Identifying Common Genetic Variants by High-Resolution Melting

Joshua G. Vandersteen,1 Pinar Bayrak-Toydemir,1,2 Robert A. Palais,3 and Carl T. Wittwer1,2

Background: Heteroduplex scanning techniques usually detect all heterozygotes, including common variants not of clinical interest.

Methods: We conducted high-resolution melting analysis on the 24 exons of the ACVRL1 and ENG genes implicated in hereditary hemorrhagic telangiectasia (HHT). DNA in samples from 13 controls and 19 patients was PCR amplified in the presence of LCGreen® I, and all 768 exons melted in an HR-1® instrument. We used 10 wild-type controls to identify common variants, and the remaining samples were blinded, amplified, and analyzed by melting curve normalization and overlay. Unlabeled probes characterized the sequence of common variants.

Results: Eleven common variants were associated with 8 of the 24 HHT exons, and 96% of normal samples contained at least 1 variant. As a result, the positive predictive value (PPV) of a heterozygous exon was low (31%), even in a population of predominantly HHT patients. However, all common variants produced unique amplicon melting curves that, when considered and eliminated, resulted in a PPV of 100%. In our blinded study, 3 of 19 heterozygous disease-causing variants were missed; however, 2 were clerical errors, and the remaining false negative would have been identified by difference analysis.

Conclusions: High-resolution melting analysis is a highly accurate heteroduplex scanning technique. With many exons, however, use of single-sample instruments may lead to clerical errors, and routine use of difference analysis is recommended. Common variants can be identified by their melting curve profiles and genotyped with unlabeled probes, greatly reducing the false-positive results common with scanning techniques.

Sequence variants can be detected as heteroduplexes after PCR amplification by separation on gels, capillaries, or columns (1). Recently, direct detection of heteroduplexes within a closed tube without separation or additions has been achieved (2). The method is based on high-resolution melting (3) of DNA duplexes in the presence of saturating fluorescent dyes (4) and appears to have an accuracy equivalent or superior to other heteroduplex scanning methods (5, 6). Applications of high-resolution melting for mutation scanning in genetics (7–10), oncology (11–17), and bacterial speciation (18, 19) have been reported.

Reports of heteroduplex scanning by high-resolution melting suggest that sequencing is required to identify all detected variants (7, 8, 11, 13). Indeed, heteroduplex scanning methods as a rule detect sequence variants but do not identify or genotype those variants. Common sequence variants that do not cause disease occur at a frequency much greater than that of disease-causing variants. These common variants lower the specificity and positive predictive value (PPV) of heteroduplex scanning methods, often resulting in a large sequencing burden to identify the variants of interest.

In many cases, high-resolution melting can distinguish between different heterozygotes. For example, hemoglobin S, hemoglobin C, and hemoglobin E heterozygotes are all differentiable (2, 20), and some RET sequence variants can be distinguished from each other (13). Indeed, in 1 study of small amplicons, all 21 random pairs of heterozygotes could be distinguished by high-resolution melting after PCR amplification (21). If common variants can be identified, then many false-positive heterozygotes

1 Department of Pathology, University of Utah Health Sciences Center, Salt Lake City, UT.
2 ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, UT.
3 Department of Mathematics, University of Utah, Salt Lake City, UT.
* Address correspondence to this author at: Department of Pathology, University of Utah Medical School, 50 N. Medical Dr., Salt Lake City, UT 84132. Fax 801-581-6001; e-mail carl.wittwer@path.utah.edu.
Received January 5, 2007; accepted April 26, 2007.
Previously published online at DOI: 10.1373/clinchem.2007.085407

4 Nonstandard abbreviations: PPV, positive predictive value; HHT, hereditary hemorrhagic telangiectasia.
could be eliminated from further consideration without sequencing.

Hereditary hemorrhagic telangiectasia (HHT) was chosen as a model disorder with disease-causing variants in 2 genes: ACVRL1 (activin A receptor type II–like 1) on chromosome 12 with 9 exons and ENG (endoglin (Osler–Rendu–Weber syndrome 1)) on chromosome 9 with 15 exons. Most HHT-related sequence variants are autosomal dominant, private, and scattered throughout these 24 exons. These variants include single-base changes (missense, nonsense, and splicing) as well as small insertions and/or deletions (22–24). Several benign variants irrelevant to HHT also occur that are detected by heteroduplex scanning methods. To eliminate these common variants from consideration, we screened DNA from normal individuals to detect variant melting curves. In exons with aberrant melting curves, we accessed public databases to identify probable sequence variants. We used unlabeled probe genotyping (25) to match the aberrant melting curves to variant sequences. Subsequent samples with unique melting profiles must be rare variants not present in the normal DNA samples. Consideration of the melting profiles of common variants can limit the need for sequencing to only rare variants not previously reported.

