Identification of an intronic single nucleotide polymorphism leading to allele dropout during validation of a *CDH1* sequencing assay: implications for designing polymerase chain reaction-based assays

Franklin M. Mullins, MD, PhD[1], Lisa Dietz, BS[1], Marla Lay, BS, MT(ASCP)[1], James L. Zehnder, MD[1], James Ford, MD[2,3], Nicolette Chun, MS[2], and Iris Schrijver, MD[1,3]

Purpose: The *CDH1* gene encodes the cell adhesion protein E-cadherin, and *CDH1* germline mutations are associated with hereditary diffuse gastric cancer. Identification of individuals at high risk of developing diffuse gastric cancer affords the opportunity for endoscopic screening or elective prophylactic gastrectomy. We set out to develop a *CDH1* sequencing assay for clinical use. **Methods:** All exons of the *CDH1* gene were amplified and sequenced with published and modified primers. **Results:** While validating the assay, we encountered a case in which a single nucleotide polymorphism located in intron 15 led to allele dropout and therefore to a false-negative result. The polymorphism leading to allele dropout was located within a primer-binding sequence, five bases away from the 3′ end of the primer. A frameshift mutation in exon 15 was detected by an alternative primer that binds away from the polymorphic site. A search of the University of California Santa Cruz single nucleotide polymorphism database revealed other polymorphisms located within primer-binding sites. A total of 12 primers in nine primer sets were modified to minimize allele dropout risk. **Conclusion:** The approach of designing primers to avoid known single nucleotide polymorphisms can be generalized to the design of any polymerase chain reaction-based assay and should be employed whenever possible. *Genet Med* 2007:9(11):752–760.

Key Words: *CDH1, allele dropout, gastric cancer, SNP, HDGC*

The cadherin 1 (*CDH1*) gene encodes the cell adhesion protein E-cadherin. *CDH1* germline mutations were first linked to hereditary diffuse gastric cancer (HDGC) in three Maori kindreds.[1] Since that time, E-cadherin germline mutations have been identified in families of various ethnic backgrounds exhibiting an elevated susceptibility to diffuse gastric cancer and lobular breast cancer. The inheritance pattern is autosomal dominant with incomplete penetrance. Mutations are associated with loss of function of the mutated E-cadherin allele. Nonsense mutations, missense mutations, deletions and insertions leading to frameshifts, and splice site mutations have all been described.[2,3]

Gastric cancers are broadly divided into intestinal and diffuse types. Overall, about 1–3% of gastric cancers occur in families with an autosomal dominant pattern of gastric cancer susceptibility. Approximately 30% of families with HDGC carry a mutation in the *CDH1* gene on chromosome 16q22.1. Current data suggest a lifetime risk of >70% for diffuse gastric cancer and up to 40% for lobular breast cancer in individuals who carry a germline *CDH1* mutation leading to loss of E-cadherin function. By the time gastric cancer becomes symptomatic, it is rarely curable with a 3-year survival of <30%. However, a high cure rate (>85% 5-year survival) is possible if the stomach is removed before tumor invasion through the gastric wall. Identification of individuals at high risk of developing diffuse gastric cancer, then, affords the opportunity for intensive early endoscopic screening, or even elective prophylactic gastrectomy.[4–6]

Germline *CDH1* mutations associated with HDGC are distributed throughout the 16 exons of the gene, and, in general, do not appear to recur in unrelated families. Therefore, the most appropriate way to identify *CDH1* mutations in families predisposed to gastric cancer is direct DNA sequencing of all 16 exons and their flanking regions. For families in which a pathogenic *CDH1* mutation has been identified, targeted sequencing of the affected exon may be appropriate. For nonsense mutations, frameshifts, and mutations affecting consensus splice

sites, the inference that a novel *CDH1* mutation is pathogenic can be relatively straightforward. However, evaluation of the significance of missense mutations may be more challenging. In the research context, cell aggregation and collagen invasion assays have been used to demonstrate that particular missense mutations lead to loss of E-cadherin function. SIFT (sorting intolerant from tolerant) or Poly Phen (polymorphism phenotyping) analysis based on evolutionary conservation may play a complementary role,[2,7,8] together with family studies and the analysis of unaffected control individuals.

The term "allele dropout" refers to the failure of a genotyping method to detect one of a patient's two germline alleles. For polymerase chain reaction (PCR)-based methods, one cause of allele dropout is a polymorphism within one of the primer-binding sites. In particular, the 3′ end of the primer must perfectly match the sequence chosen to be amplified to generate a PCR product. This feature of primer specificity has been exploited in the amplification refractory mutation system technique, which utilizes primers with 3′ ends designed to perfectly match mutations to generate products of known sizes corresponding to mutant alleles.[9] Failure to amplify both alleles is easily recognized in most PCR-based assays. However, if the two alleles differ from one another in the region of interest, dropout of a single allele may lead to incorrect assignment of homozygous wild-type or mutant status. For dominant mutations such as those described for *CDH1*, the primary danger of allele dropout is a false-negative result caused by amplification failure of a mutant allele. Depending on the analysis method, dropout of a wild-type allele may also lead to a false-negative result for an individual with a dominant mutation. For example, a partially denaturing high-performance liquid chromatography (HPLC) approach to mutation detection may fail when either the wild-type or the mutant allele drops out, if the patient specimen is analyzed in isolation. However, performing the analysis on the patient sample mixed with a reference sample should allow mutation detection in those cases in which the wild-type allele drops out. For recessive mutations, the failure to amplify a wild-type allele may lead to a false-positive result because the data may suggest homozygosity for the mutation. Conversely, allele dropout of a mutant allele may result in the identification of only one mutation when two are, in fact, present in the gene, thus precluding definitive diagnosis of the recessive condition. Finally, for some assay designs, allele dropout may also result in failure to recognize a hemizygous mutation. Of note, evaluation of possible intronic single nucleotide polymorphism (SNP)-primer mismatches was recently applied to testing for congenital long QT syndrome; exons susceptible to allelic dropout were identified, and disease-causing mutations were identified in four previously genotype-negative index cases.[10]

Here, we report a case of allele dropout leading to a false-negative result encountered during validation of a sequencing assay for the *CDH1* gene. The SNP leading to allele dropout was located five bases away from the target of the 3′ end of the primer. This finding prompted us to search SNP databases for other variants located within primer-binding sites, and a total of nine primer sets were ultimately modified so as to reduce the potential of allele dropout as much as possible. This approach can be applied to the design of any PCR-based assay.

