Germline CDH1 deletions in hereditary diffuse gastric cancer families

Carla Oliveira1,2, Janine Senz3, Pardeep Kaurah3, Hugo Pinheiro1, Remo Sanges4, Anne Haegert5, Giovanni Corso6,7, Jan Schouten9, Rebecca Fitzgerald9, Holger Vogelsang10, Gisela Keller11, Sarah Dwerryhouse12, Donna Grimmer13, Suet-Feung Chin13, Han-Kwang Yang14, Charles E. Jackson15,16, Raquel Seruca1,2, Franco Roviello6,7, Elia Stupka4, Carlos Caldas12,13 and David Huntsman3,*

1Institute of Molecular Pathology and Immunology, University of Porto (IPATIMUP), 4200-465 Porto, Portugal, 2Faculty of Medicine, University of Porto, 4200-319 Porto, Portugal, 3Hereditary Cancer Program and the Centre for Translational and Applied Genomics, British Columbia Cancer Agency, Vancouver, Canada V5Z 4E6, 4CBM S.c.r.l., AREA Science Park, Basovizza 34012, Trieste, Italy, 5Prostate Centre Microarray Facility, Vancouver, Canada V6H 3Z6, 6Department of Human Pathology and Oncology, Section of Surgical Oncology, Translational Research Laboratory, Istituto Toscano Tumori (ITT), University of Siena, 53100 Siena, Italy, 7Division of Surgical Oncology, University of Siena, 53100 Siena, Italy, 8MRC Holland, 1057 DN Amsterdam, The Netherlands, 9MRC Cancer Cell Unit, Hutchison/MRC Research Centre, Cambridge CB2 0XZ, UK, 10Department of Surgery and 11Institute of Pathology, Technische Universität München, D-81675 Munich, Germany, 12Department of Oncology, University of Cambridge, Cambridge CB2 0QQ, UK, 13Cancer Research UK Cambridge Research Institute, Li Ka Shing Centre, Cambridge CB2 0RE, UK, 14Department of Surgery and Cancer Research Institute, Seoul National University College of Medicine, Seoul 151-742, South Korea, 15Department of Surgery and 16Department of Medical Genetics, Henry Ford Hospital, Detroit, MI 48202, USA

Received December 23, 2008; Revised and Accepted January 22, 2009

Germline CDH1 point or small frameshift mutations can be identified in 30–50% of hereditary diffuse gastric cancer (HDGC) families. We hypothesized that CDH1 genomic rearrangements would be found in HDGC and identified 160 families with either two gastric cancers in first-degree relatives and with at least one diffuse gastric cancer (DGC) diagnosed before age 50, or three or more DGC in close relatives diagnosed at any age. Sixty-seven carried germline CDH1 point or small frameshift mutations. We screened germline DNA from the 93 mutation negative probands for large genomic rearrangements by Multiplex Ligation-Dependent Probe Amplification. Potential deletions were validated by RT–PCR and breakpoints cloned using a combination of oligo-CGH-arrays and long-range-PCR. In-silico analysis of the CDH1 locus was used to determine a potential mechanism for these rearrangements. Six of 93 (6.5%) previously described mutation negative HDGC probands, from low GC incidence populations (UK and North America), carried genomic deletions (UK and North America). Two families carried an identical deletion spanning 193 593 bp, encompassing the full CDH3 sequence and CDH1 exons 1 and 2. Other deletions affecting exons 1, 2, 15 and/or 16 were identified. The statistically significant over-representation of Alus around breakpoints indicates it as a likely mechanism for these deletions. When all mutations and deletions are considered, the overall frequency of CDH1 alterations in HDGC is ~46% (73/160). CDH1 large deletions occur in 4% of HDGC families by mechanisms involving mainly non-allelic homologous recombination in Alu repeat sequences. As the finding of pathogenic CDH1 mutations is useful for management of HDGC families, screening for deletions should be offered to at-risk families.

