

## Report

### Identifying Candidate Hirschsprung Disease–Associated *RET* Variants

Grzegorz M. Burzynski,<sup>1</sup> Ilja M. Nolte,<sup>2</sup> Agnes Bronda,<sup>1</sup> Krista K. Bos,<sup>1</sup> Jan Osinga,<sup>1</sup> Ivan Plaza Menacho,<sup>1</sup> Bas Twigt,<sup>1</sup> Saskia Maas,<sup>4</sup> Alice S. Brooks,<sup>5</sup> Joke B. G. M. Verheij,<sup>3</sup> Charles H. C. M. Buys,<sup>1,3</sup> and Robert M. W. Hofstra<sup>1</sup>

Departments of <sup>1</sup>Medical Genetics and <sup>2</sup>Medical Biology, University of Groningen, and <sup>3</sup>Department of Clinical Genetics, Academic Hospital Groningen, Groningen, The Netherlands; <sup>4</sup>Department of Clinical Genetics, Academic Medical Center, Amsterdam; and <sup>5</sup>Department of Clinical Genetics, Erasmus Medical Center, Rotterdam

Patients with sporadic Hirschsprung disease (HSCR) show increased allele sharing at markers in the 5' region of the *RET* locus, indicating the presence of a common ancestral *RET* mutation. In a previous study, we found a haplotype of six SNPs that was transmitted to 55.6% of our patients, whereas it was present in only 16.2% of the controls we used. Among the patients with that haplotype, 90.8% had it on both chromosomes, which led to a much higher risk of developing HSCR than when the haplotype occurred heterozygously. To more precisely define the HSCR-associated region and to identify candidate disease-associated variant(s), we sequenced the shared common haplotype region from 10 kb upstream of the *RET* gene through intron 1 and exon 2 (in total, 33 kb) in a patient homozygous for the common risk haplotype and in a control individual homozygous for the most common nonrisk haplotype. A comparison of these sequences revealed 86 sequence differences. Of these 86 variations, 8 proved to be in regions highly conserved among different vertebrates and within putative transcription factor binding sites. We therefore considered these as candidate disease-associated variants. Subsequent genotyping of these eight variants revealed a strong disease association for six of the eight markers. These six markers also showed the largest distortions in allele transmission. Interspecies comparison showed that only one of the six variations was located in a region also conserved in a nonmammalian species, making it the most likely candidate HSCR-associated variant.

Hirschsprung disease (HSCR [MIM 142623]) is caused by the absence of intrinsic ganglion cells in the myenteric and submucosal plexuses of the gastrointestinal tract, resulting in severe intestinal obstructions. HSCR is a heterogeneous disorder: a number of genes have been shown to play a role in the disease etiology (Holschneider 1982; Badner et al. 1990; Amiel and Lyonnet 2001). The *RET* gene (MIM 164761) is a main factor in HSCR development. *RET* encodes a transmembrane tyrosine kinase receptor, a protein responsible for triggering a number of downstream signaling pathways involved in crucial processes, such as neural crest cell differentiation, migration, and proliferation (Manie et al. 2001). Mutations in *RET* have been found in up to 50% of familial cases of HSCR.

For the more common sporadic form of HSCR, *RET* coding mutations have been found in not more than 20% of patients (Hofstra et al. 2000; Amiel and Lyonnet 2001). Several studies have shown, however, that the *RET* locus is linked to the disease in almost all familial cases, regardless of their mutation status (Bolk et al. 2000; Gabriel et al. 2002), and is also associated with HSCR in a large proportion of the patients with sporadic HSCR who do not have *RET* coding mutations. Furthermore, recently published articles revealed that similar haplotypes are found in the 5' region of the *RET* locus in patients from several European populations with HSCR (Borrego et al. 2003; Fitze et al. 2003; Sancandi et al. 2003; Burzynski et al. 2004), indicating the segregation of the same ancestral mutation(s) in the whole European population of patients with HSCR. The identified region of association comprises intron 1 of the *RET* gene and, depending on the markers genotyped by the different groups, different lengths of upstream genomic sequence (Borrego et al. 2003; Fitze et al. 2003; Sancandi et al. 2003; Burzynski et al. 2004). In a previous study, we found that a hap-

Received December 20, 2004; accepted for publication February 9, 2005; electronically published March 9, 2005.

