rs7267944 near *DHX35* on 20q11.22-q12 (OR=1.32, 95% 1.20-1.46, *p-value*=1.34×10^{-8}). The role of the SNPs was also assessed only in the Italian population, totaling 2,260 cases and 2,218 controls. The most significant associations were observed for rs7617304 upstream *RARRES1* on 3q25.32 (OR=1.25, 95% CI 1.12-1.39, *p-value*=4.6×10^{-5}), rs13184587 within *ARSB* on 5q14.1 (OR=1.28, 95% CI 1.15-1.43, *p-value*=8.54×10^{-6}), rs10238549 (OR=1.27, 95% CI 1.15-1.40, *p-value*=4.1×10^{-6}) and rs7800391 downstream *IMMP2L* on 7q21 (OR=1.25, 95% CI 1.14-1.38, *p-value*=5.7×10^{-6}), rs10781500 downstream *SNAPC4* on 9q34.3 (OR=1.23, 95% CI 1.12-1.36, *p-value*=3.5×10^{-5}), rs7935113 within *GALNTL4* on 11p15.3 (OR=1.36, 95% CI 1.20-1.53, *p-value*=7.41×10^{-7}), rs1220597 within *SPATA13* on 13q12.12 (OR=1.26, 95% CI 1.14-1.38, *p-value*=3.25×10^{-6}) and rs1203952 upstream *FOXA2* on 20p11 (OR=1.29, 95% CI 1.16-1.44, *p-value*=4.42×10^{-6}) (*Tab. 1*) [18-20]. The cumulative effect of the 11 independent susceptibility SNPs in 10 genes identified in the Italian population was also evaluated. A dose-dependent increase in risk of DTC was observed with an increasing number of risk alleles (OR=1.30, 95% CI 1.26-1.35, *p-trend*=3.13×10^{-47}). In particular, individuals carrying ≥14 risk alleles had 7.68 times higher risk of getting DTC as compared to those with ≤7 risk alleles (*Fig. 2*) [19].

3. Bioinformatic and eQTL analyses of the risk variants

Computational approaches were employed to functionally annotate the associated SNPs. Briefly, ENCODE Project data were quarried using the HaploReg v2 tool and eQTL analyses were performed taking advantage of data free available on GTEx Portal [14, 21]. Among the three variants that were found associated with the risk of DTC in the combined analysis of all the European populations, rs6759952 (*DIRC3*) and rs10136427 (*BATF*) were found to map in regions of weak and

strong enhancers. Both SNPs alter the binding sites for regulatory protein: rs6759952-T risk allele introduces the consensus sequence for Pou1f1, Pou2f2, TLX1 and YY1 proteins and rs10136427-C risk allele removes a FOXD3 site. Little support for the functional role of rs7267944 (*DHX35*) was found.

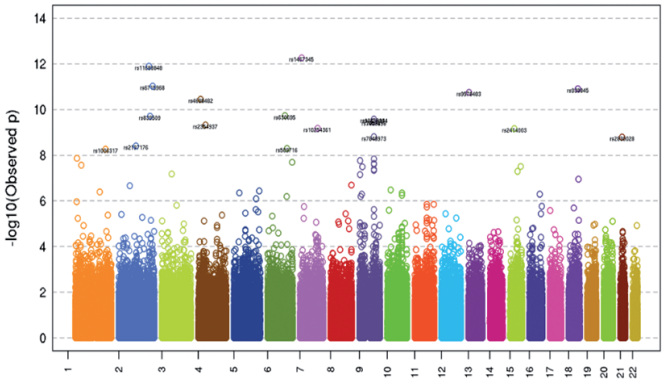

Fig. 1 – Manhattan plot of simple ^{2} allelic test of association *p-values* from the GWAS. The plot shows the –log10*p-values* for each tagSNP against chromosomal location. Values for each chromosome are reported in different colors for visual effect.

