

Development of genomic reference materials for Huntington disease genetic testing

Lisa Kalman, PhD¹, Monique A. Johnson, PhD², Jeanne Beck, PhD³, Elizabeth Berry-Kravis, MD, PhD⁴, Arlene Buller, PhD⁵, Brett Casey, MD⁶, Gerald L. Feldman, MD, PhD⁷, James Handsfield, MPH¹, John P. Jakupciak, PhD⁸, Samantha Maragh, BS⁸, Karla Matteson, PhD⁹, Kasinathan Muralidharan, PhD¹⁰, Kristy L. Richie, MA⁸, Elizabeth M. Rohlfs, PhD¹¹, Frederick Schaefer, PhD¹², Tina Sellers, MS, CGC³, Elaine Spector, PhD¹³, and C. Sue Richards, PhD²

Purpose: Diagnostic and predictive testing for Huntington disease requires an accurate measurement of CAG repeats in the HD (*IT15*) gene. However, precise repeat sizing can be technically challenging, and is complicated by the lack of quality control and reference materials (RM). The aim of this study was to characterize genomic DNA from 14 Huntington cell lines available from the National Institute of General Medical Sciences Human Genetic Cell Repository at the Coriell Cell Repositories for use as reference materials for CAG repeat sizing. **Methods:** Fourteen Huntington cell lines were selected for study. The alleles in these materials represent a large range of sizes that include important diagnostic cutoffs and allele combinations. The allele measurement study was conducted by ten volunteer laboratories using a variety of polymerase chain reaction-based in-house developed methods and by DNA sequence analysis. **Results:** The Huntington alleles in the 14 genomic DNA samples range in size from 15 to 100 CAG repeats. There was good agreement among the ten laboratories, and thus, the 95% confidence interval was small for each measurement. The allele size determined by DNA sequence analysis agreed with the laboratory developed tests. **Conclusion:** These DNA materials, which are available from Coriell Cell Repositories, will facilitate accurate and reliable Huntington genetic testing. *Genet Med* 2007;9(10):719–723.

Key Words: quality control, Huntington disease, genetic testing, reference materials, CAG repeats

Huntington disease (HD), which affects approximately 1 in 10,000 individuals, is an adult-onset, autosomal dominant

neurodegenerative disease. People affected by this fatal disorder exhibit progressive chorea, rigidity, and dementia as well as various psychotic and behavioral symptoms. A juvenile form of HD that results in more severe symptoms and rapid progression has also been observed.¹

The HD gene (*IT15*) was mapped to 4p16.3,² is composed of 67 exons spanning 180 kb³ and encodes a 348 kDa protein of unknown function.⁴ HD is caused by an expansion of an unstable polymorphic trinucleotide (CAG)_n repeat in the first exon of the HD gene.⁴ Normal alleles have 26 or fewer repeats.⁵ Mutable normal alleles have 27–35 repeats and can be meiotically unstable.⁶ Alleles with 36–39 repeats have reduced disease penetrance, whereas alleles with 40 or more repeats are fully penetrant.⁶

The discovery of the HD gene and the association of the (CAG)_n repeat length with the risk for developing HD made predictive genetic testing possible. The molecular detection of an allele associated with HD or confirmation of its absence, has

From the ¹Laboratory Practice Evaluation and Genomics Branch, Centers for Disease Control and Prevention, Atlanta, Georgia; ²Department of Molecular and Medical Genetics, Oregon Health and Science University, Portland, Oregon; ³Coriell Institute for Medical Research, Coriell Cell Repositories, Camden, New Jersey; ⁴Departments of Pediatrics, Neurological Sciences, and Biochemistry, Rush University Medical Center, Chicago, Illinois; ⁵Molecular Genetics, Quest Diagnostics, San Juan Capistrano, California; ⁶Department of Pathology, Children's and Women's Health Center of British Columbia, Vancouver, British Columbia, Canada; ⁷Clinical Genetics, Wayne State University School of Medicine/Detroit Medical Center University Laboratories, Detroit, Michigan; ⁸National Institute of Standards and Technology, Biochemical Science Division, Gaithersburg, Maryland; ⁹Department of Medical Genetics, Graduate School of Medicine, University of Tennessee Medical Center, Knoxville, Tennessee; ¹⁰Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia; ¹¹Molecular Diagnostic Laboratory, Genzyme Corporation, Westborough, Massachusetts; ¹²Molecular Genetics, Center for Genetic Testing at Saint Francis, Tulsa, Oklahoma; and ¹³Department of Pediatrics, Division of Medical Genetics, University of Colorado School of Medicine, Aurora, Colorado.

