Specific haplotypes of the RET proto-oncogene are over-represented in patients with sporadic papillary thyroid carcinoma

F Lesueur, M Corbex, J D McKay, J Lima, P Soares, P Griseri, J Burgess, I Ceccherini, S Landolfi, M Papotti, A Amorim, D E Goldgar, G Romeo

Background: Papillary thyroid carcinoma (PTC), which may be sporadic (95%) or familial (5%), has a prevalence adjusted for age in the general population of 1:100 000. Somatic rearrangements of the RET proto-oncogene are present in up to 66% of sporadic tumours, while they are rarely found in familial cases.

Purpose: In order to determine if some variants of this gene, or a combination of them, might predispose to PTC, we looked for an association of RET haplotype(s) in PTC cases and in controls from four countries matched for sex, age, and population.

Methods: Four single nucleotide polymorphisms (SNPs) across the RET coding sequence were typed and haplotype frequencies were estimated. Genotype and haplotype distributions were compared among these cases and controls.

Results: Ten haplotypes were observed, the seven most frequent of which have been previously described in sporadic Hirschsprung patients and controls. The single locus analyses suggested association of exon 2 and exon 13 SNPs with sporadic PTC. The haplotype analysis showed over-representation of one haplotype in French and Italian sporadic PTC, whereas a different haplotype was significantly under-represented in French familial PTC.

Conclusions: Our data suggest that some variants of RET and some specific haplotypes may act as low penetrance alleles in the predisposition to PTC.

The RET proto-oncogene (MIM 164761) is normally expressed in cells of neural crest origin. Its mutations are associated with a wide spectrum of disease. In Hirschsprung disease (HSCR), a congenital disorder of the enteric nervous system (MIM 142623), point mutations, insertions, or heterozygous deletions of the whole gene lead to a loss of function of the Ret protein. In multiple endocrine neoplasia type 2A and 2B (MEN 2A, MEN 2B, MIM 171400) and in medullary thyroid carcinoma (MTC, MIM 155240), point mutations in the coding sequence of RET result in a constitutive activation of the tyrosine kinase domain of the Ret receptor. Finally, in papillary thyroid carcinoma (PTC, MIM 188550), which is the most common form of thyroid malignancy, occurring sporadically in 95% of cases, the disease is very often associated with somatic rearrangements of the RET proto-oncogene, resulting in constitutive activation of the RET tyrosine kinase domain.

HSCR is characterised by genetic heterogeneity. When inherited, long segment HSCR is usually transmitted as an autosomal dominant trait with incomplete penetrance and variable expressivity, whereas short segment HSCR is transmitted in a complex way. The same definition of a complex genetic disorder can be applied to familial PTC because a variety of genes are involved in its predisposition and the mode of inheritance of familial cases is non-Mendelian. Familial PTC, which accounts for only 5% of the total PTC cases, is reported to be characterised by more aggressive behaviour than sporadic PTC. We have previously shown that RET was not linked with disease in the majority of PTC families and we recently mapped a major gene predisposing to this tumour phenotype in 42% of the families with recurrence of PTC. While a somatic rearrangement of RET (RET/PTC1) was found only once in the tumour material of a familial PTC case, RET rearrangements are observed in 2.5-35% of sporadic PTC tumours. No germline mutations of RET have ever been observed to cosegregate with PTC.

In complex disorders, common polymorphic variants can be associated with the disease phenotype, thus modifying the risk of occurrence. Over the three last years, several groups have used RET SNPs to determine whether polymorphic variants of RET might represent low penetrance alleles predisposing to RET associated disorders.

No germline mutations of RET have ever been reported so far. In PTC, as in MTC, variation in the DNA

Abbreviations: PTC, papillary thyroid carcinoma; SNP, single nucleotide polymorphism; HSCR, Hirschsprung disease; MTC, medullary thyroid carcinoma; MEN, multiple endocrine neoplasia
sequence of RET might modify expression of the Ret protein and/or lead to activation of the gene. Therefore, we looked for an association of RET haplotype(s) in patients from different countries. For this, we selected four SNPs (located in exons 2, 13, 14, and 15) on the basis of published data from the seven known SNPs encompassing the coding sequence of RET. We analysed 247 sporadic PTC cases together with 319 controls.

Since thyroid is the cancer site with the highest familial relative risk among primary cancer sites (FRR=8.60), we also studied RET haplotypes in a sample of 80 French familial PTC cases.