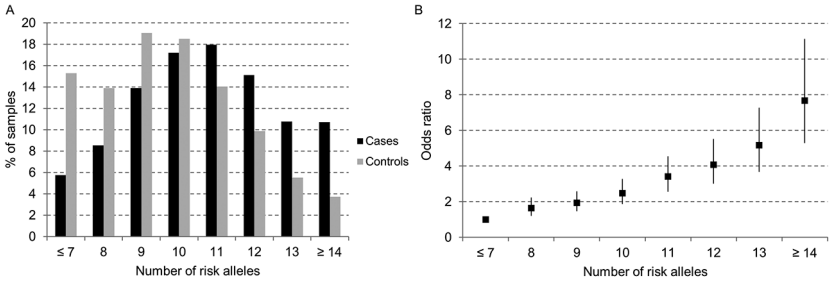

Fig. 2 – Cumulative risk assessment. (A) Sample distribution according to the number of risk alleles in eleven SNPs associated in the Italian DTC cases (black columns) and controls (grey columns). (B) Plot of the increasing ORs for DTC with increasing number of risk alleles. The category ≤7 was chosen as reference (OR=1.0); vertical bars correspond to 95% confidence intervals.

A possible regulatory consequence was identified for all the SNPs associated with the disease only in the Italian population, except for rs1220597 (*SPATA13*). The variants rs7617304 (*RARRES1*), rs10238549 and rs7800391 (*IMMP2L*) and rs7935113 (*GALNTL4*) are

Tab. 1 – Risk of differentiated thyroid cancer associated with single nucleotide polymorphisms identified in the present study. Significant results at a genome-wide level are highlighted in bold.

