Association Study of 69 Genes in the Ret Pathway Identifies Low-penetrance Loci in Sporadic Medullary Thyroid Carcinoma

Sergio Ruiz-Llorente,[1,5] Cristina Montero-Conde,[1] Roger L. Milne,[2] Christian M. Moya,[5] Arancha Cebrián,[8] Rocío Letón,[1] Alberto Cascón,[1] Fátima Mercadillo,[1] Iñigo Landa,[1] Salud Borrego,[9] Guiomar Pérez de Nanclares,[10] Cristina Álvarez-Escolá,[6] José Ángel Díaz-Pérez,[7] Ángel Carracedo,[11] Miguel Urioste,[4] Anna González-Neira,[2] Javier Benítez,[2,3,12] Pilar Santisteban,[5] Joaquín Dopazo,[12,13] Bruce A. Ponder,[8] Mercedes Robledo,[1,12] and the Medullary Thyroid Carcinoma Clinical Group

[1]Hereditary Endocrine Cancer Group, [2]Genotyping Unit (CeGen, Madrid Node), [3]Human Genetics Group, Human Genetics Cancer Programme, and [4]Familial Cancer Unit, Centro Nacional de Investigaciones Oncológicas; [5]Molecular Endocrinology Department, Biomedical Research Institute, CSIC, UAM; [6]Endocrinology Service, Hospital La Paz; [7]Endocrinology Service, Hospital Clínico San Carlos, Madrid, Spain; [8]Department of Oncology, University of Cambridge, Strangeways Research Laboratories, Cambridge, United Kingdom; [9]Unidad de Genética Médica y Diagnóstico Prenatal, Hospital Universitario Virgen del Rocío, Sevilla, Spain; [10]Endocrinology and Diabetes Research Group, Hospital de Cruces, Vizcaya, Spain; [11]Unidade de Xenetica, Instituto de Medicina Legal, Facultad de Medicina, Universidad de Santiago de Compostela, Galicia, Spain; and [12]ISCIII Center for Biomedical Research on Rare Diseases, and [13]Department of Bioinformatics, Centro de Investigación Príncipe Felipe, Valencia, Spain

Abstract

To date, few association studies have been done to better understand the genetic basis for the development of sporadic medullary thyroid carcinoma (sMTC). To identify additional low-penetrance genes, we have done a two-stage case-control study in two European populations using high-throughput genotyping. We selected 417 single nucleotide polymorphisms (SNP) belonging to 69 genes either related to *RET* signaling pathway/functions or involved in key processes for cancer development. TagSNPs and functional variants were included where possible. These SNPs were initially studied in the largest known series of sMTC cases (n = 266) and controls (n = 422), all of Spanish origin. In stage II, an independent British series of 155 sMTC patients and 531 controls was included to validate the previous results. Associations were assessed by an exhaustive analysis of individual SNPs but also considering gene- and linkage disequilibrium–based haplotypes. This strategy allowed us to identify seven low-penetrance genes, six of them (*STAT1, AURKA, BCL2, CDKN2B, CDK6*, and *COMT*) consistently associated with sMTC risk in the two case-control series and a seventh (*HRAS*) with individual SNPs and haplotypes associated with sMTC in the Spanish data set. The potential role of *CDKN2B* was confirmed by a functional assay showing a role of a SNP (rs7044859) in the promoter region in altering the binding of the transcription factor HNF1. These results highlight the utility of association studies using homogeneous series of cases for better understanding complex diseases. [Cancer Res 2007;67(19):9561–7]

Introduction

Medullary thyroid carcinoma (MTC) is a rare tumor arising from parafollicular thyroid cells or thyroid C cells and constitutes approximately 5% to 10% of all thyroid neoplasias (1). MTC generates up to 50% of local metastases to cervical and mediastinal nodal groups and ~20% of distant metastases to the lung, liver, or bone (2).

Approximately 25% of all MTCs are associated with the development of multiple endocrine neoplasia type 2 (MEN2; MIM #171400[14]), a cancer syndrome with dominant autosomal inheritance, a variable expression pattern, and almost complete penetrance. Clinical classification of MEN2 has been established based on the spectrum of germ-line mutations detected along different exons of the proto-oncogene *RET* in large series of MEN2 families (3, 4). However, the intrafamilial and interfamilial phenotypic variability concerning the age of onset and tumor spectrum suggest that genetic modifying factors exist (5), and this has been confirmed using animal models (6). However, data on possible genetic factors involved in clinical MEN2 variability are scarce (7, 8), which is partially due to the low prevalence of this syndrome (1:30,000; ref. 9).

On the other hand, 75% of MTCs appear as sporadic tumors and behave as a complex (or multifactorial) disease, probably caused by multiple common genetic variants in different susceptibility genes commonly known as "low-penetrance genes." Although linkage analysis has been very successful in the identification of genes responsible for Mendelian diseases in the last few decades, its application to complex diseases has had limited success (10). It is now widely thought that large scale genotyping methods in combination with an appropriate study design may bring the genetics of complex diseases to a point of success comparable with that of the genetics of Mendelian diseases (11). Most of the studies aiming to identify low-penetrance genes for sporadic MTC (sMTC) have been case-control in design, with a relatively small number of subjects and few genes included in the analyses (7, 12, 13).

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

S. Ruiz-Llorente and C. Montero-Conde contributed equally to this work.

Medullary Thyroid Carcinoma Clinical Group: Pilar Saavedra, Maria Concepción Blanco, Sharona Azriel, Guillermo Martínez-Guerra, Tomás Lucas-Morante, Pedro López-Mondéjar, Amparo Meoro, Antonio Picó, Pedro Iglesias, José Ignacio Lara, Alicia Loma, Francisco Pomares, Ignacio Ramos, Joaquín Serrano, Juan Girbés, Luis Arribas, María Rosa Pinedo, and Julia Sastre.

Requests for reprints: Mercedes Robledo, Hereditary Endocrine Cancer Group, Human Cancer Genetics Programme, Centro Nacional de Investigaciones Oncológicas, Calle Melchor Fernández Almagro, 28029 Madrid, Spain. Phone: 34-912246948; Fax: 34-912246923; E-mail: mrobledo@cnio.es.

doi:10.1158/0008-5472.CAN-07-1638

[14] http://www.ncbi.nlm.nih.gov/omim

We have carried out a two-stage case-control study in independent populations, focused on genes related to *RET* signaling pathway/functions or key processes for cancer development, with the aim of identifying low-penetrance genes involved in sMTC etiology.

