

## ARTICLE

# Localizing a putative mutation as the major contributor to the development of sporadic Hirschsprung disease to the *RET* genomic sequence between the promoter region and exon 2

Grzegorz M Burzynski<sup>1</sup>, Ilja M Nolte<sup>2</sup>, Jan Osinga<sup>1</sup>, Isabella Ceccherini<sup>3</sup>, Bas Twigt<sup>1</sup>, Saskia Maas<sup>4</sup>, Alice Brooks<sup>1,5</sup>, Joke Verheij<sup>6</sup>, Ivan Plaza Menacho<sup>1</sup>, Charles HCM Buys<sup>1,6</sup> and Robert MW Hofstra<sup>\*,1</sup>

<sup>1</sup>Department of Medical Genetics, University of Groningen, Groningen, The Netherlands; <sup>2</sup>Department of Medical Biology, University of Groningen, Groningen, The Netherlands; <sup>3</sup>Laboratorio di Genetica Molecolare, Istituto G. Gaslini, Genova, Italy; <sup>4</sup>Department of Clinical Genetics, Academic Medical Centre, Amsterdam, The Netherlands; <sup>5</sup>Department of Clinical Genetics, Erasmus Medical Center Rotterdam, Rotterdam, The Netherlands; <sup>6</sup>Department of Clinical Genetics, Academic Hospital Groningen, Groningen, The Netherlands

Hirschsprung disease (HSCR), a congenital disorder characterized by intestinal obstruction due to absence of enteric ganglia along variable lengths of the intestinal tract, occurs both in familial and sporadic cases. *RET* mutations have been found in approximately 50% of the families, but explains only a minority of sporadic cases. This study aims at investigating a possible role of *RET* in sporadic HSCR patients. Haplotypes of 13 DNA markers, within and flanking *RET*, have been determined for 117 sporadic HSCR patients and their parents. Strong association was observed for six markers in the 5' region of *RET*. The largest distortions in allele transmission were found at the same markers. One single haplotype composed of these six markers was present in 55.6% of patients versus 16.2% of controls. Odds ratios (ORs) revealed a highly increased risk of homozygotes for this haplotype to develop HSCR (OR > 20). These results allowed us to conclude that *RET* plays a crucial role in HSCR even when no *RET* mutations are found. An unknown functional disease variant(s) with a dosage-dependent effect in HSCR is likely located between the promoter region and exon 2 of *RET*. *European Journal of Human Genetics* (2004) 12, 604–612. doi:10.1038/sj.ejhg.5201199  
Published online 12 May 2004

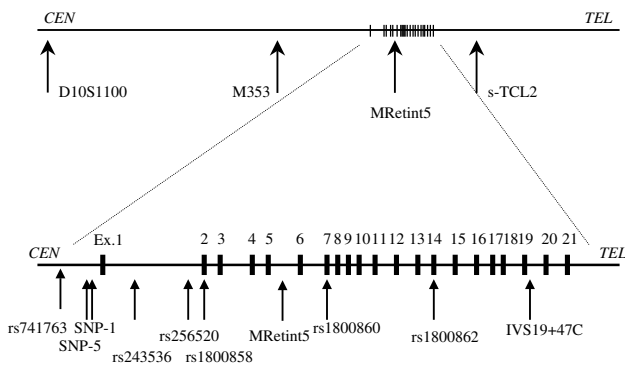
**Keywords:** Hirschsprung disease; HSCR; *RET*; haplotype reconstruction

## Introduction

Hirschsprung disease (HSCR (MIM 142623)) is characterized by the absence of intrinsic ganglion cells in the myenteric and submucosal plexuses of the gastrointestinal

tract resulting in a sometimes life-threatening intestinal obstruction. HSCR is a disorder with a complex pattern of inheritance. Differences in sex ratio with a male predominance (3:1 to 5:1), incomplete penetrance, variable expression and association with a large number of syndromes have been observed.<sup>1–3</sup> Genetic analyses of the disorder made clear that HSCR is a heterogenic disease.<sup>3</sup> To date, mutations in nine genes have been found that contribute to the disease: *RET*,<sup>4,5</sup> *GDNF*,<sup>6,7</sup> *NTN*,<sup>8</sup> *EDNRB*,<sup>9</sup> *EDN3*,<sup>10,11</sup> *ECE1*,<sup>12</sup> *SOX10*,<sup>13,14</sup> *SIP1*<sup>15,16</sup> and *PHOX2B*.<sup>17</sup> Of all of these

\*Correspondence: Dr RMW Hofstra, Department of Medical Genetics, Antonius Deusinglaan 4, 9713 AW Groningen, The Netherlands.  
Tel: +31 50 363 2925; Fax: +31 50 363 2947;  
E-mail: r.m.w.hofstra@medgen.azg.nl  
Received 2 December 2003; revised 17 February 2004; accepted 4 March 2004



**Figure 1** Schematic representation of the *RET* gene and flanking regions. In the upper part of the figure, the positions of the microsatellite markers used in this study are given. In the lower part, the *RET* gene structure is magnified and approximate positions of typed SNPs are shown.

genes, *RET* (MIM 164761) is thought to be the major gene, since almost all familial cases are linked to 10q11.2 the region in which the *RET* gene is located.<sup>18,19</sup> However, mutations in *RET* are detected in about 50% of all familial cases<sup>3,20</sup> and fail to explain 70–90% of the more commonly observed sporadic HSCR cases.<sup>3,20</sup> All other genes do explain no more than about 5–10% of the HSCR cases.<sup>21</sup>

In this study, we aim at investigating a possible role of *RET* in HSCR patients for whom no mutation had been found in the *RET* coding sequence. Therefore, we typed nine SNPs and four microsatellite markers covering the entire genomic region of *RET* (Figure 1) in a group of 117 Dutch sporadic HSCR patients and their 231 parents. Of these, 64 were selected on being investigated for mutations in *RET* and the fact that no mutations could be identified. In 53 patients, *RET* was not screened. We performed single and multilocus association and TDT analysis in order to determine in our patients a possible inherited ancestral haplotype coupled to a putative ancestral disease mutation(s). The two different subgroups of patients, that is, those screened and found negative for *RET* mutations and those not screened for mutations, were also analyzed separately to see if combining these patients was a valid procedure.