SNP	*Locus*	Closest Gene	Risk allele	Population	Number of cases/controls	Risk allele frequency (cases/controls	Allelic OR (95% CI)(*)	*p-value*(*)
rs6759952	2q35	*DIRC3*	T	GWAS	649/431	0.47/0.38	1.44 (1.21-1.72)	4.7×10^{-5}
				Italian1	1160/927	0.49/0.42	1.32 (1.17-1.49)	9.8×10^{-6}
				Italian2	322/710	0.44/0.40	1.16 (0.96-1.41)	0.12
				Polish	413/420	0.46/0.44	1.07 (0.88-1.30)	0.48
				Spanish	397/393	0.51/0.39	1.58 (1.29-1.93)	8.2×10^{-6}
				UK	497/1090	0.46/0.43	1.14 (0.98-1.32)	0.09
				Italian cohorts	2131/2068	-	1.30 (1.18-1.43)	7.3×10^{-8}
				Joint analysis	**3438/3971**	**-**	**1.25 (1.16-1.34)** P_{het}=0.03, I^2=58.5%	**6.4×10^{-10}**
rs7617304	3q25.32	*RARRES1*	A	GWAS	649/430	0.31/0.25	1.37 (1.13-1.67)	1.4×10^{-3}
				Italian1	1170/933	0.28/0.25	1.16 (1.01-1.34)	0.03
				Italian2	320/705	0.31/0.27	1.22 (1.00-1.50)	0.05
				Polish	437/431	0.19/0.21	0.92 (0.73-1.17)	0.50
				Spanish	388/394	0.26/0.25	1.09 (0.87-1.37)	0.47
				Italian cohorts	**2139/2068**	**-**	**1.25 (1.12-1.39)**	**4.6×10^{-5}**
				Joint analysis	2964/2893	-	1.17 (1.07-1.27) P_{het}=0.12, I^2=46.0%	7.0×10^{-4}
rs13184587	5q14.1	*ARSB*	G	GWAS	646/414	0.77/0.69	1.51 (1.24-1.83)	3.60×10^{-5}
				Italian1+2	1429/1541	0.76/0.72	1.19 (1.06-1.33)	4.15×10^{-3}
				Polish	442/434	0.76/0.74	1.10 (0.88-1.36)	0.41
				Spanish	382/403	0.72/0.74	0.91 (0.73-1.14)	0.42
				Italian cohorts	**2075/1955**	**-**	**1.28 (1.15-1.43)**	**8.54×10^{-6}**
				Joint analysis	2899/2792	-	1.17 (1.07-1.27) $P_{het}=6.4\times10^{-3}$, I^2=75.6%	7.16×10^{-4}
rs10238549	7q21	*IMMP2L*	C	GWAS	649/431	0.72/0.63	1.48 (1.23-1.77)	3.2×10^{-5}
				Italian1	1170/933	0.68/0.64	1.21 (1.06-1.37)	4.7×10^{-3}
				Italian2	323/695	0.71/0.68	1.16 (0.94-1.42)	0.16
				Polish	451/438	0.76/0.73	1.17 (0.94-1.45)	0.16
				Spanish	390/391	0.65/0.72	0.73 (0.59-0.91)	4.4×10^{-3}
				UK	503/1101	0.72/0.71	1.02 (0.86-1.20)	0.85
				Italian cohorts	**2142/2059**	**-**	**1.27 (1.15-1.40)**	**4.1×10^{-6}**
				Joint analysis	3486/3989	-	1.13 (1.05-1.22) $P_{het}=2.0\times10^{-4}$, I^2=79.5%	1.1×10^{-3}
rs7800391	7q21	*IMMP2L*	T	GWAS	649/430	0.43/0.34	1.45 (1.21-1.73)	5.2×10^{-5}
				Italian1	1165/934	0.40/0.39	1.22 (1.07-1.38)	2.3×10^{-3}
				Italian2	320/679	0.41/0.40	1.06 (0.87-1.28)	0.55
				Polish	444/422	0.42/0.40	1.06 (0.88-1.29)	0.53
				Spanish	407/394	0.40/0.44	0.86 (0.71-1.05)	0.14
				Italian cohorts	**2134/2043**	**-**	**1.25 (1.14-1.38)**	**5.7×10^{-6}**
				Joint analysis	2985/2859	-	1.14 (1.05-1.23) $P_{het}=5.0\times10^{-3}$, I^2=73.1%	1.3×10^{-3}
rs10781500	9q34.3	*SNAPC4*	C	GWAS	649/431	0.69/0.60	1.51 (1.23-1.86)	7.9×10^{-5}
				Italian1	1155/916	0.67/0.63	1.18 (1.04-1.34)	0.01
				Italian2	322/706	0.66/0.62	1.20 (0.98-1.46)	0.07
				Polish	404/424	0.58/0.59	0.99 (0.81-1.21)	0.91
				Spanish	386/392	0.63/0.61	1.09 (0.89-1.34)	0.42
				Italian cohorts	**2126/2053**	**-**	**1.23 (1.12-1.36)**	**3.5×10^{-5}**
				Joint analysis	2916/2869	-	1.17 (1.08-1.27) P_{het}=0.18, I^2=36.6%	1.1×10^{-4}
rs7935113	11p15.3	*GALNTL4*	C	GWAS	647/431	0.23/0.16	1.50 (1.20-1.88)	3.26×10^{-4}
				Italian1+2	1454/1572	0.20/0.16	1.28 (1.12-1.46)	2.20×10^{-4}
				Polish	452/443	0.13/0.13	0.99 (0.75-1.30)	0.93
				Spanish	352/407	0.18/0.18	1.02 (0.78-1.32)	0.90
				Italian cohorts	**2101/2003**	**-**	**1.36 (1.20-1.53)**	**7.41×10^{-7}**
				Joint analysis	2905/2853	-	1.24 (1.12-1.38) P_{het}=0.05 , I^2=61.96	2.71×10^{-5}
rs1220597	13q12.12	*SPATA13*	C	GWAS	646/414	0.48/0.39	1.42 (1.20-1.70)	7.11×10^{-5}
				Italian1+2	1458/1575	0.47/0.43	1.20 (1.08-1.33)	5.02×10^{-4}
				Polish	435/449	0.44/0.43	1.01 (0.84-1.22)	0.92
				Spanish	332/407	0.47/0.47	1.00 (0.82-1.23)	0.99
				Italian cohorts	**2104/1989**	**-**	**1.26 (1.14-1.38)**	**3.25×10^{-6}**
				Joint analysis	2871/2845	-	1.16 (1.07-1.25)	2.64×10^{-4}