MATERIALS AND METHODS

Specimens

Initially, specimens from four positive control patients previously tested on a research basis and found to be positive for *CDH1* mutations were obtained via the Stanford Cancer Genetics Clinic. A frozen tissue specimen from the patient whose sample demonstrated allele dropout and a peripheral blood specimen from an affected sibling were also tested to rule out a specimen mix-up after the initial unexpectedly negative results. A peripheral blood specimen from a negative control individual was also analyzed. All samples were collected on institutional review board-approved protocols.

DNA extraction

Peripheral blood was collected in ethylenediamine tetraacetic acid tubes. Genomic DNA was extracted and purified on a Qiagen spin column (QIAamp Blood Kit, Qiagen, Inc., Valencia, CA). DNA was extracted from a fresh tissue specimen by proteinase K digestion followed by a standardized salting out procedure (Puregene DNA Isolation, Gentra, Minneapolis, MN).

PCR and sequencing

The initial primer set employed was as described by Brooks-Wilson et al.,[2] with primers tagged with −21M13F or M13R tails at the 5′ end to facilitate forward and reverse DNA sequencing of PCR products, with universal sequencing primers. Primers that were later changed and the modified sequences are listed in Table 1. Primers were custom synthesized by Operon Biotechnologies (Huntsville, AL). PCR reactions were carried out in a volume of 50 μL containing 100 ng of genomic DNA template, 1.5 mM MgCl$_2$, 0.8 μM forward primer, 0.8 μM reverse primer, 2.5 mM dNTP mix diluted 1:20, 1X PCR buffer, and 1 μL Amplitaq Gold (Applied Biosystems, Foster City, CA). A single touchdown protocol (20 cycles of 40 seconds at 95°C, 30 seconds at 67° − (n − 1) × 1°, and 30 seconds at 72° followed by 18 cycles with an annealing temperature of 47°C) was used to amplify exons 2–15 in a thermocycler (PerkinElmer 9600). The primers for amplification of exon 1 required a separate thermocycler protocol using a primer-specific annealing temperature of 61°C, amplified for 35 cycles. An aliquot of each PCR reaction was electrophoresed on a 2% agarose gel to confirm size and purity of the products. Amplicons were purified with the QIAquick PCR purification kit (Qiagen, Inc.) before cycle sequencing in a 10-μL volume using 4 μL Big Dye v3.1 Terminator DNA Sequencing Kit (Applied Biosystems), 5 μL purified PCR product, and 1 μL of 10 μM −21M13F or M13R sequencing primer for forward or reverse sequencing, respectively. Cycle sequencing was carried out on PerkinElmer 9700 thermocyclers using 25 cycles of 96°C for 10 seconds, 50°C for 5 seconds, and 60°C for 4 min-

Table 1
Primers for the modified *CDH1* sequencing assay

Exon	Forward 5′–3′	Reverse 5′–3′	Change[b]
1	M13F gtccgcgctgctgattggctgt[a]	M13R gagcttgcggcccgaatgcgt[a]	2
2	M13F tgttggtttcggtgagcag	M13R ggtgtgggagtgcaatttct	0
3	M13F cgctctttggagaaggaatg	M13R gtaccaaggctgagaaacctg[a]	1
4	M13F gctgtctggctaggttggac	M13R ttttcccttctctctccttgg	0
5	M13F gaaagggaaaagacccagtg	M13R ggatccagcatgggttgac	0
6	M13F gcccttctcccatgttt	M13R ctttgggcttggacaacact	0
7	M13F cctcctttatccctcaggg[a]	M13R ttgtccacgggattgagc[a]	2
8	M13F ctgggctaggccaaaggt	M13R ccatgagcagtggtgacactt	0
9	M13F aatcctttagcccctgaga	M13R aggggacaagggtatgaaca	0
10	M13F aaccacagttactttgcacc[a]	M13R caaatgacaaaatgccatga	1
11	M13F agcgcttaagccgttttca	M13R agggaggggcaaggaact[a]	1
12	M13F agacttggtctggtggaag[a]	M13R attgaaaggtggggatctgg	1
13	M13F caattttattctggaatgagctttt	M13R caggaaataaacctcctccattt	0
14	M13F gctctgtgatagctgctgctt[a]	M13R gctgtttcaaatgcctacctct	1
15	M13F tctataaactgaacatagccctg[a]	M13R ctgagcttagagatgagcc[a]	2
16	M13F agacttcttgccccagatga	M13R gctgaaaaaccaccagcaa[a]	1

Forward primers have a −21M13F tail attached for sequencing: TGTAAAACGACGGCCAGT; reverse primers have a M13R tail attached for sequencing: CAGGAAACAGCTATGAC.
[a]These primers differ because they were modified. The other primers are identical to those in Brooks-Wilson et al.[2]
[b]The final column denotes whether 0, 1, or 2 primers differ for a given exon.

utes. Cycle sequencing products were treated with sodium dodecyl sulfate at 98°C for 5 minutes and purified by CentriSep-8 spin columns (Princeton Separations, Adelphia, NJ) before drying in a SpeedVac apparatus. Dried products were resuspended in 10 μL Hi-Di Formamide and sequenced by capillary electrophoresis on an ABI 3130 gene analyzer (Applied Biosystems). Bases were called using Sequence Analysis 5.2 software. Contigs were assembled and compared to the consensus reference sequence (Genbank NM 004360) and to a negative control sequence using Mutation Surveyor 2.51 software (SoftGenetics, State College, PA).