## Materials and methods

### Studied patient population

DNA samples for this study were obtained from sporadic HSCR patients and their parents with written informed consent. These patients reside throughout the Netherlands and had been referred to clinical geneticists. Almost all our families ( $n = 113$ ) consist of two parents and the affected child (trios). Of three patients, only one parent was available for the study and one patient participated with his child and spouse. Of these, 64 were screened by means

of DGGE and sequence analysis<sup>20</sup> and found negative for *RET* mutations. These 64 were selected from an initial group of 75 sporadic HSCR patients from whom 11 proved to carry a *RET* mutation. In all, 53 did not undergo the screening. As controls, we used the nontransmitted haplotypes of unaffected parents, like does the Haplotype Relative Risk method.<sup>22,23</sup> The two nontransmitted haplotypes within each trio form pseudocontrol genotypes. Under the hypothesis of random mating in our study population, these pseudocontrol genotypes are expected to be similar to genotypes of population controls. Therefore, we are allowed to use these genotypes to provide estimates of the number of heterozygotes and homozygotes and their odds ratios (ORs) in the population.

### Genotyping

In all patients and their relatives, we genotyped four microsatellite markers and nine single nucleotide polymorphisms (SNPs) in a 400 kb region flanking and containing the *RET* locus (Figure 1). These markers included two SNPs located in the promoter region of the gene (SNP-5; G>A, 202 bp upstream of the start codon, and SNP-1; C>A, 198 upstream of the start codon),<sup>24</sup> and SNP rs741763; G>C located 4 kb upstream of the *RET* gene transcription start site. Two SNPs were located in intron 1 (IVS + 6000A>C/rs2435362 and IVS1–126G>T/rs2565206 (<http://www.ncbi.nlm.nih.gov/SNP/>)) and the following SNPs were from exon 2 (c135G>A/A45A, rs1800858),<sup>25</sup> exon 7 (c1296G>A/rs1800860),<sup>25</sup> exon 14 (c2508C>T/S836S, rs1800862)<sup>25</sup> and intron 19 (IVS19 + 47C>T).<sup>26</sup> Of the four microsatellite markers, two were located upstream of *RET*, namely *D10S1100*<sup>27</sup> (300 kb upstream) and *M353*<sup>28</sup> (80 kb upstream), one located within intron 5, *MREtint5*,<sup>28</sup> and one is located 38 kb downstream the *RET* gene, *s-TCL2*<sup>28</sup> (Figure 1).

Genotyping of the microsatellite markers (*D10S1100*, *M353*, *MREtint5*, *s-TCL2*) was performed on the ABI 377 DNA sequencer machine (PE Biosystems). Data were analyzed by using the Applied Biosystem software. The SNPs, rs2565206, rs1800858, rs1800860, rs1800862 and IVS19 + 47C>T, were genotyped by restriction enzymes digestion. Additionally, two SNPs in *RET* promoter region, SNP-1 and SNP-5 were typed using the pyrosequencing method (Isogen Bioscience, Maarssen, The Netherlands). Table 1 contains characteristics of all genotyped markers.

### Statistical analyses

As a quality control on the scoring of the genotypes, all markers were tested for Hardy–Weinberg equilibrium (HWE) before analysis. For this test, we only used DNA specimens from control individuals. If resulting *P*-values were smaller than 0.05, this was regarded as a sign of low quality of the genotyping data and the corresponding marker was discarded from further analyses.

**Table 1** Characteristics of the typed markers around and within the *RET* locus

Name	Position with reference to <i>RET</i> gene and possible alleles	Genomic position according to NCBI (bp)	Heterozygosity	Primer sequences	Restriction enzyme used
D10S1100	–300 kb to ATG, CA repeat	43 047 427–43 047 448	0.668	F: 5'-TTGGGCATACTAGGCAACAGA-3' R: 5'-ACAAGTCAGAGATGGCTTTACC-3'	—
M353	–80 kb to ATG, CA repeat	43 267 787–43 267 811	0.591	F: 5'-TTAAGACTGGAACCTCTCC-3' R: 5'-CTTCCAAAATCAATACAGGC-3'	—
rs741763	–4 kb to ATG, G>C	43 343 542	0.490	F: 5'-GCCTTCCCCTGCTACTG-3' R: 5'-CTTCCCTCGCCCTCCTTCT-3'	AluI
SNP-5	–202 bp to ATG, G>A	43 347 703	0.726	Sequenced by Isogen Bioscience BV Maarsse, The Netherlands	—
SNP-1	–198 bp to ATG, C>A	43 347 707	0.342	Sequenced by Isogen Bioscience BV Maarsse, The Netherlands	—
rs2435362	IVS1+6000A>C	43 353 975	0.488	F: 5'-CACATTGCTCTGCCATCGTA-3' R: 5'-GGCATTCTGTCCACCAAGTA-3'	PstI
rs2565206	IVS1–126G>T	43 370 977	0.379	F: 5'-ATGACTTTCCGTAAAGTGC-3' R: 5'-GGAGTTTTTCATCTCTGTTC-3'	NlaIII
rs1800858	A45A, c135G>A	43 371 164	0.438	F: 5'-AGCCTTATTCTCACCATCCC-3' R: 5'-CAGTGCAGCGGCTGTGATA-3'	EagI
MRetint5	IVSV+930, CA repeat	43 378 204–43 378 245	0.345	F: 5'-CTGTAACCTCCACCCTTGG-3' R: 5'-GCTCTAGTGTGCTCTCCCA-3'	—
rs1800860	A432A c1296G>A	43 381 883	0.780	F: 5'-GCCATTACAGGCCGGTCCAGC-3' R: 5'-GAGGCCAGGCTCCAGAAGC-3'	BsmI
rs1800862	S836S, c2508C>T	43 390 290	0.27	F: 5'-GAAGACCCAAGCTGCCTGAC-3' R: 5'-GGGCTGGGTGCAGAGCCATATG-3'	AluI
rs2075912	IVS19+47C>T	43 397 413	0.117	F: 5'-TGGAGTGACCGCCATCTCT-3' R: 5'-AAGCATCACAGAGAGGAAGG-3'	MnII
s-TCL2	+38 kb downstream <i>RET</i> , CA repeat	43 439 311–43 439 352	0.309	F: 5'-CCAGACTCTCAAAGACCAGG-3' R: 5'-CATAATACTGGCCTATAG-3'	—