*To whom correspondence should be addressed. Tel: +1 6048776000 [2148]; Fax: +1 6048776178; Email: dhuntsma@bccancer.bc.ca

© 2009 The Author(s). This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
INTRODUCTION

Hereditary diffuse gastric cancer (HDGC) is an autosomal dominant disorder that accounts for <1% of all cases of gastric cancer (GC). Although uncommon, this disease constitutes an important health issue due to its severity, its high penetrance, early age at presentation and the unavailability of effective screening tools. Diffuse gastric cancer (DGC) is the most important cause of cancer lethality in mutation positive HDGC families with lobular breast cancer being a secondary medical concern (1).

Heterozygous germline point or small frameshift mutations in E-cadherin gene (CDH1) [OMIM +192090], a calcium-dependent cell-to-cell adhesion molecule and a tumor suppressor protein, are the only germline molecular defect associated with HDGC (2–4). Overall, carriers of CDH1 germline mutations have a cumulative GC risk, before age 75, of 40–67% for men and 63–83% for women and a risk for lobular breast cancer of 39–52% (5,6). Clinical expression of the disease often occurs before the third decade of life (1), but preclinical multifocal tumors, which may have the potential for metastatic spread, are already present in the stomachs of young asymptomatic CDH1 mutation carriers (7,8).

Clinical criteria have been established, in 1999, by the International Gastric Cancer Linkage Consortium (IGCLC) to select families for mutation screening in the CDH1 gene, based mainly on early age of disease presentation and diffuse histology of the tumors (9). These have been later adapted in an attempt to enlarge the population to which this screening could apply, in countries with low incidence of GC (10). Overall, 30–40% of HDGC families harbor CDH1 germline mutations, but this frequency is highly variable between countries with different incidences of GC, with low incidence countries like UK, North America and Canada, displaying nearly 50% of HDGC mutation positive families, while high incidence countries like Portugal and Italy, displaying around 10% of HDGC mutation positive families (4,6,11).

Approximately two-thirds of families, screened worldwide, remain genetically unexplained and the concern related with the management of CDH1 mutation negative gastric cancer families persists. Since the required samples for full genome linkage, in any GC family, are difficult to obtain, a candidate gene approach to identify novel susceptibility genes has been based mainly on early age of disease presentation and diffuse histology of the tumors (9). These have been later adapted in an attempt to enlarge the population to which this screening could apply, in countries with low incidence of GC (10). Overall, 30–40% of HDGC families harbor CDH1 germline mutations, but this frequency is highly variable between countries with different incidences of GC, with low incidence countries like UK, North America and Canada, displaying nearly 50% of HDGC mutation positive families, while high incidence countries like Portugal and Italy, displaying around 10% of HDGC mutation positive families (4,6,11).

RESULTS

Multiplex ligation-dependent probe amplification analysis for the CDH1 gene

We collected, from five reference centers, 93 DNA samples from families with aggregation of DGC and negative for CDH1 germline point mutations. These families were selected based on their risk to carry a germline alteration in CDH1 and the clinical criteria described in Materials and Methods section.

We found that six out of the 93 (6.5%) probands showed abnormal multiplex ligation-dependent probe amplification (MLPA) features when compared with the controls, displaying an ~70% signal reduction in one or more exons of the CDH1 gene and, suggesting the presence of large deletions affecting the CDH1 locus (Fig. 1, Table 1). No abnormalities were observed in the remaining 87 probands.

After reproducing these results in an independent MLPA reaction, we verified that four probands (families 1 to 4) presented deletions at the 5’-end of the gene: in three probands deletions encompassed at least exon 1, intron 1 and exon 2 and, in one proband a smaller deletion that encompassed one of the three probes designed for CDH1 exon 1 was found. The remaining two probands (families 5 and 6) displayed deletions at the further 3’ end of the gene: in one proband a deletion of the genomic sequence encompassing exons 14 and 16 was observed, an in the other a deletion of exon 16 only was observed (Fig. 1).

