Single-Step Scalable-Throughput Molecular Screening for Huntington Disease

Clara R. L. Teo,1 Wen Wang,2,3 Hai Yang Law,4 Caroline G. Lee,5 and Samuel S. Chong1–3*

BACKGROUND: Huntington disease (HD) is a fatal autosomal dominant neurodegenerative disorder caused by an unstable expansion of the CAG trinucleotide repeat in exon 1 of the HTT (huntingtin) gene and typically has an adult onset. Molecular diagnosis and screening for HD currently involve separate amplification and detection steps.

METHODS: We evaluated a novel, rapid microplate-based screening method for HD that combines the amplification and detection procedures in a single-step, closed-tube format. We carried out both the PCR for the HTT CAG-repeat region and the subsequent automated melting-curve analysis of the amplicon in the same wells on the plate. To establish cutoff melting temperatures (Tm,s) for each allelic class, we used a panel of reference DNA samples of known CAG-repeat sizes that represent a range of HTT alleles [normal (≤26 repeats), intermediate (27–35 repeats), reduced penetrance expanded (36–39 repeats), and fully penetrant expanded (≥40 repeats)]. We also measured well-to-well variation in Tm across the thermal block and validated cutoff Tm,s with DNA samples from 5 different populations. We also conducted a blinded validation analysis of clinical samples from an additional 40 HD-affected and 30 unaffected individuals.

RESULTS: We observed a strong correlation between CAG-repeat size and amplicon Tm among the reference DNA samples. Use of the Tm cutoffs we established revealed that 5 samples from unaffected individuals had been misclassified as affected (1.1% false-positive rate). All samples from HD-affected and unaffected individuals were correctly identified in the blinded analysis.

CONCLUSIONS: This simple and scalable homogeneous assay may serve as a convenient, rapid, and accurate screen to detect the presence of pathologic expanded HD alleles in symptomatic patients.

Huntington disease (HD)6 (OMIM 143100) is a fatal autosomal dominant neurodegenerative disorder. The disease, which typically has an adult onset, is caused by unstable expansion of a CAG trinucleotide repeat in exon 1 of the HTT (huntingtin) gene, which is located on chromosome 4p16.3 (1). Alleles with ≤26 CAG repeats are stable and considered normal. Intermediate alleles contain 27–35 triplets, which may expand into the disease range during meiosis, especially in paternal alleles. Alleles with 36–39 CAG repeats are meiotically unstable, and although such alleles have been associated with the HD phenotype, they have reduced penetrance. Individuals carrying alleles with full penetrance (≥40 CAG repeats) will display the HD pathology. Once in the disease range, repeat length is unstable during vertical transmission, with the instability appearing more pronounced in alleles of paternal origin (2–4). The age of HD onset is inversely correlated with CAG-repeat length, although other factors also play a modulating role (5, 6).

A few methods for the molecular diagnosis of HD have been proposed (7–11). The most common method for accurately sizing CAG repeats uses fluorescently tagged PCR primers that anneal to sequences flanking the CAG-repeat region and generate amplification products that can then be detected with automated genotyping systems (11, 12). The separate amplification and detection methods involved in these methods thus require the transfer of samples after the...
PCR. More recently, Gundry and Wittwer demonstrated that amplicons containing normal and expanded numbers of CAG repeats can be distinguished in the presence of SYBR Green I dye on the basis of their slightly different melting temperatures (Tm) (13). We used a 96-well plate format with a real-time PCR instrument to assess the utility of melting curve analysis (MCA) for distinguishing HTT allele classes. We used both automated capillary electrophoresis and MCA to analyze samples of known genotypes that represent a wide spectrum of CAG-repeat alleles and established Tm cutoffs for differentiating allelic classes. The assay was validated both with assumed wild-type DNA samples from 5 population groups and in a blinded analysis of clinical samples from HD-affected and unaffected individuals.

