Ancestral RET haplotype associated with Hirschsprung’s disease shows linkage disequilibrium breakpoint at −1249

R M Fernandez, G Boru, A Peciña, K Jones, M López-Alonso, G Antínolo, S Borrego, C Eng


Several studies have shown that there are specific haplotypes of the RET proto-oncogene associated with the sporadic forms of Hirschsprung’s disease (HSCR). Based on such studies, linkage disequilibrium (LD) mapping estimates predicted the presence of an HSCR locus located at 22–50 kb upstream of codon 45 (exon 2) of RET. The aim of this study was to identify a founding locus responsible for the majority of HSCR cases in a particular series.

After a systematic mutational screening upstream of exon 2 of RET, the variants found were genotyped in 117 HSCR cases and 100 controls by fluorescence resonance energy transfer (FRET) or direct sequencing. Haplotype and genotype distributions were compared among these groups using standard case-control statistical analyses.

A specific RET haplotype was identified, clearly linked to HSCR phenotype. The LD of the analysed variants was maintained along the whole region studied up to position −1249. The ancestral haplotype associated with HSCR was characterised by the presence of specific single nucleotide polymorphisms (at −200 and −196) in proximity to the transcriptional start site. Functional modelling using luciferase expression assays showed a significantly depressed activity for the HSCR linked haplotype at −200/−196 in comparison with other combinations associated with controls.

These results seem to disprove the existence of an HSCR causing mutation as conceived in the traditional sense, but strengthen the concept of a specific combination of markers conferring susceptibility to the disease in a low penetrance fashion. It is conceivable that such an “HSCR haplotype”, together with other events occurring in other genes, might give rise to the disease. This would be in agreement with a polygenic model for the disease.

Key points

- Several studies have shown that there are specific haplotypes of the RET proto-oncogene associated with the sporadic forms of Hirschsprung’s disease (HSCR).
- Based on such studies, linkage disequilibrium (LD) mapping estimates predicted the presence of an HSCR locus located at 22–50 kb upstream of codon 45 (exon 2) of RET. The aim of this study was to identify a founding locus responsible for the majority of HSCR cases in a particular series.
- After a systematic mutational screening upstream of exon 2 of RET, the variants found were genotyped in 117 HSCR cases and 100 controls by fluorescence resonance energy transfer (FRET) or direct sequencing. Haplotype and genotype distributions were compared among these groups using standard case-control statistical analyses.
- A specific RET haplotype was identified, clearly linked to HSCR phenotype. The LD of the analysed variants was maintained along the whole region studied up to position −1249. The ancestral haplotype associated with HSCR was characterised by the presence of specific single nucleotide polymorphisms (at −200 and −196) in proximity to the transcriptional start site. Functional modelling using luciferase expression assays showed a significantly depressed activity for the HSCR linked haplotype at −200/−196 in comparison with other combinations associated with controls.
- These results seem to disprove the existence of an HSCR causing mutation as conceived in the traditional sense, but strengthen the concept of a specific combination of markers conferring susceptibility to the disease in a low penetrance fashion. It is conceivable that such an “HSCR haplotype”, together with other events occurring in other genes, might give rise to the disease. This would be in agreement with a polygenic model for the disease.

Abbreviations: FRET, fluorescence resonance energy transfer; HSCR, Hirschsprung’s disease; LD, linkage disequilibrium; SNP, single nucleotide polymorphism.

METHODS
Patients and controls
In this study we included a series of 117 cases affected by clinically sporadic HSCR (20% female, 80% male), and their unaffected parents when available. More specifically, the triads (composed of the HSCR patient and both parents) were complete for 104 of the cases, while for 12 patients only one of the parents was available. There was only one patient for whom we could not collect DNA from either parent. The characteristics of our cohort of HSCR patients have been described in detail previously.15–17 In addition, we also included a group of 100 normal controls comprising...
Genomic DNA was extracted according to standard protocols.13 Mutational screening of the whole intron 1 and promoter of RET was undertaken by direct sequencing of overlapping fragments obtained by polymerase chain reaction (PCR), using an Applied Biosystems 3730 DNA sequencer, and aligning our results with the sequence provided in Genbank accession number NT 033985.

Large scale genotyping of each of the variants was done either by fluorescence resonance energy transfer (FRET) technology or by direct sequencing of the PCR products. The FRET internal probes were designed according to the sequences of interest and purchased to TIB Molbiol.

Table 1 Allelic frequencies for the RET polymorphic loci in patients with Hirschsprung’s disease (HSCR) and controls