Primers used for amplifying and sequencing the original primer-binding site for primer 15R were as follows (5′–>3′): (forward) CAATCCCGATGAAATTGGA, (reverse) TCAGGCAAGCTGAAAACATAGT.

Experiments testing the original primers for exons 15 and 16 under nontouchdown conditions utilized the following thermocycler protocol: 30 cycles of 30 seconds at 94°C, 30 seconds at an annealing temperature of 65°C, and 45 seconds at 72°C.

Identification of SNPs

SNP positions relevant to the original primer sets were initially identified via the University of California Santa Cruz genome bioinformatics browser (http://genome.ucsc.edu/). Briefly, the in silico PCR function was applied to generate the anticipated product from a primer pair. Next, the link corresponding to the chromosomal position of the product was accessed. Finally, the DNA link enabled navigation to a function called "extended case/color options," which allows differentiation of SNPs (among many other features). The identification of SNPs located within the primer-binding sites (Table 2) prompted redesign of the affected primers.

Mining of SNP data

The Ensembl Genome Browser (www.ensembl.org) was used to access the human *CDH1* genomic sequence. Hypertext links to individual SNPs were enabled by setting the "show variations" control to "all variations," allowing the identification of SNPs in previously published *CDH1* primer sets,[2,3,11] which could result in selective amplification failure due to primer mismatches. The refSNP ID links to the National Center for Biotechnology Information db-SNP Web pages (http://www.ncbi.nlm.nih.gov/projects/SNP/) yielded estimated heterozygote frequencies (summarized in Table 3), along with more detailed information with respect to population diversity.

Calculations to estimate potential impact of SNPs leading to primer mismatch

Several assumptions were made to obtain rough estimates of the possible impact of SNPs leading to primer mismatch on assay sensitivity for detecting variants in the coding regions of *CDH1* with published primer sets. It was assumed that: (1) SNPs leading to primer mismatch are independently distrib-

Table 2
Primer binding sites in the context of known SNPs

Exon amplification primers	Sequence flanking exons
1F[a] (−112, −44)	AGGG**GTCCGCGCTGCTGATTGGCTGT**GGCCGGCAGgtga**a**ccctcagccaatcagCGGTA**C**GGGG**G**GCG
1R (+24, +92)	GAGGG**ACGCATTCGGGCCGCAAGCTC**CGCGCCCCAGCCC**TGC**GCCCcttcctctcccgtcgtcaCCGCT
2F (−93, −25)	CCGGGAGCGAGGGAGGGGCGGCGCtgttggtttcggtgagcagGAGGGAACCCTCCGAGTCACCCGGTT
2R (+34, +102)	TAGGGAGGGGGTTGGAAAGGG**G**CCGagaaattgcactcccacaccCCTGGGTTGCAATGGGCAAGCTCCC
3F (−97, −29)	TTTGTGGGAGTCTTCCCACAAGTTcgctctttgggagaaggaatgCTCTTGTCTTTAATCTGTCCAATTT
3R[a] (+24, +92)	GTT**C**GCTGTTGTTTTAGTGCGCTGTCTAATCCAGGTtttctcagccttggtacc**c**gttGACATTTTG**G**GCT
4F (−107, −39)	CACTGCCCACAGAAGGCTGGGGACgctgtctggctaggttggacTGTTAGACCTGAAGTATCCGTCTTG
4R (+27, +95)	TGGGAGGGATTTGGCAGAGAAGTAccaaggagagaaagggaaaaGACCCAGTGTTGGGATCCTTCTTTA
5F (−90, −22)	TTTGGCAGAGAAGTACCAAGGAGAgaaagggaaaagacccagtgTTGGGATCCTTCTTTACTAATTCTT
5R (+34, +102)	GTAACATCCACCCAGGATTTTTTGgtcaacccatgctggatccGCAGATCAGAGGCTCTGAACACATGA
6F (−86, −18)	GGG**G**GCGCACTCTGCTCTGGCTGGgccccttctcccatgtttTCTTCCTCATCAGAGCTCAAGTCACCC
6R (+37, +105)	ACTCTTAGGTTCTTTGGACCCCAAagtgttgtcaagcccaaagGTTGTGTAA**C**TAAAGCTGATCCTTC
7F[a] (−113, −45)	GTACCACCCCCATGTCCCCTCCTTTATCCCTCAgggcagaattggattaagcaGTATTGACCCAGTCCC
7R[a] (+23, +91)	AGGGTGTGGAGGACAAATGTGTATTagctcaatcccgtggacaAAGCAAAATCCTGCTAGGCCAGTTGT
8F (−88, −20)	AATATTCTGGTTCCATGTGTTGGGctgggctaggccaaaggtGGCTAGTGTTCCTGGTCCTGACTTGGT
8R (+19, +87)	A**G**AGGGTTTCCAAAGAAAGGTCTTTTGTTGTTCATGAACTaagtgtcaccactgctcatggGCGAAG**T**C
9F (−101, −33)	TCAAAAGAACAACAAAAAAAGAGGaatcctttagccccctgagaCTCAGCTCTGCTAGCAGTCTTGGTA
9R (+23, +91)	TGCAGAAACTGG**C**ATCCTCACA**G**CtgttcataccctgtcccctGGTATCTTCTTAGAAGTAGATTCGC
10F[a] (−62, −130)	TGGCAGA**AACCACAGTTACTTTTGCACC**AACCTAATATATTA**c**caaaagcaacagttaaggaTTTAATT
10R (+57, +125)	ATATGTAAATCAATAATATGTACTtcatggcattttgtcatttgTCTGTACAAGACCATTCTCTTAATT
11F (−96, −28)	GGAAGTAA**C**CATATAA**C**TGAAGAagcgcttaagccgttttcaGCTACATGTTGTTTGC**T**GGTCCTATTC
11R[a] (+22, +90)	CTTTGCTGCCTCGACCTCCTAGCT**a**gttcagttccttgcccctcCCTTCTTTTGGAGGGAAGAGTTCAT
12F[a] (−94, −26)	GAGCCAGGACAAGATCT**AGACTTGGTCTGGTGG**aaggcaat**g**gggattcattaCTGT**T**GCCAAGCTGCC
12R (+34, +102)	TCCAACTGCCATGCTTCCCTTCCCccagatcccacctttcaatTTCCCTTCTCAACTTCTAGATGTCA
13F (−96, −28)	TTGCGGGTGTCTTTAGTTCACTAGcaattttattctggaatgagcttttTATTTTCCTCCCCTGGTCTC
13R (+17, +85)	AGTGACTCAGCCTTTGACTTAAAAaaatggaggaggtttatttcctgGAAATGATTTTTGTTCTTATAT
14F[a] (−95, −27)	CTGTCTGGTATGAGGGGT**GCTCTGTGATAGCT**gctgcttct**g**gccttctttaTCTTTGGCTCTCAACAC
14R (+42, +110)	ATTCGGAAGAAGCAATGATTAGTAagaggtaggcatttgaaacagcTCTGAATCACTTTGGATATTATC
15F[a] (−32, −100)	AAGGCATCATCCAACCATAA**TCTATAAAC**tgaacatagccctgtgtgt**a**tgACTATTTCTTTCCTACTC
15R[a] (+10, +78)	**C**CACGTGGAAAG**C**CAAAGCAT**GGCTCATCTCTAAGCT**cagga**g**gagttgtgtcaaaaaTGAGAAAAAGA
16F (−27, −95)	CATTGTCGTACCTTACATATTGCTagacttcttgccccagatgaCAGGTGTGCCCTTCCTTTCACTAAA
16R[a] (+26, +94)	ACCCATGTGCTGGGAAATGCAGaaatca**c**gttgctggtggttTTTCAGCTCCCTTCCCTTGAGATGAGT