Heterozygosity of the markers is calculated in the control genotypes, which are the pseudogenotypes comprised of the alleles not transmitted from the parents to their affected child.

Linkage disequilibrium (LD) was evaluated in order to investigate the identity-by-descent status of similar haplotypes. If LD is strong, the probability that two haplotypes are not coincidental but identical by descent is high. The strength of LD was measured separately for case and the control chromosomes using a randomization test, which permutes alleles over the haplotypes, determining the significance of  $D'$ .

For single-locus allelic association analyses, the frequencies of alleles and genotypes, respectively, were compared between patients and controls using a  $\chi^2$ -test. When this test showed a significant result, a  $Z$ -test was performed to determine as to which allele was responsible for the difference. This test assumes that patients and controls are independent samples from the population and that the numbers of specific allele genotypes follow a binomial distribution that can be approximated by a normal distribution. Under the assumption of no assortative mating in the population, transmitted and nontransmitted haplotypes meet these criteria.

Transmission distortion of each allele versus all other alleles together is tested and combined in a multiallelic transmission/distortion test.<sup>29</sup> This test evaluates whether, among heterozygous parents, one or more alleles are transmitted to their affected child more often than the 50% expectation.

The program TDTPHASE from the software package UNPHASED is used for the estimation of haplotype transmissions and determining the corresponding haplotype risks as a result of overtransmission.<sup>30</sup>

ORs and 95% confidence intervals (CIs) were calculated without correction for external variables such as age at diagnosis or type of the disease.<sup>31</sup>

All  $P$ -values and CIs were corrected for multiple testing for independent tests, that is, for 13 markers and two groups of patients (all patients and those with unknown or without a *RET* mutation), using a Bonferroni correction. Owing to LD between markers in the region, this correction is overstated and the results are, therefore, conservative.

## Results

### HWE analysis

None of the markers showed a significant deviation from HWE in the sample of nontransmitted pseudogenotypes. Hence, further analyses were performed on all 13 marker loci.

### Linkage disequilibrium

Strong LD is observed for all marker pairs both for the sample of case chromosomes and for the sample of control chromosomes (data not shown).

### Association analyses

Table 2 shows that six markers rs741763, SNP-5, SNP-1, rs2435362, rs2565206 and rs1800858, were strongly associated with the disease, revealing significant differences in frequencies of particular alleles in patients versus controls: for rs741763, the C allele is present in 81.5% of the patients chromosomes and 58.3% of the control chromosomes ( $P=2.0 \times 10^{-7}$ ) and SNP-5, the A allele is present in 66.7% of the patients chromosomes versus 22.5% of the control chromosomes ( $P=6.1 \times 10^{-24}$ ); for SNP-1, the C allele was present in 81.2% of the patients chromosomes versus 58.0% in the control chromosomes ( $P=1.1 \times 10^{-6}$ ); for rs2435362, the A allele was seen in 67.6% of the patients chromosomes versus 26.2% of the control chromosomes ( $P=1.4 \times 10^{-22}$ ); for rs2565206, allele T was found on 84.7% of the patients chromosomes versus 67.1% of the control chromosomes ( $P=1.2 \times 10^{-4}$ ); and for rs1800858, the A allele was observed on 67.1% of the patients chromosomes versus 23.1% of the control chromosomes ( $P=2.2 \times 10^{-24}$ ). Separate association analyses were performed on the groups of patients who were known to be *RET* mutation negative and those who were not screened for the mutations in order to exclude possible variations in allelic associations. Owing to the fewer number of cases, these analyses were, of course, less powerful than on the whole patients group. Nevertheless, significant associations were found for the same alleles at the same marker loci for both groups. Hence, we assumed that the group of patients not tested for the *RET* mutations was similar to the group of patients negative for a *RET* mutation and we could consider our patients' population as a whole in further analyses. Furthermore, we observed that a large proportion of our patients is homozygous for the associated alleles at the six marker loci that showed the highest association with HSCR, whereas these homozygous genotypes are hardly present in controls (Table 3a). This particularly is

the case for SNP-5 allele A homozygotes – 52.0% (53/102) among patients versus 6.4% (8/125) among controls, rs2435362 allele A homozygotes – 53.7% (58/108) versus 10.0% (13/130) and rs1800858 allele A homozygotes – 54.3% (57/105) versus 7.1% (9/126). The ORs for the homozygote genotypes at these three SNPs ranged from 16.95 to 26.50, whereas the ORs for the heterozygote genotypes ranged from 2.48 to 2.93 (Table 3a).

