Metrology for Huntington's Disease PCR Assays: Repeat Regions to Avoid

Margaret C. Kline and David L. Duewer

National Institute of Standards and Technology (NIST), Materials Measurement Laboratory, 100 Bureau Drive, Gaithersburg, MD 20899

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

Poster # G19 :
Association for Molecular Pathology 2012 Annual Meeting on Genomic Medicine
Long Beach, CA, October 25-27, 2012

Disclosure of Potential Competing Interests: *All authors are employed by the National Institute of Standards and Technology which produces and sells Standard Reference Materials*

Introduction:
Huntington's disease (HD) is an inherited neurodegenerative disease with clinical manifestations associated with expansion of a polymorphic CAG repeat. There are two repeat regions just 12 base pairs (bp) downstream of this clinically significant repeat that complicate accurate assessment of the CAG repeat number: a $[CCG]_{7\text{-}12}$ and a $[CCT]_{1\text{-}2}$ [1]. There are also two repeat regions more than 200 bp upstream of the CAG region that must be taken into account in assay design: 6 bp repeat $[GGGGGC]_{1\text{-}2}$ and 20 bp repeat $[GGCCCCGCCTCCGCCGGCGC]_{1\text{-}3}$ [2]. While none of the additional repeat regions are used for diagnostic purposes, knowledge of the existence, positions, and variability of these repeat regions remain important to avoid inadvertent inclusion of these regions in assay development. There are published PCR primer sets that limit amplification to the CAG repeat only (Figure 1, *F Primer next to CAG repeat* and *primer 11pub*) ; however, other published sets include the CCG/CCT repeats [3,4].

Use of two different primers sets to amplify CAG repeat alone and the CAG+CCG repeat together can be useful in determining the presence of point mutations in the *primer 11pub* region that could result in null alleles or "apparent homozygotes" [5,6]. There are publically available primer design software systems which suggest placement of an upstream primer (Figure 1, *F KGH primer*) that includes the 20 bp repeat.

When developing in-house assays the use of highly characterized reference materials is essential for validation purposes. The study by Kalman et al. [3] used interlaboratory consensus to assign CAG repeat designation to 14 cell line genomic DNAs available from Coriell Cell Repositories. NIST has developed a Certified Reference Material known as Standard Reference Material® (SRM) 2393 "CAG Repeat Length Mutation in Huntington's Disease" based on 6 of the 14 cell line genomic DNAs studied above. While not yet listed in the Certificate of Analysis for SRM 2393, we are aware of the sequence both downstream and upstream of the CAG repeat for all 14 of the Kalman et al. samples. We have also evaluated 190 of the 1036 population samples we maintain at NIST. Thus far, we have identified at most one novel repeat combinations.

Materials and Methods:

Sanger sequencing primers were developed to sequence an approximately 446 bp segment starting 68 bp upstream of the CAG repeat region to confirm the upstream repeat regions (Fig 1 *primers:Fhd_us_4 and Rhd_seqUS4*). After sequence confirmation, genotyping primers were developed to type 203 samples, cell lines, and normal population samples for a 282 bp segment of this upstream region using capillary electrophoresis (Fig 1 primers: *FHD_us_2 and Rhd_US4*). Typing of the CAG repeat region for these samples was accomplished using a published primer set (Figure 1, *F Primer next to CAG repeat* and *primer 11pub*) and capillary electrophoresis bp sizing calibrated to SRM 2393. A subset of the 203 samples used was genotyped using published primer sets for the CCG/CCT repeat region (Figure 1, *F Primer next to CAG repeat* and *R Primer both repeats*). Included in this subset were "apparent" homozygous samples from the CAG typing.

Table 1. PCR parameters for Genotyping assays

PCR amplifications:	10 µL reaction volumes
Master mix:	0.25 µL SpeedStar HS DNA Polymerase
	0.5 µL 10 mM dNTPs
	2 µL GC-RICH buffer (5X)
	1 µL GC-RICH resolution solution
	0.2 µL of each forward and reverse primer (10 µM).
	0.5 µL $MgCl_2$ (25 mM)
	3.6 µL Water
	2.0 µL DNA solution (≈0.5 ng/µL)
Thermocycler:	GeneAmp 9700 Applied Biosystems in 9600-emulation mode (ramp speeds of 1 °C/s)
Thermal cycling:	95 °C for 3 min
	30-35 cycles of {94 °C for 30 s, 68 °C for 30 s, 72 °C for 60 s},
	Holds: 72 °C for 10 min
	4 °C hold until removed from the thermal cycler