Materials and Methods

DNA SAMPLES
Genomic DNA from 11 HD-affected patients and 3 unaffected individuals was extracted from lymphoblastoid cell lines obtained from Coriell Cell Repositories (see Table 1 in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol54/issue6) via standard methods of phenol–chloroform extraction and ethanol precipitation. We also purchased 15 size-verified HD DNA samples from Coriell Cell Repositories. The assay was validated with anonymous DNA samples from populations of Chinese (n = 92), Malay (n = 92), Indian (n = 95), Caucasian (n = 96), and African American (n = 96) origin. The Chinese, Malay, and Indian DNA samples were extracted from discarded cord blood with phenol–chloroform extraction and ethanol precipitation, and Caucasian and African American DNA samples were purchased from Coriell Cell Repositories. CAG-repeat sizes were determined through automated analysis of capillary electrophoresis results with GeneScan® (Applied Biosystems) (11). We also conducted a blinded validation assay of 70 samples of DNA from 40 HD-affected and 30 unaffected individuals that had been isolated by 3 different methods (see Table 2 in the online Data Supplement). The samples were analyzed twice, with the samples placed randomly in different positions during each run. This study was approved by the Institutional Review Board of the National University of Singapore (NUS-IRB 07-123E).

PCR AND MCA
All PCR-MCA assays were performed on a LightCycler® 480 instrument (Roche Applied Science). We designed a primer pair, HD-F (5’-CCCTCCAGTCCCTCAAGTTCTT-3’) and HD-R (5’-GCGTGCCGGGC TGGT-3’), to amplify only across the CAG-repeat stretch of the HTT gene, and not the adjacent downstream CCG repeats, and took into account previously reported single-nucleotide polymorphisms in the region (14, 15). Each well on a 96-well reaction plate contained a 20-µL PCR reaction mixture consisting of 0.4 µmol/L of each primer, 50 ng genomic DNA, 1× Q-Solution (Qiagen), and 1× LightCycler 480 SYBR Green I Master (Roche Applied Science). An initial polymerase-activation step at 95 °C for 10 min was followed by 40 cycles of 98 °C for 45 s, 70 °C for 30 s, and 72 °C for 45 s. The PCR reaction was followed automatically with a melting program consisting of denaturation at 95 °C for 1 min, a temperature-hold step at 60 °C for 1 min, and then slow temperature ramping from 60 °C to 95 °C over approximately 27 min. We acquired 50 readings of fluorescence intensity per degree, a data-acquisition rate that we derived empirically. The same thermocycling and post-PCR melting conditions were used for amplification and MCA on a 384-well platform, with each well containing 5 µL of the PCR reaction mixture.

We used LightCycler 480 software to collect fluorescence intensities and convert the data into melting-curve and melting-peak plots. The Tm for each PCR product corresponds to the temperature at the center of the melting peak.

Selected DNA samples were used to assess run-to-run reproducibility and well-to-well uniformity in Tm. PCR-MCA runs were replicated for each DNA sample. For each run, identical aliquots of the PCR master mix containing the DNA sample were loaded into all 96 wells of the PCR reaction plate.

PCR AND GeneScan ANALYSIS
Genomic DNA was amplified in a final volume of 25 µL. Each reaction contained 0.4 µmol/L of each primer, 0.25 mmol/L of each deoxynucleoside triphosphate, 1.25 U of HotStar Taq™ DNA polymerase (Qiagen), 1× Q-Solution, 1× of the supplied PCR buffer (including 1.5 mmol/L MgCl2; Qiagen), and 50 ng genomic DNA. We used the primers described above, with the exception that we fluorescently labeled HD-R at the 5’ end with HEX (hexachlorofluorescein). An initial denaturation step at 95 °C for 15 min was followed by 30 cycles of 98 °C for 45 s, 63 °C for 60 s, and 72 °C for 90 s.

One microliter of each PCR product was mixed with 0.3 µL of the GeneScan − 500 ROX™ size calibrator and 9 µL of HiDi™ Formamide (Applied Biosystems). The mixtures were subjected to capillary electrophoresis on a 3130XL Genetic Analyzer (Applied Biosystems) for 50 min, and the electropherograms were analyzed with GeneMapper® software (version 4.0; Applied Biosystems).
STATISTICAL ANALYSIS
The Statistical Package for Social Sciences (version 15.0 for Windows; SPSS) and Microsoft Excel were used for ROC curve and correlation analyses, respectively. To determine cutoff Tm values between allele classes, we plotted ROC curves by using the sizes of the CAG repeats (obtained by GeneScan analysis) and the Tm values of the respective PCR products (obtained by PCR-MCA) as the test variables and using allele class (normal, intermediate, expanded with incomplete penetrance, or expanded with complete penetrance) as the classification variable.