### Transmission/disequilibrium test (TDT)

For the TDT, 113 case-parents triads were considered. Table 4 shows the results of the multiallelic TDT, described in detail for the specific associated alleles. Large differences are observed in transmission of alleles at marker loci on the 5' region of the *RET* gene. For rs741763, allele C is transmitted 70 times versus 20 nontransmitted ( $P=9.3 \times 10^{-6}$ ); for SNP-5, allele A is transmitted 97 times versus 14 nontransmitted ( $P=2.3 \times 10^{-13}$ ); for SNP-1, allele C is transmitted 63 versus 17 nontransmitted ( $P=1.8 \times 10^{-5}$ ); for rs2435362, allele A is 100 transmitted versus 14 nontransmitted ( $P=5.3 \times 10^{-14}$ ); for rs2565206 allele T is transmitted 48 times versus 18 nontransmitted ( $P=0.015$ ); and for rs1800858 allele A is transmitted 102 times versus 18 nontransmitted ( $P=1.2 \times 10^{-12}$ ). Allele transmissions in the two subgroups (*RET*-negative versus not *RET* screened) did not significantly differ from each other. These findings confirm the association results that we observed.

### Haplotype analyses

Based on the results of the single-locus association analyses, we estimated frequencies of transmissions for the haplotypes consisting of the six strongest associated marker loci (Table 5). Out of 23 haplotypes found, the haplotype most frequent among patients was the one consisting of the alleles that gave the highest

**Table 2** Frequencies of associated alleles in patients, subgroups and controls with, in brackets, *P*-values corrected for multiple testing using a Bonferroni correction for two independent subgroups and all independent alleles with an expected number of at least five copies

Marker name	Allele	Controls	All patients	Without mutation	Not tested
D10S1100	194	59.6	72.0 (NS)	67.5 (NS)	77.4 (0.018)
M353	166	28.1	32.6 (NS)	27.0 (NS)	38.8 (NS)
rs741763	C	58.3	81.5 ( $2.0 \times 10^{-7}$ )	79.7 ( $3.4 \times 10^{-4}$ )	83.7 ( $3.1 \times 10^{-6}$ )
SNP-5	A	22.5	66.7 ( $6.1 \times 10^{-24}$ )	63.7 ( $1.7 \times 10^{-12}$ )	69.6 ( $1.6 \times 10^{-17}$ )
SNP-1	C	58.0	81.2 ( $1.1 \times 10^{-6}$ )	81.0 ( $2.1 \times 10^{-4}$ )	81.4 ( $1.1 \times 10^{-4}$ )
rs2435362	A	26.2	67.6 ( $1.4 \times 10^{-22}$ )	64.9 ( $1.5 \times 10^{-12}$ )	70.6 ( $3.1 \times 10^{-16}$ )
rs2565206	T	67.1	84.7 ( $1.2 \times 10^{-4}$ )	83.9 (0.0058)	85.7 (0.0020)
rs1800858	A	23.1	67.1 ( $2.2 \times 10^{-24}$ )	64.5 ( $1.6 \times 10^{-13}$ )	70.0 ( $3.1 \times 10^{-17}$ )
MREInt5	243	30.3	48.5 (0.0030)	46.4 (NS)	51.1 (0.023)
rs1800860	G	84.3	85.7 (NS)	90.0 (NS)	83.3 (NS)
rs1800862	C	93.6	99.1 (0.024)	98.4 (NS)	100.0 ( $5.5 \times 10^{-4}$ )
IVS19+47C>T	T	19.2	31.4 (NS)	30.9 (NS)	32.0 (NS)
s-TLC2	171	14.7	31.6 ( $6.0 \times 10^{-4}$ )	30.7 (0.045)	32.7 (0.027)

NS = not significant.

**Table 3** ORs and 95% CIs for the six most significantly associated loci (a) separately and (b) combined in a haplotype

Name	Genotype	Controls	Cases	OR	95% CI
<i>(a) ORs and CIs separately</i>					
rs741763	G,G	23	9	1	—
	C,G	68	23	0.86	0.21–3.60
	C,C	45	79	4.49	1.17–17.25
Totals		136	111		
SNP-5	G,G	76	19	1	—
	A,G	41	30	2.93	0.99–8.68
	A,A	8	53	26.50	6.42–109.3
Totals		125	102		
SNP-1	A,A	22	10	1	—
	C,A	62	18	0.64	0.15–2.70
	C,C	42	73	3.82	1.02–14.37
Totals		126	101		
rs2435362	C,C	76	20	1	—
	A,C	41	30	2.78	0.95–8.15
	A,A	13	58	16.95	4.97–57.85
Totals		130	108		
rs2565206	G,G	15	3	1	—
	G,T	57	28	2.46	0.31–19.73
	T,T	62	80	6.45	0.85–48.93
Totals		134	111		
rs1800858	G,G	77	21	1	—
	G,A	40	27	2.48	0.84–7.32
	A,A	9	57	23.22	6.04–89.25
Totals		126	105		
<i>(b) ORs and CIs combined in a haplotype</i>					
Six-loci haplotype	Noncarriers	80	28	1	—
	Heterozygotes	22	17	2.21	0.66–7.39
	Homozygotes	6	46	21.90	4.86–98.69

The confidence intervals are corrected for multiple testing (13 loci × two subgroups).