Address for correspondence and reprints: Dr. R. M. W. Hofstra, Department of Medical Genetics, Antonius Deusinglaan 4, 9713 AW Groningen, The Netherlands. E-mail: r.m.w.hofstra@medgen.azg.nl

© 2005 by The American Society of Human Genetics. All rights reserved.  
0002-9297/2005/7605-0012\$15.00

**Table 1**

**Sequence Variations in the Common Risk Haplotype, as Compared with the Most Common Nonrisk Haplotype**

Position and Type of Variant	NCBI dbSNP Accession No.
-9557G→C	...
-9341C→T	rs3026719
-9256C→G	rs2505989
-7359A→C	rs2505990
-6347G→A	rs2505991
-6093T→C	...
-5669G→C	rs2505992
-4636G→C	rs732610
-4361C→G <sup>a,b</sup>	rs741763
-4201G→A	rs2082107
-4146C→G	...
-3820C→T	rs2505994
-3054G→A	rs2505995
-2897T→C	rs2505996
-2352T→C	rs2505997
-1782G→A	rs2505998
-1260T→C	rs2505999
-719C→T	rs2435366
-200G→A (SNP-5) <sup>a,c</sup>	rs10900296
-196A→C (SNP-1) <sup>a,c</sup>	rs10900297
IVS1+620C→A	rs3123712
IVS1+779T→C	rs2506010
IVS1+1813C→T	rs2435365
IVS1+2157C→T	rs2506011
IVS1+2820T→C	rs1864411
IVS1+2846G→T	rs1864410
IVS1+3104C→T	rs1864408
IVS1+3442A→G	rs2506013
IVS1+3460C→T	rs2435364
IVS1+3470G→A	...
IVS1+3788C→G	rs2506014
IVS1+3991del12	...
IVS1+4247C→G	rs2506015
IVS1+5094T→A	rs2506016
IVS1+5274insCT	...
IVS1+6000C→A <sup>a,b</sup>	rs2435362
IVS1+6136G→T	rs2506019
IVS1+6174G→A	rs2435361
IVS1+6294T→C	rs2506020
IVS1+6383G→C	...
IVS1+6384G→C	...
IVS1+6411G→A	...
IVS1+6751T→C	rs1897001
IVS1+6757A→G	rs1897000
IVS1+6813C→T	rs1896999
IVS1+7236G→A	rs2435359
IVS1+7436A→G	rs2435358
IVS1+7445A→C	rs2506008
IVS1+7994C→G	rs2506007
IVS1+9494C→A	rs2506004
IVS1+10371G→A	rs2435356
IVS1+10813A→G	...
IVS1+11369T→C	rs2506021
IVS1+11476C→T	rs2435342
IVS1-11490T→C	rs2506022
IVS1-11084insG	...
IVS1-10285G→T	rs2435343

(continued)

**Table 1 (continued)**

Position and Type of Variant	NCBI dbSNP Accession No.
IVS1-10032C→T	rs2435344
IVS1-10017C→T	rs2435345
IVS1-9651T→G	...
IVS1-8047C→T	rs2506023
IVS1-7593G→A	rs752975
IVS1-7525C→T	rs752977
IVS1-7477T→A	rs752978
IVS1-7298G→A	...
IVS1-6988G→A	rs2506024
IVS1-6567C→G	...
IVS1-6566G→C	...
IVS1-6393A→G	...
IVS1-6257G→C	rs3128726
IVS1-5791C→T	rs2505541
IVS1-5666C→T	rs2505540
IVS1-5127C→G	...
IVS1-5107T→C	...
IVS1-5038C→T	rs2505539
IVS1-5005A→G	rs2435346
IVS1-4925G→A	...
IVS1-4503G→A	rs2505538
IVS1-4073C→A	...
IVS1-3748C→T	rs2505537
IVS1-3702T→A	rs2505536
IVS1-2863G→A	rs2505535
IVS1-891A→T	...
IVS1-888G→C <sup>a</sup>	...
IVS1-881C→T	...
c135G→A (exon 2) <sup>d</sup>	rs1800858

NOTE.—Positions of variations localized upstream the gene refer to *RET* ATG.

<sup>a</sup> SNPs from our previously characterized six-locus ancestral haplotype.

<sup>b</sup> Previously reported in Burzynski et al. 2004.

<sup>c</sup> Previously reported in Sancandi et al. 2003.

<sup>d</sup> Previously reported in Ceccherini et al. 1994.

lotype consisting of six SNPs, spanning 27 kb (from 4 kb upstream of the gene through the whole intron 1), was transmitted from the parents in 55.6% of our patients with sporadic HSCR and was not transmitted in only 16.2% (Burzynski et al. 2004). Moreover, for individuals homozygous for the risk haplotype, the odds ratio to develop HSCR was 21.9, whereas, for the heterozygotes, it was 2.2, suggesting dose dependence of the unknown variants (Burzynski et al. 2004). Three groups tried to correlate specific haplotypes of two SNPs (which were also part of our risk haplotype) located in the basal *RET* promoter (−5 bp and −1 bp from the transcription start site) with reduced *RET* expression by performing functional assays in human cell lines (Fitze et al. 2003; Garcia-Barcelo et al. 2005; Griseri et al. 2005). However, results were not consistent, so these SNPs cannot unambiguously be considered as being pathogenic.

