A NIST Standard Reference Material for Huntington Disease Alleles

Email: margaret.kline@nist.gov
Phone: 301-975-3134

Margaret C. Kline, Amy E. Decker*, David L. Duewer and John M. Butler

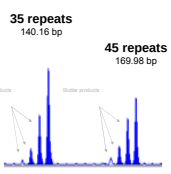
National Institute of Standards and Technology (NIST), 100 Bureau Drive MS 8312, Gaithersburg, MD 20899-8312
*Current Address: Armed Forces DNA Identification Laboratory (AFDIL), 1413 Research Boulevard, Rockville, MD 20850-3125

Huntington disease (HD) is a neurodegenerative disease of midlife onset that produces choreic movements and cognitive decline, often accompanied by psychiatric changes that affect approximately 1 in 10,000 individuals [1]. Inheritance is autosomal dominant with clinical manifestations associated with expansion of a polymorphic trinucleotide CAG repeat. Samples containing <27 CAG repeats are classified as normal while individuals with >36 repeats are affected by the disease. Thus, accurate measurement of the number of CAG repeats is critical, particularly in the intermediate diagnostic size range of 27-35 repeats.
The Applied Genetics Group at the National Institute of Standards and Technology (NIST) is working to develop an HD certified reference material. This NIST Standard Reference Material (SRM) will consist of DNA samples spanning the range of CAG repeats useful in diagnosing HD. This SRM 2393 will be helpful to clinical diagnostic laboratories wanting to ensure the accuracy and comparability of their testing results to other testing laboratories and those wishing to validate their CAG repeat sizing methods.
DNA samples were obtained from Coriell Cell Repositories representing the following Huntington alleles: 15, 17, 29, 35, 36, 40, 45, 50, and 74. These samples were amplified with a variety of PCR conditions and DNA polymerases to search for optimal conditions for reducing stutter product formation and trying to improve heterozygote peak height balance, particularly for alleles that exhibited extreme differences in size (e.g., 17, 74).

Poster #G48 at Association for Molecular Pathology 2009 Annual Meeting, Kissimmee, FL, November 19-22, 2009

Huntington Disease Classification

From American College of Medical Genetics (ACMG): Standards and Guidelines for Clinical Genetics Laboratories [1]



Repeat count	Classification	Disease status
<27	Normal	Unaffected
27–35	Intermediate	Unaffected
36–39	Variable Penetrance	+/- Affected
>39	Full Penetrance	Affected
>60	Juvenile onset	Affected

Huntington Disease Samples

from Coriell Institute for Medical Research

Coriell #	Clinical laboratory assay: Mean Allele 1 95% CI	Mean Allele 2 95% CI	Modal Allele 1 % total	Modal Allele 2 % total	Previous SEQ NIST Interlab
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
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
NA20251	39 (38.1-40.0)	50 (49.1-50.8)	39 (80%)	50 (83%)	39/50
NA20210	17 (15.4-18.2)	74 (72.0-76.6)	17 (77%)	74 (50%)	17/75

Data from Kalman et. al. (2007) [2]

The 6 samples from Coriell were chosen for SRM 2393 because the repeat counts span the various Huntington Disease classifications listed by ACMG. While these cell lines had been partially sequenced, SRM certification requires bi-directional sequencing.

SRM Certification Requirements

NIST SRM 2393 will include 6 components which are Huntington Disease cell lines Coriell maintains, previously examined by the CDC's Get-RM participants in an interlaboratory comparison.
To meet the NIST SRM certification requirements, materials obtained from Coriell have been and continue to be evaluated to be assigned a certified value at NIST.
Certified Value – Highest confidence in data accuracy and all known sources of bias have been investigated

To obtain a NIST certified value:
- Certification at NIST using a single primary method with confirmation by other method(s)
  - **DNA Sequencing**
- Certification at NIST using two independent critically-evaluated methods
  - **(1) DNA sequencing and (2) DNA sizing versus internal size standard compared to sizes of one or more sequenced alleles**
- Certification/Value-Assignment Using one Method at NIST and Different Methods by Outside Collaborating Laboratories
  - **Nominal values for candidate materials are corroborated by interlaboratory comparison involving independent typing and/or sequence analysis**

In addition, homogeneity, purity, stability and concentration issues are evaluated at NIST. **Consistency in genotyping is a MUST!**

American College of Medical Genetics

Guidelines for Huntington Disease Testing [1]

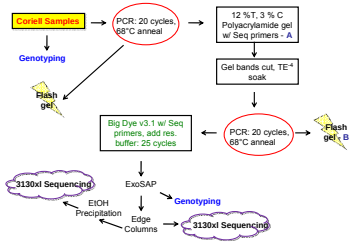
The ACMG Biochemical and Molecular Genetic Resource Committee recommends Huntington Disease alleles are sized with the following accuracy:

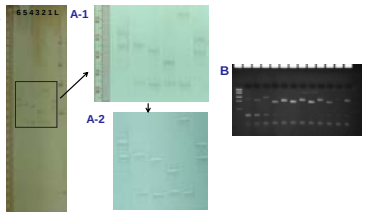
±1 repeat for alleles ≤43
±2 repeats for alleles between 44 and 50
±3 repeats for alleles between 51 and 75
±4 repeats for alleles >75

Packaging & Stability Testing for SRM 2393

Materials will be packaged in Teflon tubes based on previous stability studies.
Additional information on poster #G47

Steps to Sequence Huntington Disease Alleles

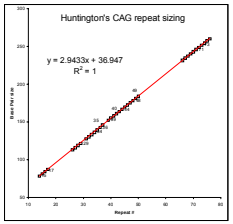




The gel pictures above display the 6 HD samples. A-(From right to left: L-ladder, 1-NA20208, 2-NA20210, 3-NA20247, 4-NA20248, 5-NA20250, 6-NA20251) A-1: zoomed in view of gel A-2: gel after bands excised B- Flash gel of samples after reamplification.

Allele Number Assignment

Accurate genotyping ([CAG] repeat number assignment) of the Huntington Disease SRM 2393 components involves more than base pair sizing of the PCR amplicons. While capillary electrophoresis base pair sizing has proven to be very reproducible when assay conditions are held constant, there is still the need to underpin these results with allele sequencing results.
The graph of "Huntington's CAG repeat sizing" shown on the right, is from accumulated genotyping data from various steps in the sequencing process (see Materials and Methods section for conditions). Sequencing results, simply counting the number of [CAG] repeats present in a sample, can then be compared to the base pair sizing results.



The graph above displays the repeat count for each of the sequenced Huntington Disease alleles. The table to the right provides the mean base pair size and standard deviation for all alleles observed for all samples. The alleles that were sequenced at NIST are listed in the light blue column. From this data, allele assignments can be made for all other base pair sizes (last column).

mean	sd	seq	est
78.26	0.11		14
81.26	0.11	15	15
84.26	0.12		16
87.18	0.22	17	17
113.11	0.01		26
116.01	0.02		27
118.98	0.10		28
121.84	0.06	29	29
127.58	0.01		31
130.57	0.07		32
133.53	0.07		33
136.52	0.09	34	34
139.55	0.13	35	35
142.81	0.10	36	36
145.96	0.07		37
152.30	0.10	39	39
155.35	0.08	40	40
158.13	0.04		41
161.06	0.06		42
163.95	0.07		43
166.86	0.10	44	44
169.76	0.09		45
172.69	0.04		46
175.53	0.03		47
178.36	0.01	48	48
181.33	0.11	49	49
184.14	0.10		50
231.66			66
234.43	0.04		67
237.27	0.09		68
240.12	0.09		69
242.99	0.08		70
245.84	0.08	71	71
248.69	0.07		72
251.55	0.09	73	73
254.38	0.06		74
257.26	0.11		75
260.18	0.07		76

Conclusions:

- SRM 2393 is a certified reference material for Huntington's Disease. It includes 6 components that span all the Huntington's Disease classification ranges, from 15-74 repeats.
- The SRM 2393 materials have been assigned certified values which enable the highest confidence in accuracy
- The procedure necessary for assigning a certified value has been demonstrated

Materials and Methods

Genotyping:

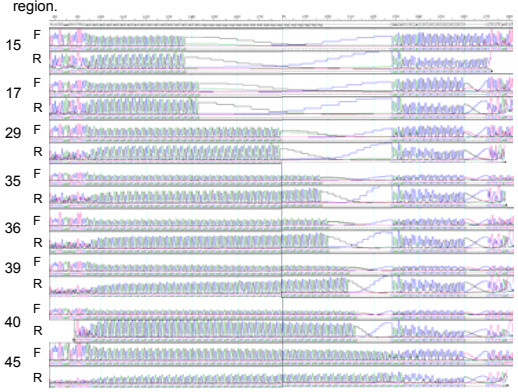
Genotyping was performed using the Roche GC-RICH PCR System [3]. PCR amplifications were performed in 10 µL reaction volumes using a master mix containing 0.4 unit of GC-RICH enzyme mix, 80 mM dNTPs, 1 M GC-RICH reaction buffer and 0.5 M GC-RICH resolution solution. The thermal cycling program was carried out on a GeneAmp 9700 (Applied Biosystems) using the following conditions in 9600-emulation mode (ramp speeds of 1 °C/s): 95 °C for 3 min; 10 cycles of 95 °C for 30 s, 65 °C for 30 s, 72 °C for 45 s; 20 cycles of 95 °C for 30 s, 65 °C for 30 s, 72 °C for 45 s + 0:05 s for each additional cycle; 72 °C for 7 min; and 4 °C hold. A 1.0 µL aliquot of the amped product was diluted in 14 µL of Hi-Di formamide and 0.4 µL GS500-LIZ internal size standard (Applied Biosystems) and analyzed on a 16-capillary ABI Prism 3130xl Genetic Analyzer using filter set G5. Separations were performed on a 36 cm array using POP-4 polymer. Data was analyzed using GeneMapper*ID* v3.2.