Materials and Methods

Subjects

Two hundred and sixty-six Spanish patients with MTC, recruited between 1996 to 2005, were classified as sporadic because germ-line mutations along exons 10 and 11 and exons 13 to 16 of the *RET* proto-oncogene were not detected. Diagnosis of MTC was based on documented pathologic examination by specialist physicians, and the mean age of diagnosis was 51.9 years (range, 14–91 years). Controls were 422 Spanish individuals free of thyroid cancer recruited between 1999 to 2005 via the following sources: 93 (22%) individuals attending the Unidade de Xenetica of the University of Santiago de Compostela for paternity testing; 131 (31%) from the Menopause Research Centre at the Instituto Palacios, Madrid, and 198 (47%) obtained from the National Blood Transfusion Centre in Madrid. These controls presented a similar age (median age, 48 years; range, 20–92 years) and sex distribution as the cases. An additional and independent group of 48 controls was used to check Taqman probe accuracy.

British cases (n = 155) were all individuals with histologically proven MTC, who were referred to the National Health Service Genetic Diagnostic Laboratory at Addenbrooke's Hospital between 1996 to 2000 for testing known *RET* mutations associated with MEN2. All these cases had tested negative and their mean age at diagnosis was 48 years (range, 14–80 years). British controls were 531 anonymous individuals selected from the European Prospective Investigation into Cancer and Nutrition study (14), a population-based cohort study of diet and health based in the East Anglia region of the United Kingdom. Median age was 60 years (range, 45–74 years), and none of them had a recorded history of malignancy or endocrine disease.

Informed consent was obtained from all participants in both studies.

Selection of Candidate Genes

A total of 69 genes were selected, 15 belonging to the RET pathway (*GFRA1, GFRA3, ARTN, GDNF, NRTN, AKT1, BRAF1, CDKN2A, CDKN2B, GRB2, NFKB1, PIK3CB, PIK3R1, PIK3R2*, and *SOS1*); 50 involved in cancer processes in which RET has been described previously to participate: genes involved in apoptotic events (*APAF1, BAX, BCL2, BCL2L1, CASP10, CASP3, CASP8, CASP9, E2F1, NFKBIA, RELA, TRAF1*, and *WRN*), signal transduction pathways (*AGTR1, ARHGDIB, BCR, CD44, COMT, EGF, EGFR, ERBB2, FOS, HIF1-α, HRAS, NFKB2, AURKA*, and *TRAF6)*, cell cycle regulation (*CCDN3, CDC25A, CDK2, CDK4, CDK6, CDKN2C, MAP2K1, MAP2K3, MAP2K4, MAP2K6, MAPK14, MAPKAPK2, MAPKAPK5, PCNA, PTEN, RB1*, and *STAT1*), and cell adhesion (*APC, IL2, IL6, PTTG1, TNFRSF6*, and *TNFSF6*); and 4 genes commonly associated with tumor development (*ATM, MDM2, TERT*, and *VEGF*).

In silico Tools and Criteria for the Selection of Polymorphisms

PupaSuite bioinformatics tool[15] (15) was used to select single nucleotide polymorphisms (SNP) considering their allele frequency and potential functional effects. In total, 417 allelic variants in the 69 genes were selected and genotyped (Supplementary Data) based on the following criteria: (*a*) taking into account linkage disequilibrium (LD) blocks defined according to the HapMap information[16] (16) for each gene, we selected tagSNPs where possible using HapMap CEU genotyping data (17) and Haploview software; (*b*) genomic areas located between LD blocks were analyzed using a SNP density of 1 per 8 to 15 kb; (*c*) when possible, putative functional SNPs were chosen; these included nonsynonymous coding SNPs, polymorphisms in intron/exon junctions, SNPs in putative transcription factor binding sites (TFBS) in the promoter regions of the genes (up to 10 kbp upstream of the transcription start site), and potential exonic splicing enhancers (ESE); (*d*) interspecies conservation (human-mouse) was used either as a criterion for reinforcing the prediction of functional motifs (TFBSs and ESEs) or as a tool for the selection of phylogenetically conserved intronic variants; (*e*) an allele frequency >10% in case of intronic SNPs and >5% in case of potentially functional SNPs; and (*f*) SNPs genotyped using the Illumina platform had to have an assay score of at least 0.6 (see details in the Illumina platform section below).

Genotyping Platforms

Two platforms were used to genotype the 417 polymorphisms: Taqman (57 SNPs from 13 genes) and Illumina (366 SNPs from 59 genes). To test the reliability of the techniques, six SNPs belonging to three genes (*BRAF, CDKN2A*, and *CDKN2B*) were studied with both platforms. Duplicates and negative controls were included on each plate as described below, as an additional quality control measure. Moreover, each genotyping assay contained both healthy controls and sMTC cases. Both Spanish and British populations were genotyped under the same experimental conditions.

Taqman platform. Probe accuracy to discriminate the two alleles at a specific locus was initially checked in a set of 48 DNA samples from controls. Both this plate and those of cases and controls subsequently genotyped included four DNA duplicates and a negative control. The assay was first done at the default annealing and extension temperature recommended by the manufacturer (a denaturing step of 15 s at 92°C, followed by annealing and extension for 1 min at 60°C for 40 cycles; see details at Applied Biosystems Web site).[17] If these conditions did not distinguish among the three possible genotype clusters, a second amplification was done using a lower annealing and extension temperature (58°C). Assays that did not give satisfactory results under the latter conditions were excluded from the analysis.

Illumina platform. Information concerning the methodologic principle and other technical characteristics of this genotyping platform is available at Illumina web page.[18] Illumina established an assay score based on the nucleotide composition of the DNA region and on the presence of duplicated or highly repetitive sequences, palindromes, and neighboring polymorphisms. Polymorphisms showing an Illumina quality score <0.6 were rejected for the final pool of genotyped SNPs. Rejected SNPs were substituted by neighboring genetic variants fulfilling the specific Illumina criteria. One DNA intra-assay duplicate, one interassay duplicate, and one negative control were included in each 96 sample assay.