<table>
<thead>
<tr>
<th>RET sequence variants</th>
<th>HSCR</th>
<th>Controls</th>
<th>χ² test with Yates correction</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1249C→T</td>
<td>0.34</td>
<td>0.43</td>
<td>χ² = 3.32, p = 0.08</td>
</tr>
<tr>
<td>-200A→G</td>
<td>0.39</td>
<td>0.83</td>
<td>χ² = 86.32, p &lt; 0.000000001</td>
</tr>
<tr>
<td>-196C→A</td>
<td>0.15</td>
<td>0.44</td>
<td>χ² = 43.7, p &lt; 0.000000001</td>
</tr>
<tr>
<td>IVS1+53G→A</td>
<td>0.15</td>
<td>0.44</td>
<td>χ² = 43.7, p &lt; 0.000000001</td>
</tr>
<tr>
<td>IVS1+620A→G</td>
<td>0.15</td>
<td>0.44</td>
<td>χ² = 43.7, p &lt; 0.000000001</td>
</tr>
<tr>
<td>IVS1+779C→T</td>
<td>0.15</td>
<td>0.44</td>
<td>χ² = 43.7, p &lt; 0.000000001</td>
</tr>
<tr>
<td>IVS1+2820C→T</td>
<td>0.15</td>
<td>0.44</td>
<td>χ² = 43.7, p &lt; 0.000000001</td>
</tr>
<tr>
<td>IVS1+2834C→T</td>
<td>0.12</td>
<td>0.31</td>
<td>χ² = 18.2, p = 0.000194</td>
</tr>
<tr>
<td>IVS1+2846T→G</td>
<td>0.39</td>
<td>0.83</td>
<td>χ² = 75.2, p &lt; 0.000000001</td>
</tr>
<tr>
<td>IVS1+3105T→C</td>
<td>0.39</td>
<td>0.83</td>
<td>χ² = 75.2, p &lt; 0.000000001</td>
</tr>
<tr>
<td>IVS1+3991del12</td>
<td>0.86</td>
<td>0.59</td>
<td>χ² = 35.8, p &lt; 0.000000001</td>
</tr>
<tr>
<td>IVS1+6136T→G</td>
<td>0.15</td>
<td>0.44</td>
<td>χ² = 43.7, p &lt; 0.000000001</td>
</tr>
<tr>
<td>IVS1+6174A→G</td>
<td>0.15</td>
<td>0.44</td>
<td>χ² = 43.7, p &lt; 0.000000001</td>
</tr>
<tr>
<td>IVS1+6383-84CG→GG</td>
<td>0.15</td>
<td>0.44</td>
<td>χ² = 43.7, p &lt; 0.000000001</td>
</tr>
<tr>
<td>IVS1+6411A→G</td>
<td>0.15</td>
<td>0.44</td>
<td>χ² = 43.7, p &lt; 0.000000001</td>
</tr>
<tr>
<td>IVS1+7236A→G</td>
<td>0.39</td>
<td>0.83</td>
<td>χ² = 75.2, p &lt; 0.000000001</td>
</tr>
<tr>
<td>IVS1+7436G→A</td>
<td>0.39</td>
<td>0.83</td>
<td>χ² = 75.2, p &lt; 0.000000001</td>
</tr>
<tr>
<td>IVS1+840del16</td>
<td>0.86</td>
<td>0.59</td>
<td>χ² = 35.8, p &lt; 0.000000001</td>
</tr>
<tr>
<td>IVS1+9202C→T</td>
<td>0.39</td>
<td>0.83</td>
<td>χ² = 75.2, p &lt; 0.000000001</td>
</tr>
<tr>
<td>IVS1+9277C→G</td>
<td>0.39</td>
<td>0.83</td>
<td>χ² = 75.2, p &lt; 0.000000001</td>
</tr>
<tr>
<td>IVS1+9494A→C</td>
<td>0.39</td>
<td>0.83</td>
<td>χ² = 75.2, p &lt; 0.000000001</td>
</tr>
<tr>
<td>IVS1+11260A→T</td>
<td>0.39</td>
<td>0.83</td>
<td>χ² = 75.2, p &lt; 0.000000001</td>
</tr>
<tr>
<td>IVS1+11369C→T</td>
<td>0.17</td>
<td>0.44</td>
<td>χ² = 35.8, p &lt; 0.000000001</td>
</tr>
<tr>
<td>Wild type T</td>
<td>0.17</td>
<td>0.44</td>
<td>χ² = 35.8, p &lt; 0.000000001</td>
</tr>
</tbody>
</table>

Statistical analysis of the allelic distribution in both groups.

Conditions for both FRET and sequencing, and the sequences of primers and probes, are available on request.

Statistical analysis

Allelic and genotypic frequencies of the RET polymorphisms analysed were calculated and then compared between patients and controls. Comparisons were made using χ² analysis with Yates’ correction, with statistical significance set at p < 0.05.

Haplotypes were constructed comprising 23 of the 42 novel variants (from -200A→C to IVS1+11637C→T) as well as the other three SNPs located in the 3’ end of the intron which had been analysed previously by our group.14 Generation of these extended haplotypes was possible and accurate because of the availability of the complete triad for 89% of the HSCR patients, which allowed us to reconstruct and compare the transmitted versus non-transmitted alleles. Based on the observation of complete linkage disequilibrium of an important part of the studied variants, and the existence of only a few combinations of these, it was plausible, and reliable, to infer the haplotypes for controls.

unselected, unrelated, race, age, and sex matched individuals from Spain.

Informed consent was obtained from all the individuals studied, in accordance with the respective committees for protection of human subjects.

Identification, characterisation, and genotyping of RET sequence variants

Genomic DNA was extracted according to standard protocols.13 Mutational screening of the whole intron 1 and promoter of RET was undertaken by direct sequencing of overlapping fragments obtained by polymerase chain reaction (PCR), using an Applied Biosystems 3730 DNA sequencer, and aligning our results with the sequence provided in Genbank accession number NT 033985.
The haplotype distribution was compared between cases and controls. In addition, parental haplotypes were examined in the context of the affected children’s haplotypes, so that transmitted and non-transmitted haplotypes were noted and their frequencies compared.

**Promoter constructs and assays**

Plasmid construction for the RET minimal promoter region

The 5’ UTR region of the RET genomic DNA that comprises the variants −200A→G and −196C→A was PCR amplified using a forward primer (5’ GGCTAGCCCCGACTGAGCTC CTACAC-3’) and a reverse primer (5’ CCGTGAAGTGGAGG GACTGGGCTA-3’). These primers span the minimum promoter region (−313 to −24 from translational start site) and both have a restriction site (XhoI and XhoI respectively) at the 5’ end.

DNA samples with GA, GC, and AC (−200, −196) polymorphic variants were PCR amplified. The PCR products and the pGL3–basic luciferase report vector (Promega) were digested with XhoI and XhoI. The fragments were cloned upstream of the XhoI site and transformed into E. coli TOPO 10 competent cells following the manufacturer’s instruction (Invitrogen). Transformants with the desired fragment were screened both by restriction digest and by direct sequencing.