Fine mapping of the deletions breakpoints

A combination of array CGH and PCR was utilized to map the putative breakpoints in five of the six probands (Fig. 2). The array CGH indicated the regions of chromosome 16 at which the breakpoints were located. PCR primers were then designed to span the breakpoints and the resulting PCR products sequenced (Table 2). Family 1 and family 2 showed identical MLPA patterns as well as virtually identical array CGH results. Sequencing of the breakpoint revealed that these two families harbored an identical 193 593 bp deletion encompassing the full sequence of CDH3 and extending to
position IVS2+57 595 in CDH1. Family 3 was found to carry a 5671 bp deletion with breakpoints at position -3809 upstream of the CDH1 transcription start site (TSS) and at position IVS2+742 in CDH1. Family 4 harbored a 150 bp deletion encompassing the TSS of CDH1, with breakpoints located 125 bp upstream and 25 bp downstream of the TSS. Array CGH was not done for this proband since the PCR encompassing CDH1 exon 1 produced both a normal sized and a shorter band that was visible by dHPLC and further used to clone the breakpoints for the deletion by direct sequencing. Family 5 carried an 8078 bp deletion ranging from IVS13-2738 to 20 bp downstream of the stop codon of CDH1, as confirmed by long-range PCR using specific primers flanking the breakpoints (Fig. 3A). Family 6 was found to carry an 828 bp deletion ranging from IVS15+3097 to 223 bp downstream of the stop codon of CDH1 (Fig. 3B).

Bioinformatics analysis of breakpoints and deletion containing sequences

Assuming that deletions are more likely to occur due to recombination events, we analyzed the genomic sequences flanking the boundaries of these five deletions, which are predicted to constitute target regions for breaks and recombinations to occur. We performed a search for Alu elements which have been reported to mediate non-allelic recombination events in several other malignancies. Evidence for non-homologous end joining (NHEJ) and non-allelic homologous recombination (NAHR) has been described in atypical NF1 microdeletions. These deletions were traced back to short sequences with similarities to Alu elements near the deletion junction (22). Moreover, deletions in the PMP22 gene have been described to be a product of homologous recombination between AluS elements present in intronic regions of the

Figure 1. MLPA output from the six HDGC families carrying germline deletions of the CDH1 locus. The first 17 bars represent signal obtained with MLPA probes within the CDH1 gene. Please note that for exon 1, three probes were used (1a, 1b and 1c). CDH1 deletions are represented, in the graph, by smaller bars marked with arrows. Bars on the right hand-side represent results obtained with control probes for MLPA (n = 15).
We also performed a search for the presence of cryptic sequences recognizable by the RAG1/2 recombinase near the breakpoints and the insertion of P and/or N nucleotides in the rejoining sites. Such sequence features, reminiscent of V(D)J recombination sites, are expected to indicate whether cleavage by the RAG1/2 recombinase may occur near these joining sites.

In three families carrying two different deletions we found that deletions overlap Alu repeats (Table 3). In families 1 and 2, bearing the same deletion at chr16:67193822–67387415, both deletion boundaries fall within annotated Alu repeats, AluSp near the start and AluSg near the end. The sequences of the two repeats are highly similar (percentage identity >75%) (Table 3, Fig. 4A). In family 3, carrying a deletion at chr16:67324886–67330557, the start of the deletion maps inside an AluSx repeat while the end map into a repeat AluSg and, the sequences of these two repeats display an identity of ~80% (Table 3, Fig. 4B).

For the deletion at chr16:67328695–67328844, present in family 4, an AluJo repeat is present 304 bp upstream from the start of the deletion, and another AluJo is present 1497 bp downstream the end of the deletion. An interesting feature of this 150 bp deletion is the presence of microsinteny/microhomology around the breakpoints due to the presence of five nucleotides sequence (CTCTC) between the upstream and downstream breakpoints.