Statistical Analysis for Individual Polymorphisms

Departure from Hardy-Weinberg equilibrium (HWE) for all SNPs was tested in controls using the *genhwi* command in STATA version 8. Associations between individual SNPs and thyroid cancer risk were assessed by comparing minor allele frequencies (MAF) between controls and cases using Pearson's χ^2 test. Genotype frequencies in cases and controls were compared and odds ratios (OR) were estimated by applying unconditional logistic regression, using homozygotes in the more frequent allele in controls as the reference group. For each SNP, the best fitting model among dominant, recessive, and multiplicative (single variable) codominant models was determined by parsimony, and this was tested against the two-variable codominant model via the likelihood ratio test. Associated nominal P values for the best fitting model were determined using the likelihood ratio test.

Two methods were considered to address the issue of multiple testing in the analysis of the Spanish series. A modified Bonferroni-corrected nominal threshold of $P = 0.05/N^*$ was used, where N^* is the "effective number of independent marker loci" after consideration of LD between SNPs on the same chromosome. N^* was calculated using the formula of Li and Ji (18), by applying the web-based program *SNPSpD*[19] (19, 20) to SNPs on individual chromosomes

[15] http://pupasuite.bioinfo.cipf.es/
[16] http://www.hapmap.org
[17] http://www.appliedbiosystems.com/
[18] http://www.illumina.com/
[19] http://genepi.qimr.edu.au/general/daleN/SNPSpD/

and summing estimates across chromosomes as described above (19). This approach, which appropriately accounts for the nonindependence of SNPs on the same chromosome due to LD, has been shown to closely approximate results from adjustment for multiple testing using permutation methods (19, 20). The distribution-free method of controlling the false discovery rate (FDR) of Benjamini and Liu (21), which is robust to the presence of nonindependent explanatory variables, was also applied.

Candidate SNPs identified in the Spanish series were tested for associations with sMTC in the British series using one-sided *P* values based on the relative frequencies in Spanish cases versus controls. Analyses of pooled data were adjusted for country as a dichotomous variable, although this was not possible for all SNPs due to genotyping failure in the British series.

Computational Inference of Haplotypes

Haplotypes for each gene were inferred using PHASE version 2.0[20] (22), a computational tool based on Bayesian methods. Case-control comparisons of haplotype distributions were carried out by applying a gene-based analysis, considering all the polymorphisms genotyped for each gene independently of LD blocks. This analysis was done for Spanish subjects with complete data for all SNPs genotyped per gene by applying the inbuilt permutation test, based on 10,000 permutations. Genes with overall Phase $P < 0.05$ were reanalyzed by LD blocks (16), where applicable, to narrow down the putative candidate region. These analyses were repeated for British data and later for data from both studies pooled, to validate results.

Associations between specific haplotypes and risk of MTC were assessed for those genes with Phase $P < 0.05$ using the Haplo.Stats package in R, which generates haplotype-specific score statistics as well as their associated *P* values. Haplotype-specific ORs were calculated by comparing the frequency of a specific haplotype with the combined frequency of the remaining haplotypes in cases versus controls.

Electrophoretic Mobility Shift Assay

Electrophoretic mobility shift assays (EMSA) were carried out to assess the biological role of the two polymorphisms (rs2106119 and rs7044859), respectively, located at putative Pax6 and HNF1α binding sites of the *CDKN2B* promoter. Purified nuclear protein extracts of the mTT human MTC cell line were obtained through routine procedures (23). Seven micrograms of these protein extracts were incubated with 200 ng of the following *Pax6*- or *HNF1α*-related dsDNA probes, which were labeled previously with γ-ATP (25 μCi):

Pax6 consensus binding region (24), rs2106119 normal allele: 5′-CTTTATTCAAGAA[T]GAGTTTCAAT-3′ and its corresponding complementary sequence; rs2106119 Polymorphic allele: 5′-CTTTATTCAAGAA[C]-GAGTTTCAAT-3′annealed to its antisense oligonucleotide.

HNF1α consensus binding region (25), rs7044859 normal allele: 5′-TATATTACAGATAGCTAATT[T]TTCAAGATTTCATGTTCACA-3′ annealed to its complementary sequence; rs7044859 polymorphic allele: 5′-TATAT-TACAGATAGCTAATT[A]TTCAAGATTTCATGTTCACA-3′ and its corresponding complementary sequence.

For competition, a 100-fold excess of the same unlabeled oligonucleotide or an unrelated (Cre1) oligonucleotide was used. EMSA conditions were similar to those described previously (23).

Results

SNP analysis. The SNP success rates of the Illumina and Taqman platforms were 90% (328 of 366 SNPs) and 86% (49 of 57 SNPs), respectively. Among the 46 SNPs that failed genotyping, 9 were monomorphic in the Spanish population, 6 exhibited more than three clusters, and 31 variants had a call rate <80%. In addition, one SNP did not fulfill HWE after correction for multiple testing ($P < 0.0002$, based on $N^* = 254$ independent comparisons) and therefore was also excluded from further analyses. More than 99% of results for the six SNPs genotyped using both platforms were concordant (data not shown). Discordant results for these six genetic variants were mainly obtained for common samples with low fluorochrome intensities. The 371 successfully genotyped polymorphisms showed 100% concordant results for interassay and intra-assay duplicates.

Comparisons of MAFs between cases and controls from the Spanish population to assess associations between individual SNPs and thyroid cancer showed unadjusted $P < 0.05$ for 22 variants (5.9% of those considered), belonging to 14 genes. Table 1 gives the reference sequence code and other characteristics of each of these SNPs. Dominant, recessive, and single variable (additive) codominant models best explained the behavior of 5, 2, and 15 polymorphisms, respectively. Two of these (rs10173099 and rs1462129 in *STAT1* and *BCL2*, respectively) remained statistically significant after more restrictive adjustment using FDR and Bonferroni-adjusted $P = 0.05$ (the latter based on $N^* = 253$ independent SNPs analyzed; Table 1). Whereas these two SNPs were the best candidates, we considered all 22 variants as potentially associated with thyroid cancer, to be validated in the British series.