Lisa Kalman, PhD, Centers for Disease Control and Prevention, 1600 Clifton Road, Mailstop G23, Atlanta, GA 30333. E-mail: LKalman@cdc.gov.

Use of trade names and commercial sources is for identification only and does not imply endorsement by the Centers for Disease Control and Prevention or the US Department of Health and Human Services.

This paper is a contribution of the US National Institute of Standards and Technology (NIST) and is not subject to copyright. Certain commercial equipment, instruments, materials, or companies are identified in this paper to specify the experimental procedure. Such identification does not imply recommendation or endorsement by NIST, nor does it imply that the materials or equipment identified are the best available for this purpose.

Disclosure: The authors declare no conflict of interest.

SAS/STAT® is a registered trademark of SAS Institute Inc. QIAquick® is a registered trademark of Qiagen GmbH. BigDye®, GeneAmp®, ABI PRISM® are registered trademarks, and POP-6™ is a trademark of Applied Biosystems. PERFORMA® DTR is a registered trademark of Edge Biosystems.

Submitted for publication: March 9, 2007.

Accepted for publication: June 8, 2007.

DOI: 10.1097/GIM.0b013e318156e8c1

an important, and potentially life-altering effect on patients and their families. Because the difference between a normal allele and one that has the potential to cause HD can be as little as one CAG repeat, it is especially important for clinical assays to be very accurate.

Accurate determination of the CAG repeat length in the HD gene is technically challenging and difficult to interpret. There are no commercially available FDA-approved test kits and, until recently, no reference materials (RMs)⁷ for HD genetic testing were available. Thus, each clinical laboratory offering this test develops its own in-house polymerase chain reaction (PCR)-based assay and RMs. Although testing for HD has been available for more than 10 years, only one Huntington cell line has been developed for use as a quality control (QC) material.⁸ QC and RMs are urgently needed by the genetic testing community to facilitate accurate size detection of HD alleles.

RMs are essential for many aspects of genetic testing. Regulatory requirements and recommendations of professional societies stipulate that positive controls should be run at least once each day in which samples are run and in the same manner as patient specimens to detect errors due to test system failure or operator error.^{6,9–18} In addition, QC and RMs are needed for test development and validation, lot-testing of new reagent batches, and for performance evaluation (proficiency testing and external quality assurance [PT/EQA]) programs.

The lack of available QC and RMs for genetic testing has been recognized as a critical need of the genetic testing community.^{19,20} To address this need, the Centers for Disease Control and Prevention (CDC), the National Institutes for Health, and the National Institute of Standards and Technology (NIST) held a series of three meetings to develop a sustainable process to collect, store, validate, and distribute RMs (mainly focused on genomic DNA) at a reasonable cost.¹⁰

Based on recommendations from the meeting participants, a new CDC-based program, the Genetic Testing Reference Material Coordination Program²¹ (GeT-RM, formerly called the Genetic Testing Quality Control Materials Program [GTQC]), was established in partnership with the genetics community. The goal of this program is to coordinate a self-sustaining community process to improve the availability of appropriate and characterized RMs for QC, proficiency testing, test development, and research. The GeT-RM Program is coordinated by the CDC, but all of the actual work, including decisions about RM priorities, mutation confirmation schemes, specimen collection, material development, and characterization, occurs through voluntary cooperation of the laboratories in the genetics community.

This article describes the collaborative development of characterized genomic DNA RMs derived from 14 Huntington cell lines by the GeT-RM program and the genetics community.