Underlined, bolded nucleotides: known SNPs.
Lowercase: binding sites of primers described in Brooks-Wilson et al.[2]
Grey highlighted boxes: binding sites for the primers utilized in our final diagnostic *CDH1* assay.
Forward primer-binding sites correspond to 5′–>3′ primer sequences as listed in Table 1; that is, the 3′ end of the primer corresponds to the last lowercase or highlighted base for forward primers.
Reverse primer binding sites are illustrated as the reverse complement of the 5′–>3′ primer sequences; that is, the 3′ end of the primer corresponds to the first lowercase or highlighted base for reverse primers.
[a]Primers modified in response to known SNPs. Note that primer 1R was modified in response to frameshift sequencing patterns within the intron which rendered the reverse sequence unreadable in approximately 50% of patients. The numbering for 16R reflects bases following the stop codon. SNPs leading to allele dropout in the present study are in small boxes (15R, 2439 + 52 G>A; 16R, 2649 + 54C>T).
F, forward; R, reverse.

Table 3
Characteristics of known SNPs within primer binding sites for three published *CDH1* primer sets[2,3,11]

Ref SNP ID	Exon	Bases	Het. frequency	Affected primer sets (ref no.)
rs28372783	1F	A/C	0.048 ± 0.148	Brooks-Wilson et al.[2]
rs34500817	1F	C/T	0.011 ± 0.075	Suriano et al.[3]
				Berx et al.[11]
rs34221314	1F	G/C	0.010 ± 0.070	Berx et al.[11]
rs33932753	3R	C/A	0.017 ± 0.090	Brooks-Wilson et al.[2]
rs35250595	4–5R	G/A	0.010 ± 0.072	Berx et al.[11]
rs7188750	5R	G/A	0.374 ± 0.217	Suriano et al.[3]
rs36102783	7F	C/T	0.010 ± 0.072	Brooks-Wilson et al.[2]
rs34374107	7R	T/C	0.010 ± 0.072	Brooks-Wilson et al.[2]
				Suriano et al.[3]
rs35423758	9R	C/T	0.010 ± 0.069	Suriano et al.[3]
				Berx et al.[11]
rs33950903	9R	G/C	0.033 ± 0.125	Suriano et al.[3]
rs33934457	10F	C/T	0.010 ± 0.070	Brooks-Wilson et al.[2]
rs34212204	11F	T/G	0.010 ± 0.069	Suriano et al.[3]
				Berx et al.[11]
rs35667437	11R	G/A	0.041 ± 0.138	Brooks-Wilson et al.[2]
				Suriano et al.[3]
				Berx et al.[11]
rs34582129	12F	G/C	0.169 ± 0.236	Brooks-Wilson et al.[2]
				Berx et al.[11]
rs35831514	12F	T/G	0.010 ± 0.070	Suriano et al.[3]
				Berx et al.[11]
rs35686369	13R	–/A	0.084 ± 0.187	Berx et al.[11]
rs34939176	13R	–/A	0.110 ± 0.207	Berx et al.[11]
rs35325124	14F	G/A	0.010 ± 0.070	Brooks-Wilson et al.[2]
rs2276327	15F	T/C/A/G	Not reported	Brooks-Wilson et al.[2]
				Berx et al.[11]
rs33965115	15R	G/A	0.075 ± 0.179	Brooks-Wilson et al.[2a]
rs34751606	15R	C/T	0.048 ± 0.147	Berx et al.[11]
rs1801026	16R	C/A/T/G	0.341 ± 0.233	Brooks-Wilson et al.[2a]

Data mining was conducted using the Ensembl Genome Viewer and the NCBI dbSNP resource. The nucleotide corresponding to the major allele is listed first.
[a]The two SNPs observed to lead to allele dropout in the present study.
Het., heterozygote.

uted with respect to each other; (2) SNPs leading to primer mismatches are equally likely to be present in mutant versus wild-type alleles; (3) there is an equal likelihood of finding a mutation in each exon; (4) SNPs leading to primer mismatch would lead to allele dropout; (5) analysis methods were designed to detect all variants from WT sequence within the al-

leles amplified. With these assumptions, a sensitivity estimate was made for each of the previously published primer sets scrutinized, according to the following formula: $\text{sens}_{est} \approx$

$$(1/16) \star \sum_{n=1}^{16} [(1 - 0.5 \star p)(1 - 0.5 \star q)],$$

where *p* and *q* are heterozygote frequencies for SNPs within primer-binding sites for exon *n*. The factor 0.5 is included to account for the chance that the minor allele is in *Cis* with the pathogenic mutation. Most exons were affected by only one SNP, in which case the calculation reduces from $[(1 - 0.5p)(1 - 0.5q)]$ to $(1 - 0.5p)$. Because it is not possible to reliably assess the likelihood that an individual SNP would lead to allele dropout, the calculations represent the most severe impact of the SNPs on overall assay sensitivity given the assumptions above.