In this study, we attempt to further investigate the re-

gion of interest by sequencing 33 kb—from 10 kb upstream of the gene through 23 kb of intron 1 and exon 2—in a patient homozygous for the common risk haplotype and in a healthy control individual homozygous for the most common nonrisk haplotype. To sequence this 33-kb region (NCBI genomic position: 42846516 bp–42880012 bp [chromosome 10]), 49 primer sets were designed. The average size of the amplification product was 750 bp, and, between amplicons, there was an overlap of ~50 bp. Primer sequences can be found on the Hirschsprung Disease Web site of the University of Groningen Department of Genetics. Samples were sequenced by BaseClear. By restricting the comparison (for economic reasons) to a single patient and a single control individual, one runs the risk of missing SNPs in the comparison. However, since the haplotypes are quite common, one might expect that a SNP responsible for the difference will have a very high probability to be found.

Sequence comparison between the two sequences (the sequence of the common HSCR haplotype and that of the most common control haplotype) revealed 86 different sequence variants (table 1). In intron 1, a deletion of 12 bp was found (IVS+3991), as were two insertions (IVS1+5274insCT and IVS1–11084insG) and two 2-base substitutions (IVS1+6383–84GG→CC and IVS1–6567CG→GC). The remaining 81 variations are all SNPs. All the variations were found in a homozygous state in the patient's DNA as well as in the control's DNA. Table 1 includes, as indicated, five of the six SNPs that form our previously identified haplotype. The sixth marker, IVS1–126G→T, is not included, since the patient's DNA and the control's DNA had the same allele at this locus. Twenty variants were located in the region upstream of the gene, 65 are located in intron 1, and 1 is located in exon 2 (table 1). The sequence-variation frequency in the upstream region is 1/536 bp, and, for the intronic region, it is 1/353 bp. This might indicate a tighter selection mechanism on the upstream sequence, as a consequence of the possible presence of more-important regulatory regions. On the other hand, the common risk haplotype observed in our patients also spans the whole of intron 1, supporting the hypothesis that this intronic region might also be involved in the regulation of *RET* expression.

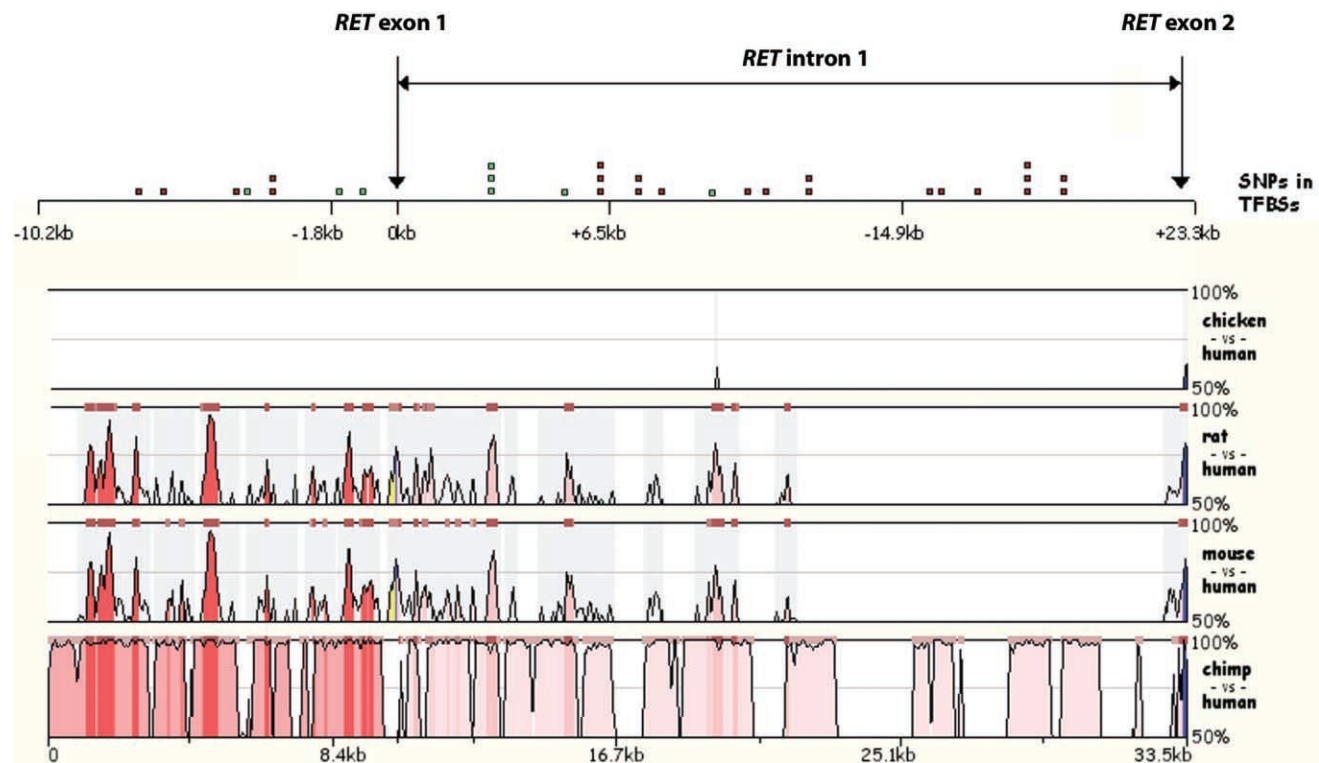
Since our aim was to delimit the HSCR-associated region by further genotyping, we selected from these 86 sequence variations the most likely candidate disease-associated variants. Variants were first selected on the basis of their location in regions that are conserved among vertebrate species. To identify evolutionarily conserved domains within this 33-kb region, we compared the human genomic sequence with the genomic DNA sequences of chimpanzee (*Pan troglodytes*), rat (*Rattus norvegicus*), mouse (*Mus musculus*), and chicken (*Gallus gallus*) (downloaded from the UCSC Genome Browser