Results
DISTINGUISHING NORMAL AND EXPANDED HD ALLELES BY PCR-MCA
The PCR-MCA assay yielded melting curves that generated different Tm peaks for normal, intermediate, and expanded HTT alleles. The reference DNA samples of known genotypes from Coriell Cell Repositories yielded PCR amplicons for normal alleles (14–23 CAG repeats) that had Tm values ranging from 84.59 °C to 87.00 °C (Fig. 1; see Table 1 in the online Data Supplement). In contrast, amplicons of intermediate alleles containing 28–30 repeats had Tm values between 87.73 °C and 88.03 °C, amplicons of expanded alleles with reduced penetrance (36–39 repeats) had Tm values between 88.37 °C and 88.72 °C, and amplicons of expanded alleles with full penetrance (40–99 repeats) had Tm values ranging from >88.72 °C to 90.88 °C. Fig. 1 shows melting peaks for selected reference DNA samples that represent a wide spectrum of HD genotypes.

We noted that NA20206 displayed only a single Tm peak for one normal allele, although this individual is heterozygous for 2 normal alleles that differ by 1 CAG repeat (Fig. 1B). Similarly, both HD-affected individuals (NA20208 and GM05542) showed only a single Tm peak for one expanded allele, although the alleles of these individuals differ by 11 repeats and 5 repeats, respectively (Fig. 1, G and H). The rest of the HD-affected individuals each carried 1 normal and 1 expanded allele and showed 2 Tm peaks, one within the normal range of Tm values and one within the expanded range (see Table 1 in the online Data Supplement; Fig. 1). The largest normal allele in the reference DNA series, a 23-repeat allele in NA20246, had a Tm of 87.00 °C, whereas the smallest expanded allele in the series, a 36-repeat allele in NA20248, had a Tm of 88.37 °C. An intermediate allele of 30 CAG repeats (in CD00022) had a Tm of 88.03 °C. Thus, a modest but distinct separation of >0.3 °C existed between the largest nonexpanded HD allele and the smallest expanded allele that we tested. These results suggest that PCR-MCA can distinguish most HD alleles with normal, intermediate, and expanded numbers of repeats and, more importantly, can clearly distinguish between pathogenic expanded HD alleles and nonpathogenic normal and intermediate HD alleles; however, the results also reveal that this system cannot distinguish normal alleles that differ by small numbers of repeats or distinguish large alleles that differ by as many as 11 repeats when both normal alleles or both expanded alleles are present in the same individual.

REPRODUCIBILITY OF THE PCR-MCA Tm
We used DNA from GM04282 (16 and 74 repeats) to assess run-to-run reproducibility and well-to-well uniformity of the Tm value and performed replicate PCR-MCA runs with the same DNA sample. For each run, we loaded identical aliquots of the PCR master mix containing GM04282 DNA into all 96 wells of the microplate. The PCR was followed immediately by MCA runs with the same DNA sample. For each run, we loaded identical aliquots of the PCR master mix containing GM04282 DNA into all 96 wells of the microplate. The PCR was followed immediately by MCA runs with the same DNA sample. The Tm peaks obtained remained generally similar across both runs, with 0.76 °C being the largest difference between runs (Fig. 2, top row, left and center panels). We observed interwell variation in Tm peaks within each run, with the Tm values of samples placed in the middle of the plate being approximately 0.6 °C lower than the Tm values of samples at the plate edges. The variation in Tm was greatest across a row and less down a column (Fig. 2, middle row, left and center panels). We observed similar trends in interwell variation across the plate for PCR-MCAs performed on a 384-well platform, but the Tm variation among wells increased slightly (SD = 0.2 °C; Fig. 2, right panels).