**Table 4** Transmissions of associated alleles in the entire sample and subgroups with *P*-values corrected for multiple testing for two independent (sub)groups and all independent alleles with a frequency of at least five copies using a Bonferroni correction

Marker name	Allele	All patients	Without mutation	Not tested
D10S1100	194	60 tr, 35 ntr (NS)	39 tr, 20 ntr (NS)	21 tr, 15 ntr (NS)
M353	166	51 tr, 33 ntr (NS)	23 tr, 14 ntr (NS)	28 tr, 19 ntr (NS)
rs741763	C	70 tr, 20 ntr ( $9.3 \times 10^{-6}$ )	36 tr, 12 ntr (0.036)	34 tr, 8 ntr (0.0041)
SNP-5	A	97 tr, 14 ntr ( $2.3 \times 10^{-13}$ )	48 tr, 3 ntr ( $2.0 \times 10^{-8}$ )	49 tr, 11 ntr ( $6.3 \times 10^{-5}$ )
SNP-1	C	63 tr, 17 ntr ( $1.8 \times 10^{-5}$ )	31 tr, 7 ntr (0.0067)	32 tr, 10 ntr (0.046)
rs2435362	A	100 tr, 14 ntr ( $5.3 \times 10^{-14}$ )	50 tr, 4 ntr ( $2.6 \times 10^{-8}$ )	50 tr, 10 ntr ( $1.6 \times 10^{-5}$ )
rs2565206	T	48 tr, 18 ntr (0.015)	30 tr, 7 ntr (0.011)	18 tr, 11 ntr (NS)
rs1800858	A	102 tr, 18 ntr ( $1.2 \times 10^{-12}$ )	50 tr, 5 ntr ( $8.8 \times 10^{-8}$ )	52 tr, 13 ntr ( $8.9 \times 10^{-5}$ )
MREInt5	243	64 tr, 30 ntr (0.030)	35 tr, 15 ntr (NS)	29 tr, 15 ntr (NS)
rs1800860	G	18 tr, 19 ntr (NS)	4 tr, 5 ntr (NS)	14 tr, 14 ntr (NS)
rs1800862	C	13 tr, 2 ntr (NS)	10 tr, 2 ntr (NS)	3 tr, 0 ntr (NS)
IVS19+47C>T	T	51 tr, 22 ntr (0.046)	24 tr, 9 ntr (NS)	27 tr, 13 ntr (NS)
s-TLC2	171	60 tr, 22 ntr (0.0018)	33 tr, 10 ntr (0.030)	27 tr, 12 ntr (0.030)

NS = not significant; tr = transmitted; ntr = nontransmitted.

single-locus associations (rs741763 – allele C; SNP-5 – allele A; SNP-1 – allele C; rs2435362 – allele A; rs2565206 – allele T; rs1800858 – allele A): it was transmitted in 55.6% and in 16.2% it was not. The corresponding risk of

overtransmission of this haplotype was estimated to be 10.31. ORs (Table 3b) for the haplotype composed of the six highest associated alleles reveal that the risk of developing HSCR is highly increased, in particular, for

**Table 5** Frequencies and relative risks (RR) of transmission of haplotypes consisting of the six most associated loci as estimated using the software package UNPHASED<sup>22</sup>

	Haplotype	Freq-T	Freq-NT	OR
1 <sup>a</sup>	CGCAGG	0.004484	0.01345	1
2	CGCATG	0.003612	0.005357	2.023
3	CGCCGG	0.1611	0.2869	1.685
4	CGCCGA	0.001933	0.002867	2.023
5	CGCCTG	0.008969	0.0449	0.5994
6	CGCCTA	0.001806	0.002678	2.023
7	CGACGG	0.003612	0.005357	2.023
8	CGACTG	0.01345	0.02242	1.8
9	CACAGG	0.001806	0.002678	2.023
10	CACAGA	0.001806	0.002679	2.023
11	CACATG	0.01377	0.01763	2.342
12	CACATA	<i>0.5557</i>	<i>0.1617</i>	<i>10.31</i>
13	CACCGG	0.001806	0.002678	2.023
14	CACCTG	0.001806	0.002678	2.023
15	CACCTA	0.02242	0	-3
16	GGCCGG	0.001807	0.00268	2.023
17	GGCCTG	0.001806	0.002678	2.023
18	GGAATG	0.003612	0.005357	2.023
19	GGAATA	0.001806	0.002678	2.023
20	GGACGG	0.008969	0.02696	0.9981
21	GGACTG	0.1704	0.359	1.424
22	GGACTA	0.004487	0.01314	1.024
23	GACATA	0.008969	0.01345	2

The risk haplotype is depicted in italics.

<sup>a</sup>Haplotype 1 is the haplotype consisting of all wild-type alleles.

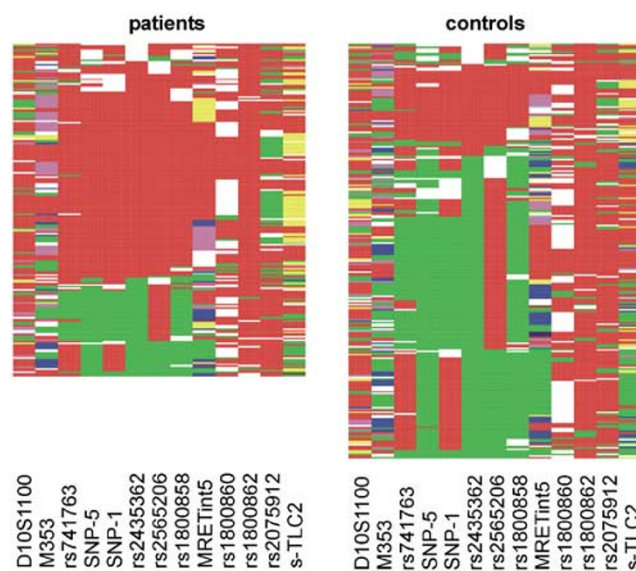
Freq-T – frequency of transmitted haplotypes; freq-NT – frequency of nontransmitted haplotypes; OR – odds ratios; LR=91.84, df=10,  $P=2.3 \times 10^{-15}$ .

homozygotes for the associated haplotype (OR=21.90; 95% CI 4.86–98.69) as compared to the risk of carriers of one associated haplotype (OR=2.21; 95% CI 0.66–7.39).

