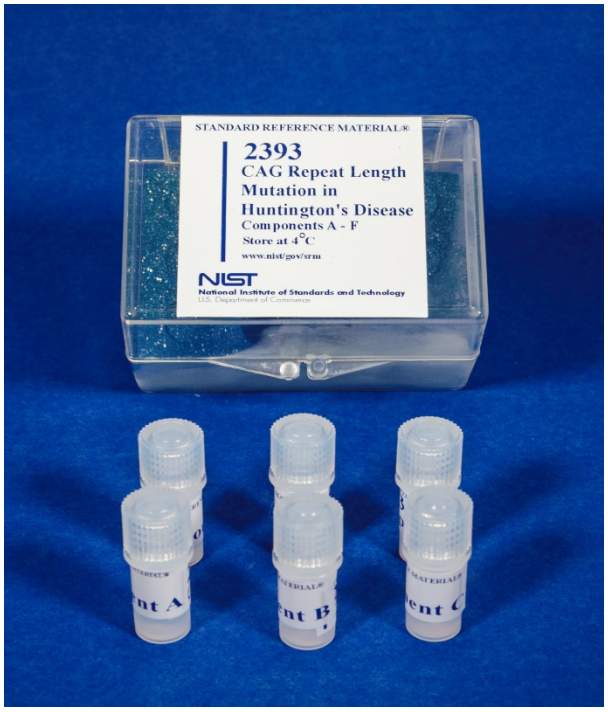
NIST Special Publication 260-203

Certification and Extension of the Period of Validity for Standard Reference Material® 2393

CAG Repeat Length Mutation in Huntington's Disease



Margaret C. Kline
Carolyn R. Steffen
David L. Duewer

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.260-203

National Institute of Standards and Technology
U.S. Department of Commerce

(This page intentionally blank)

NIST Special Publication 260-203

Certification and Extension of the Period of Validity for Standard Reference Material® 2393

CAG Repeat Length Mutation in Huntington's Disease

Margaret C. Kline
Carolyn R. Steffen
Biomolecular Measurement Division
Material Measurement Laboratory

David L. Duewer
Chemical Science Division
Material Measurement Laboratory

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.260-203

October 2020

U.S. Department of Commerce
Wilbur L. Ross, Jr., Secretary

National Institute of Standards and Technology
Walter Copan, NIST Director and Undersecretary of Commerce for Standards and Technology

National Institute of Standards and Technology Special Publication 260-203
Natl. Inst. Stand. Technol. Spec. Publ. 260-203, 58 pages (October 2020)
CODEN: NSPUE2

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.260-203

Abstract

Standard Reference Material® (SRM®) 2393 is intended for use in the value assignment of the number of Huntington's Disease CAG trinucleotide repeats contained in a human genomic material. A unit of SRM 2393 consists of six component genomic DNA solutions, labeled A to F, extracted from cell lines derived from Huntington's Disease samples. This publication documents the production, analytical methods, and statistical evaluations involved in realizing this product and ensuring its continued validity.

Key words

CAG trinucleotide repeat
clinical reference material
HHT gene locus 4p16.3;
human nuclear DNA
Huntington's Disease.

Technical Information Contact for this SRM

Please address technical questions about this SRM to srms@nist.gov where they will be assigned to the appropriate Technical Project Leader responsible for support of this material. For sales and customer service inquiries, please contact srminfo@nist.gov.

Table of Contents

List of Tables

List of Figures

Purpose and Description

Standard Reference Material® (SRM®) 2393 CAG Repeat Length Mutation in Huntington's Disease is intended for use in

- developing and validating methods for determining the number of Huntington's Disease cytosine-adenine-guanine (CAG) trinucleotide repeats contained in human genomic materials and
- qualifying control materials produced in-house and analyzed using those methods.

A unit SRM 2393 consists of six components, labeled A to F, that are DNA extracts from cell lines derived from donors with Huntington's Disease. The DNA extracts were purchased from Coriell Institute for Medical Research, (Camden, NJ), originally obtained from the National Institute of General Medical Sciences (NIGMS) Human Genetic Cell Repository, Inherited Disorders Subcollection. Each component consists of 50 µL of a dilute DNA extract in sterile 0.5 mL perfluoroalkoxy (PFA) fluoropolymer vials that have been stored in the dark between 2 °C to 8 °C. The extracted DNAs are solubilized in deionized water containing 10 mmol/L 2-amino-2-(hydroxymethyl)-1,3-propanediol hydrochloride (Tris HCl) and 0.1 mmol/L ethylenediaminetetraacetic acid disodium salt (disodium EDTA) adjusted to pH 8.0 (TE^{-4}, pH 8.0 buffer).

Warning: SRM 2393 is a Human Cell Line Source Material

SRM 2393 is a human cell line source material. Since there is no consensus on the infectivity status of extracted DNA, handle the SRM 2393 components as Biosafety Level 1 materials capable of transmitting infectious disease [1].

SRM 2393 components and derived solutions should be disposed of in accordance with local, state, and federal regulations.

Ethical Research Approval

NIST is guided by and adheres to the ethical principles set forth in the Belmont Report [2]. SRM 2393 was developed after an appropriate human subjects' research determination by the NIST Research Protection Office that the development of this product was "not human subjects research" (often referred to as research not involving human subjects) as defined in U. S. Department of Commerce Regulations, 15 CFR 27, also known as the Common Rule (45 CFR 46, Subpart A), for the Protection of Human Subjects. Therefore, the development of SRM 2393 was not subject to oversight by the NIST Institutional Review Board.

1. Introduction

Huntington Disease (HD) is a neurodegenerative disorder that produces involuntary, unpredictable body movements and cognitive decline [3,4]. HD affects ≈4 individuals per 100 000 [5]. HD is associated with CAG trinucleotide repeats at the *HHT* gene locus 4p16.3 of chromosome 4 (OMIM gene/locus code: 613004) [6]. The GenBank accession number for the reference sequence is NM_002111.

Recent authoritative Clinical Guidelines define normal allele and HD mutation categories in terms of the number of CAG repeats [5]:

- Normal alleles: ≤26 CAG repeats; non-pathogenic/stable polymorphic repeat.
- Mutable Normal alleles: (27 to 35) CAG repeats; meiotic instability range, not convincingly associated with HD phenotype but they can be meiotically unstable in sperm. Pathologic expansion of the paternal-derived allele can occur.
- HD alleles with reduced penetrance alleles: (36 to 39) CAG repeats; alleles are meiotically unstable and associated with HD phenotype, however elderly asymptomatic individuals are known.
- HD alleles with full penetrance alleles: ≥40 CAG repeats; HD pathology present. The largest HD allele currently reported has at least 250 CAG repeats.

In addition to the normal and mutation boundary categories, the Guidelines have established acceptable measurement performance criteria for HD clinical testing and grading proficiency testing surveys [5]:

- <50 CAG repeats: Consensus size ±2 repeats.
- {50 to 75} CAG repeats: consensus size ±3 repeats.
- >75 CAG repeats: consensus size ±4 repeats.

Since there is no cure for HD, only some therapies that help manage disease symptoms, it is of the utmost importance that the predictive testing of at-risk individuals or pre-natal testing be accurate. The lack of HD standards has resulted in misleading data and measurement inaccuracies among clinical laboratories. SRM 2393 CAG Repeat Length Mutation in Huntington's Disease was developed to supply the HD clinical testing laboratories with human cell line genomic materials certified for the number of Huntington's CAG repeats contained in the materials.

1.1. History of Development

The National Institute of Standards and Technology (NIST) initiated development of a HD reference material in 2005 [7]. An initial panel of 10 Epstein-Barr virus-transformed lymphoblast cell lines from the National Institute of General Medical Sciences (NIGMS) Human Genetic Cell Repository at the Coriell Cell Repositories were evaluated for use [8]. During the summer of 2006, NIST and nine other laboratories evaluated an expanded set of 14 cell lines from the same source [9]. Based on these results, it was intended to produce an SRM that delivered a clinically useful range of CAG repeats in the form of PCR amplification products. This design was ultimately rejected by the HD community as not providing materials that could be adequate analogs of clinical samples.

Recognizing the need for the SRM to deliver complete genomic DNA, in 2009 NIST purchased approximately 250 µg of DNA extracted from each of six cell lines that were evaluated in the interlaboratory study. These materials provided alleles that were considered by the participants in the study to adequately span the range of clinically significant CAG repeats [9]. Table 1 lists the SRM 2393 component designation A to F, the Coriell identification code, and the study's mean and {95 % confidence intervals} determination of the number of CAG repeats delivered by each material [9, Table 1].

Table 1. SRM 2393 Candidate Component Information

SRM 2393 Component	Coriell ID Code	Number of CAG Repeats	
		$Allele_1$	$Allele_2$
A	NA20247	15 {14.1, 15.9}	29 {28.1, 29.9}
B	NA20248	17 {15.0, 19.3}	36 {35.1, 37.3}
C	NA20250	15 {14.1, 15.8}	40 {39.2, 41.0}
D	NA20208	35 {33.4, 36.3}	45 {43.5, 46.5}
E	NA20251	39 {38.1, 40.0}	50 {49.1, 50.8}
F	NA20210	17 {15.4, 18.2}	74 {72.0, 76.6}

Following acceptance testing, these materials were diluted with TE^{-4}, pH 8.0 buffer, to approximate concentrations of 10 ng/µL. For each component, 50 µL of the dilute solution was aliquoted into sterile 0.5 mL PFA vials, labeled, and stored at 2 °C to 8 °C. The certified repeat number values for all alleles were determined by counting the number of CAG units using Sanger sequencing technology and confirmed by fragment analysis (genotyping). The complete agreement among results from independent forward and reverse sequencing and fragment analysis provided the highest confidence in the assigned integer counts. The homogeneity and stability of the solutions in the vials were confirmed through a series of studies. An interlaboratory comparison demonstrated the commutability of the materials [10]. SRM 2393 became available for purchase in March 2011.

Being a new SRM type delivered in relatively novel packaging (i.e. PFA tubes), the original Certificate of Analysis developed for SRM 2393 specified a five-year period of validity, expiring 31 January 2016. Following examination of the volume and amplifiability of the solutions in the component vials, in October 2015 the expiration date was extended to 31 January 2021. A rigorous evaluation of the materials for the next extension was initiated in early 2020 to:

- determine whether the period of validity of SRM 2393 could be further extended and
- develop cost-effective methods for characterizing materials with large number of CAG repeats, if and when replacing the original version of the SRM becomes necessary.

Based on the results of the volume, amplifiability, and concordance studies reported herein, we recommend that the expiration date be extended for another five years, to 31 January 2026.

1.2. Structure of this report

Section 2 of this report describes the procedures used in 2009 and 2010 to process and evaluate the SRM 2393 components and the re-evaluation of the material in 2015. Section 3 reports the 2020 evaluation of the SRM 2393 components. Section 4 suggests a genotyping approach for certifying the number of CAG repeats in future materials.

2. Production and Certification of SRM 2393

2.1. Material Acquisition

Genomic DNA samples with a range of HD allele sizes were purchased from Coriell Cell Repositories (Camden, NJ). Five tubes of six different cell lines each containing approximately 200 µL of 250 ng/µL DNA were received 10 July 2009. The cell-line identifier and the corresponding SRM component designation are listed in Table 1.

While the high viscosity of the solutions in the tubes made sampling difficult, about 5 µL aliquots from each tube were diluted with 245 µL of TE^{-4} buffer to approximate DNA concentrations of 5 ng/µL. The diluted materials were genotyped as described in Section 2.5 to assure the identity of all the materials. After genotyping confirmed that all five tubes of each cell line contained the same cell line, the contents in each set of five tubes were pooled and diluted with TE^{-4} pH 8.0 buffer to approximate concentrations of 10 ng/µL.

2.2. Quantification of DNA amount

The diluted components were quantified with five different PCR (qPCR) assays. These assays are locus specific; the target loci are described in Table 2.

Table 2. qPCR Target Loci Used to Quantify SRM 2393 Components

Locus Chromosome	Primers (5' – 3')	Amplicon Size (bp)	Assay Type
D4S2364 Chr. 4	F_TGTTGTCTGTAGGAGCTGAGAA R_GGTGTTTGGAGATGGCTGTT	258	In-house
hTERT Chr. 5	Quantifiler Human	140	Commercial
D10S1435 Chr. 10	F_AGTGAGCCCTCGAAGAGGTT R_GTGGTGGTGTGCACCTGTAGT	355	In-house
TH01 Chr. 11	F_TGAAAAGCTCCCGATTATCCA R_CACTCGGAAGCCCTGTGTACA	62	In-house
RPPH1 Chr. 14	Quantifiler Duo	63	Commercial

SRM 2372 Human DNA Quantitation Standard component A was used to establish the calibration curves for these assays. All assays were used previously by the Applied Genetics Group. The use of different qPCR assays in which the primers are located on different chromosomes was to evaluate the presence or absence of atypical numbers of chromosomes, most importantly chromosome four where the HD locus is located.

The manufacturer's conditions were used for the hTERT (Quantifiler Human) and RPPH1 (Quantifiler Duo) assays from Thermo Fisher (Waltham, MA). Table 3 lists the qPCR conditions that were used for the D4S2364 and D10S1435 assays. Table 4 lists the qPCR conditions that were used for the TH01 assay.

Table 3. qPCR Conditions for D4S2364 and D10S1435

PCR Master Mix:	20 µL reaction volume containing: 1.2 µL each of forward and reverse primer (10 µmol/L), 10 µL of Power SYBR Green PCR Master Mix (Applied Biosystems[1]), 7.6 µL DI Water, and 2 µL test DNA.
qPCR Instrument:	7500 Real Time PCR System Applied Biosystems in 9600-emulation mode (ramp speeds of 1 °C/s)
Thermal cycling:	Heat to 95 °C for 10 min 40 cycles of: 95 °C for 15 s 58 °C for 30 s 72 °C for 45 s
Data Collection:	During the 72 °C for 45 s thermal cycling step.

Table 4. qPCR Conditions for TH01

PCR Master Mix:	20 µL reaction volume containing: 10 µL of Power SYBR Green PCR Master Mix (Applied Biosystems[1]), 1 µL each of forward and reverse primer (10 µmole/L) 6 µL DI Water, and 2 µL test DNA.
qPCR Instrument:	7500 Real Time PCR System Applied Biosystems in 9600-emulation mode (ramp speeds of 1 °C/s)
Thermal cycling:	Heat to 95 °C for 10 min 40 cycles of: 95 °C for 15 s 60 °C for 60 s
Data Collection:	During the 60 °C for 60 s thermal cycling step

Figure 1 displays the results for the five qPCR assays; only the hTERT-based commercial assay gave outlying values. Since the hTERT locus is sometimes manipulated during cell-line immortalization [11,12], the hTERT qPCR values were not used to estimate the DNA concentration of the SRM components.

Table 5 summarizes the quantitation results for the four concordant assays. It also lists the results from the hTERT assay.

Table 5. Statistical Summary of Quantitation Results

	Four Assay Consensus			
Component	Mean ng/µL	SD ng/µL	CV %	hTert ng/µL
A	8.3	0.7	8.4	17
B	10.1	1.7	17	22
C	11.8	2.1	18	26
D	8.8	1.9	22	18
E	9.6	1.6	17	24
F	9.2	1.1	12	23

[1] Now part of Thermo Fisher Scientific, Waltham, MA USA

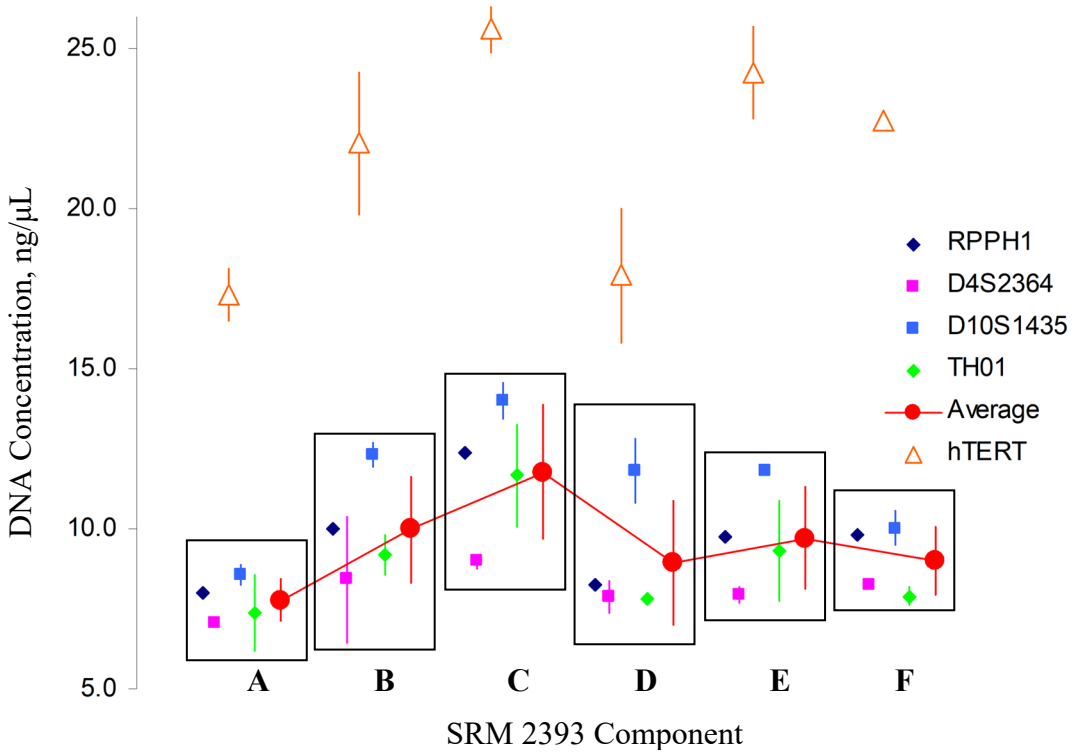


Figure 1. Quantitation Results from the Five qPCR Assays

2.3. SRM Bottling and Labeling

Microcentrifuge vials (Forensic Vial P/N 200-705-42) and closure (P/N 600-008-01) made of perfluoroalkoxy polymer (PFA) with a 500 µL volume were purchased from Savillex Corporation, Minnetonka, MN. The PFA vials were distributed into 100-unit storage boxes and autoclaved along with lids that were placed in several beakers. The autoclaving included an extended drying cycle. After autoclaving, the storage boxes and lids were stored in a laminar flow hood until use.

The six SRM 2393 components were distributed into PFA tubes over three days, two components per day, with a minimum of 511 units produced per component. A variable width eight channel pipette (LA8-300, 20µL to 300 µL; Rainin, Oakland CA) was used to fill the vials in the storage boxes. The SRM components were continuously mixed with a stir bar prior to dispensing 10 mL into a sterile disposable reagent reservoir. The material in the reagent reservoir was transferred to the PFA tubes as 50 µL aliquots using the dispensing pattern shown in Figure 2. Tube fills “1 to 12” used eight pipette tips; tube fills “13 and 14” used only two pipette tips.