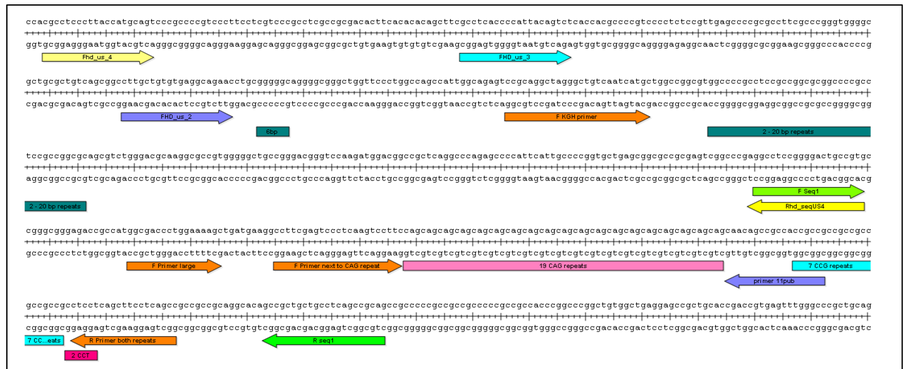

Figure 1. Huntington's annotated sequence. The 750 bp of sequence notes repeat regions and primers used in the amplification. Sequence from NCBI GenBank NT006081_1520221_1601461.

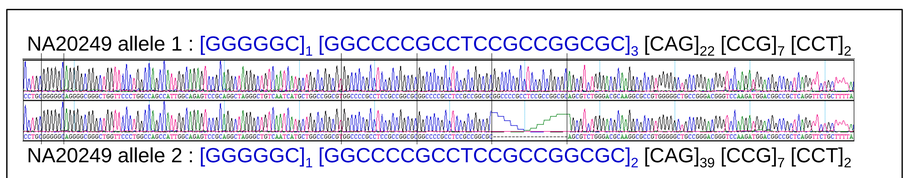

Figure 2. Sequencing data for the repeats found upstream of the CAG repeat region

Table 2. Classifications of the upstream repeat regions

α:$[GGGGGC]_1$ $[GGCCCCGCCTCCGCCGGCGC]_1$ 270 bp ±0.5 bp N=5
β:$[GGGGGC]_1$ $[GGCCCCGCCTCCGCCGGCGC]_2$ 286 bp ±0.5 bp N=184
γ:$[GGGGGC]_2$ $[GGCCCCGCCTCCGCCGGCGC]_2$ 292 bp ±0.5 bp N=29
δ:$[GGGGGC]_1$ $[GGCCCCGCCTCCGCCGGCGC]_3$ 304 bp ±0.3 bp N=12
ε:$[GGGGGC]_2$ $[GGCCCCGCCTCCGCCGGCGC]_3$ 311 bp ±0.3 bp N=3
ζ:$[GGGGGC]_?$ $[GGCCCCGCCTCCGCCGGCGC]_3$ 282 bp ±0.5 bp* N=2

* Requires additional sequencing

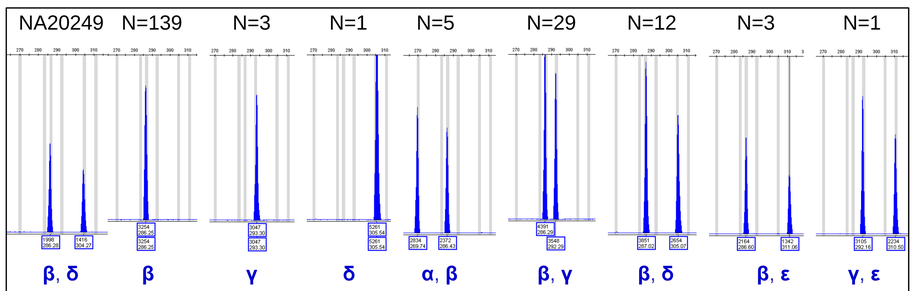

Figure 3. Example electropherograms genotyping upstream repeat regions

Table 3. Ethnicities of samples genotyped to the upstream repeats and estimated allele frequencies

	Number							Frequency					
	N	α	β	γ	δ	ε	ζ	α	β	γ	δ	ε	ζ
AfricanAm	70	2	66	27	3	3	0	0.029	0.943	0.386	0.043	0.043	0
Asian	3	0	2	0	3	0	0	0	0.667	0	1.000	0	0
Cauc	73	1	72	3	5	0	1	0.014	0.986	0.041	0.068	0	0.014
Hisp	44	2	44	2	2	0	0	0.045	1.000	0.045	0.045	0	0