							P_{het}=0.02, I^2=70.1%	
rs10136427	14q24.3	*BATF*	C	GWAS	646/414	0.88/0.81	1.62 (1.28-2.06)	5.73×10^{-5}
				Italian1+2	1515/1598	0.87/0.84	1.26 (1.09-1.45)	1.32×10^{-3}
				Polish	458/444	0.79/0.76	1.19 (0.95-1.48)	0.12
				Spanish	380/406	0.84/0.83	1.05 (0.81-1.37)	0.72
				Italian cohorts	2158/2012	-	1.40 (1.23-1.60)	4.35×10^{-7}
				Joint analysis	**2999/2862**	**-**	**1.30 (1.17-1.44)**	**9.30×10^{-7}**
							P_{het}=0.12, I^2=49.1%	
rs1203952	20p11	*FOXA2*	G	GWAS	647/429	0.28/0.21	1.49 (1.21-1.83)	1.37×10^{-4}
				Italian1+2	1453/1514	0.27/0.23	1.25 (1.11-1.41)	2.10×10^{-4}
				Polish	453/437	0.21/0.23	0.88 (0.70-1.10)	0.25
				Spanish	351/404	0.23/0.21	1.10 (0.86-1.40)	0.44
				Italian cohorts	**2100/1943**	**-**	**1.29 (1.16-1.44)**	**4.42×10^{-6}**
				Joint analysis	2904/2784	-	1.20 (1.09-1.31)	1.20×10^{-4}
							P_{het}=6.1×10^{-3}, I^2=75.80	
rs7267944	20q11.22-q12	*DHX35*	C	GWAS	646/414	0.26/0.18	1.54 (1.24-1.90)	6.60×10^{-5}
				Italian1+2	1477/1601	0.23/0.19	1.29 (1.14-1.45)	5.83×10^{-5}
				Polish	454/442	0.17/0.16	1.10 (0.85-1.41)	0.45
				Spanish	374/406	0.21/0.18	1.23 (0.96-1.57)	0.11
				Italian cohorts	2123/2015	-	1.39 (1.24-1.56)	2.13×10^{-8}
				Joint analysis	**2951/2863**	**-**	**1.32 (1.20-1.46)**	**1.34×10^{-8}**
							P_{het}=0.18, I^2=38.6%	

Tab. 2 – eQTL analysis based on Genotype-Tissue Expression (GTEx) program data. Significant results on thyroid tissues are highlighted in bold. Up=up-regulation, Down=down-regulation.