database). Sequence comparisons between these species were performed using the freely available tool MULAN (Multiple sequence Local Alignment and conservation visualization tool) and the software package Clone Manager Professional Suite (Needleman-Wunsch method). Figure 1 shows the level of evolutionary conservation between the chimpanzee, mouse, rat, and chicken, with the human sequence as a reference. Rat and mouse genomes show a comparable level of conservation with the human sequence. There are 11 regions with a similarity >70%. The distal part of the human intron 1 shows no homology with the mouse and rat counterparts, except for the short region just downstream of exon 2 and exon 2 itself. Furthermore, it has to be noticed that the rodents' intron 1 is only 12 kb long—roughly half as long as the human intron 1. Moreover, in the mouse genome, homology to the human intron 1 is found in a region 30 kb upstream of the gene and also in intron 18 of the mouse *Ret* gene (data not shown). This could mean that this region has been translocated either in rodents or in primates, since the chimpanzee's intron 1 is highly similar to that of humans.

A region that is almost identical between rodents and primates (99.5%) is located between the positions –5.7 kb and –5.2 kb, with reference to *RET* ATG (fig. 1). Between the human and chicken sequences, there is hardly any significant similarity. This might imply that conserved transcriptional regulation of *RET* is restricted to mammals. However, similarity was found in the region between positions +9.3 kb and +9.7 kb (corresponding to the human sequence), where strong homology with chimpanzee, rat, and mouse sequences is observed as well (fig. 1). This is the only region conserved among all of the species studied. The second region of moderate conservation in chicken is located in the distal part of the sequence at the beginning of exon 2.

The sequence of the chimpanzee is almost identical to the human sequence, except for a number of repeats, point mutations, and small deletions/insertions. A puzzling observation is that exon 1 is missing from the chimpanzee genome database. Instead, there is a repeated sequence in this position. This is probably a sequencing error, since the chimpanzee genome database is still not assembled well enough, and many gaps exist between contigs.

The occurrence of sequence homology between different species in noncoding regions (including promoter and intron regions) might suggest that these conserved regions contain binding sites for relevant transcription factors. We therefore checked the sequenced genomic region for the presence of specific consensus binding elements for mammalian transcription factors, and we determined whether the identified DNA variants were present in these putative transcription factor consensus binding elements. For this, we used the UCSC Genome Browser and the program GENEQUEST from the DNASTAR software package. We



**Figure 1** MULAN plots displaying the percentages of similarity of the 5' region of the *RET* locus in the chicken, rat, mouse, and chimpanzee genomic sequences, compared with the human sequence as a reference. The horizontal gray line indicates 70% identity. The top panel of the figure depicts the positions of the 31 SNPs that are located within putative transcription factor binding sites. Those indicated in green fall in one of the evolutionarily conserved regions. The upper scale refers to *RET* ATG (0 kb). The lower scale represents the length of the whole region. TFBSs = transcription factor binding sites.

aimed to find DNA elements that were lost in the patient's DNA that might result in diminished expression and might thereby explain the HSCR phenotype. We detected 31 base changes that led to the loss of such a binding site (table 2). Among the binding sites that were found, many were recognized by ubiquitous transcription factors such as SP1, AP2, or NF. We also found, however, consensus binding elements for more specifically acting factors, such as *Msx1*, that play a role in neural tube development (table 2). At the top of figure 1, the approximate positions are given of those SNPs that cause loss of the putative transcription factor binding sites. Eight of them fall in the previously mentioned evolutionarily conserved regions (−4201G→A, −1260T→C, −719C→T, IVS1+2820T→C, IVS1+2846G→T, IVS1+3104C→T, IVS1+5094T→A, and IVS1+9494C→A) (fig. 1; table 2).