**RESULTS**

Novel SNPs and other types of polymorphic loci were sought upstream of IVS1−1463 by direct sequencing of the 23 kb intron 1, exon 1, and the region 12 kb upstream of the translational start site. This revealed 42 novel SNPs and two insertion/deletion polymorphisms (available on request and posted to the JMG website). Once identified, all the variants were tested in a small group of patients and controls and the preliminary results showed a strong association with HSCR. Because our fundamental aim was the identification of the putative functional low penetrance susceptibility locus for HSCR guided by our previous statistical calculation,15 we only genotyped the variants located nearest to the 5’ end of RET in the entire series of patients and controls. Following this strategy, we observed a lack of association of specific alleles with disease state at $−1249C/T$ suggesting that linkage disequilibrium had broken (p = 0.08). Thus we subsequently concentrated on the region between −1249 and IVS1+11637,

Novel SNPs and other types of polymorphic loci were sought upstream of IVS1−1463 by direct sequencing of the 23 kb intron 1, exon 1, and the region 12 kb upstream of the translational start site. This revealed 42 novel SNPs and two insertion/deletion polymorphisms (available on request and posted to the JMG website). Once identified, all the variants were tested in a small group of patients and controls and the preliminary results showed a strong association with HSCR. Because our fundamental aim was the identification of the putative functional low penetrance susceptibility locus for HSCR guided by our previous statistical calculation,15 we only genotyped the variants located nearest to the 5’ end of RET in the entire series of patients and controls. Following this strategy, we observed a lack of association of specific alleles with disease state at $−1249C/T$ suggesting that linkage disequilibrium had broken (p = 0.08). Thus we subsequently concentrated on the region between −1249 and IVS1+11637.

**Table 2** Definition of the RET extended haplotypes

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1a</th>
<th>1b</th>
<th>1c</th>
<th>1d</th>
<th>1e</th>
<th>2a</th>
<th>2b</th>
<th>2c</th>
<th>2d</th>
<th>2e</th>
<th>3c</th>
<th>3d</th>
<th>3e</th>
<th>4c</th>
<th>4d</th>
<th>4e</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The haplotypes marked with an asterisk have a very low frequency.

$+$, presence/absence of the polymorphic variant.
which was found to contain 22 novel SNPs and two insertion/deletion polymorphisms (table 1). We observed that the distribution of each variant was significantly different in HSCR cases from controls, showing an under-representation of all SNPs and an over-representation of the deletions (table 1, p<0.0001).

Because 89% of parental pairs were available for the HSCR cases, haplotypes comprising the analysed variants and those at the 3′ end of the intron (IVS1–1463T→C, IVS1–1370C→T and IVS1–126G→T) could be constructed for 92 of the cases and inferred for the rest. Although we found 13 different haplotypes among the HSCR cases and their parents, some of these were very rare (table 2). In fact, each of the 13 extended haplotypes could be defined as different subtypes within the five fundamental intronic haplotypes previously published (0–4). Of note, haplotype 0 (previously defined as “HSCR haplotype”) is always linked to a unique combination of variants in transmitted and non-transmitted alleles within the HSCR cases and their parents, as was the frequency of particular transmitted alleles to HSCR patients (χ² = 83.64, p<0.00000001), or ν control alleles (χ² = 84.02, p<0.00000001). It is important to note that haplotype AC is in complete linkage disequilibrium with all the risk haplotypes described so far (table 2). In contrast, the frequencies for haplotypes GA (−200G→196C) and GA (−200G→196A) are considerably lower among HSCR chromosomes (25% and 15%) than among non-transmitted (44% and 40%) control chromosomes (39% and 44%). Similar conclusions can be drawn from the comparison of the distribution of pairs of haplotypes (table 2), where the differences observed are even more pronounced (fig 2B).

Luciferase reporter assay of RET promoter function
Luciferase activity was assayed in TT cells with constructs containing the three naturally existing promoter haplotypes at −200 and −196, and AC (see table 2). GC (or + +, in table 2) had the highest expression (fig 3). AC (or − −), which is over-represented in HSCR, had the lowest expression (fig 3). Interestingly, GA (− +) had an expression level less than that of GC but much more than AC, bringing its expression level much closer to that of GC than that of AC (fig 3).

DISCUSSION
It is currently well established that the RET proto-oncogene is the major gene implicated in the pathogenesis of Hirschsprung’s disease. Besides traditional germline loss of function mutations causing HSCR, several common RET

Figure 2 (A) Distribution of haplotypes comprising the RET promoter variants in transmitted and non-transmitted alleles within the Hirschsprung’s disease (HSCR) families, and in controls. (B) Distribution of pairs of haplotypes comprising the RET promoter variants in HSCR patients and controls.

Figure 3 Luciferase assay showing decreased expression from the Hirschsprung’s disease (HSCR) associated −200−196 haplotype AC (corresponding to − − in table 2) compared with the control associated GA (++ in table 2) haplotype.
proto-oncogene polymorphisms, present with different frequencies in different populations, have been associated with a variable risk of developing HSCR. In this sense, the HSCR associated variant firstly described was the silent polymorphism A45A, for which an association with the disease has been demonstrated in several cohorts of patients with different ethnic backgrounds, such as the Spanish, German, Italian, or Chinese. \(^{11,12,19–21}\)

Several hypotheses have been put forward to explain the underlying cause of such an association, including a direct action of the variant or the existence of linkage disequilibrium with some functional, still unknown allele. While no functional assays have yet confirmed the first hypothesis, the second has gained major relevance because of the results of subsequent studies. More specifically, we observed that the c.135A allele, irrespective of the 3' haplotype, was in linkage disequilibrium with a group of markers within the 3' end of the first intron (haplotype 0: IVS1–1463T, IVS1–1370C, IVS1–126G). \(^{13}\) These results strongly suggested the existence of a low penetrance locus of susceptibility for HSCR located 5' upstream all these markers.