Figure 2 is a graphical representation of the 13-loci haplotypes after clustering for maximal similarity around rs2435362 in order to visualize the observed association. The most striking difference is the much higher frequency of long haplotypes in patients than in controls, in particular, the haplotypes that appear as a large red block in the patients (left grid) as compared to the smaller red block in the controls (right grid).

## Discussion

In 70–90% of the sporadic HSCR patients, no mutations in any of the nine known HSCR susceptibility genes are identified. This raises the question whether, despite the fact that no pathogenic mutations can be detected in the majority of sporadic patients, the known HSCR susceptibility genes are involved in the development of the disease, or whether yet unknown genes are responsible. A possible involvement of *RET*, the major gene in HSCR, might be concluded from several studies. Bolk *et al*<sup>18</sup> showed that 11 of 12 families investigated were linked to *RET*. In six of them, however, no clear *RET* mutation could be identified in the coding and flanking intronic



**Figure 2** Graphical representation of the 13-loci haplotypes after clustering for maximal similarity around rs2435362. Haplotypes of patients are shown in the left half of the figure and the family-based controls in the right half. Each horizontal colored line represents one haplotype. On the x-axis, marker loci are indicated. Marker alleles are mapped to colors, white being a missing or phase-unknown allele. Different alleles at a locus have different colors. The choice of colors is such that only minimal changes in color appear from one marker locus to the next in haplotypes that are shared over a long distance. In the direction of the y-axis, haplotypes are clustered for the largest sharing around rs2435362. This algorithm causes that each block in one color represents a preserved haplotype. Clustering around marker rs2435362 was chosen because this locus was the fourth out of the six most associated markers. The same clustering algorithm was applied to patient and control haplotypes. Only patient and control haplotypes with at least five phase-known alleles are displayed. The haplotype causing the difference between patients and controls is the haplotype depicted as a red block in the middle of each grid.

sequences, as checked by single-stranded conformation polymorphism (SSCP), DGGE and sequence analysis. It suggests an involvement of *RET* although not through a clear *RET* mutation. Several association and haplotype studies also support this hypothesis.<sup>26,32,33</sup> Conserved haplotypes could be constructed using the alleles identified for different markers in and around the *RET* gene.<sup>24,34–36</sup> Carrasquillo *et al*<sup>34</sup> used the highest number of SNPs in and around the *RET* locus. They found that patients shared common haplotypes with markers in both the proximal and distal part of the gene without having a clear pathogenic mutation. This study was carried out on a Mennonite kindred, an isolated population in which HSCR occurs in one in 500 live births. Therefore, it is likely that their findings apply to this population only and that the

associated haplotype will be in LD with mutations that are most likely different from those that are present in other populations. In the recent study, Fitze *et al*<sup>36</sup> showed that the haplotype ACA comprising alleles from SNPs -5, -1 and rs1800858 (c135G>A) is overrepresented in their patients population (66.9% of 80 cases). Moreover, of 58 HSCR patients, all nonmutation carriers, 62.1% proved homozygous for a two-locus (SNP-5, SNP-1) haplotype. Sancandi *et al*<sup>24</sup> performed a haplotype analysis on a smaller group of HSCR patients (46 patients and 50 population-matched control individuals). They genotyped two SNPs in the promoter region lying close together (SNP-1 and SNP-5) and SNPs in exon 2 (A45A) and exon 13 (L769L). As relatives of the patients were not available for screening, real haplotypes could not be reconstructed. They estimated frequencies of haplotypes consisting of the markers in the promoter region and those in exons 2 and 13. Two of the haplotypes, differing only in the allele from last marker from exon 13, appeared to be much more frequent among HSCR patients (62% of the patients chromosomes *versus* 22% of control chromosomes). Assuming that these SNPs are not causative variants themselves, Sancandi *et al*<sup>24</sup> localized the unknown variant upstream of exon 2, although based on their results the region between exons 2 and 13 cannot be excluded. Borrego *et al*<sup>35</sup> described a haplotype analysis on 103 HSCR patients and their parents. They genotyped three SNPs at the end of intron one (IVS1-1463T>C; IVS1-1370C>T; IVS-126G>T) and seven in exons 2, 3, 7, 11, 13, 14 and 15. Significant haplotype frequency differences were found for the three SNPs in intron 1. From the whole pool of genotyped SNPs, they reconstructed one haplotype that was transmitted nine times and never nontransmitted. However, when they reconstructed a haplotype consisting solely of the three markers located at the end of intron 1, a haplotype spanning a region of 1.2kb was observed that was far more common in HSCR patients (59.2%) than in controls (18.5%). Furthermore, they suspected that, based on extrapolation of the strength of LD at the observed SNPs, LD would become stronger upstream to the SNPs from intron 1 and that the susceptibility variant should lie upstream but probably still in the intron 1.

All these studies suggest that the *RET* gene is involved in sporadic HSCR and that an ancestral mutation is likely to be located in the 5' region of the gene. We typed 13 markers, including four microsatellites and nine SNPs within and flanking the gene, in order to better define the region in which this ancestral mutation might be located.