See Table 2 for sequence classifications

Results:
203 samples were characterized for the variability of the upstream repeat regions. While 85 % of the 406 observed alleles were $[6\ bp]_1$ / $[20\ bp]_2$, 6 % were at least 20 bp larger or smaller. Preliminary data of the CCG/CCT repeat region, primarily for samples appearing to be homozygous for the CAG repeat, suggest repeat variability of 7 to 11 repeat units. Additional data from Sanger sequencing is needed to confirm the CCG/CCT configuration since the primers used encompassed the CAG, CCG and the CCT repeat regions. This set of primers was duplexed with the upstream repeat region primer.

Table 4. Genotyping results for the CAG, CCG and upstream repeat regions

		CAG Repeat	CCG Repeat	Upstream Repeat
SRM 2393	A	15,29	10,7	β
SRM 2393	B	17,36	7,7	β
SRM 2393	C	15,40	10,7	β
SRM 2393	D	35,45	7,7	β
SRM 2393	E	39,50	7,7	β
SRM 2393	F	17,75	7,7	β
	NA20249	22,39	7,7	β,δ
	NA20246	15,24	7,7	β,δ
Hispanic	ZT80786	17,20	9,7	β
Caucasian	WT51342	17	7,10	β
Caucasian	TT50705	17	7	α,β
AfricanAm	PT84234	15	9,10	β,γ
Caucasian	UT57293	20	7	β
Caucasian	WT51343	17	7	β,δ
AfricanAm	PT84206	17	7	β
Caucasian	WT51386	20	7,10	β
Hispanic	Y26	17	7	β
AfricanAm	OT05892	17	7	β,δ
AfricanAm	PT84251	15	9,10	β
AfricanAm	PT84241	17	7	β,δ
AfricanAm	PT84209	15	10,11	β,ε
Caucasian	MT94886	18	10	β
Asian	OT06795	17	7	β,δ
AfricanAm	TT50922	18	7,10	β
Caucasian	Y4	17	10	β
AfricanAm	Y13	15	9,10	β
Caucasian	Y5	17	7	β
AfricanAm	Y14	15	10	γ

Underline indicates the results were verified with Sanger sequencing.

No underline indicates the results are based on genotyping assays.

There are 2 cell lines from the 14 analyzed by Kalman et al.[3] that have the "**β**, **δ**" typing designation of the upstream repeats, the other 12 cell lines have just the "**β**" upstream repeat designation.

The NIST population samples listed were genotyped for the CCG repeat as well as the upstream repeats primarily for their homozygous status at the CAG repeat. The estimated CCG repeat is based on the size of the apparent homozygous allele. All typed population samples had "normal" range CAG repeats and CCG repeats.

Sanger sequencing would be needed to confirm the "true" CCG calls.

Conclusions:
Inclusion of the upstream repeat regions in the CAG repeat count may have clinical significance and should be avoided in primer design. Use of well characterized reference material can aid in assay development.

References:

[1] Andrew, S.E., Goldberg, Y.P., Theilmann, J., Hayden, M.R. A CCG repeat polymorphism adjacent to the CAG repeat in Huntington disease gene: implications for diagnostic accuracy and predictive testing. Hum Mol Genet 1994;3:65-67
[2] Coles, R., Leggo, J., Rubinsztein D. Analysis of the 5' upstream sequence of the Huntington's disease (HD) gene shows six new rare alleles which are unrelated to the age at onset of HD. J Med Genet 1997;34:371-374
[3] Kalman, L., et al. Development of genomic reference materials for Huntington disease genetic testing. Genet Med 2007;9(10):719-723
[4] Margolis, R., Ross, C. Diagnosis of Huntington Disease. Clin Chem 2003;49(10):1726-1732
[5] The American College of Medical/American Society of Human Genetics Huntington Disease Genetic Testing Working Group. Laboratory guidelines for Huntington disease genetic testing. Am J Hum Genet 1998;62:1243-1247
[6] Yu, S., Fimmel, A., Fung, D., Trent, R.J. Polymorphisms in the CAG repeat – a source of error in Huntington disease DNA testing. Clin Genet 2000;58:469-472

Acknowledgements/Disclaimer:
Certain commercial equipment, instruments and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose.

Poster available for download from NIST Applied Genetics website: http://www.nist.gov/mml/bmd/genetics/clinical_dna_pubs.cfm