Family 5 displays a deletion at chr16:67416845–67424923 that does not overlap Alu repeats, although it occurs in the proximity of this type of repetitive regions (Table 3). The closest repeats both upstream and downstream are AT rich sequences, ~300 bp distant from the deletion (Fig. 4C). Moreover, 344 bp upstream of the deletion start a FLAM_C repeat.

LBC, lobular breast cancer; PCa, pancreatic cancer; LCa, liver cancer.
exists and 442 bp downstream of the endpoint there is an AluJb sequence (Table 3). The alignment of the $-150/150$ bp surrounding the deletion ends did not show any identity.

The deletion in family 6, at chr16:67424298–67425126, is unlikely to have been produced by homology-dependent mechanisms, since a small (3 bp) sequence insertion at the junction between the deletion breakpoints was identified (Fig. 3B). For this deletion, the closest Alu repeats found on both sides of the deletion are an AluSq 50 bp upstream and an AluJb 239 bp downstream of the deletion’s start and end, which share a sequence identity $>77\%$ (Table 3).

In a range of $+/-5$ Mb flanking the five deletions reported in this study, we did not find evidence for the presence of RAG1/2 consensus sequences accompanied or not by the insertion of P and/or N nucleotides in the rejoining sites. Nevertheless, the presence of Alus overlapping 6 out of the 12 points analyzed (the ends of the six deletion sequences) is statistically significant ($z$-score 3.3, $P$-value 0.0005) as tested by performing a randomization analysis with 1000 random sampling of 12 random locations (see Materials and Methods).

Clinical features of deletion-carrying HDGC families

The proband in family 1 was a 61-year-old female of European Irish/German/Scottish/Danish ancestry presenting with lobular

---

**Table 3.** Deletion coordinates and distance from deletion ends to the closest Alu containing sequences

<table>
<thead>
<tr>
<th>Family ID</th>
<th>Deletion coordinates</th>
<th>Closest start Alu</th>
<th>Closest start Alu distance</th>
<th>Closest end Alu</th>
<th>Closest end Alu distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family 1</td>
<td>chr16:67193822–67387415</td>
<td>AluSp overlap</td>
<td>overlap</td>
<td>AluSg overlap</td>
<td>overlap</td>
</tr>
<tr>
<td>Family 2</td>
<td>chr16:67193822–67387415</td>
<td>AluSp overlap</td>
<td>overlap</td>
<td>AluSg overlap</td>
<td>overlap</td>
</tr>
<tr>
<td>Family 3</td>
<td>chr16:67324886–67330557</td>
<td>Alu5x overlap</td>
<td>overlap</td>
<td>AluSg overlap</td>
<td>overlap</td>
</tr>
<tr>
<td>Family 4</td>
<td>chr16:67328695–67328844</td>
<td>AluJo overlap</td>
<td>overlap</td>
<td>AluJo overlap</td>
<td>1497</td>
</tr>
<tr>
<td>Family 5</td>
<td>chr16:67416845–67424923</td>
<td>FLAM_C overlap</td>
<td>overlap</td>
<td>AluJb overlap</td>
<td>442</td>
</tr>
<tr>
<td>Family 6</td>
<td>chr16:67424298–67425126</td>
<td>AluSq overlap</td>
<td>overlap</td>
<td>AluJb overlap</td>
<td>239</td>
</tr>
</tbody>
</table>
breast cancer. Her daughter and son, as well as a granddaughter developed DGC at the age of 40, 37 and 28, respectively, and are already diseased. The deletion was only looked for in the proband that is currently alive.

In family 2, the proband, of unknown origin but currently living in Canada, was a female who developed DGC at 38 and later at 43 years old and the only patient tested in this family. Her two sisters also developed DGC at the ages of 30 and 35, respectively, but were not tested for the presence of the same deletion.