Previously, we reported 23 different haplotypes (Burzynski et al. 2004), of which 4 were homozygously present in our patients or controls (2 risk haplotypes and 2 non-risk haplotypes). We typed individuals homozygous for these four haplotypes by direct sequencing (used for seven SNPs) and by digestion (*HpyCH4IV*) of amplified PCR products (used for one SNP, IVS1+9494C→A). For four

of the eight markers, including the two most proximal SNPs, −4201G→A and −1260T→C, the same alleles were identified on both the risk and nonrisk haplotypes (fig. 2). Four SNPs (−719C→T, IVS1+2846, IVS1+3104, and IVS1+9494) showed perfect segregation with the disease, meaning that one allele was present on a risk haplotype and the other was on a nonrisk haplotype. Therefore, the candidate region for the causal mutation in HSCR is likely bordered by SNP−1260C→A and SNP IVS1−126G→T. As SNPs, −4201G→A and −1260T→C are not part of the candidate region; only SNPs −719C→T, IVS1+2820T→C, IVS1+2846G→T, IVS1+3104C→T, IVS1+5094T→A, and IVS1+9494C→A were subsequently genotyped in all of our families with HSCR. DNA samples used for genotyping were obtained after the receipt of written informed consent from patients with sporadic HSCR and from their parents (only those who resided in the Netherlands); the DNA samples were subsequently referred to clinical genetics centers. Almost all of the families in our study (113 of 117) consist of two parents and an affected child (trios). In three cases, only one parent was available for the study, and one patient participated with his child and spouse. Sixty-four

**Table 2****Sequence Variations in the Common Risk Haplotype with Loss of a Hypothetical Consensus Binding Element**

Position and Type of Variant	NCBI dbSNP Accession No.	Transcription Factor
–7359A→C	rs2505990	CSS(2)
–6347G→A	rs2505991	AP2/SP1, SP1erk
–4201G→A	rs2082107	SP1-YB1, ApoE-B1
–3820C→T	rs2505994	USF-APP, IgHC.4, NF-mu-E1-CS, GT-2B-RS
–3054G→A	rs2505995	HC5
–2897T→C	rs2505996	NF-E1, NF-E1.6
–1260T→C	rs2505999	MEF3_B
–719C→T	rs2435366	SP1erk, AP2/SP1, SP1-TPI(4), SP1-hsp70, hsp70.2, SP1-IE-3.3,4.5
IVS1+2820T→C	rs1864411	Uteroglobin HS-2.4-C, AP2-TBXAS1
IVS1+2846G→T	rs1864410	AP2/SP1, SP1-erk1(1), JCV
IVS1+3104C→T	rs1864408	NFI-CS3, ELP/SF1/FT2F1, SF1-CYP21
IVS1+5094T→A	rs2506016	BPV-E2.CS2, E2-RS1
IVS1+6294T→C	rs2506020	ETS-1
IVS1+6383–84GG→CC	...	SP1 complement factor, SP1-CS4, HC3
IVS1+6411G→A	...	AP2 CS4
IVS1+6751T→C	rs1897001	PEA3, IE1.2
IVS1+6813C→T	rs1896999	Site-I(2)
IVS1+7445A→C	rs2506008	HNF-5
IVS1+9494C→A	rs2506004	E1AF
IVS1+10813A→G	...	Msx1
IVS1–10032C→T	rs2435344	E2A-CAMLG, E2A, USFAPP
IVS1–10017C→T	rs2435345	SP1-TPI(4), AP-2-β
IVS1–7593G→A	rs752975	GR-MT-IIA
IVS1–7298G→A	...	AP2-CS5
IVS1–5791C→T	rs2505541	AP2-CS5
IVS1–5127C→G	...	TFIID-EIIa
IVS1–5038C→T	rs2505539	Uteroglobin HS-2.4-C, T antigen
IVS1–5005A→G	rs2435346	SP1YB1, MyoD-MCK right site, site F (4,5), AP2-APP
IVS1–3748C→T	rs2505537	AP2-β
IVS1–3702T→A	rs2505536	GR-uteroglobin(1,2), GR/Prconnexin43

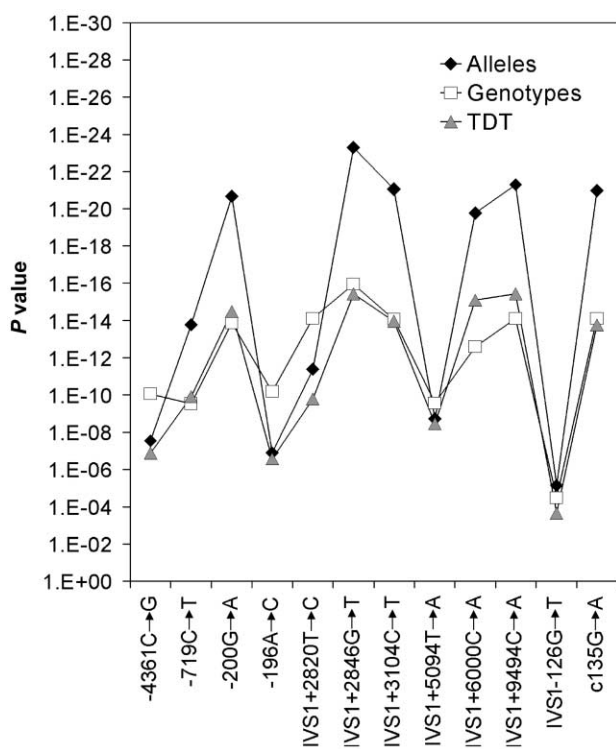
patients were screened by means of denaturing gradient gel electrophoresis and sequence analysis (Hofstra et al. 2000) and were found to be negative for *RET* mutations. These 64 were selected from an initial group of 72 patients with sporadic HSCR, of whom 8 appeared to carry a *RET* mutation. Fifty-three of the patients included in the study have not been screened for a *RET* mutation, which therefore may be present in a few of them. As controls for the transmitted haplotypes, the nontransmitted haplotypes of unaffected parents were used, as is also the case in the haplotype-relative-risk method (Falk and Rubinstein 1987; Terwilliger and Ott 1992). 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 and can be used to provide estimates of the number of heterozygotes and homozygotes in the population, as well as estimates of the odds ratios of their risk of developing HSCR. The results of the transmission/disequilibrium tests (TDTs) and the association tests for the six newly selected SNPs and