Calculation to estimate negative predictive value for published primer sets

Assuming that affected members within 30% of the families meeting criteria for testing truly carry a mutation in the *CDH1* coding region, the lower bound for the negative predictive value for published primer sets is estimated by the following formula: true negatives/(true negatives + false negatives) $\approx 0.7/(0.7 + [1 - \text{sens}_{est}] \times 0.3)$.

RESULTS

Assay validation began with the primers described by Brooks-Wilson et al.[2] Four positive control specimens were tested, along with a negative control individual. Initial results for three of the four positive control specimens were identical to those previously identified on a research basis. The one discrepant patient had been reported to carry a deletion in exon 15 (2395delC, also described as 2398delC) leading to a frameshift, but sequencing data suggested a homozygous wild-type sequence (Fig. 1A). The possibility of a specimen mix-up was considered. The original specimen from the discrepant patient was retested, as was an independently collected blood specimen and a frozen tissue specimen. Finally, a peripheral blood specimen from the patient's sibling, who was reportedly positive for the same mutation, was also tested. In all cases, the mutation was not detected (data not shown).

Once it was established that the assay was not performing adequately, an alternative primer set for exon 15 was designed (Table 1). The new exon 15 primers have target-binding sequences identical to those described by Suriano et al.,[3] but lack the long G-C-rich section used to facilitate analysis by HPLC, and contain −21M13F or M13R tails to facilitate sequencing with universal primers and retain seamlessness with the rest of the assay. The new primers successfully amplified exon 15 in the touchdown protocol, and allowed detection of the frameshift mutation in all specimens from the discrepant patient as
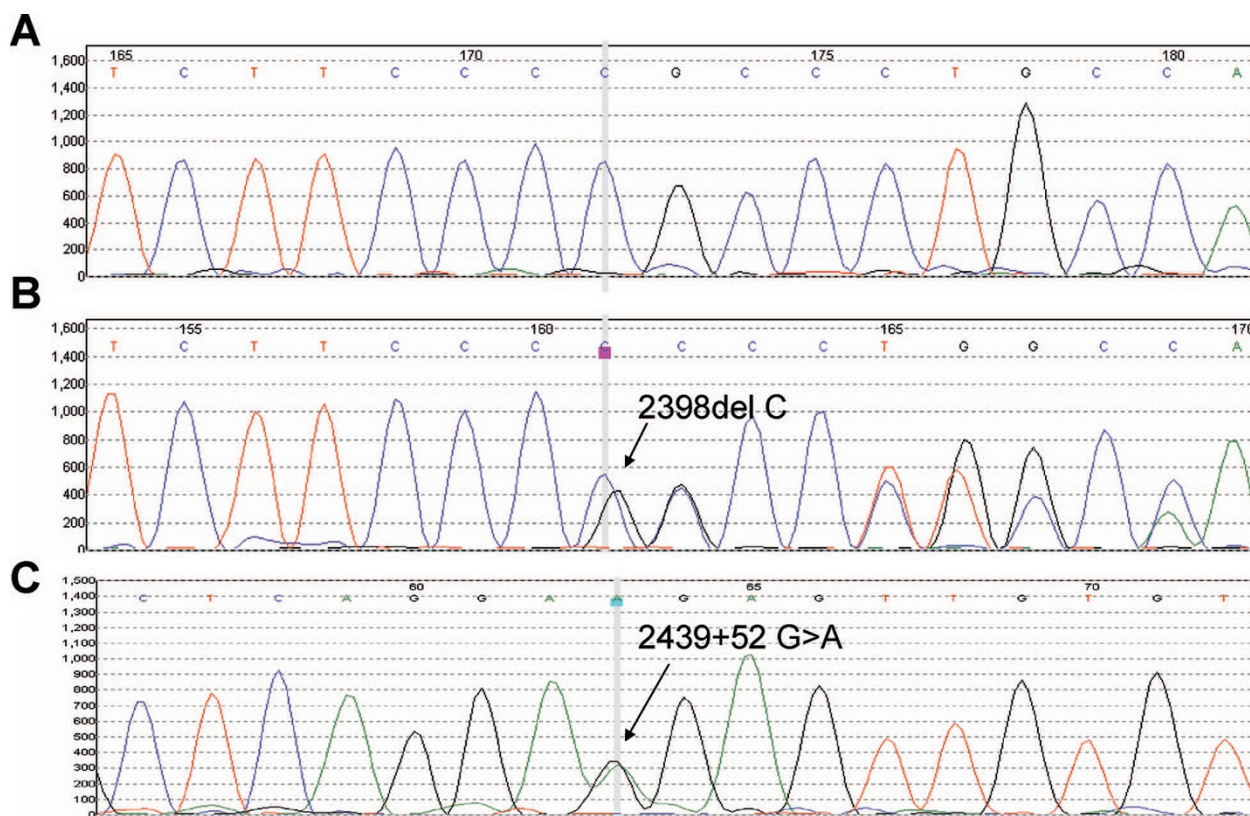
Fig. 1. Allele dropout in exon 15. A, Exon 15 sequence generated with original primers. B, Exon 15 sequence generated with modified primers, demonstrating the 2398delC mutation. C, Intronic polymorphism 2439 + 52 G>A underlying the original primer-binding site for 15R.
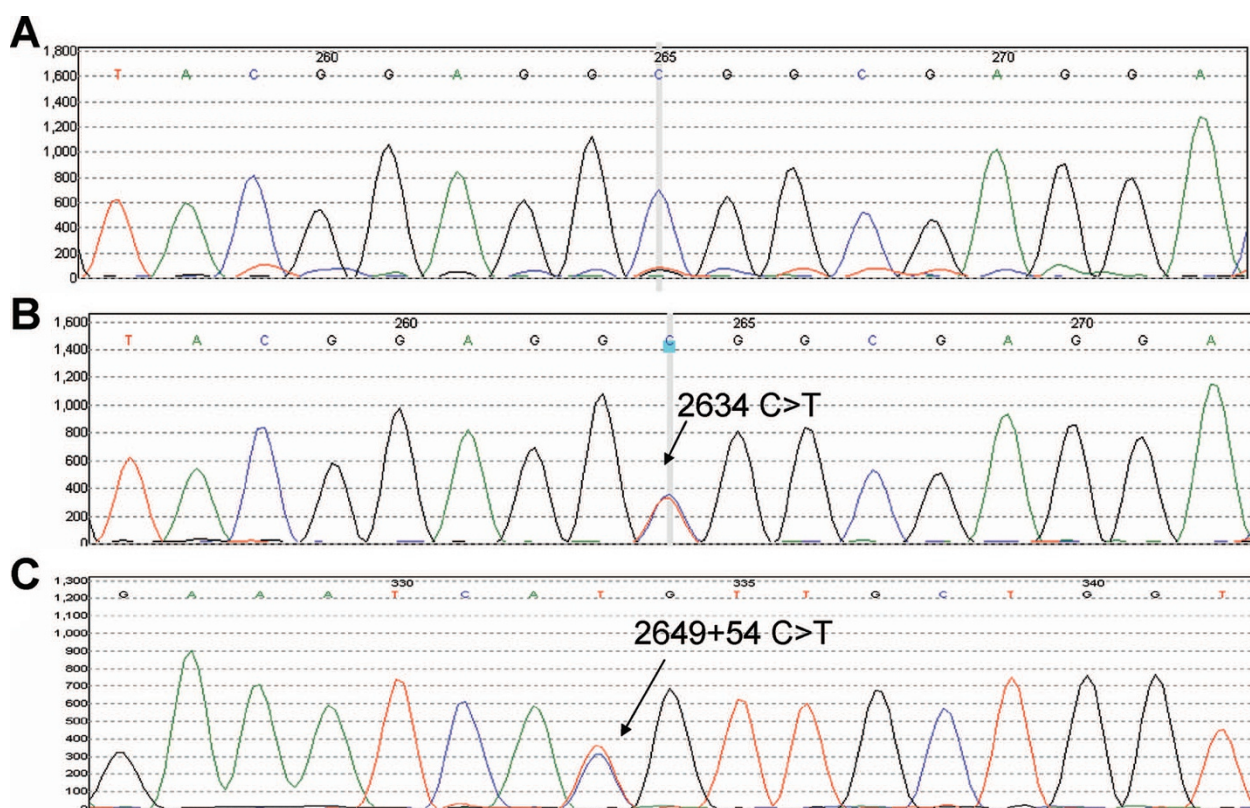