PATIENTS, MATERIALS, AND METHODS

Sporadic PTC and controls

We studied 247 subjects affected with sporadic PTC who were recruited from four countries: Italy (n=100), Portugal (n=95), France (n=26), and Australia (n=26). Healthy blood donors from Portugal (n=62), France (n=102), and Australia (n=55) served as controls. For Italy, controls (n=100) were recruited from patients attending the same hospital as PTC patients for reasons other than cancer or thyroid disease. Controls were matched for age, sex, and geographical origin and were unrelated to the PTC subjects.

DNA from PTC patients was extracted from normal thyroid tissue surrounding the tumour or from peripheral blood. Genomic DNA of controls was obtained from peripheral blood or tissue surrounding the tumour. DNA was extracted according to a standard phenol-chloroform protocol.

Familial PTC cases

Familial PTC patients came from the families collected at IARC through the International Consortium for the Genetics of Non-Medullary Thyroid Carcinoma. Eighty unrelated French patients with a family history of PTC (at least one relative with PTC) were selected. DNA samples were extracted from blood using the Puregene Kit (Gentra Systems, Minneapolis, MN).

Genotyping

Exons 2, 13, 14, and 15 of RET were amplified from genomic DNA using the primers shown in table 1. The PCR reactions were carried out in a volume of 25 µl and included 1× buffer PCR, 200 mmol/l of each deoxy-NTP, 50 ng DNA, MgCl₂ as indicated in table 1, 0.2 U Taq Platinum polymerase (Gibco BRL, Life Technologies), and 10 pmol of each primer. The reactions were carried out in a 9600 GeneAmp PCR System with the following thermal profile: denaturation at 96°C for five minutes; 30 cycles of 94°C for 30 seconds, annealing temperature as indicated in table 1 for 45 seconds, and 72°C for one minute; with a final extension at 72°C for 5 minutes.

The SNPs G135A (A45A in exon 2), T2307G (L769L in exon 13), C2508T (S836S in exon 14), and C2712G (S904S in exon 15) were genotyped by differential restriction digestion with the appropriate enzymes, as previously described for exons 2, 13, and 15 according to the manufacturer’s recommendations (MBI fermentas, BioLabs). The restriction digestion products were fractionated by electrophoresis through 3% agarose gels containing ethidium bromide and visualised under UV transillumination. For exon 14, direct sequencing of the PCR product was carried out.

Statistical analysis

Allelic frequencies for the four RET SNPs were calculated from the genotype frequencies and compared among countries by the chi-square test. Hardy-Weinberg equilibrium was tested in controls from each country. The extent of linkage disequilibrium between SNPs was expressed in terms of D’=D/Dmax or D/Dmin. Pairwise linkage disequilibrium coefficients were calculated by using Transposer macros. Association with the disease was tested by using chi-square tests when computationally feasible or Fisher’s exact test for tables with small expected cell counts.

For controls and sporadic PTC patients, haplotype frequencies were estimated by the maximum likelihood method from genotype data through the use of the E-M algorithm under the assumption of Hardy-Weinberg equilibrium (Arlequin software). In familial PTC, haplotypes frequencies were estimated using the same method because genotypic information from parents was incomplete.

Haplotype frequencies were compared using Monte Carlo simulations to estimate the significance level of the test statistic (CLUMP software). After combining haplotypes with a frequency lower than 0.05 to avoid small expected cell counts, test T1 (chi-square on the whole table) was carried out. Comparisons of each haplotype versus the others were also performed by chi-square test.

The significance level of each statistical test was taken to be 0.05; this threshold may seem quite high considering the number of tests performed, but is adequate considering the exploratory character of this study conducted in limited samples.

RESULTS

Comparison of allelic frequencies and linkage disequilibrium between populations

The four polymorphisms studied, their position, nucleotide changes, and allele frequencies are shown in table 2. All the four SNPs are in the coding sequence of RET but represent silent mutations. Allele frequencies did not significantly differ among the four populations (Italian, French, Portuguese, and Australian) (table 2). No significant deviation from Hardy-Weinberg equilibrium was observed for any polymorphism.

We also checked the pairwise linkage disequilibrium coefficients D’ in the four populations separately and did not observe any significant variation between these populations.