Table 2 summarizes the results of the validation for the SNPs genotyped in the British case-control series. Seven SNPs failed genotyping (call rate, <80%). The 15 successfully genotyped SNPs showed 100% concordant results for intra-assay and interassay duplicates. The MAF frequency in cases relative to controls was in the opposite direction to that observed in the Spanish series for seven SNPs ($P = 1$; Table 2). Among the remaining eight SNPs, two were validated (unadjusted $P < 0.05$) in the British data set (rs8-*CDK6* and rs911160-*AURKA*). Interestingly, the *AURKA* polymorphism remained significant after Bonferroni correction for 15 tests in the British population (adjusted one-sided $P = 0.02$) and fits with a dominant model both for the Spanish and British series. In addition, rs8-*CDK6* and further four SNPs (rs10173099 in *STAT1*, rs165849 in *COMT*, rs643319 in *CDKN2B*, and rs4987825 in *BCL2*) appeared as significant findings (unadjusted $P < 0.05$; data not shown) when both series were pooled but with ORs that were similar and in the same direction for the two data sets (Spanish and British) independently (Tables 1 and 2).

Haplotype analysis. Evidence of a different haplotype distribution between sMTC cases and controls ($P < 0.05$) was observed in the Spanish series for six genes (*STAT1, BCL2, HRAS, CDKN2B, AURKA*, and *AGTR1)*. In addition, haplotype estimation considering LD blocks of these genes allowed us to narrow down the candidate region for *BCL2, CDKN2B*, and *AGTR1*. Interestingly, all these genes or LD blocks were represented by significant individual SNP findings ($P < 0.05$). Three of these haplotype associations (*STAT1, AGTR1*, and *AURKA*) also appeared as significant findings in the independent British series, whereas the other three could not be evaluated due to genotyping failure.

Haplotype-specific analyses. A very similar haplotype distribution was found between cases and controls in each of the series for *STAT1* and *AURKA* (Supplementary Tables S1 and S2). Considering only statistically significant haplotypes ($P < 0.05$) and comparing those with positive ORs versus negative ORs for each gene, it was possible to identify SNPs acting as risk markers: rs10173099 and rs1547550 (both in *STAT1*), and rs911160 (in *AURKA*). With the exception of rs1547550, these were also identified as candidates in the individual SNPs analysis (Table 1), reinforcing the results. With regard to *AGTR1*, haplotype-specific

[20] http://www.stat.washington.edu/stephens/software.html

Table 1. Significant SNPs shown for the Spanish population

SNP	Gene	Location	Chromosome	Minor allele	MAF	*P* (HWE)	Unadjusted *P* (1 *df*)	Adjusted *P* (Bonferroni)	Adjusted *P* (FDR)	Best fitting model	OR (*P*)
rs10173099	*STAT1*	Intronic	2q32.2	T	0.22	0.8	0.0000024	0.00061	0.00089	Codominant	2.19 (0.000004)
rs1462129	*BCL2*	Intronic	18q21.3	T	0.50	0.5	0.00010	0.025	0.019	Recessive	0.55 (0.00009)
rs2075786	*TERT*	Intronic	5p15.33	T	0.39	0.7	0.00090	0.23	0.11	Dominant	0.56 (0.0004)
rs165849	*COMT*	3′ Dws	22q11.21	G	0.37	0.8	0.0017	0.43	0.16	Codominant	0.67 (0.002)
rs1801018	*BCL2*	ESE	18q21.3	G	0.39	0.1	0.0022	0.56	0.16	Codominant	1.48 (0.002)
rs1977971	*BCL2*	Intronic	18q21.3	G	0.49	0.7	0.0050	1	0.31	Codominant	1.39 (0.006)
rs2106119*	*CDKN2B*	TFBS	9p21	C	0.46	0.5	0.0050	1	0.27	Codominant	1.47 (0.006)
rs2069418†	*CDKN2B*	5′ Ups	9p21	C	0.41	0.8	0.0068	1	0.32	Codominant	0.68 (0.007)
rs8176332	*HRAS*	Intronic	11p15.5	A	0.09	1.0	0.0080	1	0.33	Codominant	1.63 (0.01)
rs643319†	*CDKN2B*	5′ Ups	9p21	T	0.46	0.7	0.013	1	0.48	Codominant	0.75 (0.02)
rs2850763	*BCL2*	Intronic	18q21.3	T	0.50	0.6	0.017	1	0.57	Dominant	0.59 (0.005)
rs251409	*PIK3R1*	Intronic	5q13.1	G	0.18	0.6	0.023	1	0.71	Codominant	0.68 (0.02)
rs3756668	*PIK3R1*	3′ UTR	5q13.1	A	0.47	0.9	0.024	1	0.68	Recessive	0.55 (0.008)
rs1579074†	*PTTG1*	5′ Ups	5q35.1	C	0.39	0.3	0.026	1	0.69	Codominant	0.75 (0.02)
rs8	*CDK6*	Intronic	7q21-q22	T	0.18	0.9	0.026	1	0.64	Codominant	0.69 (0.03)
rs17349‡	*PCNA*	Intronic	20pter-p12	T	0.16	1.0	0.029	1	0.67	Dominant	0.63 (0.02)
rs3730447	*PCNA*	3′ Dws	20pter-p12	A	0.16	1.0	0.029	1	0.63	Dominant	0.63 (0.02)
rs718858‡	*AGTR1*	Intronic	3q21-q25	T	0.15	0.7	0.029	1	0.60	Codominant	1.41 (0.03)
rs4987825	*BCL2*	Intronic	18q21.3	A	0.15	1.0	0.033	1	0.64	Codominant	0.68 (0.03)
rs911160	*AURKA*	Intronic	20q13.2	C	0.24	0.5	0.035	1	0.62	Dominant	1.45 (0.03)
rs3218108	*CCND3*	3′ UTR	6p21	A	0.23	0.9	0.037	1	0.62	Codominant	1.33 (0.04)
rs3731239	*CDKN2A*	Intronic	9p21	C	0.36	0.6	0.040	1	0.65	Codominant	0.77 (0.04)

NOTE: *P* (HWE), *P* value for HWE; unadjusted *P* (1 *df*), unadjusted *P* value of each SNP comparing allele frequencies in cases versus controls (1 *df*); adjusted *P* (Bonferroni), *P* values adjusted for multiple testing using the Bonferroni correction (assuming 253 independent SNPs); adjusted FDR, *P* values adjusted for multiple testing using control of the FDR; OR (*P*), OR and *P* value according to the best fitting model of each polymorphism, as specified.