Fig. 2. Allele dropout in exon 16. A, Exon 16 sequence generated with original primers. B, Exon 16 sequence generated with modified primers, demonstrating polymorphism 2634C>T (Gly878Gly, synonymous). C, Polymorphism in intervening sequence 16, 2649 + 54C>T underlying the original primer-binding site for 16R.

well as in the specimen from the sibling (Fig. 1B and data not shown).

The primer-binding sites in the discrepant patient were sequenced to investigate whether an underlying polymorphism could explain the allele dropout observed with the original primer set. Primers flanking the original exon 15 primer-binding sites were designed (described in Materials and Methods), and genomic DNA was amplified using the same touchdown protocol employed in the original assay. Products of the anticipated size were generated and sequenced. The polymorphism responsible for allele dropout was found within the sequence targeted by the original exon 15 reverse primer, but, somewhat surprisingly, was located five bases away from the 3′ end of that primer (Fig. 1C, Table 2). This polymorphism, 2439 + 52 G>A, was seen in each discrepant specimen (data not shown) and was inferred to be present on the same allele as the pathogenic deletion within the exon. Of note, the pathogenic deletion (2398delC) is located approximately 90 bases away from the site where the 3′ end of the original exon 15 reverse primer binds, and far away from the targets of all forward and reverse primers tested for exon 15; therefore, the pathogenic deletion was not thought to contribute mechanistically to allele dropout.

Once it was established that an SNP remote from the 3′ end of the primer could lead to allele dropout and an incorrect test result, the concern that other primers might be similarly vulnerable to allele dropout was raised. By accessing a public domain SNP database (http://genome.ucsc.edu/), one or more SNPs in primer-binding sites were identified for 8 of the 15 other *CDH1* exons. In total, then, known SNPs could theoretically have led to allele dropout in nine of the original 16 primer sets. Primers for these eight exons were redesigned to diminish the probability of allele dropout as much as possible (Tables 1 and 2), and the assay was revalidated by demonstrating that the modified primers were able to consistently detect mutations and polymorphisms previously seen in the control samples (data not shown).

For an additional three patients, *CDH1* sequencing was performed using both the original[2] and the redesigned primer sets (Table 1). An unequivocal additional example of allele dropout was identified in one of these samples, affecting exon 16 (Fig. 2). The original primers for exon 16 generated an apparently homozygous, wild-type sequence for this patient. The redesigned primer set, in which primer 16R had been modified to avoid a SNP six nucleotides from the 3′ end of the primer, identified a synonymous variant at base 2634 (2634C>T, Gly878Gly), located within the exon (Fig. 2B). Because the modified 16R primer was targeted to a region outside of the originally targeted location, we were also able to identify the SNP responsible for allele dropout within the sequence from the product generated by the modified primers. The SNP responsible for allele dropout with the original primers, 2649 + 54C>T, was present at the location that had been identified as potentially problematic by searching the UCSC database (Fig. 2C, Table 2).