SNP	SNP *locus*	Risk allele	Gene	Gene *locus*	*p-value*	Risk allele effect on gene expression	Tissue
rs6759952	2q35	T	*DIRC3*	2q35	2.2×10^{-7}	Up	Testis
rs7617304	3q25.32	A	*GFM1*	3q25.32	2.5×10^{-5}	Down	Pancreas
			LXN	3q25.32	7.0×10^{-5}	Up	Cells - Transformed fibroblasts
			MLF1	3q25.32	8.7×10^{-6}	Up	Heart - Left Ventricle
			RP11-379F4.4	3q25.32	1.3×10^{-6}	Down	Adipose - Visceral (Omentum)
			RP11-379F4.4	3q25.32	2.4×10^{-5}	Down	Adrenal Gland
			RP11-379F4.4	3q25.32	3.5×10^{-6}	Down	Artery - Aorta
			RP11-379F4.4	3q25.32	4.9×10^{-5}	Down	Artery - Tibial
			RP11-379F4.4	3q25.32	6.2×10^{-5}	Down	Colon - Sigmoid
			RP11-379F4.4	3q25.32	7.0×10^{-6}	Down	Esophagus - Gastroesophageal Junction
			RP11-379F4.4	3q25.32	2.6×10^{-9}	Down	Lung
			RP11-379F4.4	3q25.32	4.7×10^{-6}	Down	Nerve - Tibial
			RP11-379F4.4	3q25.32	4.7×10^{-10}	Down	Skin - Not Sun Exposed (Suprapubic)
			RP11-379F4.4	3q25.32	9.5×10^{-10}	Down	Skin - Sun Exposed (Lower leg)
			RP11-379F4.4	3q25.32	2.6×10^{-12}	Down	Testis
			RP11-	**3q25.32**	**9.3×10^{-11}**	Down	**Thyroid**
			RP11-379F4.7	3q25.32	5.2×10^{-6}	Up	Nerve - Tibial
			RP11-379F4.8	3q25.32	3.1×10^{-7}	Down	Testis
rs7800391	7q21	T	*AC003088.1*	7q31.1	2.2×10^{-20}	Down	Cells - Transformed fibroblasts
rs10781500	9q34.3	C	*CARD9*	9q34.3	8.4×10^{-7}	Up	Cells - Transformed fibroblasts
			CARD9	9q34.3	1.0×10^{-18}	Down	Whole Blood
			GPSM1	9q34.3	3.0×10^{-5}	Up	Muscle - Skeletal
			GPSM1	**9q34.3**	**2.0×10^{-5}**	**Up**	**Thyroid**
			INPP5E	9q34.3	1.9×10^{-7}	Up	Brain - Caudate (basal ganglia)
			INPP5E	9q34.3	1.7×10^{-7}	Up	Brain - Cerebellum
			INPP5E	9q34.3	2.1×10^{-5}	Up	Esophagus - Mucosa
			INPP5E	9q34.3	2.3×10^{-5}	Up	Lung
			INPP5E	9q34.3	3.4×10^{-6}	Up	Skin - Sun Exposed (Lower leg)
			INPP5E	9q34.3	1.6×10^{-9}	Up	Testis
			INPP5E	**9q34.3**	**5.3×10^{-28}**	**Up**	**Thyroid**
			PMPCA	9q34.3	7.8×10^{-6}	Down	Liver
			PMPCA	**9q34.3**	**4.1×10^{-8}**	**Up**	**Thyroid**
			SDCCAG3	9q34.3	2.1×10^{-5}	Up	Artery - Tibial
			SDCCAG3	9q34.3	1.1×10^{-5}	Up	Cells - Transformed fibroblasts
			SDCCAG3	**9q34.3**	**1.8×10^{-5}**	**Up**	**Thyroid**
			SDCCAG3	9q34.3	2.0×10^{-10}	Up	Whole Blood
			SEC16A	9q34.3	1.4×10^{-5}	Up	Pancreas
			SEC16A	9q34.3	2.2×10^{-11}	Up	Whole Blood
			SNAPC4	9q34.3	3.9×10^{-6}	Down	Whole Blood
rs1203952	20p11	G	***FOXA2***	**20p11**	**1.6×10^{-18}**	**Down**	**Thyroid**
			LINC00261	**20p11**	**3.6×10^{-11}**	**Down**	**Thyroid**
			RP4-	**20p11**	**5.0×10^{-17}**	**Down**	**Thyroid**

5p15.33, 6q14.1, 10q26.12, 10q24.33 and 15q22.33 emerged as novel DTC risk factors [22, 23]. However, current evidence from available GWASs explains only a small proportion of the disease heritability. Several explanations for the missing heritability were proposed. These included rare variants in novel pathways that are undetectable through traditional GWAS study design, structural variants, such as CNVs, that are poorly captured by existing technologies, insufficient power to detect gene-gene interactions and environmental factors. Thus, further studies on large sample sets and based on novel experimental approaches, as array-based fine-mapping, next generation sequencing and gene- environment association studies, are warranted to identify the predisposing factors that could explain a greater percentage of DTC heritability.

5. Acknowledgements

The authors are grateful to the study participants and the study investigators and coordinators for the recruitment of samples. This study was supported by the Associazione Italiana Ricerca sul Cancro (AIRC, 2008) and by the Istituto Toscano Tumori (ITT, 2010).