**Materials and Methods**

**DNA SAMPLES**

We obtained excess human genomic DNA from the samples of 19 HHT patients submitted for clinical analysis to ARUP. We performed complete exon and splice junction sequencing and identified single disease-causing variants in each sample. The samples were deidentified according to a global ARUP protocol under Institutional Review Board #7275 and blinded along with DNA from 3 healthy laboratory personnel. All DNA samples had an A<sub>260</sub>/A<sub>280</sub> ratio of 1.7 to 2.0 and were adjusted to a concentration of 50 mg/L (A<sub>260</sub> = 1.0).

**PRIMERS AND PROBES**

We designed primers for each of the 24 exons of ACVRL1 and ENG to bracket known splice site variants while avoiding benign variants within introns if possible. The number of bases separating the 3′ end of the primers from the intron:exon boundaries varied from 5 to 104 [mean (SD), 36 (26)]. Some of the larger exons (ACVRL1 exons 3, 4, and 7 and ENG exons 7 and 9a) were covered by 2 amplicons with 6 to 63 bases [45 (22)] of overlap between the 3′ ends of the primers. The amplicons varied from 126 to 356 bp in length. Primer and probe sequences are listed in Table 1 in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol53/issue7 and were synthesized by the University of Utah core synthesis facility. We designed unlabeled probes (incorporating a 3′-phosphate to prevent extension) to perfectly match common variants reported in an online database (http://137.195.14.43/cgi-bin/WebObjects/hht.woa, accessed July 1, 2006) and used them to associate common sequence variants with aberrant melting curves.

**PCR**

We performed amplification on a LightCycler 1.2 (Roche) using 10-µL volumes containing 50 ng genomic DNA in 50 mmol/L Tris, pH 8.3, 0.5 µmol/L of each primer, 0.2 mmol/L each dNTP, 2 to 4 mmol/L MgCl<sub>2</sub>, 500 mg/L BSA, 1× LCGreen I (Idaho Technology), 0.4 units KlenTaq1 (AB Peptides), and 88 ng anti-Taq antibody (TaqStart<sup>TM</sup>, Clontech). After an initial denaturation at 95 °C for 10 s, we performed 40 cycles of denaturation (95 °C for 1 s), annealing (62 to 69 °C for 1–2 s), and extension (72 to 74 °C for 3–15 s). Exact values for MgCl<sub>2</sub>, annealing, and extension are shown in Table 1 in the online Data Supplement. When unlabeled probes were used, 60 cycles were performed, the limiting primer concentration was 0.05 µmol/L, and the unlabeled probe concentration was 0.5 µmol/L. In some experiments, the probe was added after PCR as indicated in Table 1 in the online Data Supplement. All samples were heated to 95 °C and rapidly cooled to 50 °C before melting.

**MELTING ACQUISITION**

We performed high-resolution melting on an HR-1 instrument (Idaho Technology) with 24-bit acquisition of temperature and fluorescence. After PCR, denaturation, and annealing, each capillary was transferred to the HR-1 and melted from 65 to 98 °C (amplicon scanning) or 55 to 85 °C (unlabeled probes) with a slope of 0.3 °C/s, resulting in 65 points/°C.

**MELTING ANALYSIS**

Melting curves were analyzed by normalization and exponential background subtraction (26). Derivative plots of probe melting transitions were obtained by Savitsky–Golay polynomial estimation as described (27). Melting curves of PCR products were compared on difference plots of temperature-overlaid, normalized melting curves (2, 20). The normalized melting curves were adjusted to eliminate slight temperature and/or salt variation between samples by selecting a low fluorescence interval (5% to 10% fluorescence) and shifting each curve along the x axis to best overlay samples within this interval. Difference plots of temperature-overlaid, normalized curves were obtained by taking the fluorescence difference between curves at all temperature points.