Touchdown PCR selectively amplifies perfectly matched primer targets during early cycles, and the exponential nature of PCR reactions effectively overwhelms less specific targets which may not be amplified until later cycles.[12] We considered the possibility that a nontouchdown thermocycler protocol utilizing the original exon 15 and 16 primer sets[2] and a single annealing temperature (65°) might permit detection of both alleles in the two specimens found to undergo allele dropout. Even with the nontouchdown protocol, however, the deletion leading to a frameshift in exon 15 was not detected. Of note, the nontouchdown protocol did generate a subthreshold peak (not called by our basecalling software) consistent with 2634C>T in exon 16 (data not shown).

DISCUSSION

Recent studies have evaluated the frequency of human SNPs and their effect on high-throughput genotyping.[13,14] Interestingly, a large-scale analysis suggested that SNPs are clustered rather than being randomly distributed throughout the genome.[14] Although the mechanisms that have led to SNP clustering remain speculative, the implication of SNP clustering for design of clinical assays is that some genes and exons may be significantly more susceptible to allele dropout than others, especially if known SNPs are not taken into account when primers are designed. Based on our experience, *CDH1* appears to be such a highly polymorphic gene. The same large-scale SNP analysis compared genotyping failure rates observed across several platforms in three bins: mismatch due to SNP affecting the 3′ end of the primer-binding site ("SNP-in-3′-tail," operationally defined as bases one through five of the 3′ end), any mismatch due to SNP ("SNP-in-primer"), and unaffected by SNP ("Primer-seq-OK"). The "SNP-in-3′-tail" and "SNP-in-primer" bins were correlated with increased genotyping failure rates, but the lore that mismatch near the 3′ end has a greater affect was not confirmed. Instead, in some platforms the failure rate was actually lower for "SNP-in-3′-tail" than for the larger "SNP-in-primer" category.[14]

Two examples of allele dropout, affecting exons 15 and 16 of the *CDH1* gene, have been described here. The first, dropout in exon 15 due to a SNP in intron 15, led to a false-negative result observed during validation of our *CDH1* sequencing assay. Subsequently, the mutant allele (2398delC) was successfully detected with an alternative primer set, and the SNP leading to allele dropout was identified within the original reverse primer-binding site. The finding that this SNP was located relatively centrally in the primer-binding site prompted a search for SNPs in binding sites of other *CDH1* primers utilized in our assay, and ultimately to modify a total of nine primer sets with the goal of optimizing the assay and avoiding allele dropout in clinical diagnostic specimens. Of note, Kaurah et al.[15] recently characterized 2398delC as a *CDH1* founder mutation affecting a large pedigree in Newfoundland, and also described a false-negative result in their initial testing. A second example of allele dropout using the original primers, this one in exon 16, was identified in one of the initial samples, as well. Although the

variant identified by the modified primers in exon 16 is synonymous and therefore not expected to be pathogenic, the finding serves as confirmation that proactively incorporating knowledge of SNPs into primer design can reduce the likelihood of allele dropout and thereby increase the accuracy of a clinical sequencing assay.

The increased target specificity allowed by touchdown PCR protocols[12] may be viewed as a theoretical disadvantage with respect to permitting allele dropout due to polymorphisms beneath primer-binding sites. However, nontouchdown conditions did not eliminate allele dropout with the original primers in the two instances noted here, highlighting the continued existence of risk and the unpredictability of the effects of primer mismatch, regardless of protocol. Further, in the setting of the clinical laboratory, limiting the number of thermocycler protocols necessary to complete testing for one patient has significant advantages for laboratory workflow. Therefore, we have chosen to retain the touchdown approach, while redesigning primers as necessary to avoid known SNPs.

Publicly available Web-based resources (see Methods) provide estimates of SNP frequencies and facilitate calculation of the maximum expected impact of known *CDH1* SNPs on sequencing assay sensitivity for primer sets previously published in the literature. Data for SNPs affecting three such primer sets[2,3,11] are summarized in Table 3. The earliest set, published by Berx et al.,[11] was utilized in the initial study identifying E-cadherin germline mutations in HDGC[1] and in subsequent studies from New Zealand.[16,17] The other two have been utilized for research testing.[2,3] The first of these two primer sets[2] also served as the initial set for our assay, while the second set[3] has been applied in an investigational HPLC-based screening approach followed by direct DNA sequencing.

The three primer sets differ both in their overall potential susceptibility to allele dropout and in the specific exons which are susceptible (Table 3). Of note, although available data suggest that most of the potentially problematic SNPs are present at a very low frequency (close to 1%), nine appear to be present at a frequency >4%, including the two SNPs which led to allele dropout in this study. The SNP rs1801026, affecting primer 16R in the primer set reported by Brooks-Wilson et al.[2] is particularly notable because of its high heterozygote frequency (estimated at 34%) together with the observation of allele dropout due to the SNP in this study (Fig. 2).

Taking into account the SNPs and heterozygosity frequencies in Table 3, the lowest expected sensitivities for detecting exonic mutations in *CDH1* with the three published primer sets are as follows: Berx et al.,[11] 0.984; Brooks-Wilson et al.,[2] 0.977; Suriano et al.,[3] 0.984. Again, the specific exons that are susceptible vary between the primer sets.

The published primer sets characterized here with respect to vulnerability to SNPs have been used on a research basis to sequence many families affected by gastric cancer over the past 10 years. The majority of these families have received negative test results, which leave uncertainty as to the cancer susceptibility of individual family members. Retesting a subset of the 16 *CDH1* exons with primers designed to avoid known SNPs

may be of diagnostic benefit in these families. The lower limits of negative predictive value for the three published primer sets studied here (based on the lowest expected sensitivities described above) are as follows: Berx et al.,[11] 0.993; Brooks-Wilson et al.,[2] 0.990; Suriano et al.,[3] 0.993. For the primer set appearing most vulnerable to allele dropout, then (Brooks-Wilson et al.[2]), approximately 100 affected or at-risk individuals from different families would need to be resequenced (on average) to identify one *CDH1* mutation. This number is predicted to be slightly greater for families originally tested with the other primer sets.