METHODS

Cell lines and DNA preparation

Fourteen Epstein-Barr virus-transformed lymphoblast cell lines containing HD alleles with repeat sizes ranging from 15 to

100 were selected from the National Institute of General Medical Sciences Human Genetic Cell Repository at Coriell Cell Repositories. The cell lines were cultured using previously described methods.²² Approximately 2 mg of DNA was prepared from each of the selected cell lines by Coriell Cell Repositories.²³

Laboratory selection

A total of 10 clinical genetic laboratories that offer HD testing volunteered to participate in this mutation confirmation study. Nine of the laboratories are in the United States; one is in Vancouver, British Columbia, Canada.

Assay methods

Each of the 10 clinical laboratories used their own in-house developed PCR assay to determine allele sizes in all of the DNA samples. PCR primers were designed by each laboratory to hybridize just 5' and 3' of the CAG repeat region in exon 1. The location of primers for each laboratory is summarized in Figure 1.

In addition, three of the laboratories utilized a second 3' primer that allowed the amplification of the polymorphic CCG repeat region (green text, Fig. 1) that is just 3' to the CAG repeats. Use of these downstream primers allowed discrimination between alleles when the CAG repeat sizes seemed to be homozygous. In one laboratory (Laboratory 5, Fig. 1), the size of the CCG repeat region was only determined in the case of apparent homozygosity. Laboratory 5 used downstream primer 5'GGCTGAGGAAGCTGAGGAG for this indication. Two other laboratories routinely measured the CCG repeat region in each sample. One laboratory (Laboratory 1, Fig. 1) used a multiple dye multiplex PCR assay with separation on an ABI3100 capillary electrophoresis instrument. This laboratory utilized primers pairs: (1) flanking only the CAG repeat (5'CCTTCGAGTCCCTCAAGTCCTTC and 5'HEX-GGCG-GCGGTGGCGGCTGTTG), (2) flanking only the CCG repeat (5'NED-AGCAGCAGCAGCAACAGCC and 5'GGCTGAG-GAAGCTGAGGAG), and (3) flanking both repeats (5'CCTTCGAGTCCCTCAAGTCCTTC and 5'6FAM-GGCTGAG-GAAGCTGAGGAG). The other laboratory (Laboratory 2, Fig. 1) utilized a second downstream primer 5'GCGGCTGAG-GAAGCTGA.

Four of the laboratories used fluorescently labeled PCR primers, and the PCR products were analyzed using capillary electrophoresis on an automated DNA sequencer. A fifth laboratory also used fluorescently labeled PCR primers, but the PCR products were analyzed using an automated gel based sequencer. CAG repeat length of each product was determined by automated analysis. The remaining five laboratories either labeled PCR products with radioisotopes and separated the products using polyacrylamide gel electrophoresis or generated unlabeled PCR products and later hybridized with a radioisotope labeled probe specific to the CAG repeat region. CAG repeat length was determined by manual comparison to a size standard or digitizing against a standard and extrapolating repeat size.