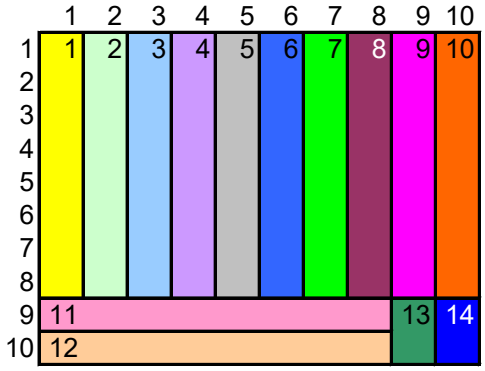


Figure 2. Fill Order per Box.

Prior to dispensing the materials, the accuracy of the volume delivered by the pipette when set to 50 µL was tested and documented. Table 6 displays the verification result as milliliters and is corrected for the 0.997 g/mL density of water near the fill temperature, 18 °C. Based on the results, the pipette was set at 51 µL to assure a minimum dispensed volume of 50 µL.

Table 6. Pipette Verification Using Gravimetric Measurement of Water.

Channel	1	2	3	4	5	6	7	8	
	0.0506	0.0502	0.0491	0.0501	0.0494	0.0497	0.0495	0.0495	
	0.0499	0.0495	0.0493	0.0492	0.0495	0.0493	0.0491	0.0491	
	0.0496	0.0495	0.0496	0.0494	0.0492	0.0494	0.0491	0.0492	
	0.0503	0.0502	0.0499	0.0492	0.0494	0.0494	0.0493	0.0493	
	0.0501	0.0495	0.0493	0.0496	0.0496	0.0496	0.0492	0.0490	
	0.0498	0.0495	0.0494	0.0492	0.0495	0.0496	0.0491	0.0498	
	0.0496	0.0496	0.0494	0.0487	0.0496	0.0493	0.0492	0.0491	
	0.0495	0.0494	0.0494	0.0499	0.0497	0.0492	0.0501	0.0483	
	0.0498	0.0499	0.0499	0.0494	0.0491	0.0498	0.0488	0.0493	
	0.0495	0.0496	0.0499	0.0494	0.0491	0.0489	0.0490	0.0490	
Avg	0.0499	0.0497	0.0496	0.0494	0.0495	0.0495	0.0493	0.0492	0.0495
sd	0.0003	0.0003	0.0003	0.0004	0.0002	0.0003	0.0003	0.0004	0.0003

After all vials in a box were capped, the box was closed and transported to a separate room for labeling. Labels furnished by the Office of Reference Materials were applied by volunteers who also checked the vials for fill and lid tightness. After labeling, the vials were stored at 2 °C to 8 °C in a locked refrigerator.

2.4. Sequencing

Both alleles for each material were Sanger sequenced in both the forward and reverse directions. Table 7 details the sequencing primers and PCR conditions used to amplify the alleles preparatory to sequencing [13].

Table 7. PCR Conditions for Amplifying Alleles

Primers (5' – 3'):	Forward Primer: GGCCTCCGGGGACTGCCGTG Reverse primer: CTGCGGCTGAGGCAGCAGCGGC
PCR Master Mix:	20 µL reaction volume containing: 0.4 µL each of forward and reverse sequencing primer (10 µmol/L), 0.5 µL dNTPs at 10 mmol/L each, 1.0 µL $MgCl_2$ solution at 25 mmol/L 4 µL of 5X Roche GC-RICH reaction buffer 0.5 µL of GC-RICH enzyme mix 2 µL GC-RICH resolution solution. 10.2 µL Water 1.0 µL DNA
Thermocycler:	GeneAmp 9700 Applied Biosystems, in 9600-emulation mode (ramp rate of 1 °C/s)
Thermal cycling:	Heat to 95 °C for 3 min 20 cycles of: 94 °C for 30 s 68 °C for 30 s 72 °C for 60 s plus 0.5 °C for each additional cycle 72 °C hold for 10 min 4 °C hold until removed from the thermal cycler

These amplified samples were separated on a polyacrylamide gel with 11 % total acrylamide and a 4 % cross link of bis-acrylamide. Figure 3 displays the separation for the six materials. Individual alleles were excised from the gel and soaked in 75 µL TE^{-4} buffer at 4 °C overnight prior to use. Two µL of each gel cut solution was reamplified using the conditions listed in Table 7.

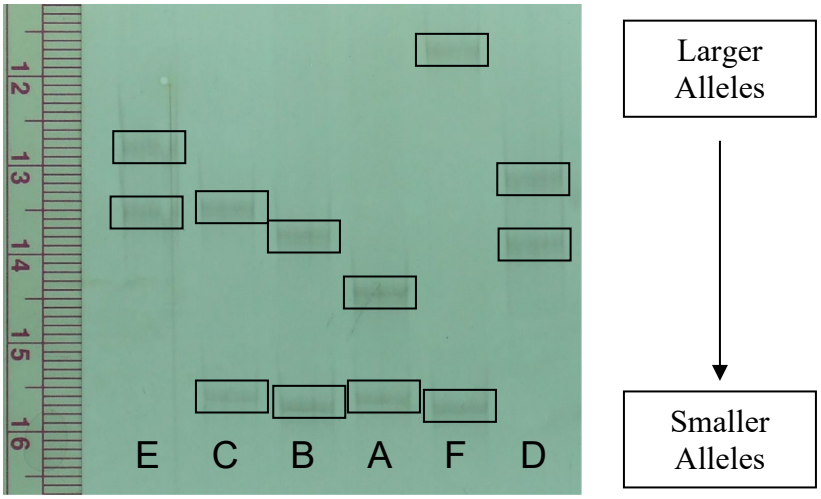


Figure 3. Gel Separation of SRM 2393 Component Alleles.
The scale to the left side indicates the distance in cm that the bands migrated from the loading well of the gel.

Separate forward and reverse sequencing reactions tubes were prepared with 5 µL of the gel cut TE^{-4} buffer and 2 µL of ExoSAP-IT P/N: 78201 L/N: 124528 (USB Corp., Cleveland, OH).

Table 8 lists the PCR conditions used.

Table 8. PCR Conditions for Forward and Reverse Sanger Sequencing

PCR Master Mix:	PCR tubes with the gel cut TE^{-4} buffer and ExoSAP-IT were heated to 37 °C for 15 min, then 80 °C for 15 min prior to the addition of the Big-Dye v3.1 sequencing reagents (Applied Biosystems[1]): Big Dye v3.1: 20 µL reaction volume using 2 µL of 5X Seq buffer, 4 µL of Big Dye Terminator, 0.16 µmol/L forward or reverse primer, and 0.5 mol/L Roche GC-RICH resolution solution.
Thermocycler:	GeneAmp 9700 Applied Biosystems, in 9600-emulation mode (ramp rate of 1 °C/s)
Thermal cycling:	25 cycles of: 96 °C for 10 s 55 °C for 5 s 65 °C for 4 min

A 2.0 µL aliquot of each purified product was diluted in 12 µL of Hi-Di formamide (Applied Biosystems[1]) and analyzed on a 16-capillary ABI Prism 3130xl Genetic Analyzer (Applied Biosystems[1]). Separations were performed on an 80 cm array using POP-7 polymer (Applied Biosystems[1]). Data were analyzed using DNASTAR Lasergene v7.1 sequence analysis software (DNASTAR Inc, Madison, WI).

The certified values for SRM 2393 components are the count of the number of CAG trinucleotide repeats within the Huntington's *HHT* locus. Figure 4 displays the locations of the sequencing and genotyping primers relative to the CAG repeat series of the symbolic map of a 21-repeat exemplar allele.

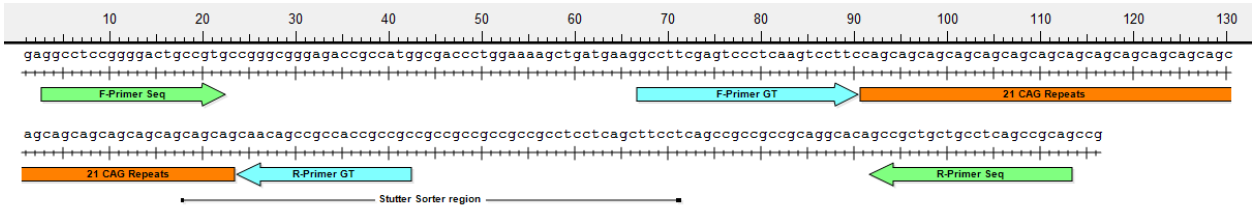


Figure 4. Sequencing Primers, Genotyping Primers, and Stutter-Sorter Locations.
Sequencing primers are green labeled arrows. Genotyping primers are the blue labeled arrows. The 21 CAG repeats are illustrated in the orange box. The gray line under the sequence marks the region where the Stutter-Sorter (explained below) is applied.

Counting the CAG repeats for alleles with more than 29 repeats is complicated by the presence of PCR amplification stutter products. HD-Stutter peaks have been described as PCR products that can be shorter by one or more repeat units than the main amplification product (parent allele) [14,15]. We refer to the stutter shorter than the parent allele by one or more repeat units as reverse stutter (*n*-1, *n*-2, …). PCR products can also be longer than the main amplification product by one or more repeat units [14,15]. We refer to these products as forward stutter (*n*+1, *n*+2, …).

Slipped strand mispairing during the PCR process has been proposed as the mechanism that produces both forward and reverse stutter. Stutter typically increases with increased repeat length and with decreased enzyme processivity [16].

Figure 5 displays the "Stutter-Sorter" tool that was developed to facilitate the interpretation of sequencing results where stutter products are visible in the generated sequences. There was complete agreement between the expected and observed sequences. Table 9 lists the number of CAG repeats for both alleles in each of the six component materials. Appendix A displays the sequencing electropherograms.

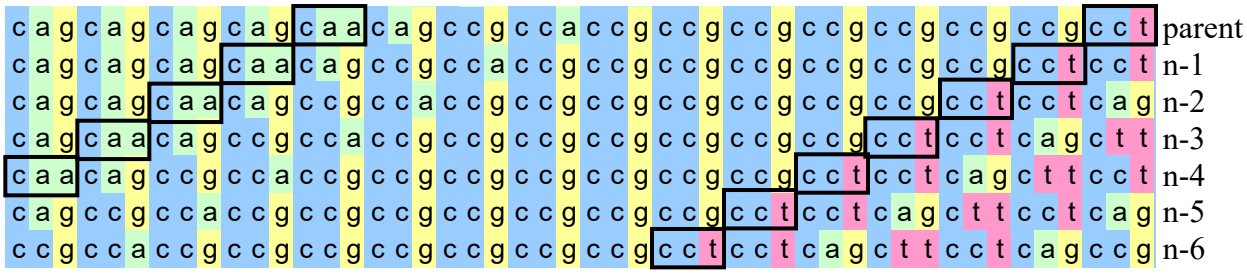


Figure 5. Stutter-Sorter Tool.

The first line of the Stutter-Sorter is the sequence of the parent allele. The CAG repeats stop prior to the "caa" shown inside the black box. The row of sequences immediately below the parent allele represents the (*n*-1) stutter product; the following rows represent (*n*-2) to (*n*-6) products. Using the appearance of the "cct" (also boxed) in the sequence helps visualize where the true "caa" is located.

Table 9. Number of CAG Repeats

Component	$Allele_1$	$Allele_2$
A	15	29
B	17	36
C	15	40
D	35	45
E	39	50
F	17	75*

*This value differs from the value in Table 1

Except for $Allele_2$ of Component F, these values are in complete accord with those determined in the 2006 interlaboratory study. The interlaboratory study determined the mean number of CAG repeats as 74 instead of 75, but with a 95 % confidence interval from (72 to 76.6) repeats. This is not unexpected since these large repeat alleles are difficult to amplify and can be associated with a great deal of both reverse and forward stutter. Recall that the Clinical Guidelines recommend a sizing accuracy of ±3 repeats at this allele size [5].

2.5. Genotyping

Genotyping was performed using the GC-RICH PCR System (Roche Applied Science[2]). Table 10 details the genotyping primers and PCR conditions used.

Table 10. PCR Conditions for Genotyping HD Alleles

Primers (5' – 3'):	Forward primer: 6FAM_GCCTTCGAGTCCCTCAAGTCCTTC Reverse primer: GCGGCGGTGGCGGCTGTTG
PCR Master Mix:	10 µL reaction volumes using a master mix containing: 0.25 µL of GC-RICH enzyme mix, 2.0 µL GC-RICH (1 mol/L) reaction buffer, 0.5 µL $MgCl_2$ stock solution (25 mmol/L) 1.5 µL GC-RICH resolution solution (0.5 mol/L), 0.2 µL each of forward and reverse primer (10 µmol/L), 0.25 µL dNTPs (80 mmol/L), 2.0 µL DNA, and 3.1 µL PCR grade water
Thermocycler:	GeneAmp 9700 Applied Biosystems, in 9600-emulation mode (ramp rate of 1 °C/s)
Thermal cycling:	Heat to 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 + 0.5 °C for each additional cycle 72 °C hold for 7 min 4 °C hold until removed from the thermal cycler

A 1.0 µL aliquot of the amplified product was diluted in 14 µL of Hi-Di formamide and 0.4 µL GS500-LIZ internal size standard (Applied Biosystems[1]) and analyzed on a 16-capillary ABI Prism 3130xl Genetic Analyzer using filter dye set G5. Separations were performed on a 36 cm array using POP-4 polymer (Applied Biosystems[1]). Data were analyzed using GeneMapper ID v3.2 (Applied Biosystems[1]). Information collected included base pair (bp) size, peak height in relative fluorescence units (RFUs) and peak area of the PCR reaction products (parent allele, reverse stutter, and forward stutter) reaching a 50 RFU peak height analytical threshold.

The sequence counts were confirmed by direct genotyping. Figure 6 displays representative electropherograms for the six SRM 2393 components in the 6FAM (blue) dye channel.

[2] Roche products are now exclusively distributed by MilliporeSigma (previously known as Sigma-Aldrich), owned by Merck KGaA (Darmstadt, Germany)

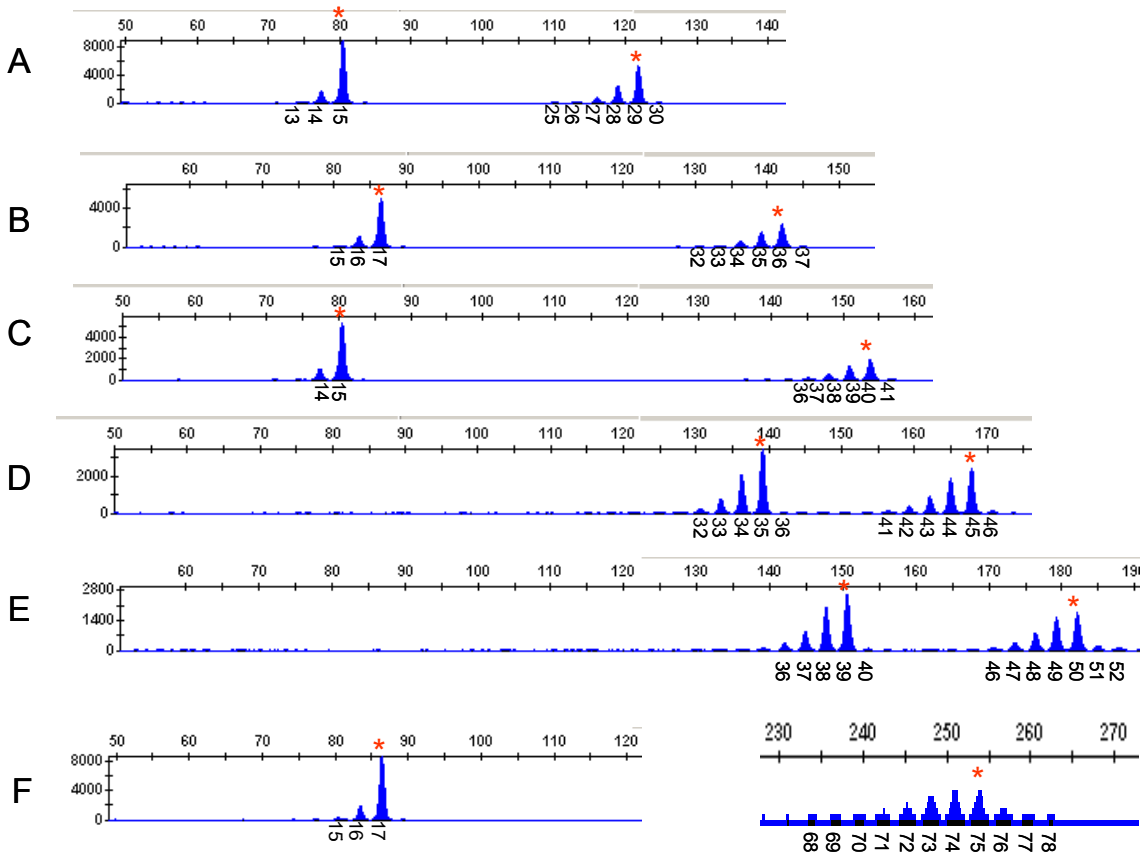


Figure 6. Genotyping results for SRM 2393 Components A to F.

The Y-axis of each panel displays relative fluorescence units (RFUs); the RFU scale (not shown) for the 75-repeat allele of Component F has been expanded from 230 bp through 270 bp to facilitate visualizing the peaks. The X-axis of each panel is shown above the peaks; they display base pair size as estimated from an internal size standard (ILS) 500 CC5 (Promega Corp, Madison WI). Values below the blue horizontal lines are the number of CAG repeats for PCR products having peak heights of 50 RFU or more. The parent allele in each set of PCR amplification products is marked with a "★".

For the relatively stutter-resistant amplification conditions described in Table 10, the parent allele amplification product provides the tallest peak for all but the large (75-repeat) allele of Component F. There are no interpretation issues with alleles having 15 through 50 CAG repeats; however, genotype interpretation with alleles larger than 50 can be complicated by the increase in both forward and reverse stutter products with increasing number of CAG repeats in the parent allele.

Figure 7 summarizes the peak heights of stutter products relative to the peak heights of three trial parent alleles as functions of the number of repeats for the parent allele. The observed peak height ratios (PHRs) for the large allele of component F are well-predicted by relatively smooth functions only when the 75-repeat allele is assigned as the parent allele.

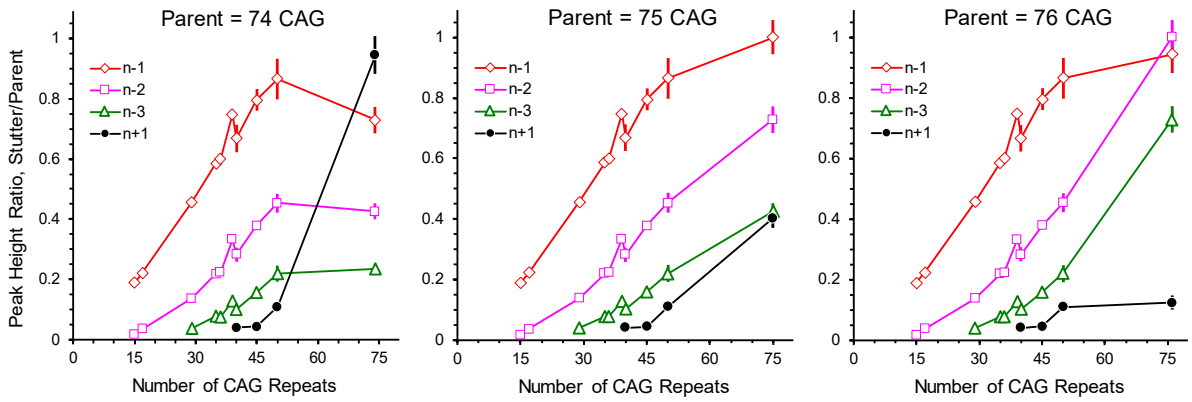


Figure 7. Stutter/Parent PeakHeight Ratios for Three Potential Parent Alleles for Component F, large allele

These three panels display the pattern of stutter-to-parent peak height ratios assuming that the parent allele of Component F's large allele is (74, 75, or 76) CAG repeats. Each symbol represents the {stutter/parent peak-height ratio, number of CAG repeats} for one allele of one of the SRM 2393 components. The open diamond, square, and triangle symbols represent the *n*-1, *n*-2, and *n*-3 reverse stutter; the solid circles represent the *n*+1 forward stutter. The parent allele for all of the HD alleles except the Component F large allele are well-established; the only changes in the three panels are the ratios for the Component F large allele.