Sequencing:

Both alleles for each sample were sequenced in both directions using the following conditions: **Initial PCR :** 20 µL reaction volume containing 0.2 µM of each sequencing primer, 250 mM dNTPs, 1.25 mM $MgCl_2$, 1X GC-RICH reaction buffer, 1U of GC-RICH enzyme mix and 0.5 M Roche GC-RICH resolution solution. Thermal cycling: 95°C for 3 min; 20 cycles of 94°C for 30 s, 68 °C for 30 s, 72°C for 1 min + 0:05 °C for each additional cycle; 72 °C for 10 min. **Big Dye v3.1**: 20 µL reaction volume using 2 µL of 5X Seq buffer, 4 µL of Big Dye Terminator, 0.16 µM F or R primer, and 0.5 M Roche GC-RICH resolution solution. Thermal cycling:; 25 cycles of 96 °C for 10 s, 55 °C for 5 s, 65 °C for 4 m. A 5.0 µL aliquot of the purified product was diluted in 10 µL of Hi-Di formamide and analyzed on a 16-capillary ABI Prism 3130xl Genetic Analyzer. Separations were performed on a 80 cm array using POP-7 polymer. Data was analyzed using DNASTAR Lasergene v7.1 sequence analysis software.

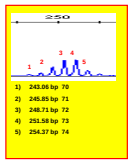
Sequencing Results:

Below are the forward and reverse sequencing results as aligned in the Lasergene, SeqMan software program. The program has inserted spaces where the [CAG] repeats are not present when comparing the smaller alleles to the larger alleles. Alleles aligned are 15, 17, 29, 35, 36, 39, 40 and 45. It should be noted that the 15 allele has a $[CCG]_{10}$ repeat region while the remaining aligned alleles have $[CCG]_7$, accounting for the second spacing region.

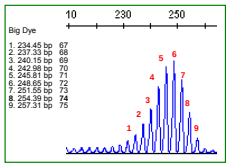


NA20210 "74" allele

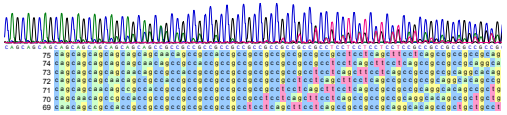
Genotyping Results **from Genomic DNA**



Genotyping Results from product used for sequencing "gel cut amp"



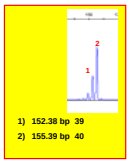
Genotyping of the largest allele from Sample NA20210 demonstrated a "population" of alleles ranging from $[CAG]_{70}$ to $[CAG]_{74}$. Amplification of the material excised from the gel yielded alleles from $[CAG]_{68}$ to $[CAG]_{75}$.



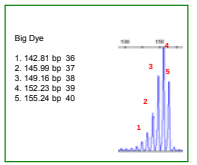
Since we were unable to obtain sequencing products without stutter alleles, we developed a "stutter sorter" that enables interpretation of the population of alleles present in the sequencing results.

NA20250 "40" allele

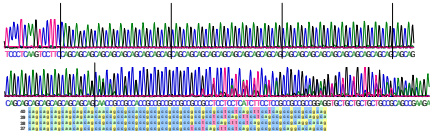
Genotyping Results **from Genomic DNA**



Genotyping Results from product used for sequencing "gel cut amp"



Genotyping of the largest allele from Sample NA20250 demonstrated a "population" of alleles ranging from $[CAG]_{36}$ to $[CAG]_{40}$. Amplification of the material excised from the gel yielded alleles from $[CAG]_{37}$ to $[CAG]_{40}$.



Sequencing results from NA20250 "40" allele. The "stutter sorter" beneath the sequencing displays the stutter
Genotyping results with Coriell sample and associated base pair sizes.
Gel cut amp after two rounds of PCR and Big Dye reactions.

References

[1] American College of Medical Genetics (2006) Technical Standards and Guidelines for Huntington Disease. Available at http://www.acmg.net/Pages/ACMG_Activities/stds-2002/HD.htm

[2] Kalman L, et al. (2007) Development of genomic reference materials for Huntington disease genetic testing. Genet Med. 9(10): 719-723.

[3] Roche Diagnostics (2008) GC-RICH PCR System Cat. No. 12 140 306 001 Mannheim Germany.

Poster available for download from NIST Applied Genetics website: http://www.nist.gov/cstl/biochemical/genetics/index.cfm