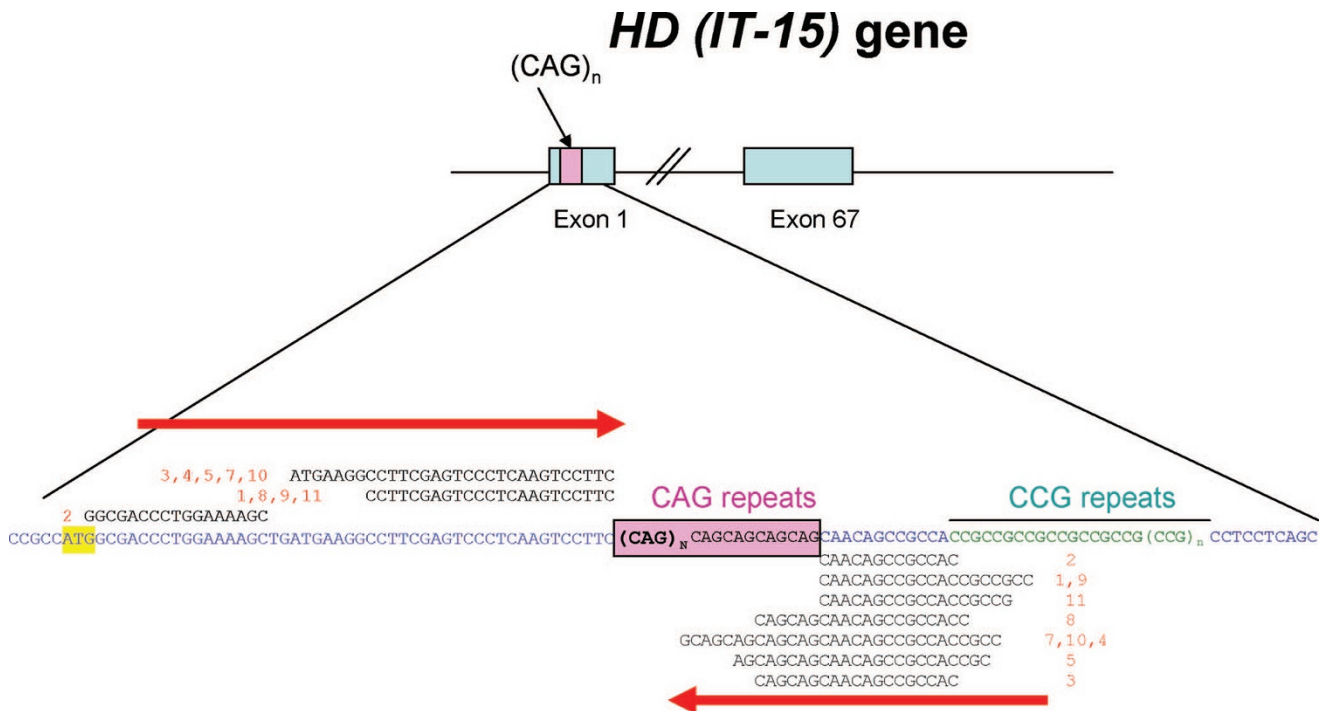


Fig. 1. Location of PCR primers flanking the CAG repeat region of the HD gene. The CAG trinucleotide repeats are indicated by shaded boxes. Calculation of the number of CAG repeats is specific to the primers used. The forward and reverse primers used by each laboratory are indicated. The numbers next to each primer correspond to individual laboratories. Primer set 11 (NIST) was used for amplification and DNA sequence analysis. The arrows indicate the direction of DNA synthesis from the primers. Primers used to measure the CCG repeats are not shown but are described in “Methods.”

Each laboratory received an aliquot of DNA from a previously characterized Huntington cell line, CD00022²⁴ with CAG repeat sizes of 18/31 for use as a size control. Laboratories were also permitted to use their own in-house validated controls. None of the laboratories reported making adjustments for abnormal migration of HD amplicons during electrophoresis.

The laboratories reported using different methods to distinguish the band of interest from other “stutter bands” depending on the method of detection. Laboratories using separation by gel electrophoresis based their selection on both the intensity (brightest) and the position of the bands. Often, but not always, the band of interest is the highest band. For laboratories using capillary electrophoresis, the prominent peak that satisfies the threshold was selected.

Protocol

Each of the 10 testing laboratories received 50- μ g aliquots of DNA from each of the 14 Huntington cell lines being tested as well as control sample CD00022. The expected CAG repeat size in each of the 14 samples was not revealed to the laboratories. The laboratories measured the CAG repeat lengths in each DNA sample in three separate assays using their in-house methods. These results were sent to the study coordinator (Sue Richards), who examined the data for quality, checked for discrepancies and compiled the data for statistical analysis.

Statistical analysis

Mean repeat values and 95% confidence intervals were calculated using SAS/STAT® PROC GLM. Modes were determined by counting the most common CAG repeat length reported from each allele.

DNA sequence analysis

DNA sequence analysis was performed to determine the CAG repeat length in each of the 14 DNA samples. PCR products were gel-purified using a filter column (QIAquick® gel extraction kit, Qiagen, Valencia, CA). Purified DNA samples were cycle sequenced (unidirectional) using BigDye® Terminator sequencing kit, version 1.1 (Applied Biosystems, Foster City, CA), using the forward primer. Cycle sequencing reactions were performed on a PE Applied Biosystems GeneAmp® PCR System 9700 thermal cycler. DNA sequencing was performed in triplicate using an ABI PRISM® Model 310 Genetic Analyzer with POP-6™ polymer system and 47 cm × 50 μm capillary. Sequencing data were analyzed with Sequencing Analysis Software, version 3.3.