The number of clusters or genotypes was determined visually from normalized and temperature-overlaid melting curves that were displayed as either melting curves or difference plots. In addition, agglomerative, unbiased
hierarchical clustering of melting curves was performed (26) as follows. Given a set of n curves, the 2 curves “closest together” are first identified. The distance between a pair of curves is taken as the mean absolute value of the fluorescence differences between the curves over all temperature acquisitions. The closest 2 curves are deleted and replaced by their mean, resulting in a new set of n – 1 curves. The next nearest pair is then replaced by the weighted mean of that pair. At each step, the weight is the number of original curves that make up each branch being averaged. This process is performed a total of n times until the last pair of curves is merged, producing a dendrogram showing the most likely clusters at each level. The process does not determine the number of clusters, that is, the number of different genotypes represented by the n curves. However, it does confirm appropriate clustering of samples at each dendrogram level.

Results

When 10 wild-type samples (without known HHT) were screened at each exon of ACVRL1 and ENG by high-resolution melting, the most common result was a single melting curve cluster, indicating that all 20 alleles were identical (observed in 17 of 24 exons). In 5 exons, 1 aberrant melting curve revealed a single common variant, and in ACVRL1, exon 3, 2 variant curves were observed. The most complex case was ENG, exon 8, where 4 distinct melting curves indicated 3 common variants (Fig. 1). Although these variants could be distinguished on normalized, overlaid melting curve plots (Fig. 1A), they were easiest to distinguish on difference plots (Fig. 1B). Although the identity of the sequence variants was initially not known, samples with melting curves identical to those in a normal population should be benign variants.

To test this method of eliminating common variants, we analyzed a blinded panel of 22 DNA samples highly enriched with known cases of HHT and an unknown number of control samples. After amplification and melting, we analyzed each exon and classified all samples as normal or as unknown variants at each exon. Normalized, temperature-overlaid melting curves were initially used for analysis rather than difference plots. Ninety percent of the melting curves matched the most common variant, whereas 6.6% matched other melting curves observed in the normal population, leaving 3.4% (18 of 528 exons) as variants of unknown significance.

Unlabeled probes were designed to identify the common variants in the normal samples using a database of known benign HHT sequence variations. Of 10 variant PCR product curves from the 10 normal individuals, all but 2 were successfully genotyped with unlabeled probes. The 2 variants that required sequencing for identification were 314→35A>G in ACVRL1, intron 3, with an allele frequency of 42%, and 524→15C>T in ENG, intron 4 (524→15C>T), with an allele frequency of 1.2% (n = 84 chromosomes). Neither variant was in the database, presumably because they flanked commonly used primers and/or were of low frequency.

Eighteen aberrant melting curves were clearly different from any curves in the normal population, indicating variants of unknown significance. These variants could be either disease causing or uncommon benign variants not identified in the screen of the 10 normal individuals. Unlabeled probes designed against less common benign variants in the database identified 2 additional previously reported benign variants, 1 in ENG exon 1 (14C>T) and 1 in ENG intron 2 (219 + 25G>T). These variants had an allele frequency of <5% and were not present in the initial screen of 20 normal chromosomes. Fig. 2 demonstrates both the original detection of a variant by difference analysis (Fig. 2A) and identification of the specific variant using an unlabeled probe matched to the variant (Fig. 2B), using the 14C>T variant of ENG, exon 1, as an example.
The remaining 16 aberrant curves suggested disease-causing variants, although the possibility of rare, previously unreported, benign variants could not be ruled out. Each of these 16 variants was in a different DNA sample, consistent with the possibility that 16 of the 22 samples may harbor disease-causing variants. When the samples were deblinded and compared with sequencing results, each of the 16 anomalous curves did correspond to a disease-causing variant, for a specificity of 100%. However, 3 disease-causing variants were missed, for a sensitivity of only 84%. On review and retesting, 2 of the 3 false negatives were attributable to clerical or omission errors in the manual single-sample analysis required with the HR-1 instrument. Fig. 3A shows derivative melting curves of the 1st false negative (ACVRL1, exon 7, 998G>T). This sample was analyzed and interpreted as aberrant; however, it was documented as negative in the final results (transcription error). Fig. 3B shows derivative melting curves of the 2nd false negative (ENG, exon 5, 586T>C). This sample was mistakenly omitted from analysis (never tested) but recorded as negative. Both samples were obviously positive upon retesting.

The 3rd and final false-negative result occurred from matching a disease-causing variant (524–2A>G) to a normal population variant (524–15C>T). Both samples were single-base heterozygotes, and their melting curves were considered the same in the blinded study when normalized fluorescence plots were used (Fig. 4A). However, the difference between these 2 variants is much easier to see on difference plots (Fig. 4B). A summary of all variants studied is shown in Table 1.

