NIST-FDA Workshop on Standards for NGS Detection of Viral Adventitious Agents in Biologics and Biomanufacturing. Sept. 18-19, 2019, Gaithersburg, MD

U.S. FOOD & DRUG ADMINISTRATION

Development of Well-Characterized Reference Virus Stocks for NGS standardization

Pei-Ju Chin, M.S., Ph.D.
Khan Lab
OVRR/CBER/FDA

Disclaimer

This presentation is an informal communication and represents my own best judgment. The material in this presentation and my comments do not represent FDA's views or policies.

Outline

- Virus selection and development of reference virus stocks
- Characterization
  - General
    - Physical properties, infectivity, genomic sequences
  - Specific
    - Determination of consensus sequences and identification of variants
    - Determination of genomic copy number
      - Standardization of ddPCR protocols
      - Discussion of factors influencing the results

In-house Efforts on NGS Standardization for Adventitious Virus Detection in Biological Products

Characterization of Reference Virus Stocks

- **Production, Storage, Distribution**
- **Characterization of viral genomes**
  - o Variant Detection
  - o Total Nucleic Acid Preparation
  - o Copy Number Determination

Viral Sequence Database

- Reference Viral DataBase (RVDB)
- Annotation of Virus-specific Sequences
- Database Clean-up

Adventitious Virus Detection Workflows/Pipelines

- Evaluation of Workflow Performance
- Optimization of Bioinformatic Pipeline

Initial Collaborative Study to Evaluate Virus Detection by Different NGS Platforms

- **Objective:** To evaluate NGS sensitivity for virus detection using different model viruses spiked in cellular background representing different biological sample types

- **Study Participants**: FDA/Khan, Sanofi, GSK

- **Study Design**:
  - Viruses for the study were selected based upon different physicochemical properties, representing potential viruses of concern in biologics, and commercial availability
  - Different sample preparation methods
  - Different sequencing platforms (454-Roche, Illumina)
  - Different virus detection pipelines for NGS data analysis

		Particle size (nm)	Envelope	Genome topology	Genome size (bp/b)	Physical chemical resistance
Epstein-Barr virus-1		122-180	YES	ds-DNA circular	172,281	Low to Medium
Feline Leukemia Virus		80-100	YES	ss-RNA dimeric	8,448	Low
Human Respiratory Syncytial Virus-A		150-300	YES	ss-RNA linear	15,158	Low to Medium
Human Reovirus-1		60-80	NO	ds-RNA segmented	1,196 3,915	Medium to High
Porcine Circovirus Type 1		16-18	NO	ss-DNA Circular	1,758	High

- **Results:** **Comparable virus detection results were obtained by the three laboratories regardless of sample processing, library preparation, sequencing platform, and bioinformatic analysis: between 0.1 and 3 viral genome copies per cell were detected for all of the model viruses used**
- **Published**: ***Khan et al., mSphere, Sept/Oct 2017*. NGS datasets are available in NCBI.**

Development of Well-characterized Virus Stocks

❖ The results from the "pilot" spiking study and AVDTIG discussions formed the basis for initiating efforts by the Khan Lab for large-scale preparation of virus reference stocks for NGS platform evaluation and standardization

❖ 5 large scale stocks were prepared by the Khan Lab through a contract with ATCC

❖ Characterization

- Infectious titer per mL (>10^6 $TCID_{50}$ per mL)
- Number of particles : TEM
- Genome copy number: ddPCR (>10^8 gc per mL)
- Residual nucleic acid: Illumina HiSeq

  (*Fabio La Neve and Davide Scaglione- MerckGroup)*
- Stability studies (24-month) have been completed *(ATCC)*
  - infectious titer; genome copy number by ddPCR

➢**Vialed individually to allow freedom for custom-mixing, as needed by user**

Availability of Virus Stocks

Virus Name	Total vials prepared*
PCV-1	392
REO	403
FeLV	503
RSV	388
EBV	490

*volume about 0.5 ml

- **Distributed for spiking studies in the AVDTIG** for broader evaluation of performance of HTS platforms and virus detection

➢ **Available for additional NGS-related evaluation/standardization: Contact arifa.khan@fda.hhs.gov**