Recently, the analysis of the promoter region of the \(RET\) proto-oncogene in three different populations of HSCR patients has revealed the existence of two new SNPs located at \(-200\) and \(-196\) from the transcription start site (\(-200\)\(\rightarrow\)G and \(-196\)\(\rightarrow\)A) in linkage disequilibrium with A45A. \(^{14–16}\) Moreover, there is a strong association with HSCR in such populations, supporting the postulate previously mentioned. In particular, the “ACA” haplotype was shown to be associated with HSCR and to present a significantly lower activity in dual-luciferase expression assays in vitro, compared with those haplotypes identified in the majority of normal controls. \(^{14}\) In agreement with these data, it has been suggested that the “ACA” combination may represent a core haplotype associated with HSCR, acting as a modifying risk allele in the development of the condition. \(^{14}\)

Based on our previous results, our group has carried out a systematic mutational screening along the 5' region of \(RET\) comprising the whole intron 1, exon 1, and the region 12 kb upstream of the translational start site, in order to identify the putative founding locus accounting for the majority of sporadic HSCR cases. \(^{13}\) Although no germline mutations were identified in our patients, we found several novel polymorphisms that are in linkage disequilibrium with the so called haplotype 0\(^+\) and A45A, and strongly associated with HSCR. This association was supported not only by the results from comparative studies between HSCR cases and normal controls, but also from the TDT analysis in the HSCR triads. The linkage disequilibrium (LD) was maintained along the entire analysed region until position \(-1249\), where the lack of association of specific alleles with the disease suggested that LD is broken. The relevance of our results resides in the finding of a putative LD breakpoint which would allow us to delimit the 5' end of the “HSCR linked” \(RET\) region at a distance of approximately 24 kb with respect to A45A, in agreement with what had been previously calculated by LD mapping estimates. \(^{13}\) This finding would be in contrast to the results of a previous study, in which the presence of a variant linked to A45A was described, located at approximately \(-4\) kb from the translational start site. \(^{16}\) It would be plausible if the differences obtained in both studies reflected the distinct genetic background of the two populations examined. Future analyses of the region in different populations may help to elucidate this.

In conclusion, our findings seem to rule out the existence of an HSCR causing mutation, as conceived in the traditional sense, but strengthen the view that there is a specific combination of markers conferring susceptibility to the disease in a low penetrance fashion. The data derived from our functional in vitro studies, in agreement with other previously reported results, \(^{14}\) suggest that the haplotype 0 (table 2) may result in a lower level of expression of the \(RET\) gene. Thus homozygous individuals for this haplotype (62% of our cases and only 2% of controls) would present with considerably lower RET levels than those with other haplotype pairs. As the pathogenesis of HSCR is generally

**Figure 4** Fluorescence resonance energy transfer (FRET) pattern for the simultaneous detection of the \(RET\) variants \(-200\)\(\rightarrow\)G and \(-196\)\(\rightarrow\)A.
ascribed to loss of function or haploinsufficiency of the RET gene, the decrease in RET transcription rate associated with this haplotype could have a functional effect linked to RET messenger RNA expression. It is conceivable that such a “HSCR haplotype”, together with other events occurring in other genes, might give rise to the disease, which would be concordance with a polygenic model of the disease. However, the precise mechanism of actuation of this RET haplotype is not completely clear. Thus further studies are warranted to elucidate the functional relevance of this haplotype in the pathogenesis of Hirschsprung’s disease.

ACKNOWLEDGEMENTS

This study was partially funded by grants IR01HD39058 from the National Institutes of Health, Bethesda, Maryland, USA (to GA, SB, and CE); Red de Centros 03/05 from the Instituto de Salud Carlos III, Spain (to RMF, GA, and SB); and CAA 56/03 from the Consejería de Salud de la Junta de Andalucía, Spain (to RMF, AP, GA, and SB). RMF is a doctoral fellow of the Fundación Ramón Areces. AP is supported by Instituto de Salud Carlos III contract CP03/00045. CE is a recipient of the Doris Duke distinguished clinical scientist award.

Authors’ affiliations

R M Fernandez, A Pecina, G Antíñolo, S Borrego, Unidad Clínica de Genética y Reproducción, Hospitales Universitarios Virgen del Rocío, Seville, Spain

G Boru, K Jones, C Eng, Human Cancer Genetics Program, Comprehensive Cancer Center, The Ohio State University, Columbus, Ohio, USA

M López-Alonso, Servicio de Cirugía Infantil, Hospitales Universitarios Virgen del Rocío, Seville, Spain

Competing interests: none declared

Correspondence to: Dr Salud Borrego, Unidad Clínica de Genética y Reproducción, Hospitales Universitarios Virgen del Rocío, Avda. Manuel Siurot s/n, 41013, Seville, Spain; salud.borrego.sspa@juntadeandalucia.es

Revised version received 2 August 2004

Accepted for publication 11 August 2004

REFERENCES


Y chromosome microdeletions are the most frequent genetic cause of severe oligozoospermia (<5 million spermatozoa/ml) and azoospermia (absence of spermatozoa in the ejaculate). Microdeletions associated with infertility occur in specific regions of the long arm of the Y chromosome, called azoosperma factor (AZF) regions. In 1996, three types of AZF deletion (AZFa, AZFb, and AZFc) were described by Vogt et al; however, after the complete physical map and sequence of the AZFb and AZFc regions was produced, it became evident that the AZFb and AZFc intervals partially overlap. The Y chromosome is extremely rich in repetitive sequences, organised in amplicons. Ampliconic sequences are characterised by sequence pairs showing nearly complete (>99.9%) nucleotide identity, organised in massive palindromes. These repeated sequences may undergo genetic exchange through gene conversion—that is, non-reciprocal transfer of sequence information occurring between duplicated sequences within the chromosome, a process that could account for the >99.9% nucleotide identity between the arms of a palindrome. Although this mechanism may serve to preserve Y chromosome genes from the gradual accumulation of deleterious mutations and thus prolong their genetic fitness, this peculiar organisation also provides the structural basis for deletions and rearrangements.