---

Fig. 2. A common DNA variant in ENG exon 1 (14C>T) detected by scanning and genotyped with an unlabeled probe. After detection of the variant on a difference plot (A), unlabeled probes were designed to match reported common variants within the amplicon. Asymmetric PCR and melting analysis in the presence of an unlabeled probe complementary to the 14T variant was performed, and the derivative melting plots are displayed (B). Wild-type DNA (black) had a single mismatch to the probe and melted in 1 transition at 70 °C, whereas the variant sample (gray) had 1 allele perfectly matched to the probe with a melting point of 75 °C.

Fig. 3. False-negative results from clerical errors. In (A), an ACVRL1, exon 7, 998G>T disease-causing variant was mistakenly transcribed as negative, although the difference plot showed an obvious sequence variant (gray) compared to wild-type (black). In (B), the sample was never tested but was mistakenly recorded as negative. Subsequent analysis of this sample after scoring the blinded study revealed an obvious ENG, exon 5, 586T>C disease-causing variant (gray) compared with wild-type (black). The PCR and melting analysis were performed on duplicate samples.
When the allele frequency of a variant is high (e.g., ACVRL1 313 + 11C>T and ACVRL1 314–35A>G), 3 genotypes are commonly present: homozygous wild-type, heterozygous, and homozygous variant. In the case of 314–35A>G, only 2 melting clusters were observed; 1 for the heterozygotes and 1 for the homozygotes (data not shown). In contrast, all 3 313 + 11C>T genotypes were separated on difference plots (Fig. 5A) and could be genotyped by unlabeled probes (Fig. 5B). When all benign variants were considered, 96% of individuals had at least 1 variant among the 24 exons. Heterozygous benign variants were present in 90% of individuals.

The PPV that a heterozygous exon predicts a disease-causing variant depends on the frequency of disease. In our highly enriched population with 19 HHT patients in 22 individuals, the PPV was 31%. With a disease frequency of 0.5, the PPV would be 20%, falling to 4.7% at a disease frequency of 0.1. However, when common variants were eliminated by normal population screening and unlabeled probe genotyping, the PPV was 100% in our study sample using HHT as a target disease.

Discussion

Heteroduplex scanning techniques for genetic variation are very sensitive (5, 6), but their utility in genetic diagnosis is compromised because both disease-causing and benign variants are detected. When many exons of a gene are scanned in an attempt to detect 1 or 2 disease-causing variants, most of the heteroduplexes are unrelated to disease. For example, 11 common variants were found in the 24 exons of ACLVR1 and ENG associated with HHT in our population, resulting in 96% of normal samples including at least 1 variant. Melting analysis of DNA from individuals unaffected by disease displays normal genetic variation as visual 2-dimensional fluorescence vs temperature plots (Fig. 1). By matching unknown heterozygotes to normal population melting clusters, common benign variants can be eliminated, even without specific genotyping (common disease alleles will also be detected in recessive disorders). If a database of common variants is available, unlabeled probes specific to the variants provide a simple means for genotyping the common variants (Fig. 2), also using high-resolution melting (25).

HHT is an autosomal dominant disease marked by recurrent bleeding and associated morbidity with a prevalence of ~1:10,000 (23). Disease-causing variants are private and highly variable and are seldom recurrent between affected families. Diagnosis is usually based on clinical manifestations, although sequencing and/or denaturing HPLC analysis (28) have been used for molecular diagnosis. High-resolution melting is an attractive alternative, because both scanning and genotyping are closed-tube methods, requiring only PCR, a saturating double-strand DNA dye (2), and a high-resolution melting instrument. By eliminating common variants through high-resolution melting, heterozygote scanning becomes much more specific. In the case of HHT, specificity was improved to 100%. Perfect specificity using these techniques on other genes is not guaranteed, but will depend on the number of normal samples screened and the completeness of the available database for benign variants. Furthermore, in the HHT samples studied here, no compound variants were identified with more than 1 variant in the same exon (e.g., 2 common variants). Such cases will be more complex and difficult to analyze. Nevertheless, HHT is an example of a typical genetic disease in which eliminating common variants by high-resolution melting was highly successful.