In correspondence with previous studies, significant associations with HSCR were found for SNPs in the 5' region of the *RET* gene. These were even stronger than those published in previous studies. Six successive SNPs, namely: rs741763, SNP-5, SNP-1, rs2435362, rs2565206 and rs1800858 (Tables 1 and 2), were found to be strongly

associated with HSCR in our population with the highest frequency increase of 23.1% among controls to 67.1% among patients. Associated alleles at marker loci genotyped also in other studies<sup>24,35,36</sup> were found to be the same (SNP-5, allele A; SNP-1, allele C; rs2565206, allele T; rs1800858, allele A). All of the alleles of genotyped markers also showed a significant transmission distortion (Table 3). Furthermore, a large proportion of our patients proved to be homozygous for the alleles at these six marker loci, whereas these homozygous genotypes had a very low frequency in controls. We were able to reconstruct haplotypes consisting of these six SNPs and observed one that was very commonly transmitted to the patients (55.6%) and significantly under-represented among controls (16.2%). Homozygosity for this ancestral haplotype was observed in 50.5% of our patients *versus* 5.5% of our controls. It appears that homozygosity for this, most likely European ancestral haplotype, in sporadic HSCR cases gave a much higher increased risk of developing HSCR than heterozygosity, that is, the disease appears dosage-dependent with respect to a mutation in this ancestral haplotype.

The six strongest associated SNPs are located in the interval spanning from the promoter region and the beginning of the exon 2 (27 kb). The region of association may not extend far 5' to the transcriptional start site of the *RET* gene, since the two microsatellites markers (*M353*, *D10S1100*) located upstream of the *RET* locus do not segregate significantly with HSCR. Nevertheless, we cannot exclude that the unknown mutation is lying upstream of the promoter region as microsatellite *M353* is located 80 kb upstream of *RET*. Alleles of marker *s-TCL2*, which lies downstream of the gene were associated when consider the patients as a whole. However, when separating the analyses on both group of patients (screened for *RET* and not tested for *RET*), no association between HSCR and an allele of the marker, as well as TDT was found.

We found very strong associations for three SNPs and, in particular, large differences in numbers of homozygotes between patients and controls giving ORs, ranging from 16.95 to 26.50 (see Table 3). This might suggest that one or more of these SNPs are themselves causally involved in HSCR. For some of the SNPs included, this has indeed been proposed. For instance the polymorphism in exon 2 might interfere with correct splicing,<sup>35</sup> but experimental proof has not been presented. Polymorphisms in the introns have been suggested to disrupt the binding sites of regulatory proteins and thereby change gene transcription.<sup>35</sup> Calculations indicated four possible binding sites for regulatory proteins. Again, it has not been proven experimentally. Fitze *et al*<sup>36</sup> performed functional study on basal *RET* promoter sequences, carrying different haplotypes at loci -5 and -1. Fitze's group has found that expression of a reporter gene in NMB and Vi-856 cell lines is reduced under the *RET* promoter carrying the 'AC' haplotype. The expression was even lower with the 'AA'

haplotype; however, this haplotype was not present in patients. They came to the conclusions that -5A variant can alter *RET* promoter activity and modulate the HSCR phenotype. It should, however, be noted that expression of a gene in different cell lines might give contradicting results. Preliminary results of the Ceccherini group indeed show such differences (Ceccherini *et al*, paper in preparation). Sequence analysis of the promoter region up to intron 1 should reveal all possible SNPs that, when typed on the entire sample, might provide sufficient statistical information to identify the true causal mutation(s).

In conclusion, we observed a very strong association for several SNPs in the promoter and in the 5' region of the *RET* gene. An increased risk of HSCR was most evident for patients homozygous for the associated alleles. The haplotype consisting of these markers showed similar results, indicating that a strong founder effect is present in our population. The alleles and consequently the haplotype found in our study is similar to that found by others who analyzed other European patient populations. The ancestral haplotype might therefore be very old making it likely that most of the European HSCR patients share the same disease-associated variant(s). We, therefore, expect that these results will allow us to identify eventually the mutation, which is obviously playing a major role in the susceptibility to HSCR.

#### Acknowledgements

This work was supported by NWO Grants 901-04-210 and 901-04-225.