2.6. Homogeneity Testing

The homogeneity of SRM 2393 components with regards to CAG repeat number was assessed based on the reproducibility of genotyping results. One randomly selected vial from each of the five full boxes of 100 vials of each of the six components were used. Two aliquots from each of the selected vials were PCR amplified as described in Table 10.

Table 11 displays the results of the observed peak area ratios calculated as the ratio of the area for a given amplification product divided by the sum of the peak areas for the parent allele and all the stutter alleles having peak height greater than 50 RFU. The mean (Mean), standard deviation (SD), and coefficient of variation expressed as a percentage (CV = 100×SD/Mean) for the *n*-1 stutter alleles, the parent allele, and (where present) the *n*+1 stutter alleles are listed.

Table 11. Peak Area Ratio Homogeneity Assessment

Component	Number Repeats	*n*-1 Stutter Allele			Parent Allele			*n*+1 Stutter Allele		
		Mean	SD	CV	Mean	SD	CV	Mean	SD	CV
A	15	0.156	0.004	2.8	0.831	0.005	0.6			
	29	0.280	0.002	0.9	0.612	0.003	0.4			
B	17	0.177	0.002	1.0	0.795	0.003	0.3			
	36	0.316	0.003	0.9	0.527	0.003	0.6			
C	15	0.158	0.004	2.5	0.833	0.009	1.1			
	40	0.319	0.011	3.3	0.478	0.015	3.1	0.019	0.001	6.8
D	35	0.311	0.002	0.8	0.531	0.004	0.8			
	45	0.332	0.007	2.2	0.418	0.010	2.4	0.018	0.001	4.9
E	39	0.335	0.003	0.8	0.448	0.006	1.4			
	50	0.317	0.013	4.0	0.366	0.015	4.0	0.040	0.002	5.1
F	17	0.177	0.004	2.3	0.794	0.007	0.9			
	75	0.257	0.007	2.8	0.243	0.008	3.3	0.103	0.004	3.8

2.7. Interlaboratory Commutability Study

Prof. Oliver Quarrell, a clinical geneticist at the Children's Hospital, Sheffield, UK working as part of the European HD Network [10], was involved in resolving discrepancies in HD sizing results from different European laboratories; he strongly advocates the use of certified reference materials. After seeing information posted on the Applied Genetics Group's website [17], he requested access to the SRM materials. Two units of the pre-release SRM were shipped to him July 6, 2010.

Results returned from his interlaboratory study where the SRM 2393 components were blinded to the study participants are concordant with the certified values except for the Component F large allele being called 74 instead of 75 [10]. As with the similar discordance with the 2006 interlaboratory study, this is not unexpected since these large repeat alleles are difficult to PCR amplify and can be associated with a great deal of both reverse and forward stutter.

2.8. 2015 Re-evaluation of Stability and Expiration Extension

Two units of the SRM components were randomly taken from the storage boxes. The samples were analyzed as detailed in Section 2.5 and Table 10 of this report. All samples were concordant with the previous data. A memo was sent to extend the expiration date by 5 years to January 2021.

3. Re-Evaluation of SRM 2393 in 2020

3.1. Confirming Amplifiability and DNA Concentration

To confirm the amplifiability of the SRM 2393 components, two droplet digital PCR (ddPCR) assays were used: NR4Q and NEIF. ddPCR methods were chosen over the qPCR methods used in the initial certification of SRM 2393 for their ability to quantify a sample without the use of a standard curve for calibration. The NEIF dPCR assay was used because of its strong performance in both our droplet (ddPCR) and chamber (cdPCR) dPCR platforms. NR4Q was chosen because its target locus is on chromosome 4, the same as HD. Comparing the results from the different assays determines whether the copy numbers are the same from both chromosomes; a normalcy check since SRM 2393 was produced from immortalized cell line DNA. These assays are defined in Table 12.

Table 12. NIST developed digital PCR assays.

Assay Target	Chromosome, Band Accession #	Primers and Probe (5' – 3') [a]	Amplicon Length, bp
NEIF Gene EIF5B	Chr 2, p11.1-q11.1 NC_000002.12	F: GCCAAACTTCAGCCTTCTCTTC R: CTCTGGCAACATTTCACACTACA P^{B+}:TCATGCAGTTGTCAGAAGCTG	67
NR4Q Gene DCK	Chr 4, q13.3-q21.1 NC_000004.12	F: TGGTGGGAATGTTCTTCAGATGA R: TCGACTGAGACAGGCATATGTT P^{B+}: GTATGAGAAACCTGAACGATGGT	83

a F: Forward primer, R: Reverse primer, P^{B+}: Blackhole Plus quencher probe (FAM labeled)

One unit of SRM 2393 was pulled from the 2 °C to 8 °C storage for analysis. The six components and non-template controls (master mix with PCR grade water replacing the DNA solution) were amplified in triplicate using both assays with the equipment and PCR conditions listed in Table 13.

Table 13. PCR Conditions for NEIF and NR4Q ddPCR Assays

Droplet Generator:	QX200 AutoDG Droplet Digital PCR System (Bio-Rad, Hercules, CA)
PCR Master Mix:	25 µL reaction volumes using a master mix containing: 12.5 µL of ddPCR Supermix for Probes (no dUTP) (Bio-Rad), 0.94 µL of 10 µmol/L each of forward and reverse primers, 0.63 µL of a 10 µmol/L probe solution, 7.5 µL PCR grade water, and 2.5 µL of (5 to 14) ng/µL DNA solution.
Thermocycler:	ProFlex (Applied Biosystems[1]) using ramp rate of 2.5 °C/s:
Thermal cycling:	Heat to 95 °C for 10 min 60 cycles of: 94 °C for 30 s 61 °C for 60 s 98 °C hold for 10 min 4 °C hold until removed from the thermal cycler
Droplet Reader:	QX200 Droplet Reader (Bio-Rad)

All components amplified as expected using both assays. The droplet size was estimated to be 0.738 nL for the conversion of the ddPCR measured copies per droplet to [DNA] as ng/µL using the formula presented in [18]. The droplet size is an estimate based on the droplet size measurements performed several years ago with the ddPCR Supermix for Probes (no dUTP) (Bio-Rad) [19].

Figure 8 demonstrates that the NEIF and NR4Q assays produced very similar values. However, the ddPCR estimates of DNA concentration differ somewhat from the consensus qPCR results listed in Table 5. Interestingly, the range in [DNA] values among the six components is unchanged. Regardless of whether the differences are measurement artifacts or result from changes in extract volume, the DNA concentration for all components remain fit for purpose.

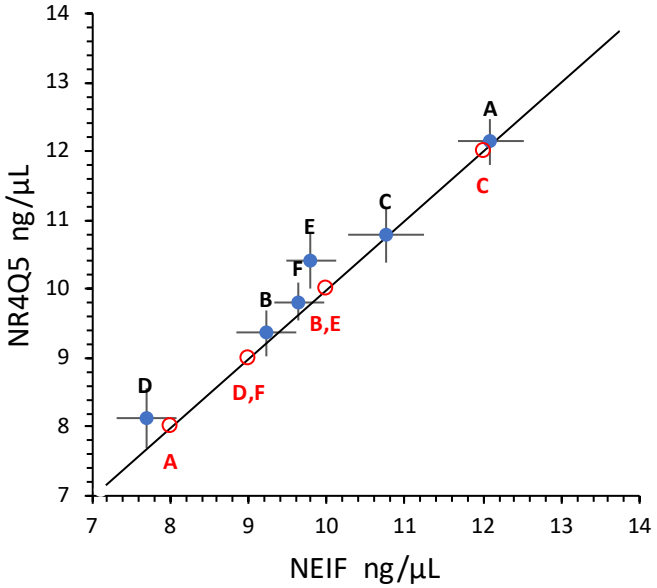


Figure 8. ddPCR of SRM 2393 components with two assays.

The vertical axis reports NR4Q assay results in ng/µL. The horizontal axis reports NEIF assay results in ng/µL. Solid blue circles display the {NEIF, NR4Q5} pairs; the bars represent one standard deviation of the technical replicates. The black diagonal line represents equality between the assays. The qPCR consensus estimates for the six components are shown as open red circles located on the equality line.

3.2. Evaluation of the Fraction of Single Stranded DNA in the Solutions

Given a dPCR assay that is extremely efficient, chamber digital PCR (cdPCR) can provide direct visualization of the proportion of single stranded DNA (ssDNA) in a predominately double stranded DNA (dsDNA) sample. A detailed description of the method can be found in NIST Special Publication (SP) 1200 27 [20].

The NEIF assay has a good PCR amplification efficiency. The proportion of ssDNA was estimated using results from our Fluidigm BioMark real time cdPCR system (South San Francisco, CA USA) as described in Table 14. Eight panels (technical replicates) were acquired for each of the six components using chip 1670102075.

Table 14. cdPCR Conditions for NEIF

cdPCR Instrument:	Fluidigm BioMark (South San Francisco, CA) real time cdPCR system with IFC controller.
cdPCR Digital Array:	BioMark 48.770, a disposable microfluidic device ("chip") that has 48 panels of 770 reaction chambers, each chamber of nominal volume 0.85 nL.
PCR Master Mix:	For each panel, 4 µL of mastermix containing: 2.5 µL Gene Expression Master mix (Thermo Fisher) 0.188 µL of 10 µmol/L each of forward and reverse primers 0.125 µL of a 10 µmol/L probe solution; 1.0 µL PCR grade water; 0.5 µL of GE Buffer (Fluidigm), and 0.5 µL of (5 to 14) ng/µL DNA solution.
Thermal cycling:	Set temperature ramp to 2 °C/s Heat to 95 °C for 10 min 60 cycles of: 95 °C for 15 s 61 °C for 60 s
Data Collection:	At the end of every cycle

Figure 9 displays the cdPCR evaluations for the SRM 2393 components using the “Poisson” component of the NIST-developed “cdPCR_OgiveMaker.xlsm” analysis system. This analysis technique has only recently been developed and validated [21]. The proportion of ssDNA in the components range between 12 % and 15 %, indicating that the ddPCR quantitation values are biased high by 6 % to 8 %. The current data can be used as the baseline for monitoring future changes to these SRM components.

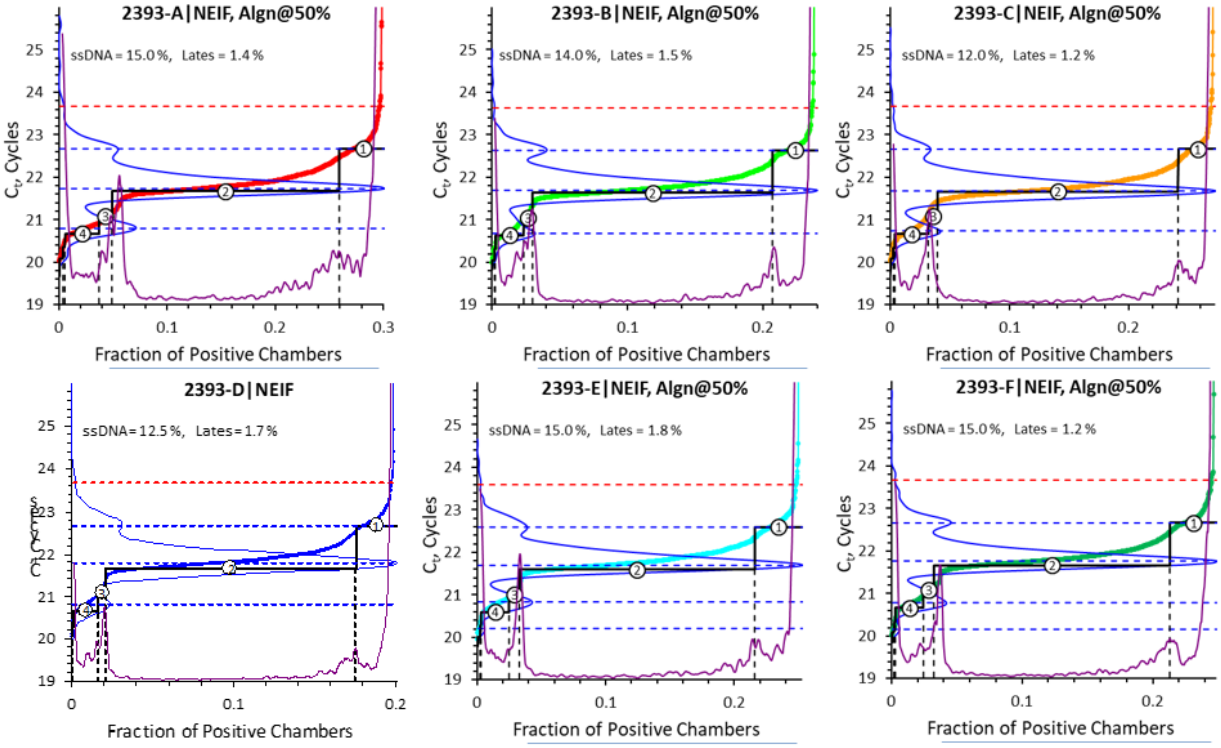


Figure 9. cdPCR estimates of ssDNA in SRM 2393 components.

The thick variously colored curves in the six panels are the observed cumulative distributions (ogives) of crossing threshold (C_t) values for six components as functions of the fraction of positive chambers (F_{pos}) for each set of technical replicates. These results were realized using the manufacturer's "linear derivative" analysis, with the ogives for the individual replicates aligned at their median (50%) value. The blue curve plotted along the C_t axis of each panel is a kernel-density representation of the number of positive chambers at given C_t. The purple curve along the F_{pos} axis estimates the ogive rate of change derivative, $d(C_t)/d(F_{pos})$. The thick black "staircase" curve is the idealized ogive for a sample consisting of the stated percentage of single-stranded DNA (ssDNA) in an extract that is otherwise double-stranded DNA (dsDNA). The middle of each tread is labeled with the number of entities per chamber that produce the tread. The horizontal blue dashed lines mark the location of kernel density peak maxima. The C_t location of the staircase is defined by alignment of the uppermost tread and the kernel density function's uppermost peak. The black vertical dashed lines connect the location of the stair risers with the horizontal axis; ideally, they bisect a peak in the derivative (purple curve). The horizontal red dotted line is one C_t above the one-entity tread; C_t values above this line are considered "late starters" and are excluded from the optimization process. The text below the title of each panel reports the estimated percentage of ssDNA and late starters in the sample. The ssDNA and "late starter" percentages are determined by matching the idealized staircase function to the empirical derivative functions using the "Poisson" component of the cdPCR_OgiveMaker software. For further information see [20].

3.3. Selection of Genotyping Primer Set

The current Clinical Guidelines indicate that the length of the CAG repeat alone is responsible for the HD phenotype [5]. However, there are two repeat regions just 12 base pairs (bp) downstream of the clinically significant CAG repeat that complicate accurate assessment of the CAG repeat number: a $[CCG]_{7\text{-}12}$ and a $[CCT]_{1\text{-}2}$ [22]. There are also two repeat regions more than 200 bp upstream of the CAG region that must be considered in assay design: 6 bp repeat $[GGGGGC]_{1\text{-}2}$ and 20 bp repeat $[GGCCCCGCCTCCGCCGGCGC]_{1\text{-}3}$ [23]. While there are published genotyping primer sets that include the CCG/CCT repeats [9,24], we use a primer set that amplifies only the region of interest displayed in Figure 4. Figure 10 displays this set along with the other repeat regions and a selection of primers used to genotype and sequence other repeat regions.

Figure 10. CAG Primer Set Mapped to a CAG19 Repeat Sample.

The predictive HD phenotype19 CAG repeats are in pink. The forward and reverse primers used are in the red-bordered boxes: the FAM-labeled forward primer "F Primer next to CAG repeat" is in orange to the left of the repeats and the reverse primer "primer 11pub"is in lavender immediately to the left of the repeats. Additional repeat motifs, not part of the HD phenotype, are highlighted in green, teal, and red. A selection of other primer sets used in HD studies are also displayed. While none of the additional repeat regions are currently used for diagnostic purposes, knowledge of the existence, positions, and variability of these repeat regions remain important to avoid inadvertently including them in assay development.

There have been customers that claim the certified numbers of CAG repeats in some of the components of SRM 2393 are incorrect. After inquiring where their primer sets were placed, we determined that they were amplifying an additional downstream repeat area that changes the apparent size. We investigated these other repeat regions and shared our results on a poster that was presented at the 2012 Association for Molecular Pathology (AMP) Annual meeting entitled "Metrology for Huntington's Disease PCR Assays: Regions to Avoid." Appendix B provides an image of this poster.

3.4. Defining the Virtual Allelic Ladder used for Genotyping

The 2011 certification and the 2015 recertification of SRM 2393 used a 16-capillary ABI Prism 3130xl Genetic Analyzer (Applied Biosystems[1]) to genotype PCR-amplified genotyped HD components. Our current capillary electrophoresis instrumentation is a 24-capillary 3500xl Genetic Analyzer (Applied Biosystems[1]). Based on the increased sensitivity of the 3500xl instrument, modifications to the previous PCR amplification methods for genotyping were required.

New virtual allelic ladders were generated to allow the GeneMapper ID-X, v1.5 (Thermo Fisher) software to assign the genotypes to the samples. Sixteen previously tested HD cell line samples, stored at -80 °C, and six previously tested population samples were used to produce the virtual allelic ladder bins. In this case the CAG repeat stutter was useful. The samples had enough diversity and stutter to limit the number of allelic ladder bins that required definition by interpolation. Table 15 lists the cell lines and population samples that were used with their predicted HD genotypes and the ladder alleles that they provide.