The ABI sequencing data were analyzed by DNASTAR Inc (Madison, WI) Lasergene6.1 SeqManII software to determine the quality scores.²⁵

RESULTS

RM needs for HD genetic testing were identified through discussions with clinical laboratory directors and other ex-

Table 1
Mean CAG repeat length in DNA from 14 HD cell lines

Coriell sample number	Clinical laboratory assay				DNA sequence Mean ^c allele 1/allele 2
	Mean ^a allele 1 (95% confidence interval)	Mean ^a allele 2 (95% confidence interval)	Mode ^b allele 1 (% total response)	Mode ^b allele 2 (% total response)	
NA20245	15 (13.7–15.8)	15 (13.7–15.8)	15 (77%)	15 (80%)	15/15
NA20206	17 (15.9–18.3)	18 (17.2–19.1)	17 (77%)	18 (87%)	17/18
NA20207	19 (17.7–20.7)	21 (20–22.1)	19 (67%)	21 (77%)	19/21
NA20246	15 (14.1–15.9)	24 (22.3–25.9)	15 (80%)	24 (77%)	15/24
NA20247	15 (14.1–15.9)	29 (28.1–29.9)	15 (80%)	29 (80%)	15/29
NA20248	17 (15.0–19.3)	36 (35.1–37.3)	17 (77%)	36 (80%)	17/36
NA20249	22 (21.1–22.8)	39 (38.3–39.9)	22 (87%)	39 (83%)	22/39
NA20250	15 (14.1–15.8)	40 (39.2–41.0)	15 (83%)	40 (80%)	15/40
NA20208	35 (33.4–36.3)	45 (43.5–46.5)	35 (80%)	45 (77%)	35/45
NA20209	45 (43.8–46.2)	47 (46.3–47.8)	45 (73%)	47 (83%)	45/46
NA20251	39 (38.1–40.0)	50 (49.1–50.8)	39 (80%)	50 (83%)	39/50
NA20252	22 (21.2–22.9)	66 (63.7–67.5)	22 (83%)	65 (37%) 66 (37%)	22/65
NA20210	17 (15.4–18.2)	74 (72.0–76.6)	17 (77%)	74 (50%)	17/75
NA20253	22 (20.3–23.3)	99 (95.8–102.7)	22 (73%)	100 (52%)	22/101

Values were rounded to nearest whole number.

^aMean repeat length calculated from 30 responses per allele (except for allele 2 of cell line NA20253 which had only 25 responses).

^bCAG repeat length reported most often out of 30 responses per allele (except for allele 2 of cell line NA20253 which had only 25 responses).

^cUnidirectional sequencing.

perts. Fourteen cell lines were selected for study. These cell lines contained a large range of allele sizes and combinations including normal alleles and alleles at important diagnostic cutoffs, e.g., 35–36 and 39–40 repeats. Additionally, some of the cell lines were included because they contained HD alleles useful for technical reasons, such as homozygous alleles, two alleles that were close to each other in repeat size or alleles with large CAG repeat sizes.

All 10 testing laboratories were able to report data from three independent size determinations for 27 of the 28 Huntington CAG repeat regions measured in this study. The 30 size measurements for each of the 27 alleles were used to determine the mean and mode CAG repeat number. The size of the remaining allele, approximately 100 CAG repeats, proved more problematic to measure. Eight laboratories were able to report CAG repeat length from three independent assays for this allele, one laboratory reported results from only one of three assays and one laboratory was unable to amplify this allele. The available data (25 measurements) were used to determine the mean and mode repeat length of this allele.

Using the composite data from the 10 clinical laboratories, the mean CAG repeat length, 95% confidence intervals, and the modal values for each cell line were determined (Table 1). There was good agreement among laboratories, as indicated by narrow 95% confidence interval ranges. In addition, except for the largest allele, the modal value was the same as the mean. This indicates that the values had a normal distribution and were not

skewed. We did not see differences between different detection methods used, suggesting that the methods are comparable.