Abbreviations: 3′ Dws, 3′ downstream; 5′ Ups, 5′ upstream; 3′ UTR, 3′ untranslated region.

*SNP located in a putative Pax6 binding sequence.

† Polymorphisms located in regulatory elements (promoter or 5′ upstream).

‡ Highly conserved intronic polymorphisms according to PupaSNP.

analysis gave inconsistent results in the two case-control series, with ORs in opposing directions (data not shown), reflecting the observed behavior of rs718858 in each data set [*P* for British series = 1 [1 degree of freedom (*df*)]; Tables 1 and 2]. We therefore deemed that there was no evidence for *AGTR1* as a low-penetrance gene. Results of the haplotype analysis for the other five significant genes are summarized in Table 3.

The haplotype distributions of *CDKN2B, BCL2*, and *HRAS* could only be analyzed in the Spanish series. There was evidence of increased risk associated with the most common haplotype in *CDKN2B* (OR, 1.48; $P = 0.01$; Supplementary Table S3). This haplotype was tagged at the first and the third nucleotides (rs2069418 and rs2106119, respectively), which were both identified as candidates in the individual SNP analysis (Table 1). The two most common haplotypes in a LD block in *BCL2* (LD block 11 according to the updated version of HapMap; Supplementary Table S3), which account for >85% of the genetic variation, were complementary (CG versus TA) and appeared as risk and protective haplotypes ($P < 0.001$), respectively (Supplementary Table S4). These haplotypes are tagged by rs1801018 and rs1462129 and were both identified as candidates in the individual SNP analysis (Table 1). Results for haplotypes also confirmed the results for individual SNPs.

SNPs as potential regulators of CDKN2B transcription. Given that the *CDKN2B* promoter was identified as a susceptibility region both by individual SNP analysis and by haplotypes, EMSA experiments were done to assess the role of certain SNPs in *CDKN2B* transcription. In this regard, rs2106119 behaves as a risk variant and potentially alters the binding of the Pax6 transcription factor (Table 1). However, the EMSAs did not reveal any binding between Pax6 protein and the nucleotide region containing this SNP (data not shown). Additional studies were carried out for rs7044859 (allele frequency, ~0.5), a PupaSNP-predicted HNF1α TFBS in complete LD with rs2106119 (Fig. 1*A*). The EMSA shows direct binding of HNF1α only to the A allele of rs7044859 (Fig. 1*B*). Taking into account the transcriptional role of HNF1α, its binding to the A allele may increase the expression of the *CDKN2B* tumor suppressor gene and therefore exert a protective role. This result is in accordance with the opposite effect observed for C allele of rs2106119 (OR, >1; Table 1; *CDKN2B* Supplementary Data), given that both alleles are located in opposite haplotypes (Fig. 1*B*).

Discussion

In this genotyping study, we have identified seven low-penetrance genes related to the development of sMTC, using the largest series of patients reported to date and validating the results with an independent series from a different population and doing functional assays to confirm the biological role of one of the

associated variants identified. For both *STAT1* and *BCL2*, an individual SNP was significantly associated with sMTC after correction for multiple testing in the Spanish series. For *AURKA*, the SNP association seen in the Spanish series was validated by a significant association after correction for multiple testing in the independent British series. For all three of these genes (although for a different SNP for *BCL2*) plus *COMT, CDKN2B*, and *CDK6*, individual SNPs were associated with sMTC ($P \leq 0.01$) in the pooled data set and gave consistent OR estimates in the two samples independently. Haplotype analyses confirmed these associations for *STAT1* and *AURKA* in both series and for *BCL2* and *CDKN2B* in the Spanish series (but could not be evaluated in the British due to genotyping failure). The seventh gene (*HRAS*) could not be genotyped in the British series but was identified as a low-penetrance gene because both an individual SNP and haplotypes were associated with sMTC in the Spanish series.

Approximately 10% of the initially selected candidate genes (7 of 69) showed association. This success in identifying low-penetrance genes is probably related with the clinical behavior of MTC tumors itself, which, unlike other complex diseases analyzed through association studies, represents a quite homogeneous entity. This fact, together with gene selection based on knowledge of the disease, may outweigh any lack of statistical power due to sample size. Furthermore, given that little evidence of population stratification has been found for European-derived populations (26, 27), we assume that these findings are probably not due to the effect of differential ancestry between cases and controls. More importantly, our validation of significant findings through a sufficiently large and independent different European series should reduce the effect of false-positive findings and the effect of population stratification, as has been described previously (28).

Biological significance of the identified low-penetrance genes. Both *STAT1* and *BCL2* are involved in apoptosis regulation and directly linked to the phosphatidylinositol 3-kinase (PI3K)-AKT-nuclear factor-κβ (NF-κβ) pathway, a signaling cascade downstream of RET (29, 30). The PI3K/AKT cascade up-regulates *BCL2* expression through a cyclic AMP response element (31). In addition, the suppression of MEN2A-RET oncogenic signaling leads to a decrease in cell viability, reduced AKT phosphorylation, and apoptosis stimulation via decreased levels of BCL2 (32). The accumulation of positive findings around the exon 2-intron 2 boundaries of *BCL2* may point to a causal variant altering the alternative splicing of *BCL2* or to a nearby nonsynonymous coding SNP located in the highly conserved exon 2. Indeed, a nonsynonymous SNP ($Thr^{43}Ala$; rs1800477) that reduces the antiapoptotic activity of the BCL2 protein has been described previously (33). However, this polymorphism was not found among 500 healthy controls (data not shown). Therefore, more functional studies are needed to define the causal variant in this gene.