The classical AZFc deletion, which removes 3.5 Mb between the b2/b4 amplicons, is the most frequent type of deletion. Taking into consideration the Y chromosome structure and the suggested deletion mechanism, a number of other possible partial deletions have been proposed in both the AZFb and AZFc regions. The frequency and the pathological significance of these partial deletions is not yet clear, although recently a partial deletion termed gr/gr has been described specifically in infertile men with varying degrees of spermatogenic failure. This deletion removes half the AZFc gene content, including two copies of the major AZFc candidate gene called DAZ. Another deletion, named b2/b3 or u3-gr/gr or g1/g3, which removes a similar quantity of AZFc genes, seems to have no effect on fertility status in association with a certain Y chromosome background common present in northern Eurasian populations (Y haplogroup N). A similar conclusion can also be drawn for the gr/gr deletion found in association with Hgr D2b, which is present in 20% of Japanese men. One of the possible explanations for the heterogeneous phenotype observed in association with complete and partial AZFc deletions is that polymorphisms or mutations are present in the autosomal homologue of DAZ, DAZL. Indeed, the finding that the human DAZ transgene is able to partially rescue the spermatogenic failure of mice homozygous for a null allele of Dazl suggests a possible interplay between DAZL and DAZ in humans. In the case of partial AZFc deletions, we can also speculate that the type and number of missing gene copies or other unknown Y chromosome related factors (for example duplications or beneficial mutations in other parts of the Y chromosome) may also influence the phenotype.

The aims of this study were: (a) to establish the clinical significance of partial AZFc deletions (that is, if any of them are specific for spermatogenic failure or can be considered a risk factor); and (b) to gain insight through a combined approach (quantitative and qualitative analysis of DAZ and CDY1 genes deleted and the Y haplogroup, and (b) carried out a mutational screen of DAZL, the autosomal homologue of DAZ.

We found that: (a) partial AZFc deletions are not specific for spermatogenic failure, and (b) gr/gr deletion can be considered as a risk factor because its frequency was significantly higher in the oligo/azooospermic group (5.3%) than in controls (0.5%) p<0.012. Gene specific analysis revealed three distinct deletion patterns, indicating that further combined studies based on gene copy and haplogroup analysis are likely to provide a means for the distinction between pathogenic and neutral (or compensated by other Y factors) types of deletion.

**Subjects**

The study population consisted of 150 infertile patients (89 idiopathic and 61 with mild abnormal andrological findings: varicocele, monolateral cryptorchidism, recurrent infections) seeking complete andrological investigation for couple infertility at the Andrology Unit of the University Hospital Careggi in Florence, Italy, and 189 normospermic controls of

**Key points**

- Deletion of the AZFc region of the Y chromosome is the most frequent molecular genetic cause of oligo/azoospermia. Within this region, partial deletions have been recently described, including the gr/gr deletion.
- A direct aetiopathogenic role has been suggested for the gr/gr deletion as it was absent in normozoospermic but present in oligospermic men.
- We studied a group of normospermic (n = 189) and oligo/azoospermic (n = 150) men using a sequence tagged site (STS) +/- method. To gain insight into the molecular basis of the heterogeneous phenotype observed in men with the deletion we: (a) defined the type of DAZ and CDY1 genes deleted and the Y haplogroup, and (b) carried out a mutational screen of DAZL, the autosomal homologue of DAZ.

**Materials and methods**

**Subjects**

The study population consisted of 150 infertile patients (89 idiopathic and 61 with mild abnormal andrological findings: varicocele, monolateral cryptorchidism, recurrent infections) seeking complete andrological investigation for couple infertility at the Andrology Unit of the University Hospital Careggi in Florence, Italy, and 189 normospermic controls of

**Abbreviations:** AZF, azoosperma factor; SFV, sequence family variant; SNV, single nucleotide variant; STS, sequence tagged site

---

**LETTER TO JMG**

The gr/gr deletion(s): a new genetic test in male infertility?


Italian origin. Cytogenetic analysis and Y chromosome microdeletion screening revealed 46,XY karyotype and the absence of AZF microdeletions in all included patients. Samples were collected using approved protocols and the informed consent of all individuals was obtained.

Screening for partial AZFc deletions

We tested eight STSs, originally described by Repping et al: sY1291, sY1191, sY1161, sY1206, sY1201, sY142, sY1258, and sY1197. In addition, we tested our samples with a complementary set of primers, o1084/o1085 and o1276/o1277.13 We identified gr/gr deletion by the following STS results: sY1291 and o1084/o85 negative; sY1161, sY1191, sY1206, sY1201, and o1276/o1277 all positive (an example of a PCR duplex result is shown in fig 1). The b2/b3 or u3-gr/gr deletion is characterised by the absence of the STS sY1191-o1276/o1277 and the presence of the rest of the STS.

Qualitative and quantitative analysis for DAZ copy and CDYa/CDYb loss was performed according to Machev et al.8 For DAZ, we chose the sequence family variant (SFV) at STS sY587 in intron 10, which distinguishes DAZ1/2 from DAZ3/4.

For CDY1 we used a C/A SFV situated 7750 bp 5’ of the CDY1 translation start codon (CDY7750). In order to quantify copy number we used a PCR based method described by Machev et al.4 The PCR products were separated on an automatic sequencer (LI-COR Gene ReadIR 4200), and quantification was performed by ONE-Dscan.

