Model for Production of Nucleic Acid-Based NIST Standard Reference Materials for Clinical Diagnostics

Ross J. Haynes, Marcia J. Holden, Margaret C. Kline, and John M. Butler

Marcia.Holden@nist.gov
301-975-4162
Ross.Haynes@nist.gov
301-975-4469

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

Poster #ID62 at Association for Molecular Pathology Annual Meeting, Long Beach, CA October 25-27, 2012

Abstract

Introduction: Certified reference materials are critical for standardization of assays for the detection and quantification of nucleic acids involved in infectious disease, genetic disease and cancer. The first viral DNA Standard Reference Material® (SRM) produced at the National Institute of Standards and Technology (NIST) "SRM 2366 Cytomegalovirus for DNA Measurements" was released in 2011. The work preceding the release helped develop a model for future viral DNA SRMs; particularly, material production, packaging, and certification.

Materials and Methods: The material that later became SRM 2366 was acquired as the Towne strain of CMV in a bacterial artificial chromosome (Towne Δ147 BAC). At NIST the BAC was propagated in *Escherichia coli* and the DNA was isolated and purified. CMV DNA in buffer was packaged in perfluoroalkoxy fluoropolymer (PFA) tubes to minimize absorption of DNA onto tube walls. An innovation for the certification of the concentration of the viral DNA was the use of digital PCR (dPCR). The use of dPCR allows the direct counting of molecules, so quantification can be achieved without the use of calibrants. In the past the number of replicate PCR reactions needed limited the accuracy and precision of digital PCR. However, new microfluidic technology makes the technique viable for measuring the concentration of DNA with reasonable measurement uncertainty. In addition to quantification, the DNA sequence of relevant regions of the CMV genome was verified by Sanger sequencing.

Results: SRM 2366 has three components A, B, and C; by digital PCR the concentrations of each are 420, 1702 and 19641 copies per microliter, with standard uncertainties of 56, 130, and 365 copies per microliter, respectively. Sequence information is given for 14555 base pairs across nine regions of interest – UL34, UL54, UL55 to 56, UL80, UL83, UL97, UL122 to 127, UL132 and US17. All the regions of interest match the Towne sequence as deposited in GenBank accession number AY315197.2 with one exception at position number 78651 (UL54) is described as a Y [C or T] at NIST and a C in GenBank.

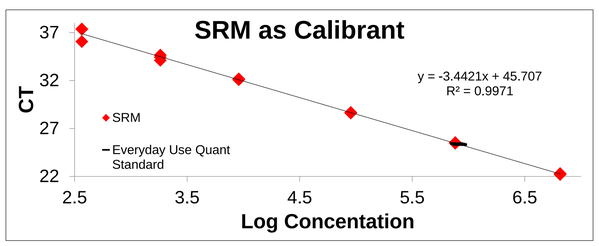
Conclusions: SRM 2366 Cytomegalovirus for DNA Measurements is the model by which NIST will produce additional Standard Reference Materials. The second viral DNA SRM is currently in development for BK virus. This SRM will also be DNA in buffer, sequenced, and certified for concentration with digital PCR.

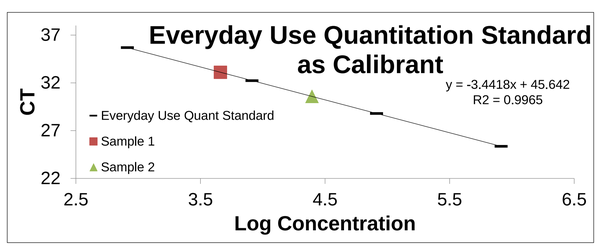
Certification of SRM 2366

- ***Materials***: DNA from Towne strain in a bacterial artificial chromosome (BAC) containing all of the viral genome except for regions IRS1, US1-15, and UL147; BAC provided by Dr. Hua Zhu at the University of Medicine and Dentistry of New Jersey
- BAC format used to give consistent product and prevent truncated genomes which can result from culturing virus
- ***Certification***:
  - Concentration determined via digital PCR
- ***Information Values***:
  - DNA sequence of genes that are targets for qPCR
  - Covering published qPCR primers and commercial CMV detection assays

How to Use SRM 2366 to Calibrate Everyday-Use-Quantitation-Standard (EUQS)

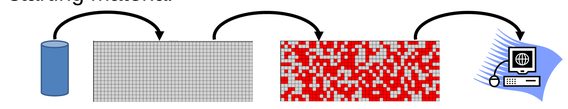
1. Make dilution series of SRM and EUQS
2. Optional: Process if necessary (e.g. if your EUQS, then extract both SRM and EUQS)
3. Quantify via qPCR, treating the SRM as the true value
4. Assign value to EUQS (Note: this value is now traceable to a NIST Standard)
5. Use EUQS with NIST-traceable value to quantify patient samples