The CAG repeat length for each DNA sample was also determined by NIST using DNA sequence analysis (Table 1). The repeat lengths determined by DNA sequencing and laboratory analysis were identical for 25 of the 28 alleles. For the remaining three alleles, the value determined by DNA sequence analysis fell within the 95% confidence limits of the mean values determined by the 10 laboratories. Quality scores were generated from the data for each of the alleles sequenced. The scores were in the acceptable range (>40) for all alleles, except for an allele in sample NA20207 with 21 repeats. It is likely that the score for this allele (37.8) was low because of difficulty physically separating the PCR products for the two alleles (19/21) before sequence analysis. It is also possible that there is some minor contamination or mosaicism in sample NA20207, which could account for both the lower modal value and quality score for this allele.

CONCLUSION

This study showed that there were no significant differences in the analytic values obtained for the 14 samples tested using different HD assays or methodologies among the 10 laboratories. There was also very good agreement between the CAG repeat sizes obtained by the laboratories and the NIST DNA sequence analysis. All three of the alleles in which a discrepancy

was found between the results of DNA sequencing and clinical assays were larger than 45 repeats, which are more difficult to size or sequence accurately. Because these three alleles were in the fully penetrant allele range, the interpretation of alleles in this size range would be unaffected.

The CAG repeat length of 22 of the 28 HD alleles measured in this study using the laboratory developed PCR methods agreed with the allele size designated by the original submitter of the cell lines to Coriell. However, the CAG repeat size of six alleles (21%) differed. This result suggests the need for users to confirm the mutations or alleles in genomic DNA materials from any source using sequencing or other less equivocal methods, and also underscores the need for characterization studies of potential RMs.

For most genetic tests and mutations, there is no publicly available source of DNA or cell lines that can be used as RMs for QC, PT/EQA, genetic test development/validation or research. In lieu of publicly available materials, clinical laboratories and test developers use residual patient specimens when available, or synthesized oligonucleotides. The lack of available materials also affects the ability of proficiency test providers to produce sufficient and varied challenges and also limits the development and validation of new tests.

There is evidence that the availability of RMs improves the accuracy of HD repeat sizing. Data collected by the National European External Quality Assessment Service indicated a marked decrease in the variance of HD repeat sizing after a 35 repeat HD RM was provided to the participating laboratories.²⁶ Results of recent CAP proficiency surveys for HD testing indicate that laboratories generally are able to accurately measure the CAG repeat length. In each challenge, however, there are a variety of repeat lengths reported for a given allele. At times, the variation is 10–20% or more around the consensus value, suggesting that the availability of HD RMs may help to improve analytical performance.

The genomic DNA materials characterized in this project will be useful for QC, proficiency testing, test development and research, and should help to ensure the accuracy of HD genetic testing. This is especially important for alleles that are close to the diagnostic cutoffs because inaccuracy in sizing by even one repeat could impact the clinical interpretation. In addition, these studies validate our voluntary community-based approach to RM development, and may serve as a model for similar projects in the future.

DNA samples purified from these cell lines, as well as other materials developed by GeT-RM, are publicly available from the Coriell Cell Repositories.²⁷ More information about the program and available QC and RMs can be found on the GeT-RM website.²¹