To evaluate the risk of misdiagnosing expanded and intermediate alleles as normal alleles caused by interwell variation in Tm, we performed similar runs with NA20248 and NA20250 DNA samples as templates and determined the SD for Tm among wells to be approximately 0.2 °C for both the 36-repeat and 39-repeat alleles in NA20248 and NA20250, respectively (Fig. 3). Similarly, for DNA from an anonymous cord blood sample with HD alleles of 16 and 27 repeats (based on GeneScan analysis), we measured the Tm SD for the 27-repeat allele across a plate. The limited amount of this DNA sample allowed us to test only 14 well positions distributed evenly across the 96-well plate, and we obtained a mean (SD) Tm of 87.8 °C (0.3 °C) for the 27-repeat allele (Fig. 3).

DETERMINATION OF Tm CUTOFFS SEPARATING NORMAL, INTERMEDIATE, AND EXPANDED ALLELE CLASSES
We conducted an ROC curve analysis of the results obtained with DNA samples supplied by Coriell Cell Repositories to identify the optimal Tm cutoff values separating the 4 allele classes and obtained Tm cutoffs of 87.3 °C for the boundary between normal and intermediate alleles, 88.2 °C for the boundary between intermediate alleles and expanded alleles with reduced...
Fig. 1. GeneScan electropherograms (left) and melting-peak plots (right) of HD CAG-repeat amplicons from 9 reference DNA samples.

GeneScan results are presented as the intensity of the fluorescence peak (y-axis) vs amplicon size (x-axis), and melting-peak plots are presented as the negative derivative of fluorescence (−dF/dT) (y-axis) vs temperature (x-axis). The sizes of CAG repeats (rpts) are indicated next to the corresponding fluorescence peaks. The GeneScan 500 ROX size calibrators are the smaller peaks. Amplicon Tₘ,s are indicated above the corresponding melting peaks. The Tₘ cutoffs of 87.3 °C, 88.2 °C, and 88.5 °C separate the Tₘ ranges for normal, intermediate, reduced penetrance expanded, and fully penetrant expanded alleles.
penetrance, and 88.5 °C for the boundary between expanded alleles with reduced penetrance and expanded alleles with full penetrance (Fig. 1).

None of the Tms for the 36-repeat and 39-repeat expanded alleles of NA20248 and NA20250 (Fig. 3) fell below the calculated Tm cutoff of 88.2 °C separating intermediate alleles from expanded alleles, regardless of well position. Similarly, none of the Tms for the 27-repeat intermediate allele from the anonymous DNA sample (Fig. 3) fell below the calculated Tm cutoff of 87.3 °C separating normal alleles from intermediate alleles. Thus, there was minimal risk of mistaking an intermediate or expanded allele for a nonexpanded allele due to between-well variation in Tm.

VALIDATION OF PCR-MCA SCREENING FOR HD

To test the feasibility of PCR-MCA screening for HD with the 96-well plate format, we genotyped DNA samples from anonymous donors from 5 different populations (Caucasian, African American, Chinese, Malay, and Indian). In parallel, we precisely sized the CAG-repeat region of each sample via the PCR and GeneScan analysis (see Figs. 1–5 in the online Data Supplement). When we excluded samples in which the melting peaks of 2 alleles of different sizes were too close and merged to form a single peak (producing a single Tm that was the average of those of the 2 unresolved melting peaks), we found CAG-repeat size and amplicon Tm to be strongly correlated in all 5 populations (Fig. 4). Only individuals with 2 normal alleles that differed by \( \pm 6 \) repeats generated 2 distinct Tm peaks with this PCR-MCA assay. Including the Tms of merged peaks produced greater within-population variation and poorer correlation between CAG-repeat size and amplicon Tm (Fig. 4). Of the 471 samples from unaffected individuals that we screened, 20 samples (4%) had the Tm of the larger allele exceed their allele class cutoffs. Of these 20 samples, only 5 of the samples from unaffected individu-
uals were incorrectly called as being from HD-affected individuals (1% false-positive rate). This result was due to the presence of an intermediate allele with a Tm ≥88.2 °C (3 Caucasians and 2 African Americans with CAG-repeat lengths of 27, 29, 31, and 34) (Fig. 4; see Figs. 1 and 2 in the online Data Supplement). The remaining 15 samples had a large normal allele (23–26 repeats) with a Tm ≥87.3 °C, the Tm cutoff separating the normal and intermediate allele classes. None of the normal alleles had Tms ≥88.2 °C, the Tm cutoff separating the intermediate and expanded allele classes. Of the 11 samples carrying intermediate alleles, only 1 sample with a 27-repeat allele had a Tm that fell below the 87.3 °C cutoff separating normal and intermediate alleles. There were no expanded alleles in the populations used in this validation study.