REFERENCES

[1] Sipos JA, Mazzaferri EL, Thyroid cancer epidemiology and prognostic variables, *Clin Oncol (R Coll Radiol)*, 2010:22:395-404.

[2] Grande E, Diez JJ, Zafon C, Capdevila J, Thyroid cancer: molecular aspects and new therapeutic strategies, *J Thyroid Res*, 2012:2012:847108.

[3] 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.

[4] Fallah M, Pukkala E, Tryggvadottir L, Olsen JH, Tretli S, Sundquist K, Hemminki K, 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.

[5] Landa I, Robledo M, Association studies in thyroid cancer susceptibility: are we on the right track?, *J Mol Endocrinol*, 2011:47:R43-58.

[6] Figlioli G, Elisei R, Romei C, Melaiu O, Cipollini M, Bambi F, Chen B, Kohler A, Cristaudo A, Hemminki K, Gemignani F, Forsti A, Landi S, A Comprehensive Meta-analysis of Case-Control Association Studies to Evaluate

Polymorphisms Associated with the Risk of Differentiated Thyroid Carcinoma, *Cancer Epidemiol Biomarkers Prev*, 2016:25:700-713.

[7] Kruglyak L, The road to genome-wide association studies, *Nat Rev Genet*, 2008:9:314-318.

[8] Hirschhorn JN, Daly MJ, Genome-wide association studies for common diseases and complex traits, *Nat Rev Genet*, 2005:6:95-108.

[9] Hosking FJ, Dobbins SE, Houlston RS, Genome-wide association studies for detecting cancer susceptibility, *Br Med Bull*, 2011:97:27-46.

[10] Cooper GM, Shendure J, Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data, *Nat Rev Genet*, 2011:12:628-640.

[11] Consortium EP, The ENCODE (ENCyclopedia Of DNA Elements) Project, *Science*, 2004:306:636-640.

[12] Rockman MV, Kruglyak L, Genetics of global gene expression, *Nat Rev Genet*, 2006:7:862-872.

[13] Stranger BE, Forrest MS, Clark AG, Minichiello MJ, Deutsch S, Lyle R, Hunt S, Kahl B, Antonarakis SE, Tavare S, Deloukas P, Dermitzakis ET, Genome-wide associations of gene expression variation in humans, *PLoS Genet*, 2005:1:e78.

[14] Consortium GT, Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans, *Science*, 2015:348:648-660.

[15] Gudmundsson J, Sulem P, Gudbjartsson DF, Jonasson JG, Sigurdsson A, Bergthorsson JT, He H, Blondal T, Geller F, Jakobsdottir M, Magnusdottir DN, Matthiasdottir S, Stacey SN, Skarphedinsson OB, Helgadottir H, Li W, Nagy R, Aguillo E, Faure E, Prats E, Saez B, Martinez M, Eyjolfsson GI, Bjornsdottir US, Holm H, Kristjansson K, Frigge ML, Kristvinsson H, Gulcher JR, Jonsson T, Rafnar T, Hjartarsson H, Mayordomo JI, de la Chapelle A, Hrafnkelsson J, Thorsteinsdottir U, Kong A, Stefansson K, Common variants on 9q22.33 and 14q13.3 predispose to thyroid cancer in European populations, *Nat Genet*, 2009:41:460-464.

[16] Takahashi M, Saenko VA, Rogounovitch TI, Kawaguchi T, Drozd VM, Takigawa-Imamura H, Akulevich NM, Ratanajaraya C, Mitsutake N, Takamura N, Danilova LI, Lushchik ML, Demidchik YE, Heath S, Yamada R, Lathrop M, Matsuda F, Yamashita S, The FOXE1 locus is a major genetic determinant for radiation-related thyroid carcinoma in Chernobyl, *Hum Mol Genet*, 2010:19:2516-2523.