the six SNPs from our previous study are shown in figure 3. All SNPs were strongly associated. For six of them (–200G→A [SNP-5], IVS1+2820T→C, IVS1+2846G→T, IVS1+3104C→T, IVS1+5094T→A, and IVS1+9494C→A), the association was stronger than for the other six. Corresponding allele frequencies were all in the range of 66.7%–68.6% among patients and 21.8%–25.4% among pseudocontrols. Haplotype analysis showed that these six risk alleles are present on the same risk haplotype, CTACCTTAAATA (table 3). Its frequency is estimated to be 53.3% among patients and 15.9% among controls (table 3). Another haplotype, CCACCTTAAATA, which only differs from the risk haplotype at the second SNP (which also contains the risk alleles of our previous six-locus risk haplotype), was present in 6.01% of patients and in 1.5% of controls. Since this haplotype also confers an increased risk, it does not contribute to further restriction of our candidate region.

The question remains: which of the identified variants is contributing to the disease development? Several candidates have already been put forward. SNPs –200G→A

														Results from previous study			
														haplotype	freq T	freq NT	comment
-4361C→G	-4201G→A	-1260C→A	-719C→T	-200G→A	-196A→C	<b>IVS1+2820T→C</b>	<b>IVS1+2846G→T</b>	<b>IVS1+3104C→T</b>	<b>IVS1+5094T→A</b>	<b>IVS1+6000C→A</b>	<b>IVS1+9494C→A</b>	IVS1-126G→T	c135G→A	CACATA	55.6%	16.2%	risk
C	A	C	T	A	C	C	T	T	A	A	A	T	A	CACCTA	2.2%	0.0%	risk
C	A	C	C	G	C	C	G	C	A	C	C	G	A	CGCCGG	16.1%	28.7%	non-risk
G	G	T	C	G	A	T	G	C	T	C	C	T	G	GGACTG	17.0%	35.9%	non-risk

**Figure 2** Haplotypes consisting of 14 SNPs—6 from our previous risk haplotype and 8 newly selected ones—of four homozygous individuals. Individuals homozygous for one of the haplotypes identified in our previous study (the corresponding SNPs are indicated in *plain typeface*) were typed for eight new SNPs (indicated in *bold typeface*). They were also homozygous for all eight new SNPs. Risk alleles are in dark gray boxes; nonrisk alleles are in light gray boxes. freq T = frequency of transmitted haplotype; freq NT = frequency of nontransmitted haplotype.



**Figure 3** Results of allelic and genotypic association tests and TDT for the 5' *RET* locus SNPs. Results presented regard markers characterized in the previous study (Burzynski et al. 2004) as well as six new loci. For the single-locus allelic and genotypic association analyses, the frequencies of the alleles and the genotypes, respectively, were compared between patients and pseudocontrols by use of a  $\chi^2$  test. Transmission distortion of each allele versus all other alleles together was tested and combined in a multiallelic TDT (Spielman and Ewens 1996). All *P* values and CIs were corrected for multiple testing for independent tests by use of a Bonferroni correction.