Table 1 Primers, PCR conditions, and restriction enzymes used for the genotyping

<table>
<thead>
<tr>
<th>Exon</th>
<th>Primers sequence (5’ &gt; 3’)</th>
<th>Annealing temp (°C)</th>
<th>Mg conc (mmol/l)</th>
<th>Product size (bp)</th>
<th>Restriction enzyme</th>
<th>Size of fragments</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Fwd: TCACCTCCATCCCTACTTCC</td>
<td>60</td>
<td>1.2</td>
<td>263</td>
<td>Eagl</td>
<td>207+87</td>
</tr>
<tr>
<td></td>
<td>Rv: CTATGCGGACACTTGAGC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Fwd: CTCCTCGTGAACCTGGG</td>
<td>65</td>
<td>1.2</td>
<td>238</td>
<td>Taql</td>
<td>139+99</td>
</tr>
<tr>
<td></td>
<td>Rv: TACCCCTGCACTGGCCTTA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Fwd: AAGACCCAAGCTGCGTGA</td>
<td>66</td>
<td>1.5</td>
<td>328</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Rv: GTGGTGCTGAGGGTTGG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Fwd: GACTCGTGCTATTTTCTCTC</td>
<td>59</td>
<td>1.5</td>
<td>235</td>
<td>RsaI</td>
<td>106+129</td>
</tr>
<tr>
<td></td>
<td>Rv: TATCITCTCCAGGCTTCCAAA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
(data not shown). These results allowed us to analyse the data of the four populations pooled together. The pairwise linkage disequilibrium coefficients $D'$ in the pooled populations are given in table 3.

### Association of individual polymorphic loci with PTC (single locus analyses)

In order to test the impact that polymorphisms of the RET proto-oncogene might have on the aetiology of PTC, we investigated the genotype distribution of the four SNPs in exons 2, 13, 14, and 15 in patients with sporadic PTC. Table 4 presents the results of single locus analyses in the four populations considered independently and in the pooled populations. Of the four loci, only the two SNPs in exons 2 and 13 showed significant association with PTC; homozygotes for G G of exon 2 were over-represented in Italian PTC patients. The same trend was observed in the Portuguese and French populations, but not in the Australian population; homozygotes for G G of exon 13 were over-represented in the PTC patients when the four populations were pooled together. This trend was observed in each population taken independently, except in the Australian population.

Allele G of exon 2 is in weak association with allele A of exon 13 ($D'=0.42$, table 2), thus the effect observed for exon 262 Lesueur, Corbex, McKay, et al www.jmedgenet.com

<table>
<thead>
<tr>
<th>Exon</th>
<th>Amino acid change</th>
<th>Nucleotide substitution</th>
<th>Less frequent allele frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>A45A</td>
<td>G135A</td>
<td>Italy 0.30, Portugal 0.33, France 0.29, Australia 0.24</td>
</tr>
<tr>
<td>13</td>
<td>L769L</td>
<td>T2307G</td>
<td>Italy 0.24, Portugal 0.19, France 0.27, Australia 0.25</td>
</tr>
<tr>
<td>14</td>
<td>S836S</td>
<td>C2508T</td>
<td>Italy 0.06, Portugal 0.05, France 0.07, Australia 0.05</td>
</tr>
<tr>
<td>15</td>
<td>S904S</td>
<td>C2712G</td>
<td>Italy 0.21, Portugal 0.25, France 0.26, Australia 0.18</td>
</tr>
</tbody>
</table>

*Frequency of the rarer allele is given. Bold indicates significant results at the alpha=0.05 level.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Frequency†</th>
<th>Odds ratio‡</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PTC Controls</td>
<td>Model All† Italy Portugal Australia France</td>
</tr>
<tr>
<td>Exon 2</td>
<td>0.26 0.29</td>
<td>Codom 0.83 0.70 0.78 1.31 0.72</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dom 0.73 0.54* 0.68 1.19 0.71</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rec 0.94 0.85 1.10 0.44 2.13</td>
</tr>
<tr>
<td>Exon 13</td>
<td>0.22 0.24</td>
<td>Codom 1.17 1.26 1.48 0.92 1.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dom 1.04 1.08 1.32 0.95 0.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rec 2.22* 3.57 5.56 0.71 2.44</td>
</tr>
<tr>
<td>Exon 14</td>
<td>0.05 0.06</td>
<td>Codom 0.89 0.90 0.86 1.29 1.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dom 0.85 0.90 0.74 1.30 1.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rec – – – – – –</td>
</tr>
<tr>
<td>Exon 15</td>
<td>0.19 0.19</td>
<td>Codom 1.01 0.91 0.75 0.48 1.84</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dom 1.03 0.96 0.76 0.45 1.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rec 1.07 1.53 2.11 – –</td>
</tr>
</tbody>
</table>

*Significant p value (less than 0.05).
†Frequency of the rare allele in the pooled population.
‡Odds ratio is given for each model: codominant (comparison of allelic frequencies), dominant (heterozygotes and rare homozygotes have the same risk), and recessive (heterozygotes and frequent homozygotes have the same risk), respectively.
13 G is not the result of linkage disequilibrium with exon 2 G (and vice versa). The results suggest that each of the G alleles has a separate effect on risk, or that the G G haplotype carries a functional at risk allele.