References

1. GeneReviews. Huntington Disease. Available at: <http://www.genereviews.org>. Accessed December 1, 2006.

2. Gusella JF, Wexler NS, Conneally PM, Naylor SL. A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* 1983;306:234–238.
3. Ambrose CM, Duyao MP, Barnes G, Bates GP. Structure and expression of the Huntington's disease gene: Evidence against simple inactivation due to an expanded CAG repeat. *Somat Cell Mol Genet* 1994;20:27–38.
4. The Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 1993;72:971–983.
5. Nance MA, Seltzer W, Ashizawa T, Bennett R, et al. ACMG/ASHG Statement. Laboratory guidelines for Huntington disease genetic testing. *Am J Hum Genet* 1998;62:1243–1247.
6. American College of Medical Genetics. Technical Standards and Guidelines for Huntington Disease, 2006. Available at: http://www.acmg.net/Pages/ACMG_Activities/stds-2002/HD.htm Accessed December 4, 2006.
7. Emons H, Fajgelj A, van der Veen AMH, Watters R. New definitions on reference materials. *Accred Qual Assur* 2006;10:576–578.
8. Bernacki SH, Beck JC, Stankovic AK, Williams LO. Genetically characterized positive control cell lines derived from residual clinical blood samples. *Clin Chem* 2005;51:2013–2024.
9. Code of Federal Regulations. The Clinical Laboratory Improvement Amendments (CLIA). 42 CFR Part 493. (1256). Available at: http://www.access.gpo.gov/nara/cfr/waisidx_03/42cfr493_03.html Accessed December 4, 2006.
10. Chen B, O'Connell CD, Boone DJ, Amos JA, et al. Developing a sustainable process to provide quality control materials for genetic testing. *Genet Med* 2005;7:534–549.
11. The Clinical and Laboratory Standards Institute (CLSI). MM01–A2 Molecular Diagnostic Methods for Genetic Diseases; Approved Guideline, 2nd ed. Wayne, PA: CLSI, 2006.
12. American College of Medical Genetics. American College of Medical Genetics Standards and Guidelines for Clinical Genetic Laboratories. Available at: http://www.acmg.net/Pages/ACMG_Activities/stds-2002/g.htm Accessed December 4, 2006.
13. The European Molecular Genetics Quality Network. Best Practice Guidelines. Available at: <http://www.emqn.org/emqn/BestPractice.html> Accessed December 4, 2006.
14. Washington State Medical Test Site Rules (Chapter 246–338 WAC). <http://apps.leg.wa.gov/WAC/default.aspx?cite=246-338>. Accessed December 4, 2006.
15. New York State Clinical Laboratory Evaluation Program (CLEP) Laboratory Standards. Available at: <http://www.wadsworth.org/labcert/clep/clep.html> Accessed January 23, 2007.
16. American College of Medical Genetics. American College of Medical Genetics Technical Standards and Guidelines for CFTR Mutation Testing. Available at: http://www.acmg.net/Pages/ACMG_Activities/stds-2002/cf.htm Accessed January 24, 2007.
17. College of American Pathologists. College of American Pathologists Molecular Pathology Checklist. Available at: <http://www.cap.org> Accessed December 4, 2006.
18. Association for Molecular Pathology statement. Recommendations for in-house development and operation of molecular diagnostic tests. *Am J Clin Pathol* 1999;111:449–463.
19. Williams LO, Cole EC, Lubin IM, Iglesias NI, et al. Quality assurance in human molecular genetics testing- Status and recommendations. *Arch Pathol Lab Med* 2003;127:1353–1358.
20. Williams LO, Cole EC. General recommendations for quality assurance programs for laboratory molecular genetic tests. Final Monograph. 1999. Available at: <http://www.phppo.cdc.gov/dls/pdf/genetics/dyncor.pdf> Accessed December 4, 2006.
21. Centers for Disease Control and Prevention. Genetic Testing Reference Material Coordination Program. Available at: <http://www.cdc.gov/dls/genetics/qcmaterials/default.aspx> Accessed June 5, 2007.
22. Beck JC, Beiswanger CM, John EM, Satraiano, West, D. Successful transformation of cryopreserved lymphocytes: a resource for epidemiological studies. *Cancer Epidemiol Biomarkers Prev* 2001;10:551–554.
23. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 1988;16:1215.
24. Coriell Cell Repositories. HD cell line CD00022. <http://ccr.coriell.org/cdc/cdc.cgi/display.cgi?CD00022>.
25. Alex CF, Shavlik JW, Blattner FR. Neural network input representations that produce accurate consensus sequences from DNA fragment assemblies. *Bioinformatics* 1999;15:723–728.
26. Dequeker E, Ramsden S, Grody WW, Stenzel T, et al. Quality control in molecular genetic testing. *Nat Rev* 2001;2:717–723.
27. The Coriell Cell Repositories. NIGMS Human Genetic Cell Repository. <http://ccr.coriell.org/nigms/>. Accessed January 25, 2007.