Y chromosome haplotyping

Y chromosome haplotyping was performed as previously published for the YAP, M9, SRY4064, and 92R7 polymorphisms.17 The Tat polymorphism18 was assayed by PCR digestion with primers o1489: atgtatatagtacgtctgtagg and o1490: gtaaacatgtgaggtgctc (annealing 54°C). The 12I2 assay was performed as a duplex with 5RY primers 3’5RYT15 and 3’5RYT16 and the12I2 primers 12I2D and 12I2F.3 These primer pairs were used at a concentration of 300 nmol/l for each 5’RY primer and 600 nmol/l for each 12I2 primer, and the annealing temperature used was 52°C. Polymorphisms were visualised by restriction enzyme digest for M9 (HindI), SRY4064 (BsrBI), Tat (HpyCH4IV), and 92R7 (HindIII).

All men were haplotyped for YAP, 12I2, and 92R7 polymorphisms (table 1), and individuals with partial AZF deletions were further genotyped with the Tat, M9, and SRY-4064 SNPs, defining five haplogroups, E, J, and K (xN3, P), N3, and P, and one paragroup Y* (xD, E, J, K).

Statistical analysis

Statistical analysis was performed using the statistical package SPSS for Windows (version 11; SPSS, Chicago, IL, USA). We tested the significance of the observed difference in the incidence of gr/gr deletion between our infertile and control populations using Fisher’s exact test. Our null hypothesis was that incidence is the same in infertile and control populations. All variables were checked for normal distribution by Kolmogorov-Smirnov one sample test. For comparisons of means between groups of different genotypes, Student’s t test for independent samples, when normal distribution was observed, was applied. Logarithmic transformation of data was performed in order to normalise the distribution when the presence of log normal distribution was checked. Finally, in case of non-normalised distribution, the non-parametric Mann-Whitney U test was applied to achieve the same objective. A p value <0.05 was considered statistically significant for each test.

RESULTS

Frequency and type of partial AZFc deletions in patients and normospermic men

Based on the STS +/- analysis16 it was possible to distinguish between different types of partial deletions (fig 1). We found two types of partial AZFc deletions: the gr/gr (8/339) and the b2/b3 (3/339) deletion, whereas b1/b3 was absent in our study population.

The gr/gr and b2/b3 deletions were found in both the infertile and normospermic groups, although at different frequencies. The frequency of gr/gr deletions was significantly higher in the patient (8/150; 5.3%) than in the normospermic group (1/189; 0.5%) (p<0.012; odds ratio (OR) 10, 16; 95% confidence interval (CI) 1.28 to 80.3). In contrast, the frequency of the less common b2/b3 (u3-gr/gr) deletion was not different between the two groups (1.3% v 0.5%; p = 0.41).

DAZ and CDY1 gene copy definition

In order to further characterise the deletions, we defined the type of missing DAZ (DAZ1/DAZ2/DAZ3/DAZ4) and CDY1 (CDY1a/CDY1b) gene copies. In the gr/gr deletion group, we found three types of deletion pattern: DAZ1/DAZ2+CDY1a, DAZ3/DAZ4+CDY1a, and DAZ3/DAZ4+CDY1b (fig 2). In the b2/ b3 deletions group, we found two types of deletion pattern: DAZ3/DAZ4+CDY1a and DAZ3/DAZ4+CDY1b. The presence of different gene copy deletions patterns (five combinations found in our study) indicates that each type of partial deletion may be further divided into subtypes, and although the number of gene copies removed is the same, the missing copy type is different.

Interestingly, DAZ1/DAZ2 and CDY1a were deleted only in the patient group, whereas DAZ3/DAZ4 and CDY1b were deleted in both groups. However, the deletion of CDY1a was not consistently associated with the absence of DAZ1/DAZ2 copies and thus can be deleted with either DAZ gene pair. The absence of the DAZ and CDY1 copies was further confirmed by a densitometric analysis, according to Machev et al.19 In all the gr/gr and b2/b3 deletion cases, we found a reduction of gene dosage for DAZ and CDY1. We have therefore no case of b2/b4 duplication among the men with deletions.

Genotype–phenotype association

Both types of deletion were associated with a wide range of sperm count, from azoospermia to normozoospermia (table 1). Although the mean values of the three principal sperm parameters were lower in patients with gr/gr deletion than in patients without, the difference was not significant: 2.1 v 4.5 million spermatozoa/ml for mean sperm concentration, 12.8% v 20.8% for sperm motility, and 11.6% v 15.3% for morphology. As CDY1a deletion was a specific feature of the patient group, we calculated the frequency of gr/gr deletions with missing CDY1a in both patients and controls; the difference between the two groups was highly significant (p<0.003). It is therefore possible that only specific patterns of partial deletions are deleterious for spermatogenesis.

We also sequenced the entire coding sequence of DAZL to test whether allelic forms of DAZL might underlie the phenotypic variation associated with the partial deletions. We found no new mutations. Only one single nucleotide variant (SNV) was found, at nucleotide position 260.19 Two patients were heterozygous for this SNV; however, its incidence in normospermic and infertile men is similar19 and thus it is unlikely to have any significant effect on the phenotype.

Genotype–Y chromosome haplogroup and phenotype relationship

Recent studies reported that b2/b3 (u3-gr/gr or g1/g3) deletions primarily arise in one family of closely related
Y chromosomes (hgrN). The three b2/b3 (g1/g3) deletions found in our study were observed in three different Y chromosome backgrounds (hgr J and DE in two oligospermic patients and N3 in a normospermic man). In previous studies, gr/gr deletions were found to occur on different Y chromosome backgrounds, in our cases on hgr P, J, DE, and Y*. The only normozoospermic man with this type of deletion (CS111) carries an hgrDE Y chromosome, which was also found in one severe oligozoospermic man (A234). However, these men have independent gr/gr deletions; both are deleted for DAZ3/DAZ4 but subject CS111 has a deletion of CDY1b and subject A234 a deletion of CDY1a.