Digital PCR (dPCR) to Determine Concentration of DNA

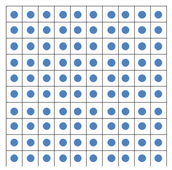
1. Create a PCR master mix as in qPCR with template DNA (some manufacturer specific reagents may be required)
2. Aliquot across hundreds or thousands of reactions; either using microfluidics or microemulsions
3. Thermal cycle as in qPCR and count wells with amplification
4. Use Poisson statistics to determine the concentration of starting material



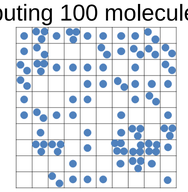
1) Create a PCR mastermix as if for qPCR
2) Aliquot across 100s or 1000s of PCR reactions
3) Thermal cycle as if for qPCR & count wells with amplification
4) Use Poisson statistics to determine concentration of starting material

Random vs. Even Distribution

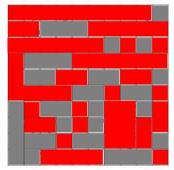
- dPCR and Poisson statistics rely on random distribution
- Consider distributing 100 molecules over 100 wells



Even distribution
Arithmetic division
All 100 well have exactly one molecule



Random distribution
Statistical distribution
Wells have 0, 1, 2 or more molecules



Heat Map
37 wells with 0 molecules &
63 wells with ≥1 molecule

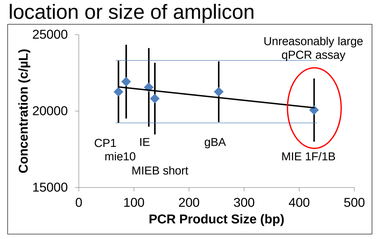
The heat map is representative of the data you get from dPCR; if you could count each molecule of DNA individually you would not need Poison statistics

Poisson Statistics

1. Poisson equation for estimating total number of copies across all wells $N\lambda = N \ln(N/N\text{-}x)$; where λ = average copies per well, x = number of positive wells, and N = number of wells [1].
2. Divide estimated number of copies across all wells ($N\lambda$) by the total volume aliquoted; $N\lambda$/volume(µL) = copies/µL
3. Correct for dilution of DNA into mastermix and any other dilutions
4. Uncertainty follows binomial $\hat{p} \pm z_c \sqrt{\frac{\hat{p}(1-\hat{p})}{N}}$; p = x/N [2].

Concentration vs. Size of Amplicon

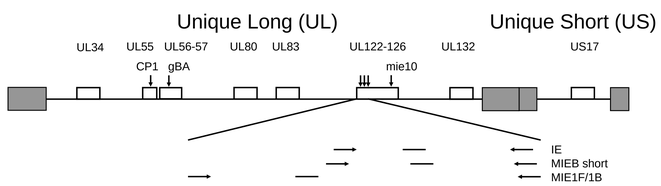
Six targets spread across the genome – three of which were overlapping varying by size of amplicon – were tested to verify concentration determination would not be skewed by any one target location or size of amplicon



ANOVA p-value = 0.43
Correlation p-value = 0.07

Without MIE 1F/1B:
ANOVA p-value = 0.81
Correlation p-value = 0.70

No significant p-values (not enough evidence to declare a difference)



Concentration of SRM Components

- Three components (A, B, & C) were certified for concentration using digital PCR.
- One TaqMan assay – CP1 – was used for the certification
- Three randomly selected tubes from each component were measured with 99 panels using Fluidigm 12.765 digital arrays (i.e. 75,735 PCR reactions). Each panel was treated as a separate measurement and data was combined analytically [3]

Component	A	B	C
Value copies/µL	420	1702	19641
Standard uncertainty copies/µL	±56	±130	±365
Relative uncertainty	13.30%	7.60%	1.80%
95 % confidence interval copies/µL	301 to 523	1446 to 1959	18924 to 20359

Sequenced Regions

- Towne sequence Genbank AY315197.2 was used as the reference for sequence and nucleotide numbering.
- There is complete agreement between the NIST sequence and the reference with one exception noted below
- Sanger sequenced with two independent analysts

Reference Sequences	Rational	Nucleotide range	# bases
UL34	CT	43202 to 44971	1770
UL54 (except at 78651)*	PT	77695 to 79992	2298
UL55 to 56	PT	80848 to 82731	1884
UL80	CT	114401 to 116793	2393
UL83	PT	118890 to 119937	1048
UL97	Mut	140784 to 142090	1307
UL122 to 126	PT	170525 to 173182	2658
UL132	CT	176380 to 177192	813
US17	PT	198929 to 199312	384