Given these numbers, one might reasonably question whether resequencing affected individuals from families previously testing as negative for *CDH1* mutations would be an effective use of resources. Such patients have already undergone a genetic test of, a priori, uncertain benefit in the past, and may be reluctant to revisit that experience. Further, counseling patients that they are very likely to receive the same negative result with the resequencing of a subset of exons may be dissuasive. Nevertheless, the natural history of *CDH1*-related disease supports that retesting should be considered and offered. Specifically, in light of the absence of effective screening methods for diffuse gastric cancer, identification of a familial mutation offers options for a proactive approach to disease.

In a large kindred recently treated at our center, six carriers of a *CDH1* mutation were exhaustively screened with stool occult blood testing, upper gastrointestinal endoscopy with random gastric biopsies, endoscopic ultrasonography, computed tomography, and positron emission tomography scans to evaluate the stomach for occult cancer. Although all preoperative evaluations were normal, all six chose to undergo prophylactic gastrectomy. Although their stomachs appeared normal at the time of surgery, they were each found to have multiple foci of T1 invasive diffuse adenocarcinoma on histopathological examination of the entire stomach. No lymph node spread or distant metastases were observed. All have recovered from their prophylactic/therapeutic gastrectomies and are currently healthy. They have been spared radiation and chemotherapy. The women among the group are taking tamoxifen for breast cancer risk reduction and undergo regular screening with mammography and magnetic resonance imaging.[6] The impact of identifying a *CDH1* mutation on this family has been profound, and the opportunity to impact other families similarly is a great motivation both for counseling and (re-)testing efforts and for work toward the identification of other genes responsible for HDGC.

To offer appropriate retesting requires knowledge of the primer set originally used in sequencing an individual and knowledge of the exons which may have been vulnerable to allele dropout with that primer set (Table 3). For primers vulnerable only to very low-frequency SNPs restricted to a single ethnic group, patient ethnicity might also be taken into consideration in customizing an appropriate panel for retesting.

Allele dropout cannot always be predicted or prevented, but scrutinizing the relevant primer regions for SNPs is a rational approach toward optimal sensitivity of this sequencing assay

for detecting *CDH1* variants. Known SNP data should be incorporated into the design of PCR-based assays whenever possible, but especially in cases of highly polymorphic genes for which an incorrect test result may have devastating consequences for patient care. *CDH1* testing certainly falls into this category, and we therefore make the very specific recommendation that primers binding to regions with known SNPs should not be incorporated into clinical *CDH1* sequencing assays.

References

1. Guilford P, Hopkins J, Harraway J, McLeod M, et al. E-cadherin germline mutations in familial gastric cancer. *Nature* 1998;392:402–405.
2. Brooks-Wilson AR, Kaurah P, Suriano G, Leach S, et al. Germline E-cadherin mutations in hereditary diffuse gastric cancer: assessment of 42 new families and review of genetic screening criteria. *J Med Genet* 2004;41:508–517.
3. Suriano G, Yew S, Ferreira P, Senz J, et al. Characterization of a recurrent germ line mutation of the E-cadherin gene: implications for genetic testing and clinical management. *Clin Cancer Res* 2005;11:5401–5409.
4. Huntsman DG, Carneiro F, Lewis FR, MacLeod PM, et al. Early gastric cancer in young, asymptomatic carriers of germ-line E-cadherin mutations. *N Engl J Med* 2001;344:1904–1909.
5. Ford JM. Inherited susceptibility to gastric cancer: advances in genetics and guidelines for clinical management. Alexandria, VA: ASCO Educational Sessions; 2002: 116–125.
6. Norton JA, Ham CM, Dam JV, Jeffrey RB, et al. *CDH1* truncating mutations in the E-cadherin gene: an indication for total gastrectomy to treat hereditary diffuse gastric cancer. *Ann Surg* 2007;245:873–879.
7. Suriano G, Seixas S, Rocha J, Seruca R. A model to infer the pathogenic significance of *CDH1* germline missense variants. *J Mol Med* 2006;84:1023–1031.
8. Ng PC, Henikoff S. Predicting the effects of amino Acid substitutions on protein function. *Annu Rev Genomics Hum Genet* 2006;7:61–80.
9. Newton CR, Graham A, Heptinstall LE, Powell SJ, et al. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res* 1989;17:2503–2516.
10. Tester DJ, Cronk LB, Carr JL, Schulz V, et al. Allelic dropout in long QT syndrome genetic testing: a possible mechanism underlying false-negative results. *Heart Rhythm* 2006;3:815–821.
11. Berx G, Cleton-Jansen AM, Nollet F, de Leeuw WJ, et al. E-cadherin is a tumour/invasion suppressor gene mutated in human lobular breast cancers. *EMBO J* 1995; 14:6107–6115.
12. Don RH, Cox PT, Wainwright BJ, Baker K, et al. 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res* 1991;19:4008.
13. Miller RD, Phillips MS, Jo I, Donaldson MA, et al. High-density single-nucleotide polymorphism maps of the human genome. *Genomics* 2005;86:117–126.
14. Koboldt DC, Miller RD, Kwok PY. Distribution of human SNPs and its effect on high-throughput genotyping. *Hum Mutat* 2006;27:249–254.
15. Kaurah P, MacMillan A, Boyd N, Senz J, et al. Founder and recurrent *CDH1* mutations in families with hereditary diffuse gastric cancer. *JAMA* 2007;297:2360–2372.
16. Humar B, Toro T, Graziano F, Muller H, et al. Novel germline *CDH1* mutations in hereditary diffuse gastric cancer families. *Hum Mutat* 2002;19:518–525.
17. More H, Humar B, Weber W, Ward R, et al. Identification of seven novel germline mutations in the human E-cadherin (*CDH1*) gene. *Hum Mutat* 2007;28:203.