Table 15. Samples Used to Produce Virtual Allelic Ladder Bins and Resulting Alleles

Sample	Genotype
CD00022c	18,31
GT37862	21,22
JT51471	12,31
NA20206	17,18
NA20207	19,21
NA20208	35,45
NA20209	45,46
NA20210	17,75
NA20212	19,34
NA20245	15,15
NA20246	15,24
NA20247	15,29
NA20248	17,36
NA20249	22,39
NA20250	15,40
NA20251	39,50
NA20252	22,65
NA20253	22,101
PT84236	14,15
TT51435	10,28
WT51362	17,26
ZT80786	17,20

Sample	Ladder Allele
TT51435	10
JT51471	12
PT84236	14
NA20245	15
NA20245	15
NA20246	15
NA20247	15
NA20250	15
PT84236	15
NA20206	17
NA20210	17
NA20248	17
WT51362	17
ZT80786	17
CD00022c	18
NA20206	18
NA20207	19
NA20212	19
ZT80786	20
GT37862	21
NA20207	21
GT37862	22
NA20249	22
NA20252	22
NA20253	22
NA20246	24
WT51362	26
TT51435	28
NA20247	29
CD00022c	31
JT51471	31
NA20212	34
NA20208	35
NA20248	36
NA20249	39
NA20251	39
NA20250	40
NA20208	45
NA20209	45
NA20209	46
NA20251	50
NA20252	65
NA20210	75
NA20253	101

The new virtual allelic ladders were built using two commercial kits designed for amplification of GC-rich sequences such as the CAG repeat sequences. The two kits are:

- Roche GC-RICH PCR System, dNTPack (MilliporeSigma[2]).
  This system, assigned the short name "Roche", is composed of a special enzyme blend of thermostable Taq DNA Polymerase and Tgo DNA Polymerase, a thermostable enzyme derived from *Thermococcus gorgonarius* with a proofreading (3′-5′ exonuclease) activity. This polymerase mixture outperforms Taq DNA Polymerase in respect to yields, fidelity and specificity beside the possibility to amplify fragments up to 5 kb in length. The GC-RICH PCR reaction buffer in combination with the included GC-RICH resolution solution allows to amplify very efficiently difficult templates like GC-rich targets [25].
- PrimeSTAR HS DNA Polymerase with GC Buffer (Takara Bio USA).
  This system, assigned the short name "PrimeSTAR", is designed for high-fidelity PCR amplification of GC-rich templates (75% or greater GC content) and is based on a unique, high-fidelity PCR polymerase, PrimeSTAR HS DNA Polymerase. The GC buffer supplied with PrimeSTAR HS facilitates robust, efficient, and accurate extension through even highly GC-rich template regions. It offers both maximal accuracy and better amplification efficiency of high-GC templates than regular Taq polymerase [26].

Table 16 details the PCR amplification conditions used for the Roche system; Table 17 details the conditions used for the PrimeSTAR system.

Table 16. PCR Conditions for Roche GC-RICH PCR System, dNTPack

Primers (5' – 3'):	Forward primer: 6FAM-GCCTTCGAGTCCCTCAAGTCCTTC Reverse primer: GCGGCGGTGGCGGCTGTTG
PCR Master Mix:	15 µL reaction volumes using a master mix containing: 3 µL of 5X GC-Rich PCR reaction buffer 0.375 µL each of Forward and Reverse primer (10µmol/L) 0.375 µL PCR Grade Nucleotide Mix 1 µL GC-Rich resolution solution (5 mol/L) 1 µL $MgCl_2$ stock solution (25 mmol/L) 0.375 µL GC-Rich Enzyme mix 5.5 µL Water, PCR Grade 3 µL DNA
Thermocycler:	Proflex (Applied Biosystems[1]) using ramp rate of 2.0 °C/s:
Thermal cycling:	Heat to 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 plus 5 s for each additional cycle 72 °C hold for 7 min 4 °C hold until removed from the thermal cycler

Table 17. PCR Conditions for PrimeSTAR HS DNA Polymerase with GC Buffer

Primers (5' – 3'):	Forward primer: 6FAM-GCCTTCGAGTCCCTCAAGTCCTTC Reverse primer: GCGGCGGTGGCGGCTGTTG
PCR Master Mix:	15 µL reaction volumes using a master mix containing: 7.5 µL 2X PrimeSTAR GC Buffer (Mg^{2+} plus) 0.375 µL each of Forward and Reverse primer (10µmol/L) 1.2 µL of dNTP mixture (2.5 mmol/L each) 0.20 µL of PrimeSTAR HS DNA Polymerase 1.5 µL of DNA sample 3.85 µL of PCR grade water
Thermocycler:	Proflex (Applied Biosystems[1]) using ramp rate of 2.0 °C/s:
Thermal cycling:	30 cycles of: 98 °C for 30 s 65 °C for 5 s 72 °C for 60 s 4 °C hold until removed from the thermal cycler

After amplification, 1 µL of the amplicon was mixed with 9 µL of an injection mastermix consisting of 8.7 µL of HiDi formamide (Thermo Fisher) and 0.3 µL GeneScan (GS) 500 Liz size ladder (Thermo Fisher) in a 96-well sample plate. The plate was analyzed on an Applied Biosystems 3500xL Genetic Analyzer for Human Identification instrument equipped with a 36 cm 24-capillary array, filled with POP-4 polymer and run at 60 °C. The instrument applied a DS-33 matrix called "G5 dye set".

Multiple runs using the same amplified products were injected for (24, 15, 8, or 5) s. Changing the injection duration produced on-scale peak heights. The preferential amplification of the smaller alleles results in those small alleles being off scale in order to clearly distinguish the larger HD alleles. Peaks from various runs that were on-scale, were sorted by size and combined into size bins with the average mean and standard deviations calculated.

Table 18 lists the number of on-scale determinations of alleles of with CAG repeats from 9 to 107, their mean value, and when there were at least two determinations, the standard deviation and the relative standard deviation expressed in percentage form for both amplification systems.

Table 18. Allele Sizes Used for the Virtual Allelic Ladders

	Roche GC Rich PCR System [a]				PrimeSTAR HS with GC buffer [a]				
Allele	*n*	Mean, bp	SD, bp	CV, %	*n*	Mean, bp	SD, bp	CV, %	Δ[b], bp
9	1	62.58			7	61.12	0.07	0.108	1.46
10	1	65.94			10	64.48	0.30	0.460	1.46
11	1	69.01			11	67.55	0.04	0.065	1.46
12	5	72.22	0.12	0.161	16	70.76	0.15	0.209	1.46
13	5	75.12	0.10	0.130	22	73.79	0.06	0.085	1.34
14	9	78.01	0.06	0.073	24	76.81	0.15	0.190	1.20
15	11	80.78	0.18	0.224	30	79.77	0.16	0.196	1.01
16	9	83.94	0.11	0.128	28	82.85	0.04	0.052	1.09
17	9	86.81	0.18	0.209	23	85.81	0.05	0.057	1.00
18	7	89.86	0.16	0.179	27	88.79	0.06	0.066	1.07
19	7	92.95	0.14	0.147	26	91.79	0.09	0.096	1.16
20	7	95.81	0.12	0.122	20	94.73	0.10	0.105	1.08
21	7	98.73	0.13	0.133	19	97.68	0.12	0.124	1.06
22	5	101.60	0.07	0.072	18	100.61	0.13	0.131	0.99
23	8	104.50	0.11	0.103	17	103.49	0.16	0.151	1.01
24	4	107.26	0.17	0.156	9	106.26	0.14	0.130	1.00
25	3	110.29	0.08	0.073	11	109.13	0.11	0.100	1.15
26	4	113.21	0.05	0.041	11	112.02	0.09	0.078	1.18
27	4	116.00	0.16	0.134	11	114.95	0.12	0.103	1.05
28	4	118.88	0.18	0.151	11	117.85	0.08	0.066	1.02
29	4	121.83	0.17	0.140	11	120.78	0.09	0.077	1.05
30	4	124.90	0.06	0.047	8	123.75	0.09	0.071	1.14
31	3	127.82	0.09	0.068	9	126.73	0.12	0.094	1.09
32	5	130.85	0.06	0.049	11	129.66	0.10	0.081	1.19
33	3	133.82	0.05	0.039	9	132.66	0.11	0.085	1.16
34	3	136.81	0.05	0.033	10	135.68	0.12	0.091	1.13
35	5	139.97	0.12	0.085	11	138.70	0.12	0.089	1.27
36	4	143.14	0.02	0.015	12	141.91	0.14	0.098	1.23
37	4	146.32	0.04	0.027	10	145.15	0.14	0.095	1.17
38	3	149.50	0.06	0.039	8	148.35	0.12	0.082	1.15
39	3	152.60	0.06	0.038	10	151.48	0.12	0.079	1.12
40	2	155.59	0.07	0.045	10	154.46	0.10	0.062	1.13
41	3	158.46	0.05	0.031	6	157.43	0.12	0.076	1.03
42	2	161.36	0.00	0.000	6	160.34	0.06	0.035	1.02
43	2	164.28	0.01	0.009	5	163.27	0.08	0.049	1.01
44	2	167.20	0.01	0.008	5	166.14	0.12	0.070	1.06
45	2	170.12	0.03	0.017	5	169.04	0.11	0.065	1.08
46	2	173.04	0.09	0.053	6	171.93	0.12	0.072	1.11
47	2	175.93	0.04	0.024	4	174.87	0.12	0.067	1.06
48	1	178.82			4	177.75	0.15	0.083	1.07
49	1	181.69			2	180.65	0.14	0.078	1.04
50	1	184.55			2	183.52	0.20	0.108	1.03
51	1	187.42			2	186.35	0.19	0.102	1.07
52		*190.29*			2	189.22	0.25	0.135	
53		*193.14*			2	191.84	0.06	0.033	
54		*195.99*			2	194.61	0.17	0.087	
55		*198.84*			2	197.53	0.19	0.097	
56		*201.69*			2	200.35	0.21	0.102	
57		*204.54*			2	203.19	0.08	0.038	
58		*207.39*			2	206.04	0.08	0.041	
59	1	210.20			3	208.90	0.28	0.135	1.30
60	1	213.02			3	211.77	0.26	0.124	1.25
61	1	215.85			3	214.60	0.27	0.126	1.26
62	1	218.68			3	217.43	0.29	0.135	1.25
63	1	221.53			3	220.28	0.30	0.136	1.26

	Roche GC Rich PCR System [a]				PrimeSTAR HS with GC buffer [a]				
Allele	*n*	Mean, bp	SD, bp	CV, %	*n*	Mean, bp	SD, bp	CV, %	Δ[b], bp
64	1	224.38			3	223.13	0.32	0.143	1.25
65	1	227.24			3	226.00	0.26	0.115	1.25
66	1	230.11			3	228.82	0.30	0.131	1.30
67	1	232.98			3	231.75	0.29	0.123	1.23
68	1	235.87			3	234.59	0.33	0.139	1.29
69	1	238.58			3	237.43	0.23	0.096	1.15
70	1	241.47			3	240.29	0.21	0.087	1.19
71	1	244.28			2	243.32	0.34	0.139	0.96
72	1	247.19			2	246.18	0.32	0.129	1.01
73	1	250.01			2	249.09	0.37	0.148	0.92
74	1	252.94			2	251.96	0.35	0.138	0.99
75	1	255.78			2	254.79	0.39	0.153	1.00
76	1	258.62			2	257.67	0.37	0.143	0.95
77	1	261.58			2	260.62	0.42	0.160	0.96
78		*264.55*			2	263.53	0.39	0.148	
79		*267.48*			1	266.07			
80		*270.41*			1	269.01			
81		*273.34*			1	271.85			
82		*276.27*			1	274.81			
83		*279.20*			1	277.67			
84		*282.13*			1	280.54			
85		*285.06*				*283.49*			
86		*287.99*				*286.44*			
87		*290.92*				*289.38*			
88		*293.85*				*292.33*			
89		*296.78*				*295.28*			
90		*299.71*				*298.22*			
91		*302.64*				*301.17*			
92		*305.57*				*304.12*			
93		*308.50*				*307.06*			
94	1	311.45			1	310.01			1.44
95	1	314.29			2	313.25	0.28	0.090	1.04
96	1	317.35			2	316.20	0.33	0.103	1.15
97	1	320.29			2	319.15	0.21	0.064	1.15
98	1	323.24			2	322.10	0.24	0.075	1.14
99	1	326.07			2	325.06	0.29	0.089	1.01
100	1	329.01			2	327.95	0.25	0.078	1.06
101	1	331.95			2	330.90	0.31	0.094	1.05
102	1	334.90			2	333.80	0.28	0.085	1.10
103	1	337.74			2	336.70	0.25	0.076	1.04
104					1	339.44			
105					2	342.45	0.40		
106					1	345.00			
107					1	347.83			
								n:	72
								Mean, bp:	1.13
								Standard Deviation, bp:	0.13

a *n*: number alleles with on-scale allelic heights.
Mean: mean size, in bp, of the alleles as provided by GeneMapper ID*X* software.
Values in *red italic* are interpolated from the available experimentally determined allelic sizes.
SD: standard deviation, in bp.
CV: relative standard deviation, in %: 100 SD/Mean.

b Δ: Difference in allelic sizes produced by the Roche and PrimeSTAR amplification kits, in bp.

The relatively constant difference in the amplicon size between the two PCR amplification kits, 1.13 bp ± 0.13 bp, may be related to the different DNA polymerases in each kit. The linear relationships between the allele size in bp and the number of CAG repeats for the two kits is shown in Figure 11. The structure of the residuals is related to the number, spacing, and electrophoretic mobility of the components of the GS500-LIZ internal size standard used to define the bp scale.

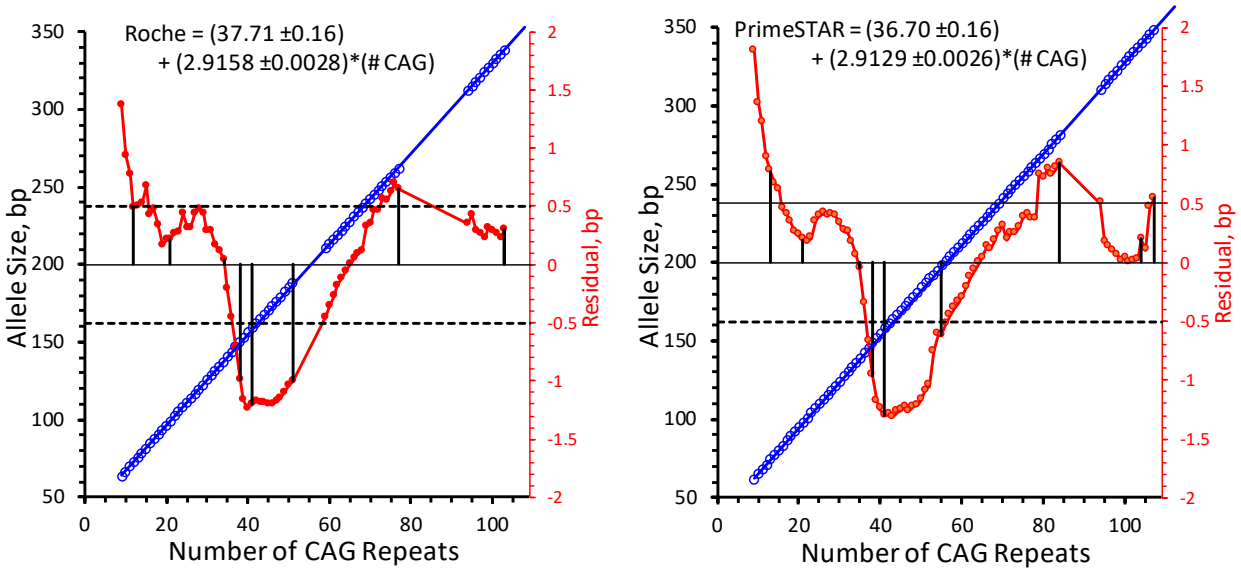


Figure 11. Allele Size as Functions of Number of CAG Repeats.

The panel to the left displays the linear relationship between allele size in basepairs (bp) and the number of CAG repeats in the allele for the Roche GC-RICH (Roche) PCR system. The panel to the right displays the relationship for the PrimeSTAR HS DNA Polymerase (PrimeSTAR) system. The open blue circles represent observed alleles; the diagonal blue lines through the data represent linear least squares regression fits to the data; the coefficients of the fits are displayed at the top of each panel. The solid red circles and connecting red lines display the residuals from the linear relationships. The horizontal solid black lines mark zero residual; the dashed black lines bracket the (-0.5 to 0.5) bp interval on the residuals. The vertical black lines mark the locations of the GS500-LIZ internal sizing ladder components.

3.5. Documenting Stutter

In general, the PCR amplification of large alleles is not as efficient as the amplification of smaller alleles. There is great disparity in the peak heights between the smaller HD alleles and the larger HD alleles within a sample. Figure 12 compares the electropherograms of CD00022 (18,31) amplified with both the Roche and PrimeSTAR kits. Figure 13 likewise compares the electropherograms of NA20210 (17,75), the SRM 2373 Component F material, amplified with both the Roche and PrimeSTAR kits.

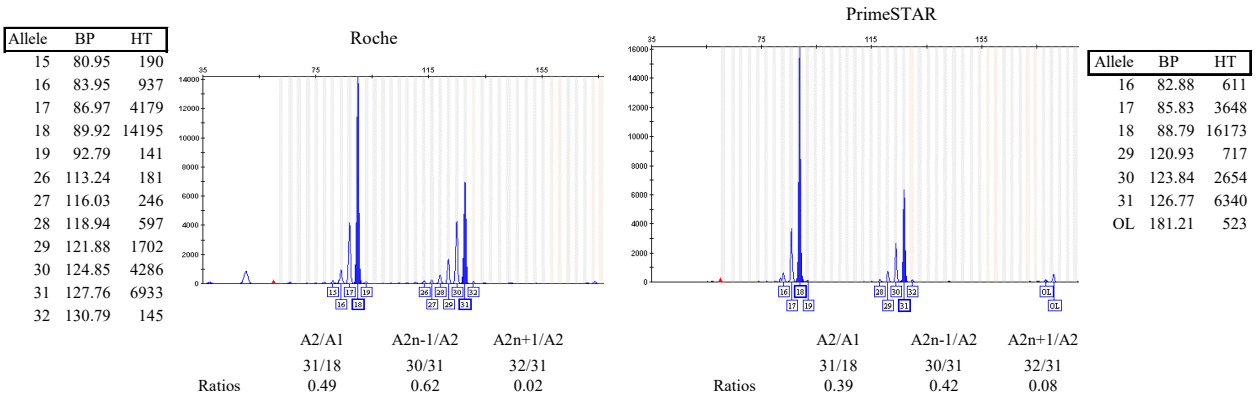


Figure 12. Comparison of Roche and PrimeSTAR Small Allele Amplification.

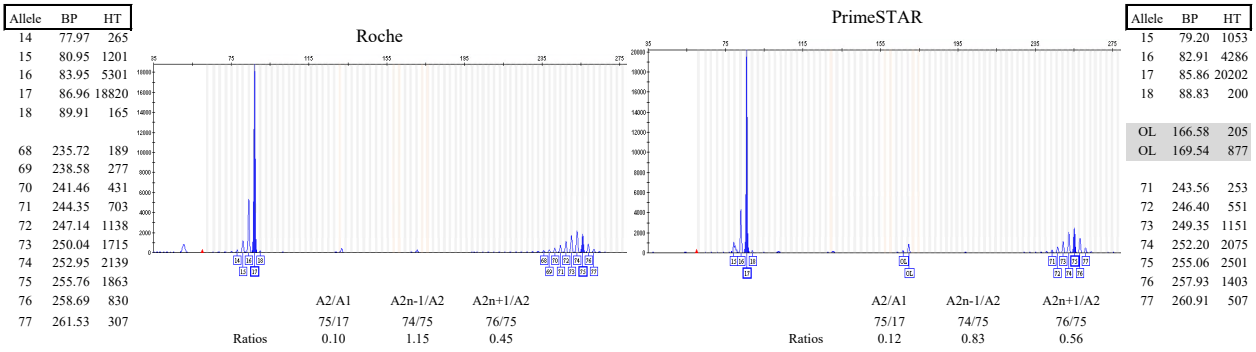


Figure 13. Comparison of Roche and PrimeSTAR Large Allele Amplification.