The presence of the same deletion in two different families could be explained two ways. Either of these families had a common ancestor or the deletions arose independently. To clarify this issue, we performed haplotype analysis with seven microsatellite markers surrounding the CDH1 gene and verified that an allele for each of the microsatellite markers was common between the two families, demonstrating the existence of a common ancestor carrying this deletion (Fig. 5).

The proband in family 3, was a 34-year-old female of European Lithuanian origin, with DGC. Her mother was also affected by the same disease at uncertain age, and her grandfather developed pancreatic cancer at the age of 43 and none of them was tested for the deletion. The proband in family 4, was a 40-year-old male of Hispanic origin who developed DGC at the age of 40. His 24-year-old brother died from exactly the same disease and his father died at 42 with a pancreatic cancer, neither has been tested for the deletion. Carrier tests have been performed in ten other family members and one of which was positive.

In family 5, the proband was a patient of European ancestry, who developed DGC at 33 years. His two sisters developed DGC at 33 and 35 years. At least one second degree relative developed GC, but no more information was available. So far, only the proband was tested, but five other family members are currently being tested for the presence of the deletion.

As for family 6, the proband was a 51-year-old patient of European ancestry who developed DGC. His mother died of GC of unconfirmed histology at age of 52. Two sisters were recently diagnosed with DGC by surveillance endoscopy, although only one of them was tested for the deletion and proved to be a carrier. This sister’s offspring, constituted of two asymptomatic individuals, also tested positive for the deletion. Four other second degree relatives died of cancer, one with DGC, two of GC of unknown histology and one of lobular breast cancer.

Clinical criteria, frequency of CDH1 alterations and GC incidence

In the present study, we analyzed 93 families based on two clinical criteria: (i) three or more DGC in first degree relatives diagnosed at any age or; (ii) two or more GC in first degree relatives with at least one DGC diagnosed before age 50;
and found that 6.5% of these families carry large \textit{CDH1} germ-line deletions.

With the purpose of clarifying the relative contribution of germline \textit{CDH1} point mutations and deletions in a large series of families fulfilling the aforementioned criteria, we collected data not only on the 93 negative families, herein analyzed for genomic rearrangements, but also on 67 HDGC families carrying \textit{CDH1} point mutations, from five different reference centers, and originated from countries with low, moderate and high GC incidence rates.

From the series of 160 probands, 73 (45.6%) were proved to carry germline \textit{CDH1} alterations: 41.9% point mutations and 3.8% large deletions (Fig. 6).

In low incidence countries namely in the North America, Canada, UK and Holland, the frequency of germline \textit{CDH1} alterations was 51.6%, while in moderate (Germany) and high (Portugal and Italy) incidence countries the frequency of alterations was 25 and 22.2%, respectively (Fig. 6). Large deletions were only identified in countries with low incidence of GC and account for 9% of all \textit{CDH1} alterations in HDGC families from such countries (Table 1).

\textbf{DISCUSSION}

In this report, we demonstrate that germline large deletions affecting the \textit{CDH1} locus are one of the mechanisms underlying HDGC. Our data supports the addition of MLPA based screening for such deletions to clinical screening for \textit{CDH1} mutations.

The \textit{CDH1} locus, like the \textit{hMLH1}, \textit{hMSH2}, \textit{BRCA1}, \textit{APC} and \textit{CDKN2A} loci, is a region susceptible to genomic rearrangements leading to increased cancer susceptibility (18,19,21,26).