To test the accuracy of the PCR-MCA assay for distinguishing between samples from HD-affected and unaffected individuals, we performed an additional blinded analysis of 70 clinical samples, 40 from HD-affected individuals and 30 from unaffected individuals. The analysis was performed twice. In the first run, we placed samples randomly, starting from the center of the block and moving toward the periphery. In the second run, samples were placed randomly, beginning at the periphery and moving toward the center of the block (see Figs. 6 and 7 in the online Data Supplement). The results showed that this assay and the use of an 88.2 °C Tm cutoff allowed HD-affected individuals to be distinguished from unaffected individuals with 100% sensitivity and specificity in both of the runs (see Table 2 in the online Data Supplement). No expanded alleles were misdiagnosed as normal. Only 1 sample from an HD-unaffected individual with an intermediate allele (23 and 29 repeats) was classified as having normal alleles only. This result was due to merging of the melting peaks of the large normal allele and the small intermediate allele for this sample.

Discussion

Molecular diagnosis of HD is important in confirming the clinical diagnosis of individuals displaying HD-like symptoms and in presymptomatic adults at risk of developing the disease (16). A rapid closed-tube method for identifying HD CAG-repeat expansions has been described. This method combined PCR and MCA in the presence of the SYBR Green I dye on a capillary-based LightCycler instrument (13). We have adapted and validated this homogeneous method for screening for HD alleles in symptomatic patients for use on a microplate-based real-time PCR instrument that uses the SYBR Green I dye and a rate of fluorescence data acquisition that is comparable to that of a high-resolution melt program.

The results of the validation study indicate that the PCR-MCA assay is able to distinguish between most normal and intermediate alleles and can differentiate all normal and expanded HD alleles with no
dropout of any of the expanded alleles tested, up to 99 CAG repeats. The PCR-MCA assay appears to be more robust than capillary electrophoresis for detecting large expanded alleles in the presence of a normal allele, possibly because of the tendency of the SYBR Green I dye to migrate from the melted shorter amplicon and intercalate with the larger amplicon (Fig. 1, D–F and I).

The assay was able to distinguish different normal alleles in the same individual that differ by ≥6 CAG repeats. Normal alleles that differ by 5 or fewer CAG repeats have similar T_m values that cause the melting peaks to merge and form a single but slightly broader peak. Similarly, the melting peaks of 2 different expanded alleles in homozygous affected individuals could not be separated when such alleles differed by up to as many as 10
CAG repeats. However, this inability to distinguish clearly between 2 alleles in the fully penetrant range of expansion should not lead to misdiagnosis or a false-negative result.

CAG-repeat length and amplicon $T_m$ were strongly correlated among the DNA samples from Coriell Cell Repositories (Fig. 4). Normal alleles had $T_m$ $\leq$ 87.00 °C, intermediate alleles had $T_m$s between 87.73 °C and 88.03 °C, and expanded alleles had $T_m$s $\geq$ 88.37 °C. We used these results to establish $T_m$ cutoffs separating the different allele classes and validated these cutoffs both in an analysis of DNA samples from anonymous donors from 5 populations and in a blinded analysis of DNA from clinical samples.

None of the normal alleles were classified as expanded alleles among the 471 samples from the 5 populations. Fifteen large normal alleles (23–26 repeats) were misclassified as intermediate alleles, and 5 intermediate alleles were misclassified as expanded alleles with incomplete penetrance (see Fig. 4 and Figs. 1–5 in the online Data Supplement for details). The single misclassification of an intermediate allele (27 repeats) as normal was due to the co-presence of a normal allele that was 5 repeats shorter than the intermediate allele, which produced merged melting peaks with an apparent $T_m$ that was the average of the $T_m$s for the 2 alleles.