With sample CD00022, the PHR of the 31 CAG repeat allele to the 18 allele is 0.49 in the Roche kit and 0.39 with PrimeSTAR. With sample NA20210, the PHR of the 75 allele to the 17 allele is 0.10 with Roche and 0.12 with PrimeSTAR. This suggests that the Roche kit may provide better between-allele balance than PrimeSTAR for the smaller sized alleles but the kits are about equally efficient for larger sized alleles.

However, the PHR of the *n*-1 30-stutter allele relative to the parent 31 allele is 0.62 with Roche and 0.42 with PrimeSTAR; the PHR of the *n*-1 74 stutter allele relative to the parent 75 is 1.15 with Roche and 0.83 with PrimeSTAR. This suggests that there is relatively less stutter with PrimeSTAR relative to Roche. This reduced *n*-1 stutter becomes very important in determining the genotype as the allele sizes get larger since it is generally assumed that "the parent is the allele with the largest peak height" may not be appropriate [9,27].

Figure 14 compares the PHRs of the *n*-1, *n*-2, *n*-3, and *n*+1 stutter alleles for all of the alleles provided by the 22 samples used to define the virtual ladder, using all alleles with peak heights greater than 50 RFU. Not only is the *n*-1 stutter reduction provided by the PrimeSTAR kit confirmed, the *n*-2 and *n*-3 stutter is also much reduced. The *n*+1 stutter appears to be about the same with both kits.

Figure 15 provides a direct comparison of the results presented in Figure 14, plotting the ratio of the PHR results for the Roche kit relative to the PrimeSTAR. The Roche kit generates progressively more *n*-1, *n*-2, *n*-3 stutter than PrimeSTAR. PrimeSTAR generates about the same to slightly more *n*+1 stutter than does Roche.

These kit-specific differences in stutter are the likely explanation for the consensus call for the large allele in sample NA20210 being 74 rather than 75 in both the 2006 and 2012 interlaboratory studies since the Roche kit was used [9,10].

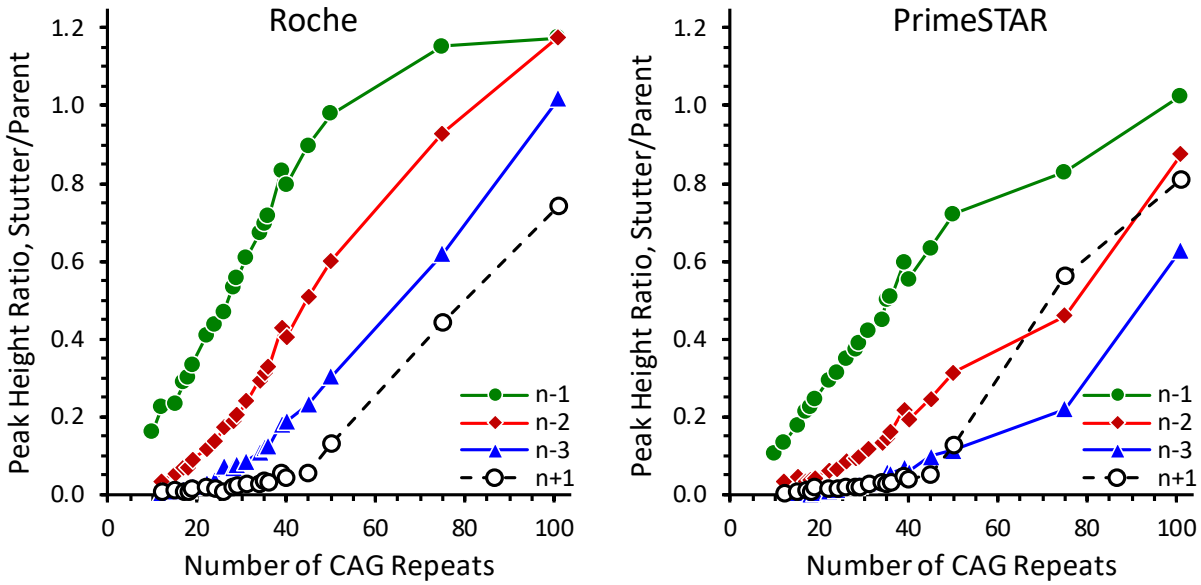


Figure 14. Peak Height Ratios of the *n*-1, *n*-2, *n*-3, and *n*+1 Stutter Alleles.

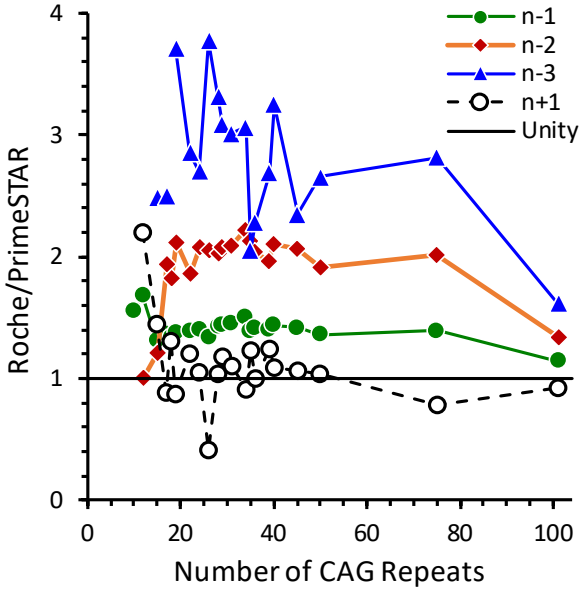


Figure 15. Comparison of Peak Height Ratios for the *n*-1, *n*-2, *n*-3, and *n*+1 Stutter Alleles.

3.6. Confirming the Alleles

The very different "minus" and "plus" stutter tendencies of the Roche and PrimeSTAR kits suggests a mechanism for visualizing the identification of the true parent of large alleles. The PHRs of the *n*+1, *n*+2, … repeats (to the right of the parent) should be approximately the same for the two kits while the PHRs of the *n*-1, *n*-2, … repeats (to the left of the parent) should progressively diverge. Figure 16 displays this expected pattern for the parent 31 allele. Appendix C displays the patterns for all the alleles investigated in this study.

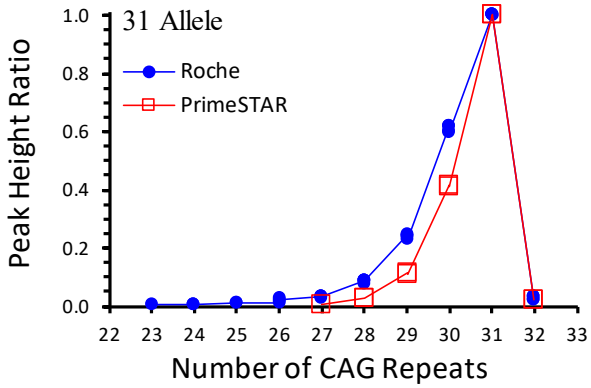


Figure 16. Comparison of Roche and PrimeSTAR Peak Height Ratios for Parent 31 Allele

The solid blue circles and blue connecting line denote the peak height ratios (PHR) for Roche-amplification of two samples delivering the parent 31 allele. The open red squares and red connecting line denote the PHR for PrimeSTAR amplification.

Figure 17 displays the patterns for the large allele of the NA20210 sample assuming the parent allele is either the 74 allele or the 75 allele. The expected forward-allele pattern (the alignment of the *n*+1, *n*+2, and *n*+3 repeats for the two kits) is realized only if the parent is the 75 allele.

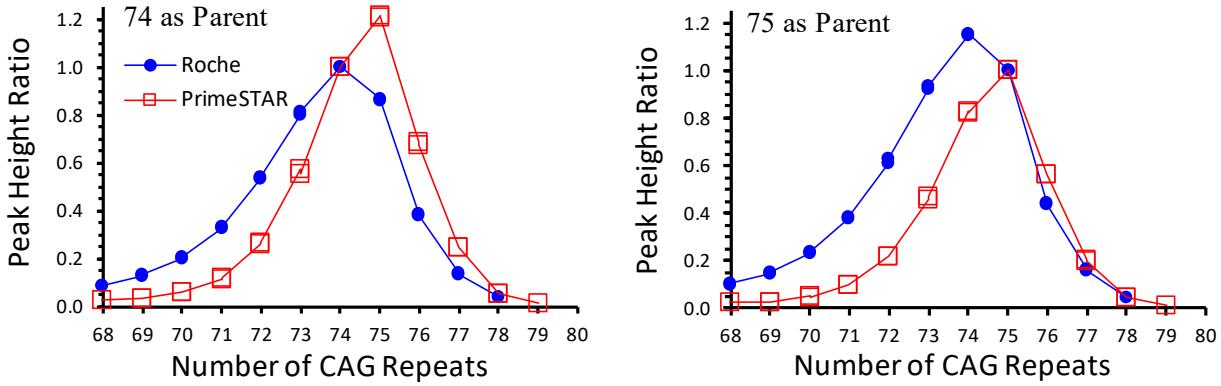


Figure 17. Comparison of Peak Height Ratios for NA20210 Large Allele

The panel to the left compares the Roche and PrimeSTAR peak height ratios (PHR) for the large allele of sample NA20210 assumes that the 74 allele is the parent. The panel to the right compares the PHRs assumes that the 75 allele is the parent. The graphical elements of the panels are defined in the legend to Figure 16.

While sample NA20253 is not a component of SRM 2393, the parent of its large allele has been asserted to be either the 100 or the 101 allele [9,28,29,30]. The pattern of differences between the Roche and PrimeSTAR amplifications is compatible with the expected pattern only when the parent is the 101 allele. However, the differences between the patterns for the two assumptions are small compared to those or the NA20210 75 allele. This suggests that this PHR analysis method may not be applicable to alleles containing many more than 101 CAG repeats.

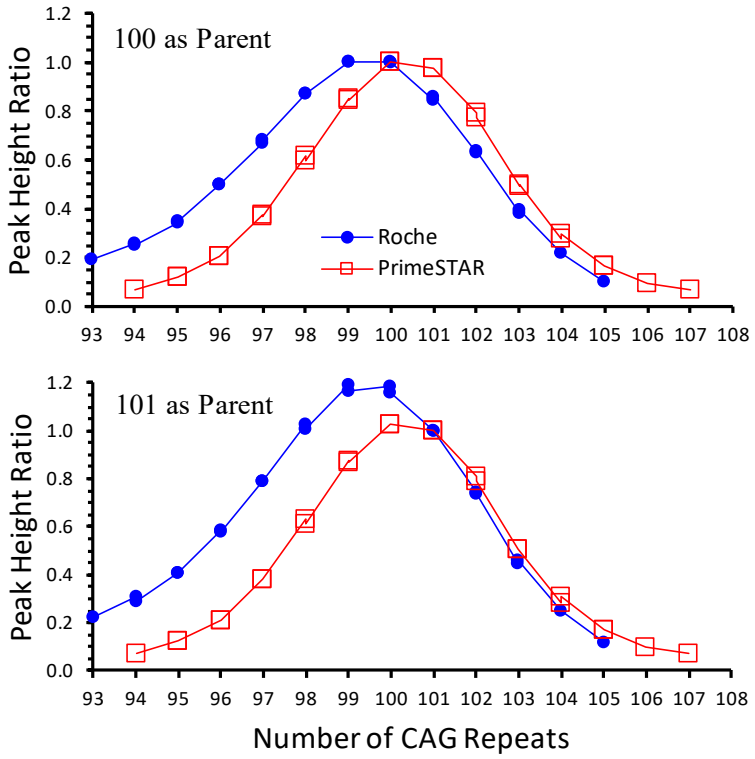


Figure 18. Comparison of Peak Height Ratios for NA20253 Large Allele

The upper panel compares the Roche and PrimeSTAR peak height ratios (PHR) for the large allele of sample NA20253 assumes that the 100 allele is the parent. The lower panel compares the PHRs assumes that the 101 allele is the parent. The graphical elements of the panels are defined in the legend to Figure 16.

4. Evolved Genotyping Process for Determining CAG Repeat Counts

The tallest peak within the (*n*-3, *n*-2, *n*-1, *n*, *n*+1, *n*+2) peak cluster for the large allele of Component F is the 75 allele only with the PrimeSTAR amplification system (Figure 13). If stable, this pattern would simplify confirmation of the identity of the true parent alleles. This pattern did not change during optimization studies using the originally purchased PrimeSTAR kit. However, under the same amplification conditions off ladder (OL) alleles and spurious peaks increased in number and height with newly purchased Prime STAR amplification kits (Figure 19).

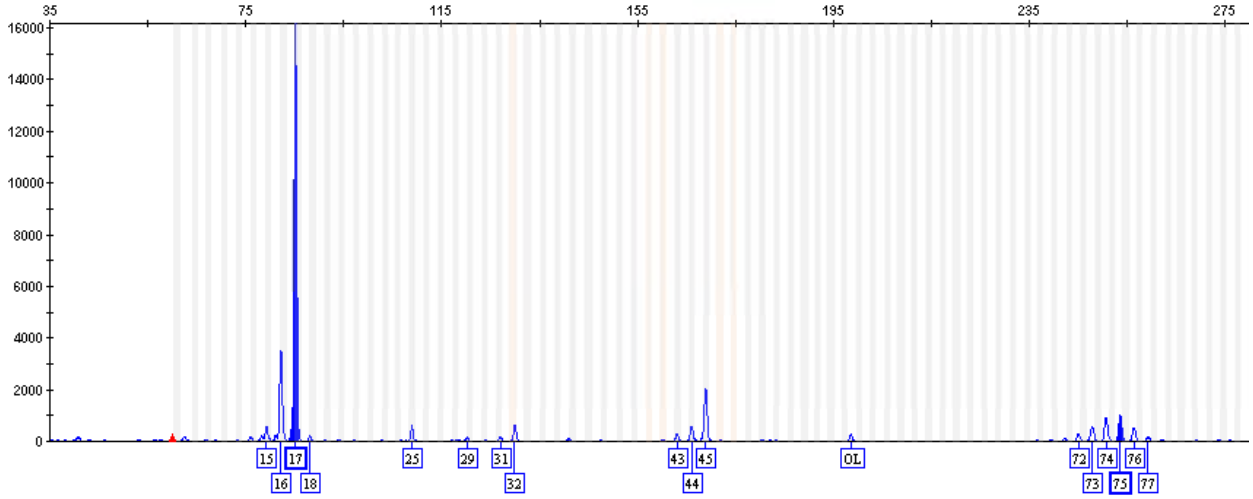


Figure 19. SRM 2393 Component F Amplified with Newly Purchased PrimeSTAR kit.
Parent HD alleles are highlighted in blue. Unfilled alleles marked as 25, 29, 31, 32, 43, 44, 45, and OL are spurious alleles. The spurious 45 allele has a greater peak height than the 75 allele.

While the few OL alleles visible in Figure 13 do not confuse genotyping, the increased prevalence and intensity of the OL and spurious peaks with the new purchased kits renders the current PrimeStar system of questionable utility. Therefore, the Roche Kit has been used to support the extension of the period of validity for SRM 2393.

4.1. Genotyping Parameters

Two additional units of SRM 2393 were pulled from the 2 °C to 8 °C storage refrigerator, for analysis. Components from each unit of SRM 2393 were amplified using the master mix and amplification parameters specified in Table 16.

After amplification, duplicate aliquots of each amplified component were prepared as follows: 1 µL of the amplicon was mixed with 9 µL of an injection mastermix consisting of 8.7 µL of HiDi formamide (Thermo Fisher) and 0.3 µL GeneScan (GS) 500 Liz size standard (Thermo Fisher) in a 96-well sample plate.

The samples were then further denatured by heating the plate to 96 °C for 3 minutes then immediately snap-cooled (placed on ice for 3 minutes).

The plate was analyzed with the 3500xL Genetic Analyzer instrument equipped with a 36 cm 24-capillary array, filled with POP-4 polymer and run at 60 °C. Two injections of each prepared sample were performed; one injection at 8 s and the other injection at 5 s, both with an injection voltage of 1.2 kV. The samples were run at 15 kV.

The instrument applied a DS-33 matrix called "G5 dye set". The data was then analyzed with GeneMapper ID-X, v1.5 with the prepared HD parameters.

4.2. Final Results

The 2020 CE genotyping results confirmed amplifiability and stability of components as they did for the original certification in 2011. Figure 20 shows the agreement between the allele calls (stutter product and parent alleles) and the base pair (bp) assignments made for the SRM 2393 components run in 2020. The linear agreement between the smaller alleles such 15 and 17, with the largest alleles 70 to 77, demonstrates the system is in control. However, new bins and panels were created to adjust for the electrophoretic mobility shifts created by the five-year difference of instrumentation and reagents.

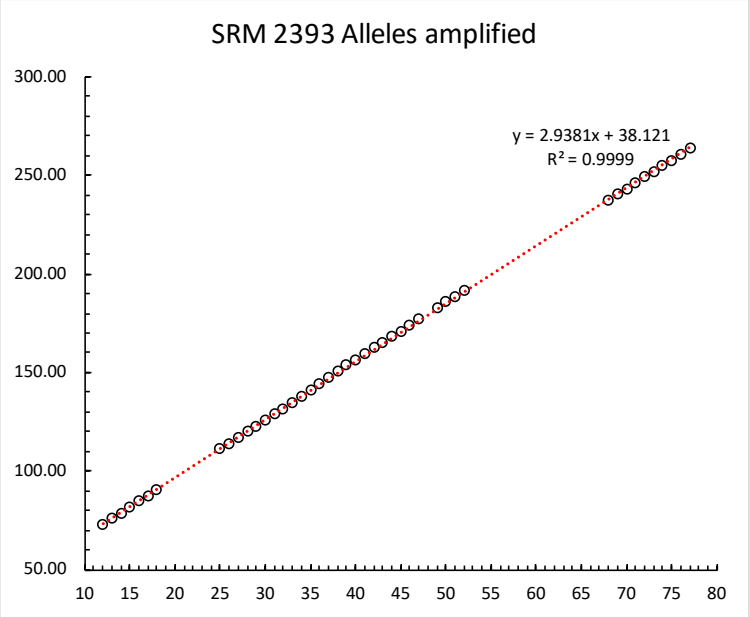


Figure 20. Allele calls and BP sizes of SRM 2393 components.

Exemplar electropherograms with the genotypes are shown in Figure 21. The 15 allele (components A and C) and 17 allele (components B and F) peaks are taller than the *n*-1, *n*-2, *n*-3, and *n*+1 stutter products. However, the peak height of Component E's 49 allele is 92 % of the parent 50 allele. In component F, the parent 75 allele peak is shorter than the 74 peak. This electropherogram appears the same as it did in the original certification where Sanger Sequencing confirmed the 75 allele as the parent allele. See Figure 6 for exemplar electropherograms from 2011 to compare.