In contrast to the perfect specificity of this study, sensitivity was compromised, as 3 of 19 disease-causing variants were not identified. However, prior reports with
high-resolution melting suggest a scanning sensitivity of 100% for amplicons <400 bp (5, 9, 10). In retrospect, each of the 3 false positives is easily explained and suggests methods to prevent similar errors in the future. Two of the missed results were identified as clerical errors, and repeated testing showed each variant as easily identifiable (Fig. 3). Not surprisingly, sample tracking can be a problem when many exons and samples are analyzed on an instrument that is limited to a single analysis at a time, such as the HR-1. Such errors should be much less frequent in 96- or 384-well array format, now available on the LightScanner™ instrument (Idaho Technology) (9, 10). The final false-positive sample would have been identified if difference plots were used for the blinded analysis. Difference plots focus on the difference between curves, allowing easier visualization of genotype clusters (Fig. 4). Visual inspection is convenient when the number of samples is limited. However, when there are many curves to compare, intuitive visual clustering becomes less attractive than automatic clustering. Hierarchal clustering confirmed the groupings shown in Figs. 1 and 4. However, hierarchal methods do not automatically determine the cluster level or number of genotypes present. The possibility of determining the number of genotypes by the ratio of distances between consecutive cluster levels is being investigated.

Hieroduplex scanning by high-resolution melting is a simple and sensitive technique. The specificity for disease-causing variants can be greatly improved by screening normal samples, displaying any genetic variation as difference plots of melting curves, and using gene-specific databases to identify known variants with unlabeled probes. With the recent availability of high-resolution melting instruments that can analyze 96 or 384 samples in parallel, it is now feasible to scan 96 normal samples on a plate to establish a variation map for each exon. Subsequent unknown samples could then be compared against such variation maps to eliminate common variants and focus on rare variants that are most likely disease causing.

Table 1. HHT disease-causing variants and common variant frequencies identified by high-resolution melting analysis in the ACVRL1 and ENG genes.*

<table>
<thead>
<tr>
<th>Gene</th>
<th>Exon</th>
<th>No. melting clusters*</th>
<th>Disease-causing variants</th>
<th>Common variants (allele frequency)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACVRL1</td>
<td>2</td>
<td>2</td>
<td>632G&gt;A</td>
<td>1–38C&gt;T (6.0%)</td>
<td>(29)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
<td>998G&gt;T</td>
<td>313 + 11C&gt;T (46.4%)</td>
<td>(30)</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2</td>
<td>1000_1005delCGCAATinsG</td>
<td>314–35A&gt;G (41.7%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>2</td>
<td>1023C&gt;G</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>4</td>
<td>1112insG</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>3</td>
<td>1204G&gt;A</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>2</td>
<td>1445C&gt;T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENG</td>
<td>1</td>
<td>2</td>
<td>14C&gt;T (1.6%)</td>
<td>207 G&gt;A (6.0%)</td>
<td>(32)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4</td>
<td>157T&gt;C</td>
<td>219 + 25G&gt;T (3.6%)</td>
<td>(34)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2</td>
<td>360 + 1G&gt;A</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3</td>
<td>388C&gt;T</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>6</td>
<td>511C&gt;T*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>2</td>
<td>524–2A&gt;G</td>
<td>524–15C&gt;T (1.2%)</td>
<td>(23)</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>4</td>
<td>586T&gt;C</td>
<td>229 + 25G&gt;T (3.6%)</td>
<td>(32)</td>
</tr>
<tr>
<td></td>
<td>9a</td>
<td>2</td>
<td>596,597insCG</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>3</td>
<td>640,644delGGCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Coding DNA sequence numbering according to NM_000020.1 (ACVRL1) and BC_014271.2 (ENG).

* ACVRL1 exons 5 and 9 and ENG exons 6, 9b, and 11–14 are not listed because only 1 melting cluster was observed and no sequence variants were identified.

* Three melting clusters were observed at 1 single-nucleotide polymorphism, corresponding to the 3 possible genotypes (see Fig. 5).

* Two related HHT patients had the ENG 511C>T disease-causing variant.
Grant/funding support: This work was supported by National Institutes of Health Grants GM072419 and GM073396, a Center of Excellence Grant from the State of Utah, and ARUP.

Financial disclosures: Aspects of high-resolution melting analysis are licensed from the University of Utah to Idaho Technology. C.T.W. holds equity interest in Idaho Technology.

Acknowledgments: We thank ARUP for providing the deidentified clinical samples.

References