The analysis of deletions found in our HDGC patients, revealed that two out of five deletions showed extensive sequence identity between the two breakpoints, strongly suggesting that NAHR was the mechanism of their formation (Families 1, 2 and 3). The deletion breakpoints of these two deletions fall within Alu repeats similar to what has been previously reported for \textit{hMLH1}, \textit{hMSH2}, \textit{BRCA1} and infrequently for \textit{APC} (18,19,26). Alu repetitive sequences have long been believed to give rise to genomic deletions promoting recombinational instability (27). PMP22 deletions have been described to be a product of homologous recombination between intronic AluS elements in hereditary neuropathies (23), but the type of Alu sequences involved in deleterious deletions underlying cancer-associated syndromes have not been systematically explored in the literature. Nevertheless, at least one of the Alu sequences located at large deletion breakpoints in \textit{hMLH1} gene are described to be of the AluSx type (28). In our study, we analyzed the type of Alu sequences associated with \textit{CDH1} large deletions in HDGC, and found that six breakpoints overlap AluSx elements: three overlap AluSg, two overlap AluSp and one overlaps an AluSx (Table 3).

One of the deletions herein identified occurs in a context of microhomology and could have arisen by microhomology-mediated recombination, a mechanism of NHEJ. This mechanism is associated with very short stretches of sequence identity (a few bp) between the two ends of the breakpoint junctions (29,30), rather than with long stretches of sequence identity at these locations (31,32). Moreover, it involves the double strand breakage of DNA followed by end joining in the absence of extensive sequence homology. A similar microhomology-mediated intra-strand repair mechanism has been previously described to occur either in pathological conditions such as trisomy in a family displaying an inverted duplication/deletion of 2p25.1–25.3 (33) or in recombination events underlying genomic variation during evolution (34).
with specific chemicals or radiation during late spermatogenic stages, generates very short deletions, duplications, or inversions with microhomology around breakpoints. The ligation by NHEJ mediated by microhomology, is raised in the aforementioned report, as the major path of repair (35), and provides insight on how this specific CDH1 deletion may have initially arisen.

Another of our deletions is unlikely to have been produced by homology-dependent mechanisms, since a small sequence insertion (3 bp) at the junction between the deletion breakpoints was identified. Since no extensive homology exists in the vicinity of breakpoints, it is likely that this deletion has been mediated by NHEJ. Identical deletions involving small insertions (3, 4, 6 bp), of unknown origin at the site of the breakpoints, have been previously identified within copy number variants (CNVs) breakpoint-region sequences, in normal population (32,34).

In contrast to all other CDH1 deletions described so far in this report, one does not seem to be caused by NAHR, microhomology-mediated recombination or small sequence insertions at the junction site. This deletion is flanked by upstream and downstream low complexity AT rich sequences, as well as FLAM_C repeat and an AluJb sequence. Similar features surrounding deletions have been previously described to occur in CNVs described in apparently normal individuals (34), but the mechanisms involved in their generation is not known. The AT rich sequences could mediate a microhomology-mediated deletion mechanism based on their base composition bias, though further evidence would be needed to prove this.

Deletions found in HDGC families mainly fall into three separate classes: products of NAHR, in which recombination between Alu containing sequences leads to deletion of the sequence between them; products of microhomology-mediated recombination, a mechanism of NHEJ; and results of small insertions of unknown origin at the site of the breakpoint junction. These deleterious deletions in CDH1 appear as the pathogenic counterpart of apparently normal evolution generated CNVs (32,34).

The probands studied (n = 160) fulfilled clinically used criteria for HDGC (see Materials and Methods). We found that 45.6% (73/160) of them displayed CDH1 alterations (Fig. 6). In comparison, only 7/123 (5.7%) of the probands of families who did not meet the above testing criteria, had point mutations or small frameshift mutations, and none had deletions detectable through MLPA (data not shown). Based on this analysis, we have demonstrated that the clinical criteria herein used (three or more DGC in first degree relatives diagnosed at any age or two or more GC in first degree relatives with at least one DGC diagnosed before age 50) will identify most families with HDGC associated germline CDH1 alterations, with a mutation pick up rate of ~46%. Based on these criteria, ~4% of HDGC patients display large CDH1 germline deletions (Fig. 6).