* The sequence at position 78651 is described as a Y (C/T) at NIST and a C in Genbank

CT: commercial target; PT: published target; Mut: mutation confers resistance to ganciclovir

BK Virus

- Polyomavirus closely related to JC virus and simian virus 40
- Genome is circular dsDNA approximately 5 kb
- Six common subtypes 1a, 1c, 3, 4, 5, & 6
- The Plan:
  1. Ideally the SRM would contain a panel of virus subtypes
     - Viral DNA in plasmids and isolated viral DNA
  2. Sequence of all components would be given
  3. One concentration per subtype
  4. Concentrations would aim high so dilutions can be made to calibrate everyday use quantitative standard (see "How to Use SRM 2366 to Calibrate EUQS").
  5. Concentrations would be certified by dPCR

Packaging

- Perfluoroalkoxy fluoropolymer (PFA) (e.g. Teflon) screw cap vials (0.5 mL) were used to package SRM 2366.
- PFA used to minimize absorption of DNA into tube walls
- BK virus infections can have high viral load, therefore the concentration of the SRM will be higher
- Currently investigating if non-PFA tubes will have insignificant absorption of DNA into tube walls considering high concentration of BK virus DNA

Material Acquisition

- Obtaining material to create a SRM is a difficult part of the process
- If you have access to material and would be willing to sign a Material Transfer Agreement, please contact Ross Haynes Ross.Haynes@nist.gov 301-975-4469
- Any collaboration – intellectual or material – is welcome to help facilitate the best possible clinical reference material

Viruses Under Consideration

- Epstein Barr virus (2 major strains; Type A & Type B)
- Adenovirus

Clinical DNA Information Resource

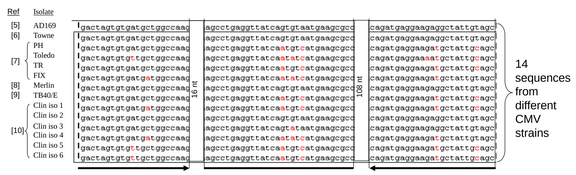
http://www.nist.gov/mml/bmd/genetics/clinical_dna.cfm

Clinical DNA Information Resource contains:

- Recent publications and presentations – including this one
- Information on DNA-based Standard Reference Materials (SRMs) and candidate SRMs
- Information to aid scientists in their research

Sequence alignments

The qPCR assay below [4] has a large number of mismatches that cause a false negative rate of 24%. A little more time in the library, probably would have lead this researcher to redesign the primer and probe binding sites.



Published Assays

- 73 unique qPCR assays have been collected, the primers and probes have been documented and put into one spreadsheet
- Errors have been corrected and traced to:
  1. Transcribing from paper to paper and
  2. Changes in GenBank sequences
- Each assay has a key to which strains of CMV [5] – [9] match that assay, which is a rough indicator of the sensitivity of the assay

References:
[1] Heyries et. al. "Megapixel digital PCR." Nature Methods. 8 (8), 649-651 (2011).
[2] Dube et al. "Mathematical analysis of copy number variation in a DNA sample using digital PCR on a nanofluidic device." PLoS One. 2008 Aug 6;3(8):e2876.
[3] Toman and Possolo. "Laboratory effects models for interlaboratory comparisons." Accred Qual Assur (2009) 14:553-563.
[4] Tanaka et al. "Quantitative analysis of cytomegalovirus load using a real-time PCR assay." J Med Virol. 2000 Apr;60(4):455-62.
[5] Chee et al. "Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169." Curr Top Microbiol Immunol. 1990;154:125-69. Genbank X17403
[6] Dunn et al. "Functional profiling of a human cytomegalovirus genome." Proc Natl Acad Sci U S A. 2003 Nov 25;100(24):14223-8. Genbank AY315197 "Towne"
[7] Murphy et al. Coding potential of laboratory and clinical strains of human cytomegalovirus. Proc Natl Acad Sci U S A. 2003 Dec 9;100(25):14976-81. Genbank AC146904 - AC146907 "PH-BAC", "Toledo BAC", "TR-BAC", "FIX-BAC"
[8] Dolan et al. "Genetic content of wild-type human cytomegalovirus." J Gen Virol. 2004 May;85(Pt 5):1301-12. Genbank NC_006273 "Merlin"
[9] Sinzger et al. "Cloning and sequencing of a highly productive, endotheliotropic virus strain derived from human cytomegalovirus TB40/E." J Gen Virol. 2008 Feb;89(Pt 2):359-68. Genbank EF999921 "TB40-BAC4"
[10] Chou. "Effect of interstrain variation on diagnostic DNA amplification of the cytomegalovirus major immediate-early gene region." J Clin Microbiol. 1992 Sep;30(9):2307-10. Genbank M95634 - M95639

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