Table 2. Validation of the 22 SNPs in the British data set

SNP	Gene	Location	Chromosome	MAF	*P* (HWE)	*P* (1 *df*)	Model	OR (*P*)
rs10173099	*STAT1*	Intronic	2q32.2	0.32	0.08	0.1	Codominant	1.21 (0.1)
rs1462129	*BCL2*	Intronic	18q21.3	—	—	—	—	—
rs2075786	*TERT*	Intronic	5p15.33	—	—	—	—	—
rs165849	*COMT*	3′ Dws	22q11.21	0.32	0.2	0.3	Codominant	0.90 (0.3)
rs1801018	*BCL2*	ESE	18q21.3	—	—	—	—	—
rs1977971	*BCL2*	Intronic	18q21.3	0.55	0.04	1	—	—
rs2106119*	*CDKN2B*	TFBS	9p21	—	—	—	—	—
rs2069418 †	*CDKN2B*	5′ Ups	9p21	—	—	—	—	—
rs8176332	*HRAS*	Intronic	11p15.5	—	—	—	—	—
rs643319 †	*CDKN2B*	5′ Ups	9p21	0.46	0.5	0.1	Codominant	0.87 (0.1)
rs2850763	*BCL2*	Intronic	18q21.3	0.44	0.9	0.4	Dominant	0.96 (0.6)
rs251409	*PIK3R1*	Intronic	5q13.1	0.15	0.6	1	—	—
rs3756668	*PIK3R1*	3′ UTR	5q13.1	0.45	0.001	1	—	—
rs1579074 †	*PTTG1*	5′ Ups	5q35.1	0.41	0.9	0.4	Codominant	0.97 (0.8)
rs8	*CDK6*	Intronic	7q21-q22	0.24	0.1	0.01	Codominant	0.67 (0.02)
rs17349 ‡	*PCNA*	Intronic	20pter-p12	—	—	—	—	—
rs3730447	*PCNA*	3′ Dws	20pter-p12	0.12	0.8	1	—	—
rs718858 ‡	*AGTR1*	Intronic	3q21-q25	0.19	0.003	1	—	—
rs4987825	*BCL2*	Intronic	18q21.3	0.08	0.07	0.06	Codominant	0.67 (0.1)
rs911160	*AURKA*	Intronic	20q13.2-q13.3	0.22	0.02	0.0013	Dominant	1.66 (0.002)
rs3218108	*CCND3*	3′ UTR	6p21	0.31	0.0001	1	—	—
rs3731239	*CDKN2A*	Intronic	9p21	0.37	0.4	1	—	—

NOTE: MAF (minor allele as defined among Spanish controls; see Table 1); *P* (HWE), *P* value for HWE; *P* (1 *df*), unadjusted one-sided *P* value of each SNP comparing allele frequencies in cases versus controls (1 *df*); OR (*P*), OR and one-sided *P* value associated to the best fitting model of each polymorphism (based on results for the Spanish data set) for British data set.

Abbreviations: 3′ Dws, 3′ downstream; 5′ Ups, 5′ upstream; 3′ UTR, 3′ untranslated region.

*SNP located in a putative Pax6 binding sequence.

† Polymorphisms located in regulatory elements (promoter or 5′ upstream).

‡ Highly conserved intronic polymorphisms.

Table 3. *P* values obtained using the Phase program for the haplotype distribution between sMTC cases and controls

	Spanish Phase *P* value	British Phase *P* value	Pooled *P* value
STAT1	0.0001	0.0326	0.0001
BCL2	0.0461 (0.0002)	NA	NA
HRAS	0.0093	NA	NA
CDKN2B	0.0321 (0.0178)	NA	NA
AURKA	0.0397	0.0032	0.0011

NOTE: Phase values for LD blocks (where defined) are shown in parentheses: *BCL2* (LD block 11, rs1462129 and 1801018) and *CDKN2B* (LD block 2, rs2069418, rs573687, rs2106119, and rs643319). *AGTR1* Phase *P* values are not shown in this table because individual haplotypes had opposing effects in the two series.
Abbreviation: NA, data not available due to genotyping failure.

With respect to *STAT1*, it has been described recently that acetylated forms of its protein bind to NF-κβ, thus decreasing the expression of antiapoptotic NF-κβ targets and regulating carcinogenic events (34). This well-known transcription factor is not only implicated in the RET pathway but is also constitutively expressed in C cell carcinoma and required for RET-mediated malignant transformation (35). *STAT1* also independently activates the NF-κβ and mitogen-activated protein kinase (MAPK) pathways.

The variant rs911160 in *AURKA*, one of the other five genes identified in this study, is located ~4 kb upstream of two nonsynonymous polymorphisms (rs1047972, pVal57Ile; rs2273535, pIle31Phe). Various different studies have shown the role of this last SNP in the development and invasiveness of neoplasias both in human and in mouse (36, 37). Moreover, using HapMap data and covering all the gene sequence, we confirmed that all these variants were included in the same LD block. The role of both nonsynonymous variants in sMTC development remains to be elucidated.

Functional studies. The analyzed 8.1 kb of the CDKN2B promoter showed positive association using both approaches and included the putative TFBS SNP (rs2106119). Although this SNP potentially alters the binding of PAX6 (PupaSuite prediction), a transcription factor participating in the development of neuroendocrine tissue (38), molecular studies showed that it was not functional. However, we showed that the transcription factor HNF1 differentially binds to the alleles of rs7044859, as predicted by PupaSuite. This finding pinpoints the SNP rs7044859 as a functional variant that may explain the association observed in this study with the marker rs2106119 because these two SNPs are in complete LD according to HapMap data.

Different studies have assumed that most SNPs that increase the susceptibility to developing complex human genetic diseases are located at regulatory genomic regions (39). Despite that this TFBS SNP is located 8.5 kb upstream of the transcription start of *CDKN2B*, the putative relation between the transcription factor HNF1 and *CDKN2B* expression is supported by the existence of transcription enhancers located at even greater distances, as occurs with the *β-globin* gene (40, 41). Moreover, no other genes overlap with the region around rs7044859. All these data reinforce the role of this functional SNP of *CDKN2B* in sMTC susceptibility, given that it may be modifying the expression levels of this well-known tumor suppressor gene. In fact, the decreased expression of *CDKN2B* as a consequence of promoter hypermethylation has been described to be involved in the development of hematologic malignancies, melanoma, and ovarian cancer (42, 43). Inversely, *CDKN2B* expression is increased in response to HRAS-Raf-MAPK/extracellular signal-regulated kinase (ERK) kinase signaling (44), a cascade involved in RET signaling. Another fact that supports the relation between *CDKN2B* and sMTC involves an intronic *CDKN2B* polymorphism (rs974336), associated previously not only with the development of papillary and medullary thyroid cancer but also with a higher prevalence of extensive lymph node metastasis (45).