(SNP-5) and  $-196A \rightarrow C$  (SNP-1), which were subjects of functional studies (Fitze et al. 2003; Garcia-Barcelo et al. 2005; Griseri et al. 2005) and are part of the risk haplotype identified in our previous study (Burzynski et al. 2004), were not present among the eight variants we selected. These two candidate variations did not fulfill our current criteria, which are based on interspecies sequence conservation and changes in transcription factor binding sites. Nevertheless, Fitze et al. (2003) claimed that SNP  $-200G \rightarrow A$ , in particular, caused reduced *RET* promoter activity. Griseri et al. (2005) found rather contradictory results, as they observed strong cell line dependency and no consistent results in promoter activity assays. These two SNPs are localized in close proximity to a transcription start site. They are part of the *RET* basal promoter, which is conserved in human, mouse, and rat. Because of these contradictory results and because of our findings presented here, it is likely that they are in linkage disequilibrium with mutation(s) lying nearby. The closest polymorphism that we found upstream of SNP-5 is SNP  $-719C \rightarrow T$ , one of our eight selected polymorphisms. It causes the loss of several putative sites recognized by the *SP1* and *AP2* transcription factors. It is likely that this SNP is still a part of the basal promoter and, thus, is a good candidate for future promoter activity studies. Three other SNPs (*IVS1+2820T* $\rightarrow$ *C*, *IVS1+2846G* $\rightarrow$ *T*, and *IVS1+3104C* $\rightarrow$ *T*) lie close together and are located in the proximal part of intron 1, in a region of strong sequence conservation between species (fig. 1, peak from  $-5.7$  kb to  $-5.2$  kb). The most distal SNP of the eight, *IVS1+9494C* $\rightarrow$ *A*, lies in an interesting region, since all species, including chicken, show homology with the human genome at this position (fig. 1, peak from  $+9.3$  kb to  $+9.7$  kb). The occurrence of evolutionarily conserved regions

**Table 3**

**Haplotype Frequencies over 12 SNPs**

ALLELE AT MARKER												% OF	% OF
-4361C→G	-719C→T	SNP-5	SNP-1	<i>IVS1+2820T→C</i>	<i>IVS1+2846G→T</i>	<i>IVS1+3104C→T</i>	<i>IVS1+5094T→A</i>	IVS1+6000C→A	<i>IVS1+9494C→A</i>	IVS1-126G→T	c135G→A	PATIENTS (n = 230)	CONTROLS (n = 290)
C	<i>C</i>	G	C	<i>C</i>	<i>G</i>	<i>C</i>	<i>A</i>	C	<i>C</i>	G	G	11.04	20.88
C	<i>C</i>	G	C	<i>C</i>	<i>G</i>	<i>C</i>	<i>A</i>	C	<i>C</i>	G	G	.00	3.43
G	<i>C</i>	G	A	<i>T</i>	<i>G</i>	<i>C</i>	<i>T</i>	C	<i>C</i>	T	G	11.05	24.41
G	<i>T</i>	G	A	<i>T</i>	<i>G</i>	<i>C</i>	<i>T</i>	C	<i>C</i>	T	G	.91	4.64
C	<i>C</i>	A	C	<i>C</i>	<i>T</i>	<i>T</i>	<i>A</i>	A	<i>A</i>	T	A	6.05	1.46
C	<i>T</i>	A	C	<i>C</i>	<i>T</i>	<i>T</i>	<i>A</i>	A	<i>A</i>	T	A	53.31	15.94

NOTE.—Frequencies were estimated by means of an expectation-maximization algorithm implemented in our own software. Newly selected SNPs are in bold italics.

located in an intronic sequence supports the idea that intron 1 is involved in transcriptional regulation. Downstream of the last SNP, IVS1+9494C→A, there is only one more short region of moderate homology in the human and rodent sequences (fig. 1, peak from +11.5 kb to +11.6 kb). Intron 1 in the mouse and rat genomes is only 12 kb long, whereas, in humans, it is 23 kb. This likely results either from a translocation or deletion of this sequence in rodent genomes or from an insertion in primate genomes, since, in the mouse, we found regions of partial homology to the missing sequence 30 kb upstream of *Ret* and also in intron 18 (data not shown).

On the basis of the statistical analyses, no single SNP could be identified as being the causal one (fig. 3). Since the region around SNP IVS1+9494 is the best-conserved region—not only in mammals but also in chicken—this DNA variation is a good candidate for disease susceptibility, although we cannot fully exclude the possibility that a nearby DNA variation(s) that we might have missed could actually be responsible for HSCR or could act in combination with the candidate proposed here. Marker IVS1+9494C→A causes the loss of a binding site of the *ETV4* enhancer (ETS variant gene 4 *E1A* enhancer binding protein, E1AF [DNA-binding sequence TGGACGT]). *ETV4* is a member of the ETS transcription factor family and is a downstream nuclear target of RAS-MAP kinase signaling. Its disturbed functioning was implicated in malignant tumors, such as ovarian carcinoma and breast cancer (Davidson et al. 2004). Transcription factors with an evolutionarily conserved ETS domain play (in general) important roles in cell development, cell differentiation, cell proliferation, apoptosis, and tissue remodeling (Oikawa 2004), which suggests that *ETV4* can influence the HSCR phenotype. Functional studies of *ETV4* and its interaction with *RET* regulatory sequences should follow, to determine the importance of this enhancer in *RET* transcription.

## Acknowledgments

This study was supported by the Nederlandse organisatie voor Wetenschappelijk Onderzoek (grants 901-04-210 and 901-04-225).