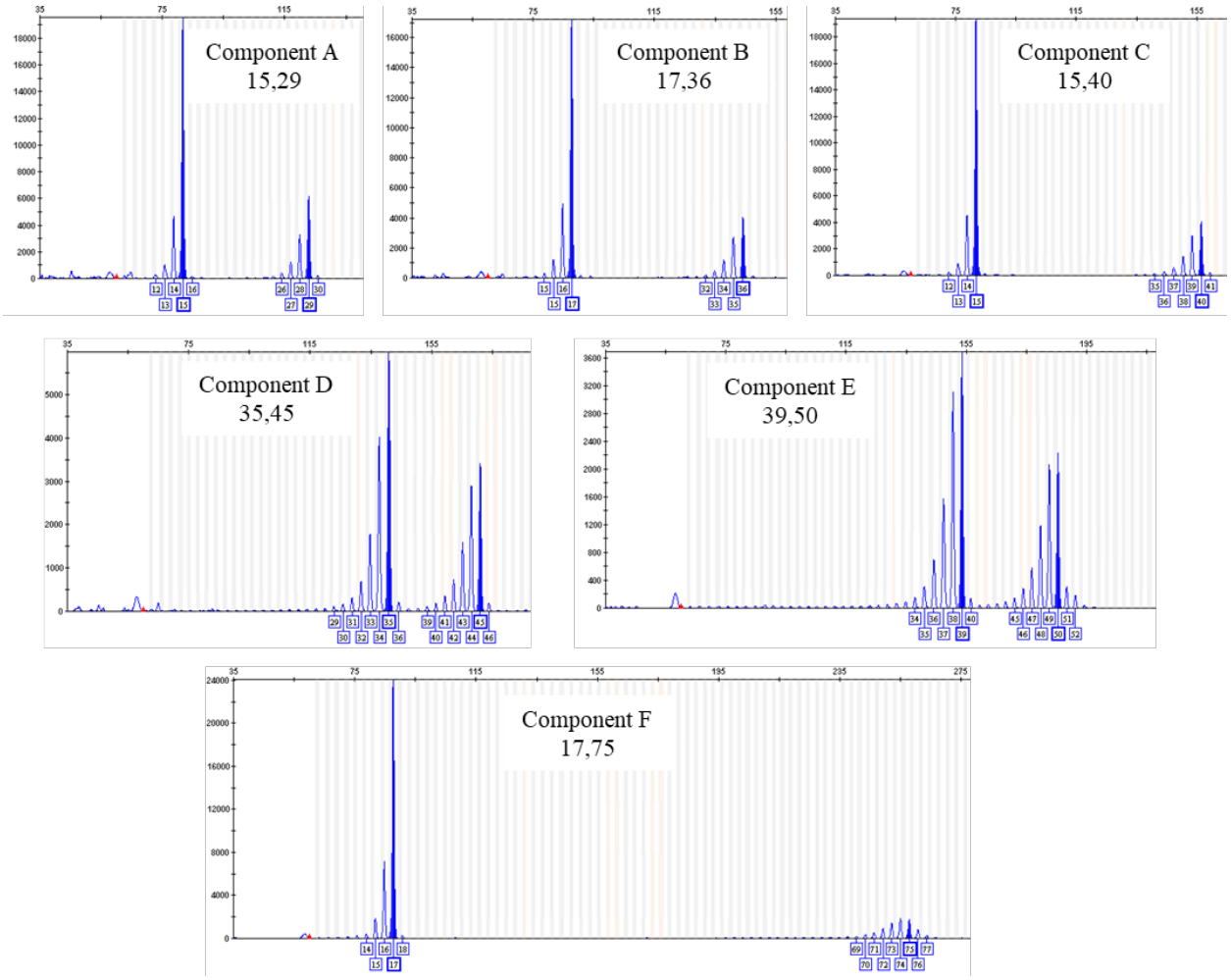


Figure 21. Exemplar Electropherograms of SRM 2393 Components A to F.

X-axis labels are allele designations; Y-axis scale is instrument assigned Relative Fluorescence Units. Solid blue filled peaks are the components genotype "parent alleles." Unfilled peaks are "stutter" alleles.

References

1 CDC/NIH (2007) Biosafety in Microbiological and Biomedical Laboratories, 5th ed.; Richardson J, Barkley WE, Richmond J, McKinney RW, Eds. Centers for Disease Control and Prevention. https://www.cdc.gov/biosafety/publications/bmbl5/index.htm

2 National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, Department of Health, Education and Welfare (1978) Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research. DHEW Publication No. (OS) 78-0012 http://videocast.nih.gov/pdf/ohrp_belmont_report.pdf

3 Walker FO; Huntington's Disease; Lancet 2007;369(9557):218-228. https://doi.org/10.1016/S0140-6736(07)60111-1

4 Huntington's Disease Society of America; http://www.hdsa.org/ (accessed May 2020).

5 Bean L, Bayrak-Toydemir P. American College of Medical Genetics and Genomics Standards and Guidelines for Clinical Genetics Laboratories 2014 edition: technical standards and guidelines for Huntington disease. Genet Med 2014;16:e2 https://doi.org/10.1038/gim.2014.146

6 Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, Tanzi RE, Watkins PC, Ottina K, Wallace MR, Sakaguchi AY, Young AB, Shoulson I, Bonilla E, Martin JB. A polymorphic DNA marker genetically linked to Huntington's disease. Nature 1983;306(5940):234-238. https://doi.org/10.1038/306234a0

7 Levin BC, Richie KL, Jakupciak JP. Advances in Huntington's disease diagnostics: development of a standard reference material. Expert Rev Mol Diagn 2006;6(4):587-596. https://doi.org/10.1586/14737159.6.4.587

8 Jakupciak JP, Richie KL. Analytical validation of quality control materials for Huntington's disease examination. Accred Qual Assur 2007;12:139-145. https://doi.org/10.1007/s00769-006-0226-5

9 Kalman L, Johnson MA, Beck J, Berry-Kravis E, Buller A, Casey B, Feldman GL, Handsfield J, Jakupciak JP, Maragh S, Matteson K, Muralidharan K, Richie KL, Rohlfs EM, Schaefer F, Sellers T, Spector E, Richards CS. Development of genomic reference materials for Huntington disease genetic testing. Genet Med 2007;9(10):719-723. https://doi.org/10.1097/gim.0b013e318156e8c1

10 Quarrell OW, Handley O, O'Donovan K, Dumoulin C, Ramos-Arroyo M, Biunno I, Bauer P, Kline M, Landwehrmeyer GB. European Huntington's Disease Network. Discrepancies in reporting the CAG repeat lengths for Huntington's disease. Eur J Hum Genet 2012;20(1):20-26. https://doi.org/10.1038/ejhg.2011.136

11 Bodnar AG, Ouellette M, Frolkis M, Holt SE, Chiu CP, Morin GB, Harley CB, Shay JW, Lichtsteiner S, Wright WE. Extension of life-span by introduction of telomerase into normal human cells. Science 1998;279(5349):349-352. https://doi.org/10.1126/science.279.5349.349

12 Lee KM, Choi KH, Ouellette MM. Use of exogenous hTERT to immortalize primary human cells. Cytotechnology 2004;45(1-2):33-38. https://doi.org/10.1007/10.1007/s10616-004-5123-3

13 Kline MC, Hill CR, Decker AE, Butler JM. STR sequence analysis for characterizing normal, variant, and null alleles. Forensic Sci Int Genet 2011;5(4):329-332. https://doi.org/10.1016/j.fsigen.2010.09.005

14 Olejniczak M, Kozlowski P, Sobczak K, Krzyzosiak WJ. Accurate and sensitive analysis of triplet repeat expansions by capillary electrophoresis. Electrophoresis 2005;26:2198-2207. https://doi.org/10.1002/elps.200410197

15 Olejniczak M, Krzyzosiak WJ. Genotyping of simple sequence repeat factors implicated in shadow band generation revisited. Electophoresis 2006;27(19):3724-34. https://doi.org/10.1002/elps.200600136

16 Walsh PS, Fildes NJ, Reynolds R. Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. Nucleic Acids Res 1996;24(14):2807-2812. https://doi.org/10.1093/nar/24.14.2807

17 Kline MC, Decker AE, Duewer DL, Butler JM. A NIST Standard Reference Material for Huntington Disease Alleles. Poster given at Association for Molecular Pathology 2009 Annual Meeting (AMP 2009), Kissimmee, FL, November 19-22, 2009. https://www.nist.gov/system/files/documents/mml/bmd/genetics/AMP2009-HDposter.pdf

18 Duewer, D. L.; Kline, M. C.; Romsos, E. L.; Toman, B. Evaluating droplet digital PCR for the quantification of human genomic DNA: converting copies per nanoliter to nanograms nuclear DNA per microliter. Anal Bioanal Chem 2018, 410(12), 2879-2887. https://doi.org/10.1007/s00216-018-0982-1

19 Dagata JA, Farkas N, Kramer JA. Method for measuring the volume of nominally 100 µm diameter spherical water in oil emulsion droplets. NIST Special Publication 260 184. 2016. https://doi.org/10.6028/NIST.SP.260-184

20 Kline MC, Duewer DL. Evaluation of Methods for Assessing the Proportion of Single Stranded Nuclear DNA in Human Blood Extracts. NIST Special Publication (NIST SP) 1200 27, 2019. https://doi.org/10.6028/NIST.SP.1200-27

21 Kline MC, Duewer DL. Evaluating Digital PCR for the Quantification of Human Nuclear DNA: Determining Target Strandedness. Anal Bioanl Chem 2020;412(19):4729-4760. https://doi.org/10.1007/s00216-020-02733-2

22 Andrew SE, Goldberg YP, Theilmann J, Zeisler J, Hayden MR. A CCG repeat polymorphism adjacent to the CAG repeat in the Huntington disease gene: implications for diagnostic accuracy and predictive testing. Hum Mol Genet 1994;3(1):65-67. https://doi.org/10.1093/hmg/3.1.65

23 Coles R, Leggo J, Rubinsztein D. Analysis of the 5' upstream sequence of the Huntington's disease (HD) gene shows 6 new rare alleles which are unrelated to the age at onset of HD. J Med Genet 1997;34:371-374. https://doi.org/10.1136/jmg.34.5.371

24 Margolis RL, Ross CA. Diagnosis of Huntington Disease. Clin Chem 2003;49(10):1726-1732. https://doi.org/10.1373/49.10.1726

25 MilliporeSigma. GC-RICH PCR System, Product 12140306001. https://www.sigmaaldrich.com/catalog/product/roche/12140306001?lang=en&region=US

26 TaKaRa. PrimeSTAR HS with GC buffer, Catalog # R044B https://www.takarabio.com/products/pcr/gc-rich-pcr/primestar-hs-with-gc-buffer

27 Jama M, Millson A, Miller CE, Lyon E. Triplet Repeat Primed PCR Simplifies Testing for Huntington Disease. J Mol Diagn 2013;15:255e262. https://dx.doi.org/10.1016/j.jmoldx.2012.09.005

28 Thibert JR, Nicholson K, Wisotsky J, Le V, Latham GJ, Statt S. A Robust PCR/CE Assay Using AmplideX Technology for Rapid and Accurate Genotyping of CAG Repeat Expansions in *HTT*. Poster presented at 2018 ACMG Annual Clinical Genetics Meeting, Charlotte, NC Apr 10-14, 2018. https://asuragen.com/wp-content/uploads/2018/04/ACMG18_Julie-Thibert.pdf

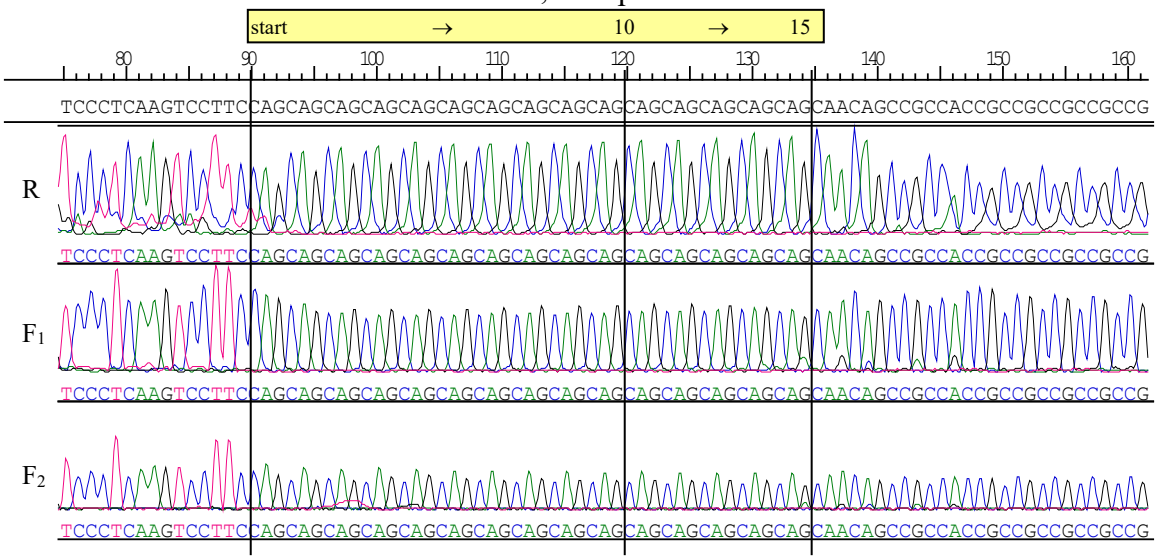
29 Jama M, Millson A, Miller CE, Lyon E. Triplet Repeat Primed PCR Simplifies Testing for Huntington Disease. J Mol Diagn 2013;15(2):255-262. https://doi.org/10.1016/j.jmoldx.2012.09.005

30 Zhao M, Cheah FSH, Chen M, Lee CG, Law HY, Chong SS. Improved high sensitivity screen for Huntington disease using a one-step triplet-primed PCR and melting curve assay. PLoS One 2017;12(7):e0180984. https://doi.org/10.1371/journal.pone.0180984

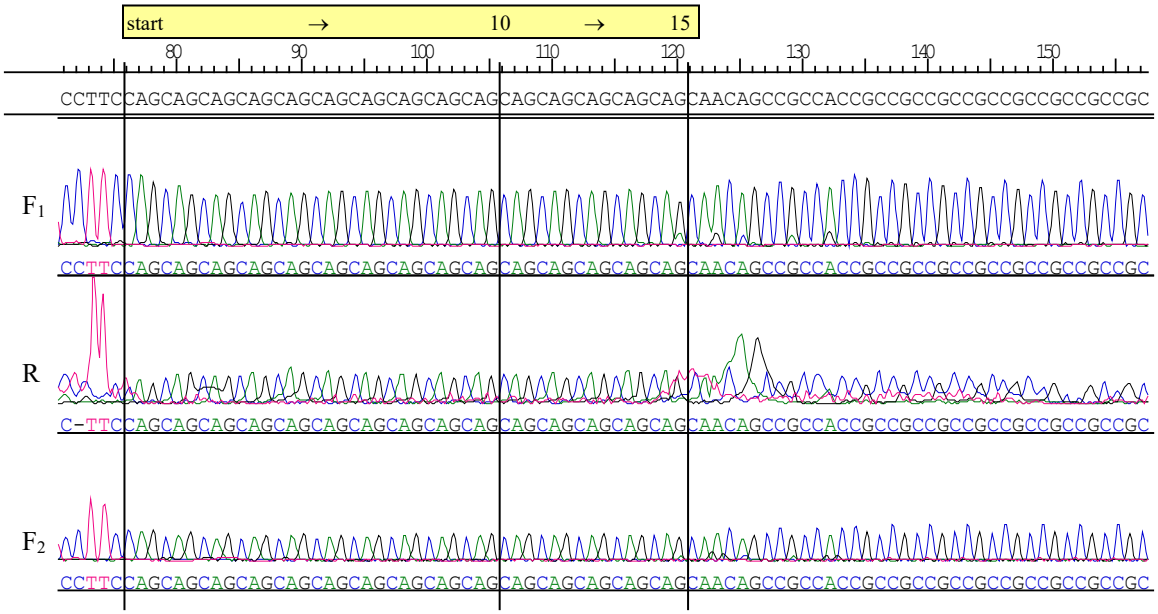
Appendix A. Sanger Sequencing Results

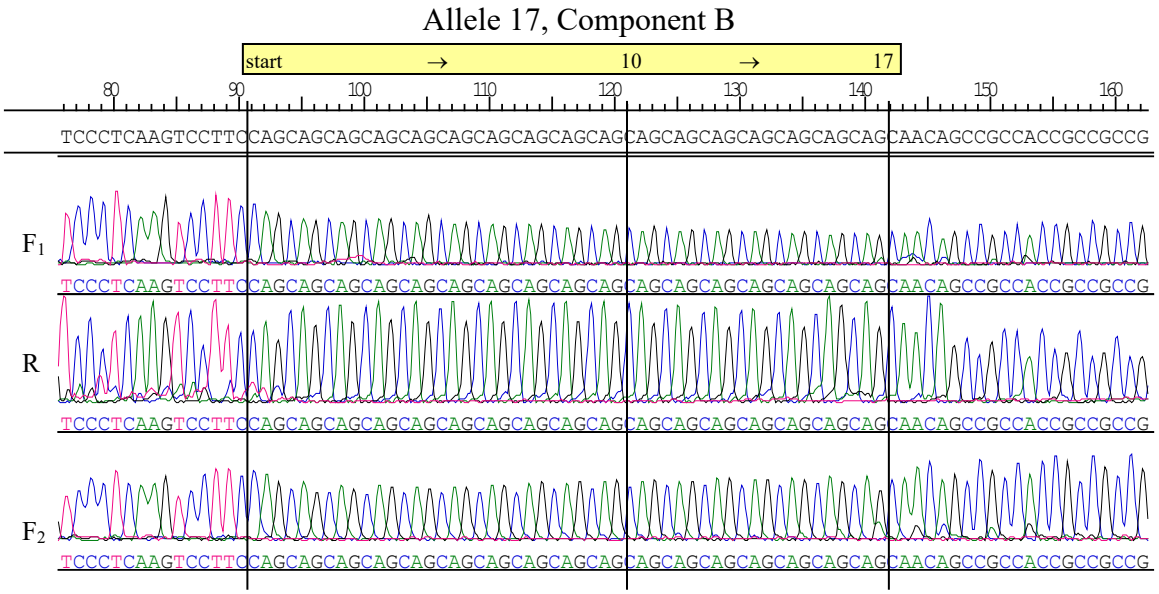
Vertical lines are placed at the beginning of the CAG repeat region, every 10 repeats and at the end of the repeat region to facilitate counting. Where helpful, the Stutter-Sorter alignment tool has been added to clarify the results.