HRAS and *CDK6* are related to *CDKN2B*, the former through signaling as mentioned above (29, 30) and the latter through its involvement in cell cycle regulation. This offers biological support for our finding that both are low-penetrance genes for sMTC. Little is known about how the function of *COMT* might be related to the

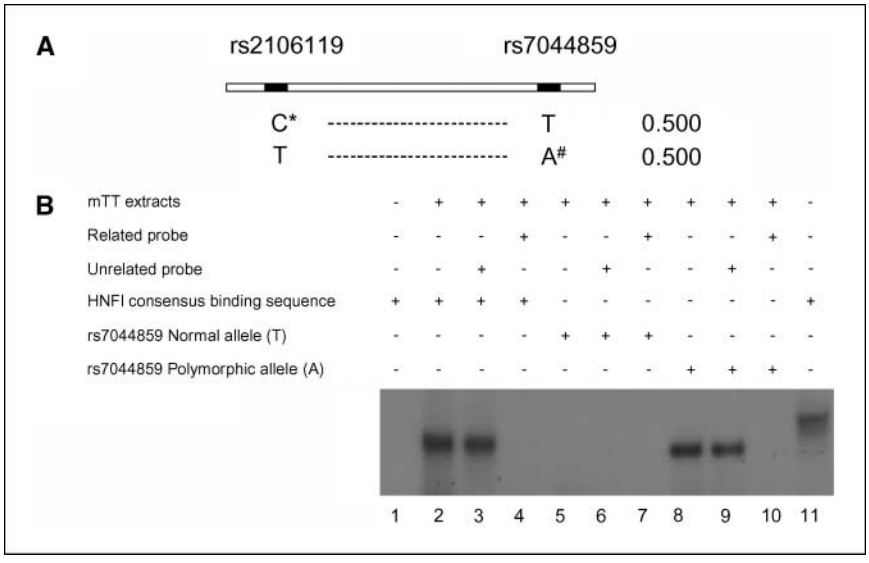

Figure 1. *A,* schematic representation of the LD between rs2106119 and rs7044859. The C allele (marked with an asterisk) of rs2106119, and therefore the C-T haplotype, was shown to act as a risk factor through individual SNP and haplotype analysis. *B,* EMSA using protein extracts from mTT cells (derived from a MTC) and dsDNA probes (HNF1 consensus binding sequence and normal or polymorphic rs7044859 allele). A 100-fold excess of unlabeled related (*lanes 4, 7*, and *10*) or unrelated (Cre1; *lanes 3, 6*, and *9*) oligonucleotides were used for competition. HEPG2 nuclear protein extracts were used as positive controls for the expression of HNF1 (*lane 11*). The EMSA shows binding exclusively for the polymorphic allele of rs7044859 (A allele), which therefore potentially could act as a protective SNP (marked with square in *A*) due to an increased transcription of *CDKN2B*, a well-known tumor suppressor gene.

development of MTC, but its identification in this study may open new avenues for exploring the etiology of this complex disease.

Conclusions. To date, *RET* and its ligands and coreceptors have been linked to sMTC development. However, no association studies have been done for the RET downstream effectors. Our study identified seven low-penetrance genes (*BCL2, STAT1, HRAS, CDKNB2B, CDK6, COMT*, and *AURKA*), five of which are directly and indirectly involved in RET signaling either through the PI3K cascade or through the RAS-RAF-ERK cascade (29, 30). The potential role of *CDKN2B* was confirmed by a functional assay showing the role of a SNP in the promoter region in altering the binding of the transcription factor HNF1. The elucidation of the global effect of such genetic risk variants acting in combination will be essential to better understand the behavior of sMTC and to verify the potential modifier role of these variants in the development of MEN type 2–associated MTC. This will be an important first step in the application of low-penetrance gene discovery to the clinical management of these patients. Further studies involving a larger international series will be necessary to analyze the interactions between the identified low-penetrance genes and how they may influence the phenotype in familial forms.

Acknowledgments

Received 5/7/2007; revised 6/29/2007; accepted 7/20/2007.

Grant support: Spanish MTC study was supported by "Fondo de Investigaciones Sanitarias" (FIS) Projects no. PI052509 and PI041313 (M. Robledo), PI041216 and RD06/0020/0060 (P. Santisteban) ISCIII CIBER-ER from the Spanish Ministry of Health, the Spanish Ministry of Education and Science grant BIO2005-01078, and National Institute of Bioinformatics (http://www.inab.org) a platform of Genome, Spain. UK MTC study was funded by Cancer Research UK. S. Ruiz-Llorente has a FIS postdoctoral contract (CD05/00055). C. Montero-Conde and I. Landa were supported by predoctoral fellowships of Basque Country Government and Spanish National Health Fund (PI052509; FIS12/2005).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