In the blinded validation analysis, samples were randomly distributed across the 96-well plate to determine if well position and DNA-extraction method affected the diagnostic accuracy of this assay (see Table 2 and Figs. 6 and 7 in the online Data Supplement). All samples from HD-affected individuals were detected with 100% sensitivity, regardless of DNA-extraction method and well position.

The observation of interwell variation in $T_m$ for the same sample reflects an inherent property of all block-based regular and real-time thermal-cycling instruments—uneven heating across the reaction plate. Therefore, block-based instruments may not be ideal for precisely classifying alleles with the $T_m$ cutoffs we have defined; however, the assay does represent an effective and accurate method for screening patients with clinical symptoms of HD, i.e., to differentiate those who carry a disease-causing expanded allele (and thus are truly affected with this disorder) from those who display HD-like symptoms but are unaffected by HD. Nonetheless, including control samples of known CAG-repeat sizes at strategic locations around the plate in the same run is necessary to detect appreciable instrument-related $T_m$ shifts in the heating block over time. When available, we recommend that positive-control samples carry alleles with 27 and 36 repeats, which are the smallest repeat sizes for intermediate and expanded alleles, respectively.

Recent studies have shown that the reproducibility of results of single-nucleotide polymorphism genotyping across a reaction plate with high-resolution melting analysis, a recent derivation of standard amplicon melting analysis, can be greatly improved by including 2 temperature calibrators that melt above and below the expected $T_m$ of the amplicon (17, 18). Well-to-well variation in $T_m$ can be corrected by aligning the melting peaks of the temperature calibrators for all the samples in a reaction plate. Such temperature calibrators could potentially be applied to HD PCR-MCA to correct for well-to-well variation in $T_m$ across a thermal block; however, unlike the single invariant amplicon size involved in genotyping of single-nucleotide polymorphisms with high-resolution melting analysis, appropriate temperature calibrators would have to be developed to bracket the entire $T_m$ range of CAG-repeat amplicon sizes in HD. This goal may be difficult to achieve, given the 99 °C thermal limit of the LightCycler 480 block and the need for the upper limit of an amplicon $T_m$ to be several degrees below the thermal limit because of the requirement that the fluorescence be at baseline before the final temperature cutoff. This consideration is further complicated by the fact that the $T_m$ of a 99-repeat allele is already 90.88 °C, and the $T_m$s of larger alleles are unknown. As with existing PCR-based diagnostic assays for HD, there is a risk of failure to detect the hyperexpanded allele with the PCR-MCA method. Hyperexpanded alleles are extremely rare and usually present atypical phenotypes (19, 20). If such hyperexpansion is suspected, because of a very early age of disease onset for example, we recommend that a negative test or screening result be followed up with Southern blotting and pedigree analysis.

Despite some limitations, the most distinctive advantage of this assay over existing molecular diagnostic assays for HD involving single-sample, capillary-based approaches is its technical simplicity and potential for scaled-up screening. This simple highly automated assay requires no probe or fluorescently tagged primer (except for CAG-sizing follow-up) and no sample transfer. Moreover, the single-step, closed-tube format reduces the potential for the sample mix-ups inherent in multistep analyses. PCR-MCA is thus attractive as a rapid first-line screening assay for HD because size confirmation will be required only for the small fraction of samples that test positive, as indicated by the $T_m$ values. Any false-positive calls will be detected at this stage and will not lead to misdiagnosis. The results of preliminary trials with the new generation of DNA-sat-
urating dyes (such as those used in high-resolution melting analysis) in the place of SYBR Green I suggest that better $T_m$ (and allele) resolution can be achieved in some situations (data not shown), and these findings warrant further evaluations. The potential of adapting PCR-MCA to the molecular diagnosis of other trinucleotide-repeat disorders is also worth exploring.

Grant/Funding Support: This research was supported by National Medical Research Council of Singapore grant NMRC/1079/2006 to S.S.C. and C.G.L.

Financial Disclosures: None declared.

Acknowledgment: We thank Shen Liang, Biostatistics Unit, Yong Loo Lin School of Medicine, for assistance with statistical analyses.

References