**RET haplotype analysis in PTC and unrelated controls**

A total of 10 haplotypes was observed in the case-control groups (table 5). The six most frequent haplotypes exist in the four populations. The additional rare haplotypes were pooled in one class to allow comparison between PTC and controls.

Although a chi-square test performed on the whole table (test T1, see Patients, materials, and methods) was not significant, some of the haplotypes displayed different frequencies between cases and controls, indicating putative associations with the disease.

In the French population, the G T C G haplotype was significantly over-represented in PTC. This was not observed in the three other populations. This association was the result of the fact that in our French control group, the frequency of this haplotype was lower than in the three other control groups. This indicated that the association was probably not a genuine one.

In the Italian population, the G G C C haplotype was over-represented in PTC patients. This was consistent with the result obtained with the single locus analysis (over-representation of the homozygous G G in exon 2, table 4). This haplotype was also over-represented in French PTC, although the difference between cases and controls was not statistically significant (p=0.12). When the four populations were pooled (table 5), the G G C C haplotype was significantly over-represented (p=0.03). In the Portuguese and Australian populations, the frequencies of this haplotype were not different in cases and controls. The G G C C haplotype may therefore reflect the presence of a low penetrance predisposing allele for PTC in the Italian and French populations. However, owing to the small sample set and the number of tests performed, this result should be confirmed on larger samples.

**RET genotype and haplotype analyses in familial PTC**

When comparing the four SNP allele frequencies in the 80 French familial PTC and the 102 French controls (table 6), we observed an under-representation of the rare T allele of the SNP in exon 14 (S836S) which did not exist in sporadic PTC.

In familial PTC, we observed the seven haplotypes already observed in the French population (table 6) but one of them was under-represented in familial PTC (the G G T C haplotype). This under-representation of G G T C was not observed in the sporadic PTC. This haplotype was the only one in the French population comprising the rare T allele of S836S in exon 14. This result was consistent with the single locus analysis.

**DISCUSSION**

It is now well established that polymorphisms may have a relatively strong effect on disease susceptibility. This has been illustrated by a number of reported examples, the last one in chronological order being represented by variants in the NOD2 gene associated with susceptibility to Crohn’s disease.**

In the present study, we observed that some variants within RET could represent low penetrant alleles for the PTC phenotype. Of the four SNPs studied, the strongest association with
PTC was found for polymorphism A45A (G135A) in exon 2 and L769L (T2307G) in exon 13. They were also the ones found to be associated with sporadic HSCR. However, the A allele of exon 2 was found to be at risk for HSCR whereas it was found to be a protective allele for PTC in our study. This observation is consistent with the fact that mutations of RET occurring in HSCR disease and in thyroid cancer have opposite effects. In HSCR, known mutations lead to a loss of function of the RET protein, whereas in thyroid cancer, known mutations lead to a constitutive activation of the receptor (germinal point mutations in MTC and somatic RET/PTC rearrangements in sporadic PTC).

With the four SNPs studied, a total of 10 haplotypes of RET was observed when samples from the four countries were pooled. The seven most frequent haplotypes have been described by Borrego et al in controls of Spanish and German origin. Distribution of some of these haplotypes differed between cases and controls. The G G C C haplotype is over-represented in both Italian and French sporadic PTC and this over-representation is significant when all four populations are pooled (OR 1.68, 95% CI 1.04-2.71). The G G C C haplotype includes the G allele of exon 2 and the G allele of exon 13, the two alleles that are associated with an increasing risk in the single locus analysis.

There is only one other haplotype that includes the two Gs, G G T C, but it is very rare owing to the very low frequency of the exon 14 T allele (6.0%, table 3). To confirm the association and to ascertain if the increased risk is the result of independent effects of exon 2 and exon 13 or to some exon 2-exon 13 containing haplotypes, larger studies are required.

Another interesting finding raised by this study is the differential association with RET in sporadic and familial PTC. The haplotype G G T C is found in only 1.9% of the French familial PTC and this over-representation is significant when all four populations are pooled (OR 1.68, 95% CI 1.04-2.71). This under-representation is not observed in sporadic PTC. Interestingly, an under-representation of the exon 14 T allele is observed in the French familial PTC while an over-representation of the T allele has been reported in MTC patients. This T allele is even more represented in MTC patients with the somatic mutation M918T of RET. These results suggest an opposite role played by the exon 14 T allele in the aetiology of thyroid carcinoma of follicular and parafollicular cell origin.