References

1. Randolph GW, Maniar D. Medullary carcinoma of the thyroid. Cancer Control 2000;7:253–61.
2. De Lellis RA, Lloyd RV, Heitz PU, Eng C. World Health Organization classification of tumours: pathology and genetics of tumours of endocrine organs. Chapter 2: tumour of the thyroid and parathyroid. Lyon: IARC Press; 2004. pp. 86–91.
3. Mulligan LM, Kwok JB, Healey CS, et al. Germ-line mutations of the RET proto-oncogene in multiple endocrine neoplasia type 2A. Nature 1993;363:458–60.
4. Ponder BA, Smith D. The MEN II syndromes and the role of the ret proto-oncogene. Adv Cancer Res 1996;70: 179–222.
5. Ponder BA. The phenotypes associated with ret mutations in the multiple endocrine neoplasia type 2 syndrome. Cancer Res 1999;59:1736–41s; discussion 1742s.
6. Cranston AN, Ponder BA. Modulation of medullary thyroid carcinoma penetrance suggests the presence of modifier genes in a RET transgenic mouse model. Cancer Res 2003;63:4777–80.
7. Robledo M, Gil L, Pollan M, et al. Polymorphisms G691S/S904S of RET as genetic modifiers of MEN 2A. Cancer Res 2003;63:1814–7.
8. Vanhorne JB, Andrew SD, Harrison KJ, et al. A model for GFR α4 function and a potential modifying role in multiple endocrine neoplasia 2. Oncogene 2005;24:1091–7.
9. Marsh D, Zori R. Genetic insights into familial cancers—update and recent discoveries. Cancer Lett 2002;181:125–64.
10. Risch NJ. Searching for genetic determinants in the new millennium. Nature 2000;405:847–56.
11. Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for Mendelian disease, future approaches for complex disease. Nat Genet 2003;33 Suppl:228–37.
12. Borrego S, Fernandez RM, Dziema H, et al. Evaluation of germline sequence variants of GFRA1, GFRA2, and GFRA3 genes in a cohort of Spanish patients with sporadic medullary thyroid cancer. Thyroid 2002;12: 1017–22.
13. Cebrian A, Lesueur F, Martin S, et al. Polymorphisms in the initiators of RET (rearranged during transfection) signaling pathway and susceptibility to sporadic medullary thyroid carcinoma. J Clin Endocrinol Metab 2005; 90:6268–74.
14. Day N, Oakes S, Luben R, et al. EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. Br J Cancer 1999;80 Suppl 1:95–103.
15. Conde L, Vaquerizas JM, Dopazo H, et al. PupaSuite: finding functional single nucleotide polymorphisms for large-scale genotyping purposes. Nucleic Acids Res 2006; 34:W621–5.
16. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. Science 2002;296:2225–9.
17. Consortium TIH. The International HapMap project. Nature 2003;426:789–96.
18. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity 2005;95:221–7.
19. Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. Am J Hum Genet 2004;74:765–9.
20. Nyholt DR. Evaluation of Nyholt's procedure for multiple testing correction—author's reply. Hum Hered 2005;60:61–2.
21. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. Behav Brain Res 2001;125:279–84.
22. Stephens M, Donnelly P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet 2003;73:1162–9.
23. Moya CM, Perez de Nanclares G, Castano L, et al. Functional study of a novel single deletion in the TITF1/NKX2.1 homeobox gene that produces congenital hypothyroidism and benign chorea but not pulmonary distress. J Clin Endocrinol Metab 2006;91: 1832–41.
24. Zhou YH, Zheng JB, Gu X, Saunders GF, Yung WK. Novel PAX6 binding sites in the human genome and the role of repetitive elements in the evolution of gene regulation. Genome Res 2002;12:1716–22.
25. Usui T, Kuno T, Ueyama H, Ohkubo I, Mizutani T. Proximal HNF1 element is essential for the induction of human UDP-glucuronosyltransferase 1A1 by glucocorticoid receptor. Biochem Pharmacol 2006;71:693–701.
26. Ardlie KG, Lunetta KL, Seielstad M. Testing for population subdivision and association in four case-control studies. Am J Hum Genet 2002;71:304–11.
27. Freedman ML, Reich D, Penney KL, et al. Assessing the impact of population stratification on genetic association studies. Nat Genet 2004;36:388–93.
28. Pharoah PD, Dunning AM, Ponder BA, Easton DF. Association studies for finding cancer-susceptibility genetic variants. Nat Rev Cancer 2004;4:850–60.
29. Manie S, Santoro M, Fusco A, Billaud M. The RET receptor: function in development and dysfunction in congenital malformation. Trends Genet 2001;17:580–9.
30. Takahashi M. The GDNF/RET signaling pathway and human diseases. Cytokine Growth Factor Rev 2001;12: 361–73.
31. Pugazhenthi S, Nesterova A, Sable C, et al. Akt/protein kinase B up-regulates Bcl-2 expression through cAMP-response element-binding protein. J Biol Chem 2000;275:10761–6.
32. Drosten M, Hilken G, Bockmann M, et al. Role of MEN2A-derived RET in maintenance and proliferation of medullary thyroid carcinoma. J Natl Cancer Inst 2004; 96:1231–9.
33. Komaki S, Kohno M, Matsuura N, et al. The polymorphic 43Thr bcl-2 protein confers relative resistance to autoimmunity: an analytical evaluation. Hum Genet 1998;103:435–40.
34. Kramer OH, Baus D, Knauer SK, et al. Acetylation of Stat1 modulates NF-κB activity. Genes Dev 2006;20: 473–85.
35. Ludwig L, Kessler H, Wagner M, et al. Nuclear factor-κB is constitutively active in C-cell carcinoma and required for RET-induced transformation. Cancer Res 2001;61:4526–35.
36. Ewart-Toland A, Briassouli P, de Koning JP, et al. Identification of Stk6/STK15 as a candidate low-penetrance tumor-susceptibility gene in mouse and human. Nat Genet 2003;34:403–12.
37. Ewart-Toland A, Dai Q, Gao YT, et al. Aurora-A/STK15 T+91A is a general low penetrance cancer susceptibility gene: a meta-analysis of multiple cancer types. Carcinogenesis 2005;26:1368–73.
38. Chi N, Epstein JA. Getting your Pax straight: Pax proteins in development and disease. Trends Genet 2002;18:41–7.
39. Mackay TF. Quantitative trait loci in *Drosophila*. Nat Rev Genet 2001;2:11–20.
40. Szutorisz H, Dillon N, Tora L. The role of enhancers as centres for general transcription factor recruitment. Trends Biochem Sci 2005;30:593–9.
41. Vieira KF, Levings PP, Hill MA, et al. Recruitment of transcription complexes to the β-globin gene locus *in vivo* and *in vitro*. J Biol Chem 2004;279:50350–7.
42. Kusy S, Larsen CJ, Roche J. p14ARF, p15INK4b, and p16INK4a methylation status in chronic myelogenous leukemia. Leuk Lymphoma 2004;45:1989–94.
43. Liu Z, Wang LE, Wang L, et al. Methylation and messenger RNA expression of p15INK4b but not p16INK4a are independent risk factors for ovarian cancer. Clin Cancer Res 2005;11:4968–76.
44. Malumbres M, Perez De Castro I, Hernandez MI, Jimenez M, Corral T, Pellicer A. Cellular response to oncogenic ras involves induction of the Cdk4 and Cdk6 inhibitor p15(INK4b). Mol Cell Biol 2000;20: 2915–25.
45. Goretzki PE, Gorelov V, Dotzenrath C, Witte J, Roeher HD. A frequent mutation/polymorphism in tumor suppressor gene INK4B (MTS-2) in papillary and medullary thyroid cancer. Surgery 1996;120:1081–8.