[17] Gudmundsson J, Sulem P, Gudbjartsson DF, Jonasson JG, Masson G, He H, Jonasdottir A, Sigurdsson A, Stacey SN, Johannsdottir H, Helgadottir HT, Li W, Nagy R, Ringel MD, Kloos RT, de Visser MC, Plantinga TS, den Heijer M, Aguillo E, Panadero A, Prats E, Garcia-Castano A, De Juan A, Rivera F, Walters GB, Bjarnason H, Tryggvadottir L, Eyjolfsson GI, Bjornsdottir US, Holm H, Olafsson I, Kristjansson K, Kristvinsson H, Magnusson OT, Thorleifsson G, Gulcher JR, Kong A, Kiemeney LA, Jonsson T, Hjartarson H, Mayordomo JI, Netea-Maier RT, de la Chapelle A, Hrafnkelsson J, Thorsteinsdottir U, Rafnar T, Stefansson K, Discovery of common variants associated with low TSH levels and thyroid cancer risk, *Nat Genet*, 2012:44:319-322.

[18] Kohler A, Chen B, Gemignani F, Elisei R, Romei C, Figlioli G, Cipollini M,

Cristaudo A, Bambi F, Hoffmann P, Herms S, Kalemba M, Kula D, Harris S, Broderick P, Houlston R, Pastor S, Marcos R, Velazquez A, Jarzab B, Hemminki K, Landi S, Forsti A, Genome-wide association study on differentiated thyroid cancer, *J Clin Endocrinol Metab*, 2013:98:E1674-1681.

[19] Figlioli G, Chen B, Elisei R, Romei C, Campo C, Cipollini M, Cristaudo A, Bambi F, Paolicchi E, Hoffmann P, Herms S, Kalemba M, Kula D, Pastor S, Marcos R, Velazquez A, Jarzab B, Landi S, Hemminki K, Gemignani F, Forsti A, Novel genetic variants in differentiated thyroid cancer and assessment of the cumulative risk, *Sci Rep*, 2015:5:8922.

[20] Figlioli G, Kohler A, Chen B, Elisei R, Romei C, Cipollini M, Cristaudo A, Bambi F, Paolicchi E, Hoffmann P, Herms S, Kalemba M, Kula D, Pastor S, Marcos R, Velazquez A, Jarzab B, Landi S, Hemminki K, Forsti A, Gemignani F, Novel genome-wide association study-based candidate loci for differentiated thyroid cancer risk, *J Clin Endocrinol Metab*, 2014:99:E2084-2092.

[21] Ward LD, Kellis M, HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants, *Nucleic Acids Res*, 2012:40:D930-934.

[22] Gudmundsson J, Thorleifsson G, Sigurdsson JK, Stefansdottir L, Jonasson JG, Gudjonsson SA, Gudbjartsson DF, Masson G, Johannsdottir H, Halldorsson GH, Stacey SN, Helgason H, Sulem P, Senter L, He H, Liyanarachchi S, Ringel MD, Aguillo E, Panadero A, Prats E, Garcia-Castano A, De Juan A, Rivera F, Xu L, Kiemeney LA, Eyjolfsson GI, Sigurdardottir O, Olafsson I, Kristvinsson H, Netea-Maier RT, Jonsson T, Mayordomo JI, Plantinga TS, Hjartarson H, Hrafnkelsson J, Sturgis EM, Thorsteinsdottir U, Rafnar T, de la Chapelle A, Stefansson K, A genome-wide association study yields five novel thyroid cancer risk loci, *Nat Commun*, 2017:8:14517.

[23] Mancikova V, Cruz R, Inglada-Perez L, Fernandez-Rozadilla C, Landa I, Cameselle-Teijeiro J, Celeiro C, Pastor S, Velazquez A, Marcos R, Andia V, Alvarez-Escola C, Meoro A, Schiavi F, Opocher G, Quintela I, Ansede-Bermejo J, Ruiz-Ponte C, Santisteban P, Robledo M, Carracedo A, Thyroid cancer GWAS identifies 10q26.12 and 6q14.1 as novel susceptibility loci and reveals genetic heterogeneity among populations, *Int J Cancer*, 2015:137:1870-1878.