**DISCUSSION**

Classical AZF deletions, defined by a few well designed markers, are clinically relevant because a clear cause–effect relationship between these deletions and spermatogenic failure has been established. Owing to the absence of such deletions in normospermic men, the screening for AZF deletions has become part of the routine diagnostic investigation of the infertile male. The complete absence of the AZFa interval is associated with the most severe phenotype (complete absence of germ cells in the testis), whereas deletion of the entire AZFb region is associated with spermatogenic arrest (in general at the stage of primary spermatocytes). Deletion of the AZFc region can lead to different degrees of spermatogenic failure, with a significant variability between individuals, ranging from the absence of germ cells in the testis to the presence of spermatozoa in the ejaculate (oligospermia). These genotype–phenotype correlations refer to the three classical AZF deletions, and further data are needed to better define any clinically relevant difference between the so called classical AZFb (P5/proximal P1) and the two newly identified AZFb+c (P5/distal P1 and P4/distal P1) deletions.

---

**Figure 1** Schematic representation of the proposed two step analysis for the detection of partial AZFc deletions. (A) the first step based on a PCR duplex with sY1191/1291 (alternatively, with o1084/5 and o1276/7) shows the four different possible outcomes of the analysis: b1/b3, gr/gr, b2/b3 (or u3-gr/gr or g1/g3), or absence of deletion; (B) the second step defines the breakpoints of the deletions.
In contrast to the situation for the complete AZFc deletion, genotype–phenotype correlations for the partial AZFc deletion seem to be more complicated. Using an STS $+/-2$ approach\(^9\) we found two types of partial AZFc deletion, gr/gr and b2/b3 (u3-gr/gr or g1/g3) in our study population. Both deletion types were present in the normospermic group, indicating that these deletions, in contrast to b2/b4 deletions, are not specific for spermatogenic failure. The gr/gr deletion seems to be more frequent than b2/b3 (or u3-gr/gr) in our Italian cohort (73% of the deletions were gr/gr). The similar frequency of the b2/b3 deletion in patients (1.3%) and controls (0.5%) does not support a pathogenic role for this

<table>
<thead>
<tr>
<th>Code</th>
<th>Aetiology</th>
<th>DAZL (exons from 1 to 11)</th>
<th>Deleted gene copies</th>
<th>Semen parameters</th>
<th>Number [n/ml*10^6]</th>
<th>Motility (A+B %)</th>
<th>Morphology (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A49*</td>
<td>Idiopathic</td>
<td>No mutations</td>
<td>3/4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A170</td>
<td>Idiopathic</td>
<td>No mutations</td>
<td>3/4</td>
<td></td>
<td>0.38</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>A202</td>
<td>Idiopathic</td>
<td>No mutations</td>
<td>1/2</td>
<td></td>
<td>0.9</td>
<td>20</td>
<td>16</td>
</tr>
<tr>
<td>A239</td>
<td>Idiopathic</td>
<td>No mutations</td>
<td>1/2</td>
<td></td>
<td>10</td>
<td>30</td>
<td>13</td>
</tr>
<tr>
<td>A286</td>
<td>Idiopathic</td>
<td>SNP 260</td>
<td>3/4</td>
<td></td>
<td>4.2</td>
<td>14</td>
<td>16</td>
</tr>
<tr>
<td>A 353*</td>
<td>Idiopathic</td>
<td>No mutations</td>
<td>3/4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A 46</td>
<td>Cryptorchidism sx</td>
<td>No mutations</td>
<td>1/2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A 184</td>
<td>Varicocele sx</td>
<td>No mutations</td>
<td>1/2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A 234</td>
<td>Varicocele sx</td>
<td>SNP 260</td>
<td>3/4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS 64*</td>
<td>—</td>
<td>No mutations</td>
<td>3/4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS111</td>
<td>—</td>
<td>No mutations</td>
<td>3/4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2 Schematic representation of the five distinct deletion patterns (based on the type of DAZ and CDY copies deleted) found in the 11 subjects with partial AZFc deletion. The gr/gr and b2/b3 (g1/g3) deletions were defined using a two step analysis based on STS $+/-2$ described by Repping et al\(^9\) and Machev et al\(^16\). The orientation of the amplicons/sequences is not indicated because a number of possible inversion events may take place, leading to different deletion patterns. Open and filled boxes indicate the absence or presence of a given marker or gene, respectively.
deletion; however, it has been found at high frequency in the general population of northern Europe in association with hgrN, indicating that hgrN may contain compensating Y linked factors.12,13 Our normospermic man with b2/b3 (or u3-gr/gr) deletion was hgrN3, whereas the two oligospermic men are hgrJ and hgrDE. Although the number is low, we can speculate that b2/b3 (u3-gr/gr; g1/g3) deletions are pathogenic only in association with certain haplogroups.

The frequency of gr/gr deletions was significantly higher in the infertile group with respect to controls, suggesting a possible deleterious effect of this polymorphism on spermatogenic efficiency. Similarly to the Repping study,7 the significant OR (10.1) indicates that gr/gr deletions can be considered as new risk factors for oligozoospermia. Although the mean sperm parameters (concentration, motility, and morphology) were not significantly different in the infertile group between men with gr/gr deletions and those without, there was a clear trend toward lower values in the former.

However, the most intriguing phenomenon observed in association with this type of deletion is the extreme heterogeneity of the phenotype, ranging from azoospermia to normozoospermia. In order to gain more insight into this phenomenon, we performed mutational analysis of DAZL, an autosomal homologue of DAZ. DAZ was copied to the Y chromosome relatively recently, in the old world primate lineage, and is 90% identical to its autosomal ancestor DAZL.22

We found no new mutations in the entire coding region of DAZL, and the polymorphic Thr12-Ala change (T12A) does not seem to have any modulating effect, probably because of its relatively high frequency in the normospermic group. However, as our mutational analysis was focused exclusively on the coding regions, we cannot exclude promoter variations that might affect the level of expression of DAZL.