## Electronic-Database Information

The URLs for data presented herein are as follows:

Hirschsprung Disease, University of Groningen Department of Medical Genetics, <http://www.rug.nl/med/faculteit/disciplinegroepen/medischegenetica/hereditarydiseases/hirschsprungDisease> (for additional information about HSCR)

MULAN, <http://mulan.dcode.org/>

NCBI Home Page, <http://www.ncbi.nlm.nih.gov/>

Online Mendelian Inheritance in Man (OMIM), <http://www.ncbi.nlm.nih.gov/Omim/> (for HSCR and *RET*)  
UCSC Genome Browser, <http://genome.ucsc.edu/>

## References

- Amiel J, Lyonnet S (2001) Hirschsprung disease, associated syndromes, and genetics: a review. *J Med Genet* 38:729–739
- Badner J, Sieber W, Garver K, Chakravarti A (1990) A genetic study of Hirschsprung disease. *Am J Hum Genet* 46:568–580
- Bolk S, Pelet A, Hofstra RMW, Angrist M, Salomon R, Croaker D, Buys CHCM, Lyonnet S, Chakravarti A (2000) 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* 97:268–273
- Borrego S, Wright F, Fernandez R, Williams N, Lopez-Alonso M, Davuluri R, Antinolo G, Eng C (2003) 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* 72:88–100
- Burzynski GM, Nolte IM, Osinga J, Ceccherini I, Twigt B, Maas S, Brooks A, Verheij J, Plaza Menacho I, Buys CHCM, Hofstra RMW (2004) 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. *Eur J Hum Genet* 12:604–612
- Ceccherini I, Hofstra RM, Luo Y, Stulp RP, Barone V, Stelwagen T, Boccardi R, Nijveen H, Bolino A, Seri M (1994) DNA polymorphisms and conditions for SSCP analysis of the 20 exons of the *ret* proto-oncogene. *Oncogene* 9:3025–3029
- Davidson B, Goldberg I, Tell L, Vigdorichik S, Baekelandt M, Berner A, Kristensen GB, Reich R, Kopolovic J (2004) The clinical role of the *PEA3* transcription factor in ovarian and breast carcinoma in effusions. *Clin Exp Metastasis* 21:191–199
- Falk CT, Rubinstein P (1987) Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. *Ann Hum Genet* 51:227–233
- Fitze G, Appelt H, König I, Gorgens H, Stein U, Walther W, Gossen M, Schreiber M, Ziegler A, Roesner D, Schackert HK (2003) Functional haplotypes of the *RET* proto-oncogene promoter are associated with Hirschsprung disease (HSCR). *Hum Mol Genet* 12:3207–3214
- Gabriel SB, Salomon R, Pelet A, Angrist M, Amiel J, Fornage M, Attie-Bitach T, Olson JM, Hofstra R, Buys C, Steffann J, Munnich A, Lyonnet S, Chakravarti A (2002) Segregation at three loci explains familial and population risk in Hirschsprung disease. *Nat Genet* 31:89–93
- Garcia-Barcelo M, Ganster RW, Lui VC, Leon TY, So MT, Lau AM, Fu M, Sham MH, Knight J, Zannini MS, Sham PC, Tam PK (2005) TTF-1 and *RET* promoter SNPs: regulation of *RET* transcription in Hirschsprung's disease. *Hum Mol Genet* 14:191–204
- Griseri P, Bachetti T, Puppo F, Lantieri F, Ravazzolo R, Devoto M, Ceccherini I (2005) A common haplotype at the 5' end of the *RET* proto-oncogene underlies genetic susceptibility to



- HSCR development through reduced gene expression. *Hum Mutat* 25:189–195
- Hofstra RM, Wu Y, Stulp RP, Elfferich P, Osinga J, Maas SM, Siderius L, Brooks AS, vd Ende JJ, Heydendael VM, Severijnen RS, Bax KM, Meijers C, Buys CH (2000) RET and GDNF gene scanning in Hirschsprung patients using two dual denaturing gel systems. *Hum Mutat* 15:418–429
- Holschneider A (1982) Hirschsprung's congenital megacolon: the concept of physiopathology and therapy. *Med Welt* 33: 210–213
- Manie S, Santoro M, Fusco A, Billaud M (2001) The RET receptor: function in development and dysfunction in congenital malformation. *Trends Genet* 17:580–589
- Oikawa T (2004) ETS transcription factors: possible targets for cancer therapy. *Cancer Sci* 95:626–633
- Sancandi M, Griseri P, Pesce B, Patrone G, Puppo F, Lerone M, Martucciello G, Romeo G, Ravazzolo R, Devoto M, Ceccherini I (2003) Single nucleotide polymorphic alleles in the 5' region of the *RET* proto-oncogene define a risk haplotype in Hirschsprung disease. *J Med Genet* 40:714–718
- Spielman RS, Ewens WJ (1996) The TDT and other family-based tests for linkage disequilibrium and association. *Am J Hum Genet* 59:983–989
- Terwilliger JD, Ott J (1992) A haplotype-based “haplotype relative risk” approach to detect allelic associations. *Hum Hered* 42:337–346