Recently, we mapped a major predisposing gene on chromosome 2q21 involved in 42% of 80 families with recurrence of PTC. The influence of modifier genes in these families is suggested by the intrafamilial heterogeneity of the phenotype (different variants of PTC and presence of both benign and malignant tumours in the same family). Using the French subset of the PTC families used for mapping of locus 2q21, we showed in the present study that the G G C C haplotype is also more represented in the French familial PTC (OR 1.94, 95% CI 0.83-4.56) and could therefore act as a modifier or low predisposing allele.

In a previous study, we tested by linkage the involvement of RET as a major predisposing gene in a subset of the 80 PTC families used to map a new gene on chromosome 2. The proportion of families linked to RET could not be estimated, owing to the small number of families and to the high genetic heterogeneity between these families. The aim of the present study was to evaluate if common alleles of RET could represent a low risk for PTC. The magnitude of the effect between the RET SNPs/haplotypes and PTC is quite modest at best, that is, ORs of around 2 for a haplotype with a frequency of 5%. By itself, this locus would not be expected to contribute much to familial clustering with a predicted familial risk of 1.04 (about 2% of the overall familial effect). Thus, it is not surprising then that there would be no evidence of linkage between RET and PTC in a family based linkage analysis.

The mechanism by which the silent polymorphisms may act in the development of PTC or other RET related diseases is unknown. Assuming a direct effect of one of the four RET polymorphisms considered, a series of possible effects should be taken into account, including transcript stability, RNA splicing, and DNA protein binding and protein folding. Hypotheses have been made regarding the possible mechanisms. For instance, it has been proposed that the silent sequence variant at codon 45 in exon 2 could lead to aberrantly spliced products, resulting in a protein with a 21 amino acid deletion in the extracellular domain, altering a part of the extracellular signal peptide sequence. The C>T transition of polymorphism S836S in exon 14 generates a putative binding site (ACTcaAGT) for a thyroid specific factor not yet identified. As an attempt to explain a possible role of the T allele of S836S over-represented in MTC, Griseri et al carried out gel retardation assays and did not find any difference between normal (C) and variant (T) oligonucleotide binding profiles. A possible interference of this SNP with correct RET mRNA processing was also experimentally excluded.

To test the possible interference of the other three SNPs independently and of some combinations of all of them (corresponding to the haplotypes sharing different breaking down between cases and controls, G G C C, G T C C, G G T C) with correct splicing, we have used appropriate gene predictions that are included in the NIX software (see Data access). The exon prediction analyses performed could not detect any abnormal alternative splicing. Alternatively, the over-represented haplotypes might be in linkage disequilibrium with another nucleotide variant, of either the RET gene or flanking regions, displaying a low penetrant causative effect by itself. If our results are confirmed in other populations, a search for mutations should be undertaken in subjects sharing the specific haplotypes (G G C C in Italian PTC patients and G T C C in French PTC patients) and functional analyses will also be needed to elucidate the precise mechanism of the haplotypes associated with the PTC phenotype.

ACKNOWLEDGEMENTS

We are grateful to Suzanne Pauly and Cyril Lafaye for their technical assistance. Fabienne Lesueur is recipient of a Special Training Award from the International Agency for Research on Cancer. Data access. URLS for data in this report are as follows: Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim (for Hirshsprung disease (MIM 142623), for multiple endocrine neoplasia type 2 (171400), for medullary thyroid carcinoma (155240), for papillary thyroid carcinoma (MIM 188550), for RET proto-oncogene (MIM 164761). NIX software is available on http://www.hgmp.mrc.ac.uk/NIX/.

Authors’ affiliations
F Lesueur, M Corbex, J D McKay, D E Goldgar, G Romeo, International Agency for Research on Cancer, Lyon, France
J Lima, P Soares, A Amorim, IPATIMUP, Porto, Portugal
G Griseri, I Cecherini, Laboratorio di Genetica Molecolare, Istituto Giannina Gaslini, Genova, Italy
J D McKay, J Burgess, Department of Endocrinology, Royal Hobart Hospital, Hobart, Tasmania, Australia
S Landolfi, M Papotti, Dipartimento di Anatomia Patologica, University of Turin, Italy
P Soares, Faculty of Medicine, University of Porto, Portugal
A Amorim, Faculty of Sciences, University of Porto, Portugal
G Romeo, Laboratorio di Genetica Medica, Dipartimento di Medicina Interna, Cardioangiologiopa, Epatologia, University of Bologna Medical School, Italy

REFERENCES