#### References

- Holschneider A: Hirschsprung's congenital megacolon. The concept of physiopathology and therapy. *Med Welt* 1982; **33**: 210–213.
- Badner J, Sieber W, Garver K, Chakravarti A: *Am J Hum Genet* 1990; **46**: 568–580.
- Amiel J, Lyonnet S: Hirschsprung disease, associated syndromes, and genetics: a review. *J Med Genet* 2001; **38**: 729–739.
- Ederly P, Lyonnet S, Mulligan L *et al*: Mutations of the *RET* proto-oncogene in Hirschsprung's disease. *Nature* 1994; **367**: 378–380.
- Romeo G, Ronchetto P, Luo Y *et al*: Point mutations affecting the tyrosine kinase domain of the *RET* proto-oncogene in Hirschsprung's disease. *Nature* 1994; **367**: 377–378.
- Angrist M, Bolk S, Halushka M, Lapchak PA, Chakravarti A: Germline mutations in glial cell line-derived neurotrophic factor (GDNF) and *RET* in a Hirschsprung disease patient. *Nat Genet* 1996; **14**: 341–344.
- Salomon R, Attie T, Pelet A *et al*: Germline mutations of the *RET* ligand GDNF are not sufficient to cause Hirschsprung disease. *Nat Genet* 1996; **14**: 345–347.
- Doray B, Salomon R, Amiel J *et al*: Mutation of the *RET* ligand, neurturin, supports multigenic inheritance in Hirschsprung disease. *Hum Mol Genet* 1998; **7**: 1449–1452.
- Puffenberger E, Hosoda K, Washington S *et al*: A missense mutation of the endothelin-B receptor gene in multigenic Hirschsprung's disease. *Cell* 1994; **79**: 1257–1266.
- Ederly P, Attie T, Amiel J *et al*: Mutation of the endothelin-3 gene in the Waardenburg–Hirschsprung disease (Shah–Waardenburg syndrome). *Nat Genet* 1996; **12**: 442–444.
- Hofstra R, Osinga J, Tan-Sindhunata G *et al*: A homozygous mutation in the endothelin-3 gene associated with a combined Waardenburg type 2 and Hirschsprung phenotype (Shah–Waardenburg syndrome). *Nat Genet* 1996; **12**: 445–447.
- Hofstra R, Valdenaire O, Arch E *et al*: A loss-of-function mutation in the endothelin-converting enzyme 1 (ECE-1) associated with Hirschsprung disease, cardiac defects, and autonomic dysfunction. *Am J Hum Genet* 1999; **64**: 304–308.
- Pingault V, Bondurand N, Kuhlbrodt K *et al*: SOX10 mutations in patients with Waardenburg–Hirschsprung disease. *Nat Genet* 1998; **18**: 171–173.
- Paratore C, Eichenberger C, Suter U, Sommer L: Sox10 haploinsufficiency affects maintenance of progenitor cells in a mouse model of Hirschsprung disease. *Hum Mol Genet* 2002; **11**: 3075–3085.
- Cacheux V, Dastot-Le Moal F, Kaariainen H *et al*: Loss-of-function mutations in SIP1 Smad interacting protein 1 result in a syndromic Hirschsprung disease. *Hum Mol Genet* 2001; **10**: 1503–1510.
- Wakamatsu N, Yamada Y, Yamada K *et al*: Mutations in SIP1, encoding Smad interacting protein-1, cause a form of Hirschsprung disease. *Nat Genet* 2001; **27**: 369–370.
- Amiel J, Laudier B, Attie-Bitach T *et al*: Polyalanine expansion and frameshift mutations of the paired-like homeobox gene PHOX2B in congenital central hypoventilation syndrome. *Nat Genet* 2003; **33**: 459–461.
- Bolk S, Pelet A, Hofstra RM *et al*: A human model for multigenic inheritance: phenotypic expression in Hirschsprung disease requires both the *RET* gene and a new 9q31 locus. *Proc Natl Acad Sci USA* 2000; **97**: 268–273.
- Gabriel SB, Salomon R, Pelet A *et al*: Segregation at three loci explains familial and population risk in Hirschsprung disease. *Nat Genet* 2002; **31**: 89–93.
- Hofstra RMW, Wu Y, Stulp RP *et al*: *RET* and *GDNF* gene scanning in Hirschsprung patients using two dual denaturing gel systems. *Hum Mut* 2000; **15**: 418–429.
- Chakravarti A: Endothelin receptor-mediated signaling in Hirschsprung disease. *Hum Mol Genet* 1996; **5**: 303–307.
- Falk CT, Rubinstein P: Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. *Ann Hum Genet* 1987; **51**: 227–233.
- Terwilliger JD, Ott J: A haplotype-based 'haplotype relative risk' approach to detect allelic associations. *Hum Hered* 1992; **42**: 337–346.
- Sancandi M, Griseri P, Pesce B *et al*: Single nucleotide polymorphic alleles in the 5' region of the *RET* proto-oncogene define a risk haplotype in Hirschsprung disease. *J Med Genet* 2003; **40**: 714–718.
- Ceccherini I, Hofstra RM, Luo Y *et al*: DNA polymorphisms and conditions for SSCP analysis of the 20 exons of the *ret* proto-oncogene. *Oncogene* 1994; **9**: 3025–3029.
- Griseri P, Pesce B, Patrone G *et al*: A rare haplotype of the *RET* proto-oncogene is a risk modifying allele in Hirschsprung disease. *Am J Hum Genet* 2002; **71**: 969–974.
- Carlson KM: Three dinucleotide repeat polymorphisms closely linked to the *RET* protooncogene D10S1098, D10S1099 and D10S1100. *Hum Mol Genet* 1994; **3**: 1207.
- Pasini B, Hofstra RM, Yin L *et al*: The physical map of the human *RET* proto-oncogene. *Oncogene* 1995; **11**: 1737–1743.
- Spielman RS, Ewens WJ: The TDT and other family-based tests for linkage disequilibrium and association. *Am J Hum Genet* 1996; **59**: 983–989.
- Breslow N, Day N: *The analysis of case-control studies Statistical methods in cancer research I*, IARC Sci. Publ. No 32. Lyon, France: IARC; 1980, p. 134.
- Woolf B: On estimating the relationship between blood group and disease. *Ann Hum Genet* 1995; **19**: 251–253.
- Fitze G, Schreiber M, Kuhlisch E, Schackert H: Association of *RET* protooncogene codon 45 polymorphism with Hirschsprung disease. *Am J Hum Genet* 1999; **65**: 1469–1473.



- 33 Borrego S, Ruiz A, Saez M *et al*: *RET* genotypes comprising specific haplotypes of polymorphic variants predispose to isolated Hirschsprung disease. *J Med Genet* 2000; **37**: 572–578.
- 34 Carrasquillo M, McCallion A, Puffenberg E, Kashuk C, Nouri N, Chakravarti A: Genome-wide association study and mouse model identify interaction between *RET* and EDNRB pathways in Hirschsprung disease. *Nat Genet* 2002; **32**: 237–244.
- 35 Borrego S, Wright F, Fernandez R *et al*: A founding locus within the *RET* proto-oncogene may account for a large proportion of apparently sporadic Hirschsprung disease and a subset of cases of sporadic medullary thyroid carcinoma. *Am J Hum Genet* 2003; **72**: 88–100.
- 36 Fitze G, Appelt H, Konig I *et al*: Functional haplotypes of the *RET* proto-oncogene promoter are associated with Hirschsprung disease (HSCR). *Hum Mol Genet* 2003; **12**: 3207–3214.