Two of the families shared the same large deletion that included the whole of the CDH3 gene and the first two exons of CDH1. Individuals with homozygous germline truncating mutations in the CDH3 gene present with hypertrichosis with juvenile macular dystrophy (HJMD) syndrome (36,37). Heterozygous carriers of such mutations are phenotypically normal and have not been reported to have increased GC risk. Haplotype analysis supports a shared ancestor as the source of the (chr16:67193822–67387415) in these two families. We have previously used this approach to demonstrate that recurrent HDGC associated CDH1 germline point mutations can arise from a common ancestor (6).

In conclusion, 6.5% of probands from strictly defined HDGC families who previously tested negative for CDH1 mutations carried heterozygous germline deletions. As these can be readily identified using MLPA, an assay frequently used in clinical cancer genetics laboratories, we suggest that MLPA be added to the testing strategy for all families who meet clinical criteria for HDGC.

**MATERIALS AND METHODS**

**Patients and families**

The collection of clinical data and peripheral blood samples was approved by the appropriate Ethics Committee from each of the centers participating in this work: University of British Columbia, Vancouver, Canada; Institute of Molecular Pathology and Immunology, University of Porto, Porto, Portugal; Department of Human Pathology and Oncology, Section of Surgical Oncology, Translational Research Laboratory, University of Siena, Italy; Institute of Pathology, Technische Universität München, Munich, Germany; Department of Oncology, University of Cambridge, Cambridge, UK. A series of 160 probands from families with aggregation of DGC was analyzed in this study, based on the following clinical criteria: (i) three or more DGC in first degree relatives diagnosed at any age or; (ii) two or more GC in first degree relatives with at least one DGC diagnosed before age 50 (see Table 1). From these 160 probands, 67 showed point mutations in the CDH1 gene and were used only for analysis of general frequency of CDH1 mutations within HDGC families. The majority of these families have been published elsewhere (3,6,10–13,38–46) except for 15 families from University of British Columbia, Vancouver, Canada (unpublished data DGH), six families from Department of Oncology, University of Cambridge, Cambridge (unpublished data CC) and two families from Technische Universität München, Munich, Germany (unpublished data GK). The remaining 93 probands belonged to families that tested negative for CDH1 germline point mutations. These were analyzed for CDH1 genomic rearrangements.

Peripheral blood leukocytes from the 93 negative probands were referred for analysis at the different centers. High molecular weight genomic DNA was isolated from all patients using standard methodologies.

**Multiplex ligation-dependent probe amplification**

A series of 93 probands, negative for CDH1 germline point mutations, were tested for the presence of large genomic alterations at the CDH1 locus using MLPA. MLPA is a PCR-based technique for identifying gene dosage alterations, which has been previously described in detail (47). All reagents were provided by the manufacturer in a kit (SALSA P085 CDH1 MLA
kit; MRC-Holland, Amsterdam, Holland), and testing was performed according to the manufacturer’s recommendations. The assay contains probes for all but one of the CDH1 exons. For technical reasons exon 8 has been excluded and exon 1 has three different probes. The assay was done according to manufacturer’s instructions. Briefly, the probes were hybridized overnight at 60°C with 100 ng of germline DNA and subsequently ligated. Ten microliters of the ligation reaction was used in the PCR which contained FAM labeled primers. The PCR products were analyzed on an ABI 3130 x L Genetic Analyzer. The output for each patient displayed a peak that corresponded to the amount of amplified ligated probe present for each exon of the CDH1 gene, along with several peaks representative of extragenic regions used for ligation and PCR monitoring controls. Peak heights for fragments corresponding to specific exons and control regions were binned, appended to a table, and then saved as a text file, all within the Applied Biosystems Genotyper software. The resulting data were analyzed using the Cofalyzer software downloaded from MRC-Holland. Any exon with a decrease or increase in peak height of 70% was scored as a deletion or duplication, respectively. MLPA assays which showed a variation between peak ratios were repeated.