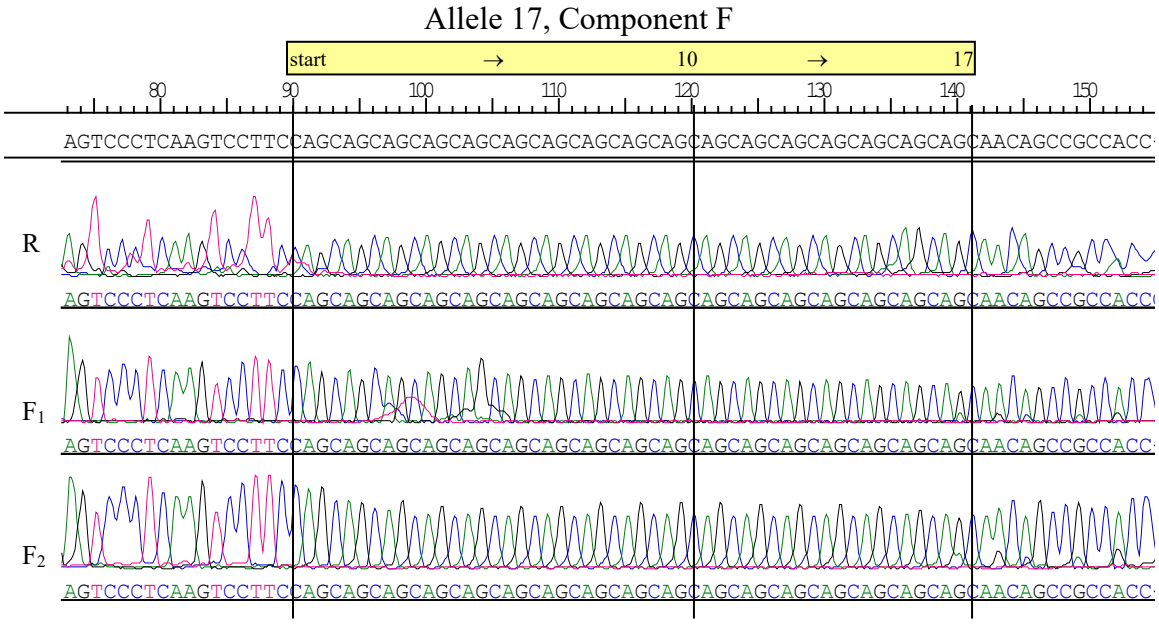
Allele 15, Component A

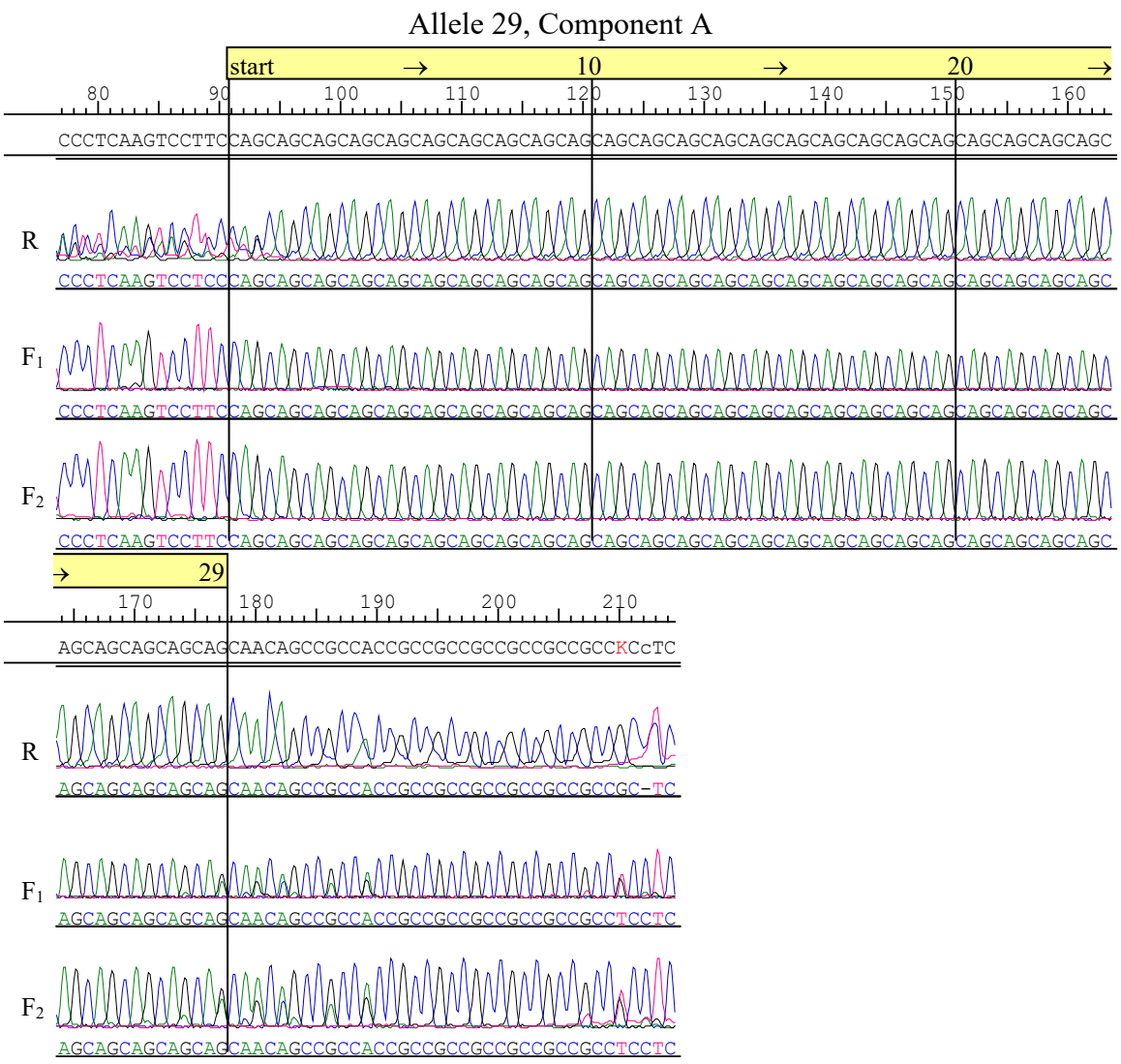


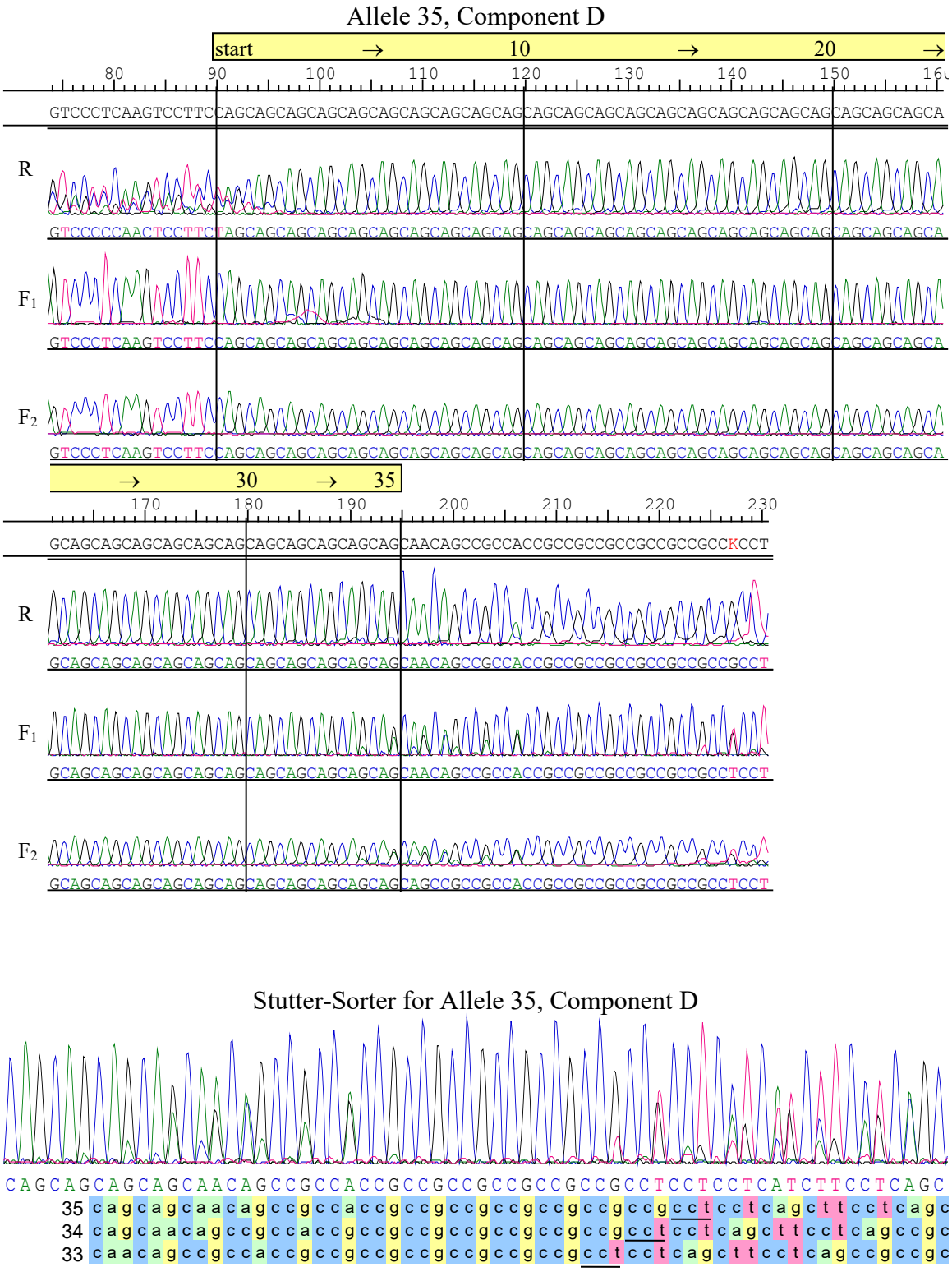
Allele 15, Component C

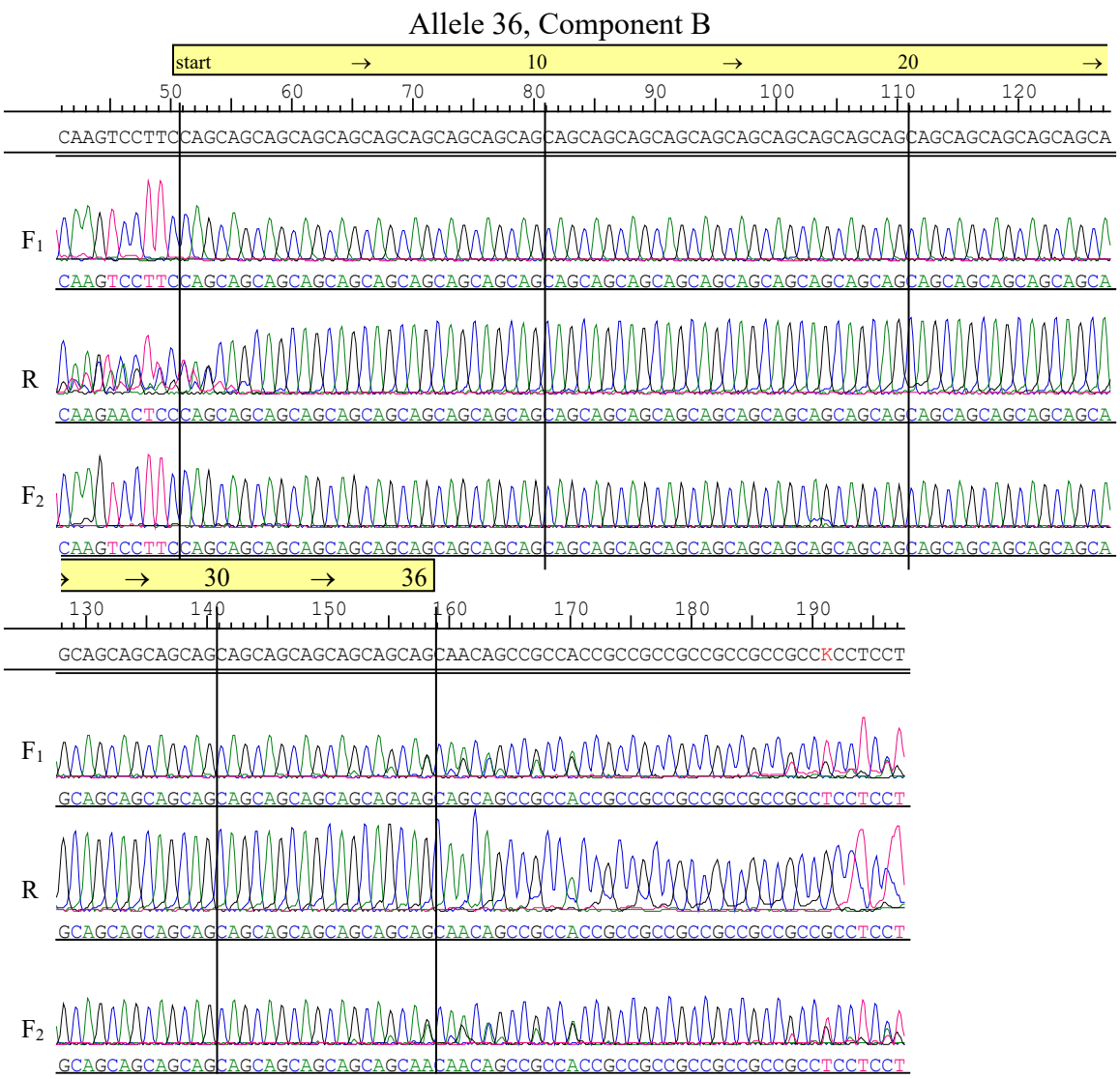


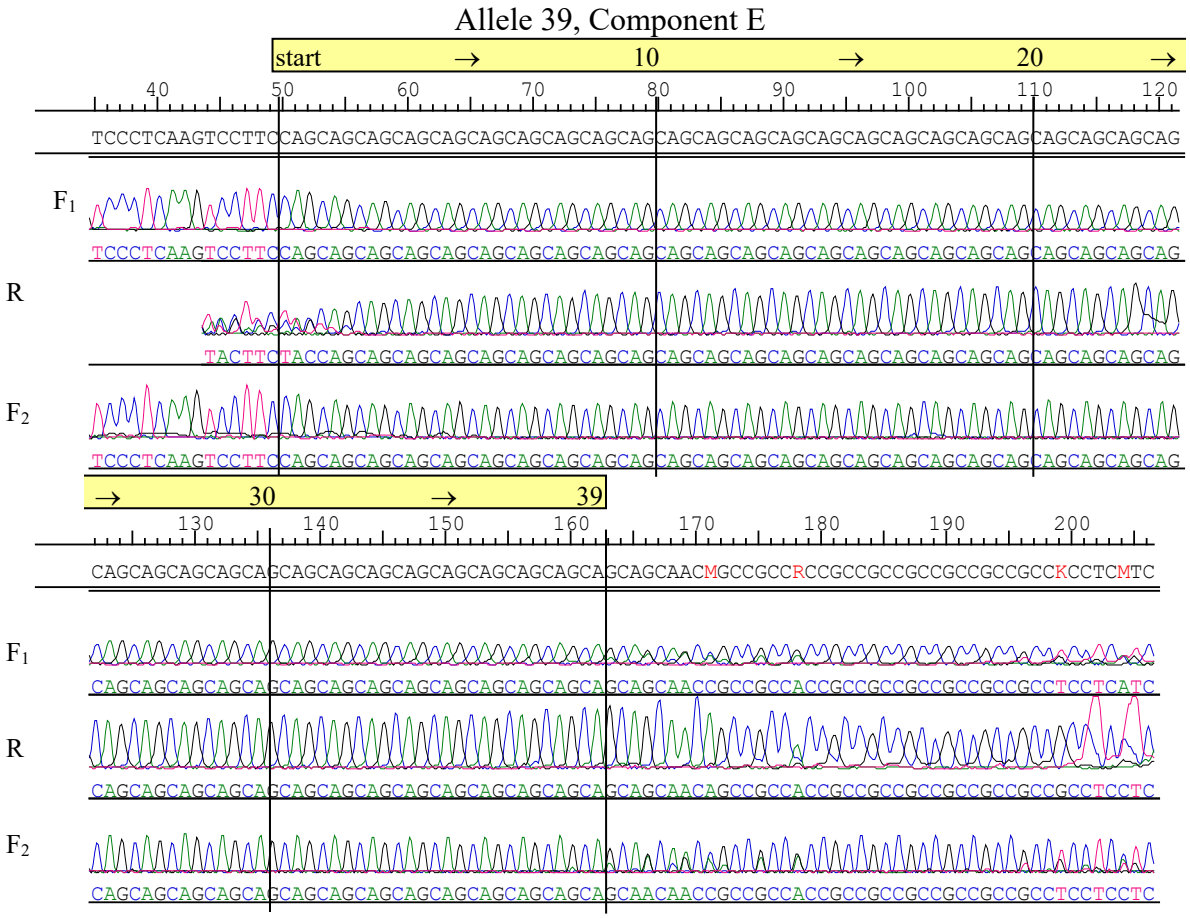


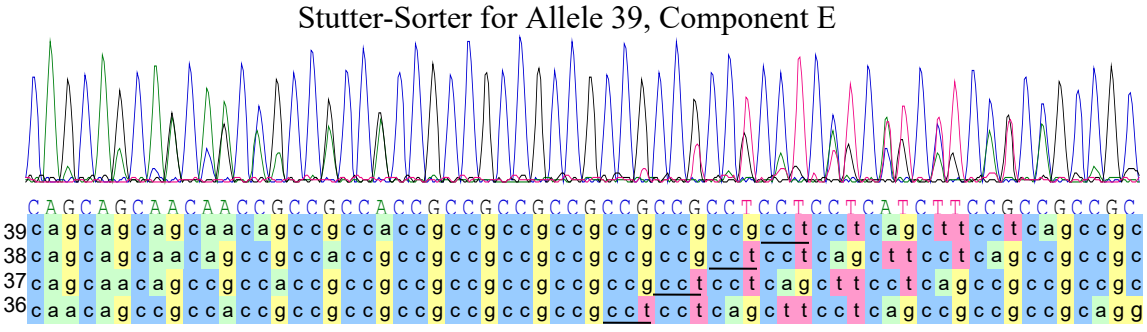


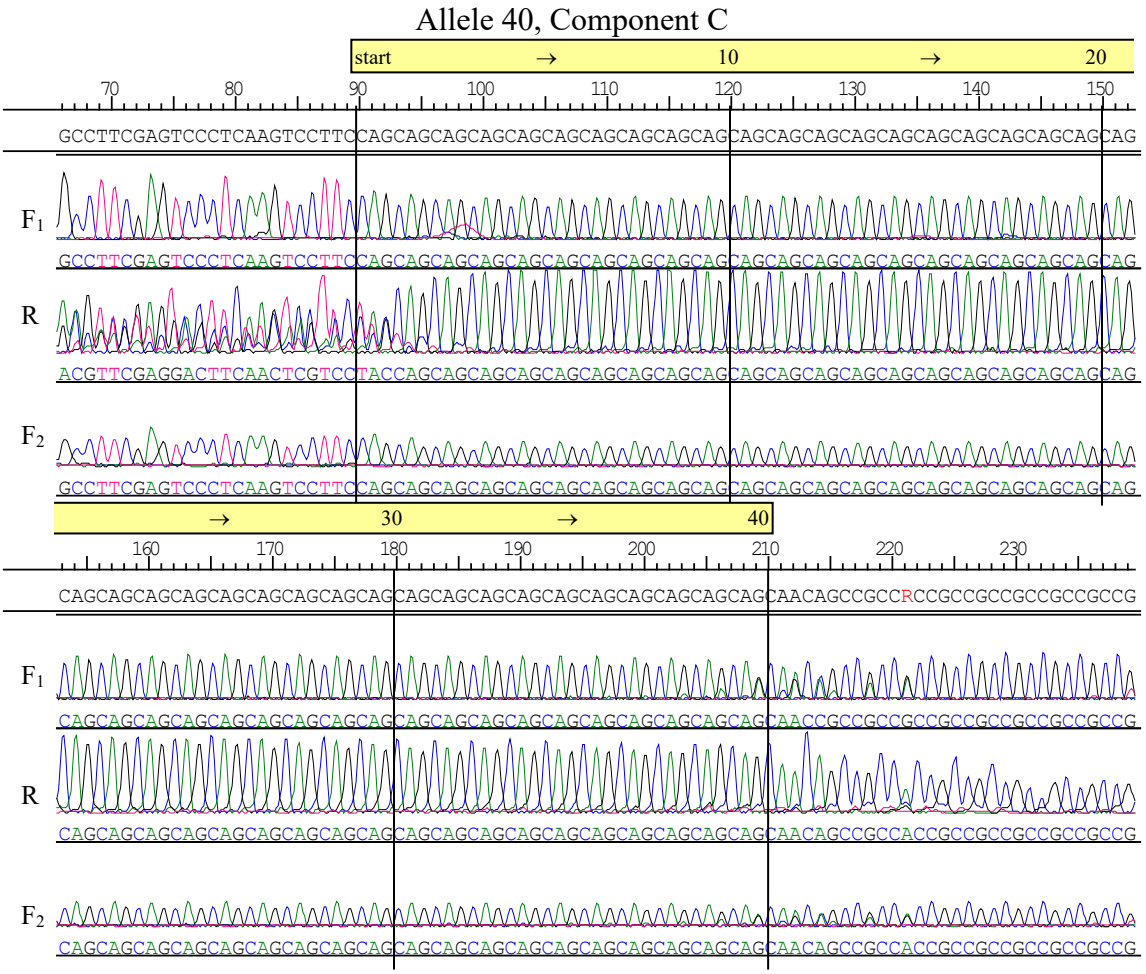


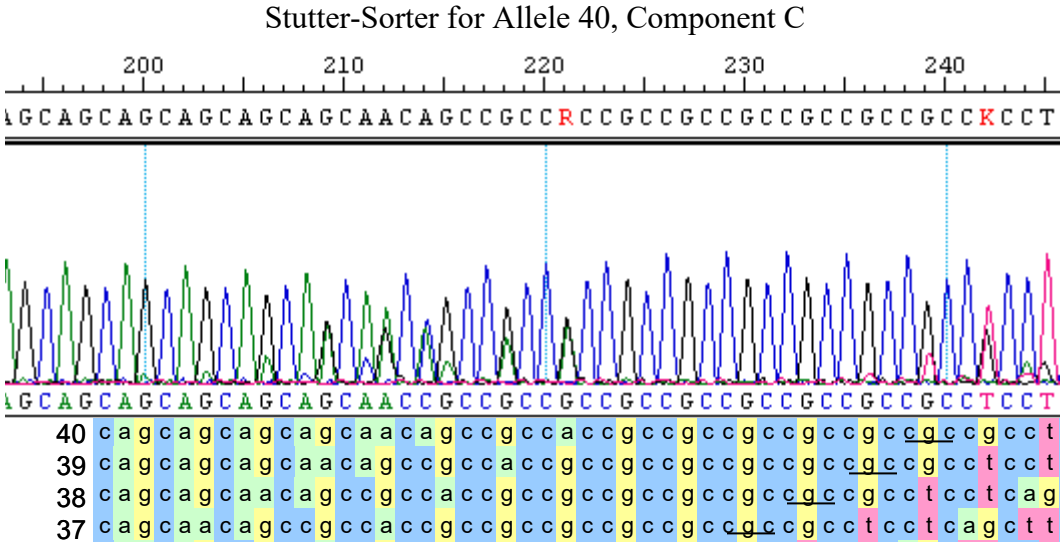


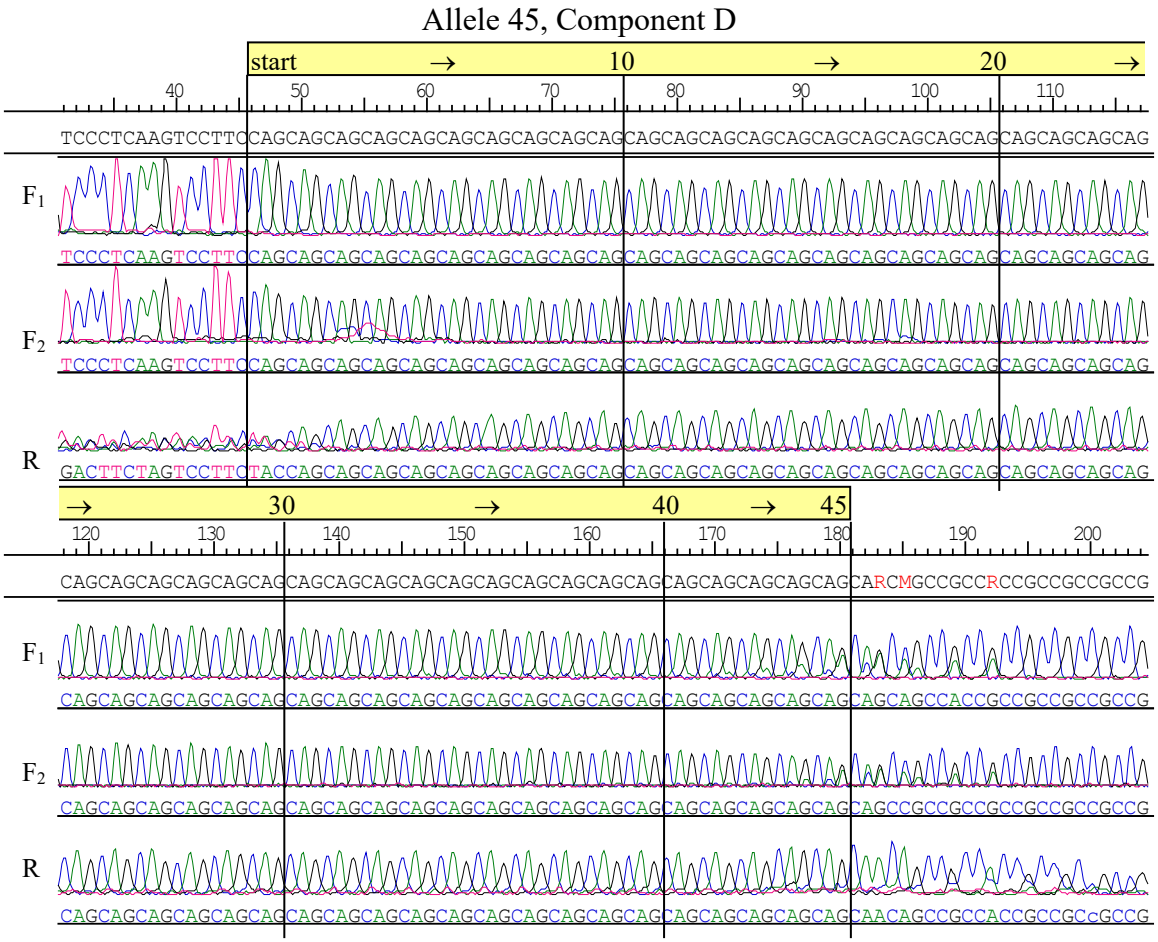


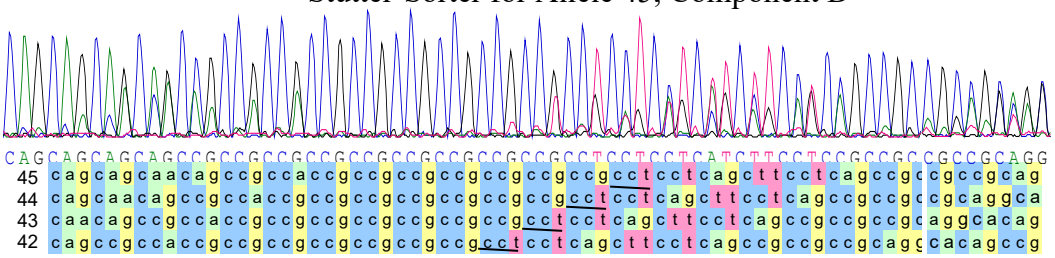




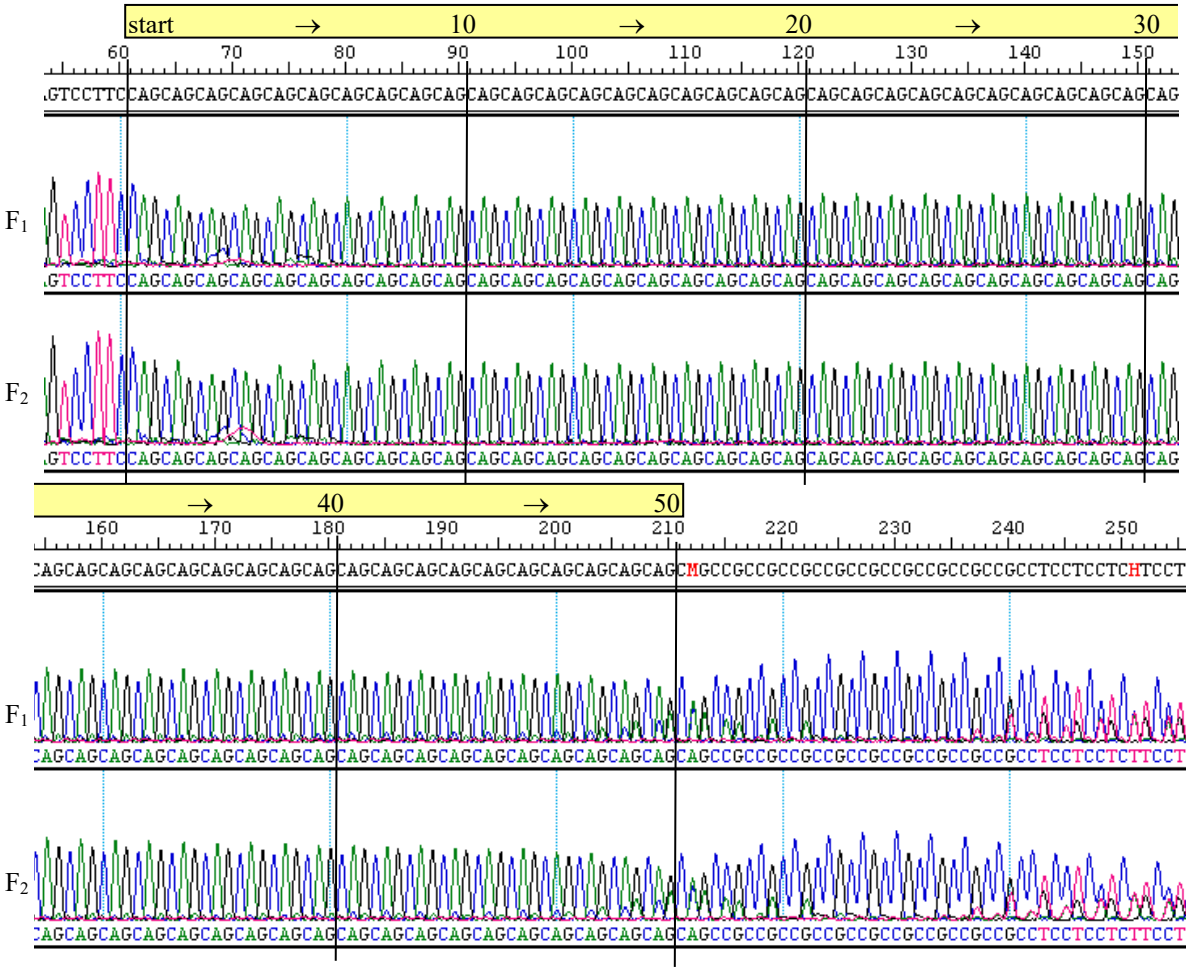




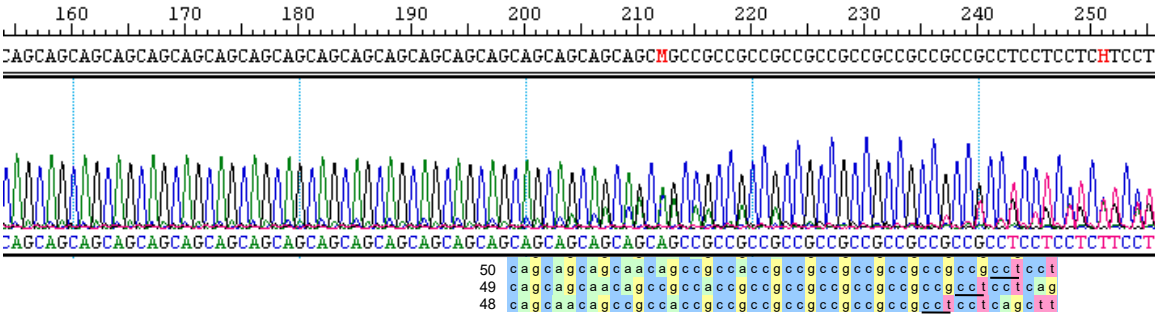


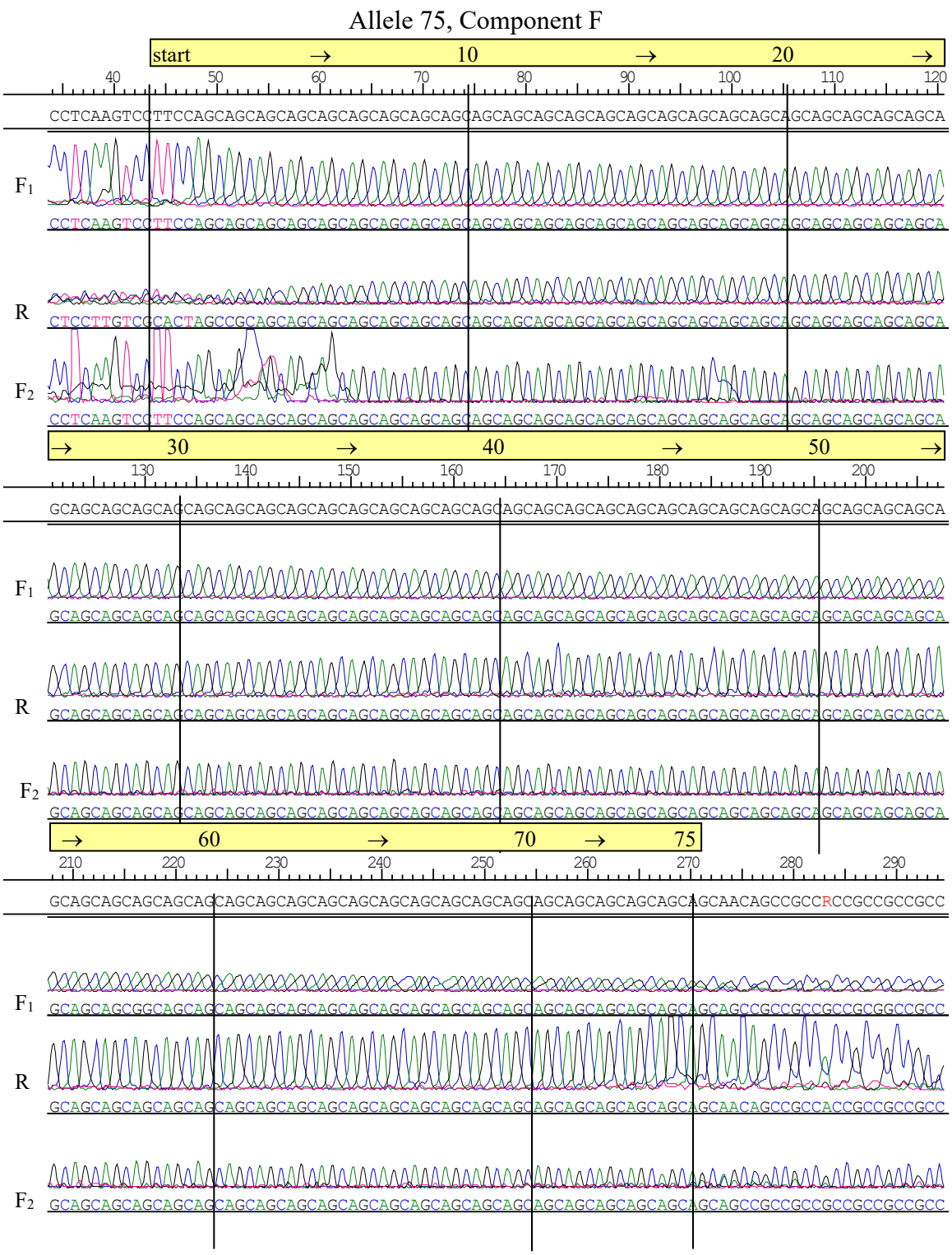


Allele 50, Component E



Stutter-Sorter for Allele 50, Component E





Stutter-Sorter for Allele 75, Component F

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 $[CCG]_{7-12}$ and a $[CCT]_{1-2}$ [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 $[GGGGGC]_{1-2}$ and 20 bp repeat $[GGCCCCGCCTCCGCCGGCGC]_{1-3}$ [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.

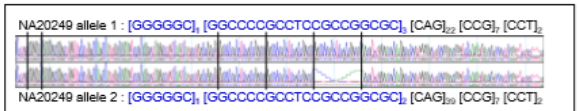


Figure 2. Sequencing data for the repeats found upstream of the CAG repeat region

Table 2. Classifications of the upstream repeat regions

α:[GGGGGC]$_1$ [GGCCCCGCCTCCGCCGGCGC]$_1$ 270 bp ±0.5 bp N=5