Interestingly, deletion of the CDF1a copy was found only in the patient group, providing an even more significant difference (p<0.003) between the infertile and normospermic group. A similar phenomenon has also been observed in another study population from Mediterranean France.35 The different combinations of loci found deleted in cases of partial AZFc deletion indicate that a number of possible inversion events must have preceded these deletions.4,11

Clearly, the type of missing gene copies and the Y chromosome structure (hgr) on which the deletion arises are of fundamental importance for the understanding of a potential cause–effect relationship. Our data seem to support the hypothesis that DAZ1/DAZ2 copies are functionally more important than DAZ3/DAZ4,23 as the former were missing only in the patient group (p = 0.037); however, a recent study found no difference in DAZ1/DAZ2 deletion frequency between infertile men and fertile controls (with unknown sperm count), which appears to contradict this observation.11 On the other hand, our observation that deletion of CDF1a is strongly associated with infertility is consistent with the findings of this same study.33 This now requires further confirmation.

Although it is clear that the STS +/− analysis alone does not provide information about the type of missing gene copies and the hgr, we propose a flow chart (fig 1) in order to carry out partial AZFc deletion analysis in a cost effective and relatively simple manner. At this point, it is difficult to decide whether routine screening for partial AZFc deletions will be worthwhile. Clearly, gr/gr deletions, especially those resulting in the loss of CDF1a, can be considered a risk factor. However, the ideal situation would be to define a genetic profile specific for spermatogenic failure, although this may not exist.

Future studies in larger, well selected groups of subjects (patients without interfering abnormal andrological findings and normospermic controls) should focus on the combined definition of the type and copy number of the AZFc genes deleted in men with partial deletions and the haplogroup of the Y chromosome. This will probably provide a means of distinguishing between pathogenic and neutral (or compensated by other Y factors) deletion types. If we consider that, together with the gr/gr deletion, other Y related factors (protective or negative) will be transmitted to the male offspring, we can eventually propose screening for partial AZFc deletion prior to assisted reproductive techniques. However, until a clear definition of pathogenic and non-pathogenic deletions is established, the prediction of the testicular phenotype of the offspring will remain rather vague.

**ACKNOWLEDGEMENTS**

We are grateful to Dr M Rotondi for his valuable contribution to the statistical analysis of our data. We thank the clinical staff of the Andrology Unit and Reproductive Physiology for patient’s recruitment: Professor M Maggi, Dr G Corona, Dr M Magini, Dr L Petrone, Dr A Cilotti, Professor E Cecchi, Professor I Noci. The study was supported by a grant from the University of Florence.

Authors’ affiliations
C Giachini, E Guarducci, S Deg’Innocenti, L Becherini, G Forti, C Krausz, Andrology Unit, Department of Clinical Pathophysiology, Center for Research, Transfer and Higher Education on Chronic, Inflammatory, Degenerative and Neoplastic Disorders, University of Florence, Viale Pieraccini, 6, Florence 50139, Italy
G Longepeid, M J Mitchell, Inserm U.491, Faculté de médecine, 13385 Marseille, France; Institut de Médecine de la Reproduction, 13008 Marseille, France

Competing interests: none declared

The first two authors have contributed equally to the study and they should be regarded as joint first authors.

Correspondence to: Dr C Krausz, Andrology Unit, Viale Pieraccini, 6, Florence, 50139, Italy; c.krausz@afc.unifi.it

Received 13 October 2004

Revised 29 November 2004

Accepted 1 December 2004

**REFERENCES**


www.jmedgenet.com


CORRECTION

doi: 10.1136/jmg.2004.23960corr1

The Letter to JMG titled, Ancestral RET haplotype associated with Hirschsprung’s disease shows linkage disequilibrium at – 1249 (J Med Genet 2005;42:322–27) should have been published as a short report and the following abstract was omitted:

Background: Hirschsprung disease (HSCR) is a complex disorder with traditional germline mutations in RET in up to 30% of familial cases and in 3% of sporadic cases in a population-based series. We have previously demonstrated that an ancestral haplotype at the 5′ end of RET (haplotype 0) was strongly associated with a large subset of isolated HSCR cases and that a putative low penetration susceptibility locus was encompassed within this ancestral haplotype, anchored by exon 2 SNP A45A.

Objective: To determine the 5′ extent of the HSCR-associated ancestral haplotype by defining the linkage disequilibrium breakpoint in search for the low penetrance susceptibility locus.

Methods: Systematic screening of the region upstream of the anchoring A45A SNP, comprising RET intron 1, exon 1, and promoter sequences in 117 population-based HSCR cases and 100 controls. Dual luciferase assay to determine differential activities between SNP combinations near a transcriptional start site.

Results: New SNP’s were found which formed upstream haplotypes, anchored by A45A, in linkage disequilibrium with HSCR (X2 = 76.96, p<0.00000001). Linkage disequilibrium appeared to break at the –1249C/T SNP. Further, the HSCR-associated genotype (00) was found in >60% of HSCR but only 2% of controls. Because only 2 variants, -200A>G and -196C>A, lie within the promoter region and are in proximity to the transcriptional start site (at –195), we modelled these combinations into constructs for luciferase reporter assay. The HSCR-associated SNP combination showed the lowest activity and the control-associated combination, the highest.

Conclusions: Our observations seem to discard the existence of a HSCR-causing mutation as it is conceived in the traditional sense, but strengthen the idea of a specific combination of variants conferring susceptibility to the disease in a low penetrance fashion. The data derived from our functional “in vitro” studies would suggest that the HSCR-associated haplotype 0 may result in a lower level of expression of the RET gene.

The journal apologises for this error.