**Comparative genomic hybridization microarray**

A CGH array strategy was done to confirm the deletions identified by MLPA and to identify the location of deletion breakpoints. Custom CGH arrays were designed using Agilent’s eArray program. An 8×15K array format was generated with Agilent HD CGH Database Catalog Probes for coverage of chromosome 16: 066880834–067879823. Genomic DNA (1.5 µg) was labeled using the Agilent Genomic DNA Labeling Kit PLUS (Agilent Technologies, Santa Clara, CA). Samples labeled with Cy3 were co-hybridized with Cy5 labeled gender-matched normal reference gDNA (Promega, Madison WI) to custom 8×15K Agilent human CGH microarrays. Microarrays were washed with Agilent CGH wash buffers and scanned on an Agilent DNA Microarray Scanner. Images were analyzed using Agilent Feature Extraction (version 8.5) and data were analyzed using Agilent CGH Analytics (version 3.4.27).

**Breakpoint identification**

Based on the information obtained from the CGH array, primers were designed spanning the putative breakpoints for each case and used in long-range PCR. A table with primer sequences is supplied as supplementary material (Supplementary Material, Table S1).

**Haplotype type, Table**

Analysis was performed using the following microsatellite markers surrounding the CDH1 gene: D16S318, D16S3107, D16S3025, D16S496, D16S3067, D16S3095 and D16S752. Each marker was amplified and then analyzed on an ABI 3130 x L Genetic Analyzer followed by analysis using Gene mapper V7.0 software. Primer sequences for the microsatellite markers D16S3025, D16S496, D16S3095 and D16S752 have been previously reported (11). Sequences for the remaining markers, D16S318, D16S3107 and D16S3067 are listed in the genome database (http://www.gdb.org).

**Analysis of breakpoints sequence context**

Breakpoints were defined as a set of coordinates on the genome spanning the genomic sequence of the deletions. The coordinates have been loaded on the UCSC genome browser as a BED file accessible on http://biodev.cbm.fvg.it/remo/DEL.txt. Both automatic and manual bioinformatics analysis have been carried out in order to analyze the genomic context of these regions.

For each deletion we analyzed the closest repeats in a range of 1500 bp flanking the start/end points. Repeats were obtained from the Ensembl Homo sapiens core database version 50 and UCSC genome browser Human Mar. 2006 Assembly. Sequences in a range of −150/+50 bp flanking the start-point and −50/+150 bp flanking the end-point of the deletions have been collected in order to test the similarity around the breakpoints by using both clustalw and blast2 with default parameters. An interval of 5 Mb around the deletion start/end points has been analyzed in order to look for the presence of the RAG1 binding site on both strands by using a custom perl script and known consensus sequences (CACTGTG, [11,13]GGTTTTTGT, CACAGTG, ACAAAGAC [11,13]CACAGTG, ACAATAAC [11,13]CAACAGTG, [22,24]ACAGAAC [22,24]AACAGT, [22,24]CACAGTG, [22,24]ACAGAAC). In order to perform randomization analysis we collected 150 000 random points on the human genome and calculated for each of them the distance from the closest Alu. The distribution of the distances has been trimmed in order to avoid considering points in poorly annotated regions and/or in gene deserts and/or in genomic gaps, which resulted nonetheless to be very far from Alus. We trimmed out all the points retaining a distance bigger than 4519 bp (the 1.5*IQR of the distribution of all the distances) from the closest Alu.

**SUPPLEMENTARY MATERIAL**

Supplementary Material is available at HMG online.

**ACKNOWLEDGEMENTS**

We gratefully acknowledge patients, families and their caregivers for their willing participation in this research project and who provided consent regarding use of the information obtained from the study.

**Conflict of Interest statement.** None declared.

**FUNDING**

This work was supported by the following grants: The National Cancer Institute of Canada (Canadian Cancer Society); The Michael Smith Foundation for Health Research to D.H.; The Portuguese Foundation for Science and Technology—FCT (grant numbers POCTI/SAU-OBS/58111/2004,