β:[GGGGGC]$_1$ [GGCCCCGCCTCCGCCGGCGC]$_2$ 288 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]$_3$ [GGCCCCGCCTCCGCCGGCGC]$_2$ 282 bp ±0.5 bp* N=2

* Requires additional sequencing

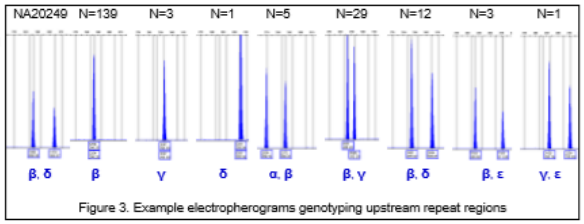


Figure 3. Example electropherograms genotyping upstream repeat regions

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
	NA20245	22,35	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	Y25	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	MT34886	18	10	β
Asian	OT06795	17	7	β,δ
AfricanAm	TT50322	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

Appendix C. Stutter Ratios for Roche and PrimeSTAR Systems

The following panels display the (Allele)/(Parent Allele) peak height ratios for all alleles near the parent that have peak heights greater than 50 RFU. The vertical axis in each panel displays the ratios; the horizontal axis displays the number of CAG repeats in allele. Solid blue circles and blue lines represent results from the Roche system. Open red squares and red lines represent results from the PrimeSTAR system. Each panel displays all the available results for a given allele, regardless of sample. All ratios are normalized such that the ratio for the parent allele is 1.0.