Characterization of Reference Genomes

Genomic Characterization Virus Stocks

- Illumina HiSeq 2500
- Paired End: 2X101bp
- Flowcell V3 - 1 LANE/virus
- Total yield: 232 Gb (over 5 lanes)
- >Q30: 85%

Fabio La Neve and Davide Scaglione. 2018. High-Throughput Sequencing as a method to qualify virus stocks. Parenteral Drug Association (PDA) Europe Conference Virus Forum

Read Mapping Profiles for Reference Genome in Virus Stocks

Virus Name	Total Reads	% Ref Genome covered
RSV	533,643,230	100
REO	436,235,370	100
EBV	392,018,864	100
FeLV	505,153,300	100
PCV	408,596,882	100

EBV/SMRV	392,018,864	100
PCV/PERV	408,596,882	100

SMRV: Squirrel Monkey Retrovirus
PERV: Porcine Endogenous Retrovirus

Characterization of Viral Genome Copy Number

Background and Approaches

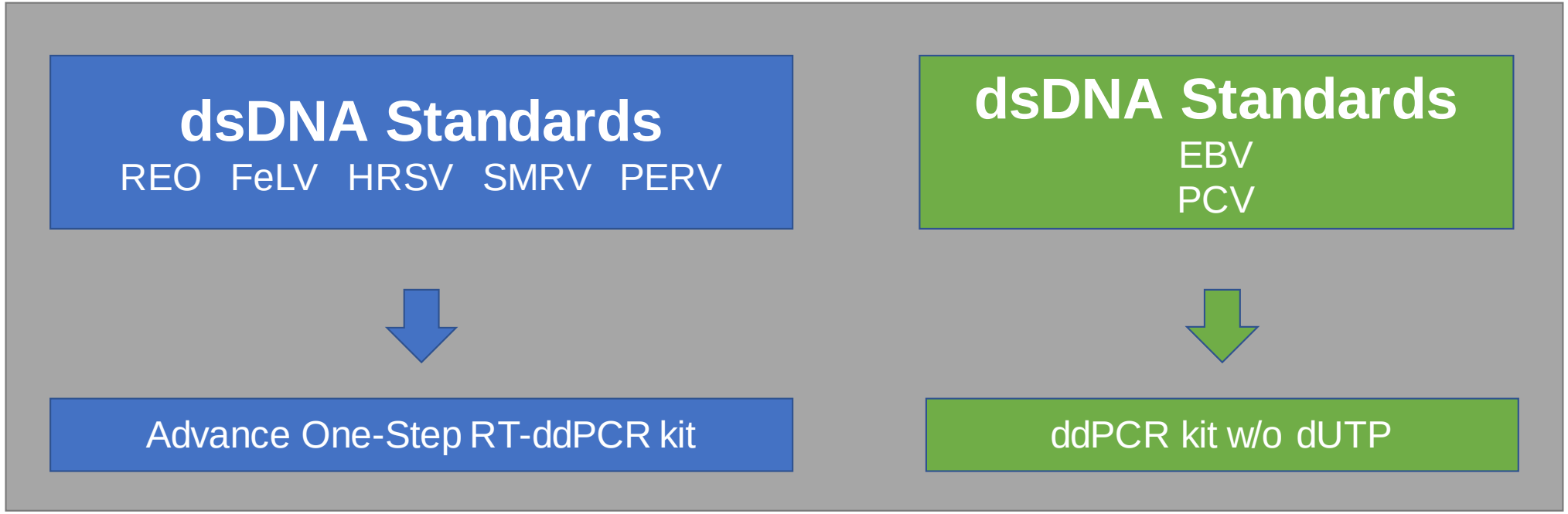
- The virus stocks were generated for performing spiking studies to evaluate the sensitivity of next-generation sequencing (NGS) technology for adventitious virus detection in different matrices mimicking samples relevant to biologics and biotherapeutics
- A dependable and robust technology is crucial for obtaining an accurate genome copy number, which is critical for the experimental design and evaluation of spiking studies
- Protocols for Digital Droplet PCR (ddPCR) were developed and evaluated *in-house*
  - Evaluation of sensitivity and accuracy by dsDNA standards
  - Two-Step (*in-house*) versus One-Step (ATCC) RT-ddPCR
  - Pretreatment of samples with large genome
    - Restriction enzyme digestion

ddPCR Primer/Probe Panel

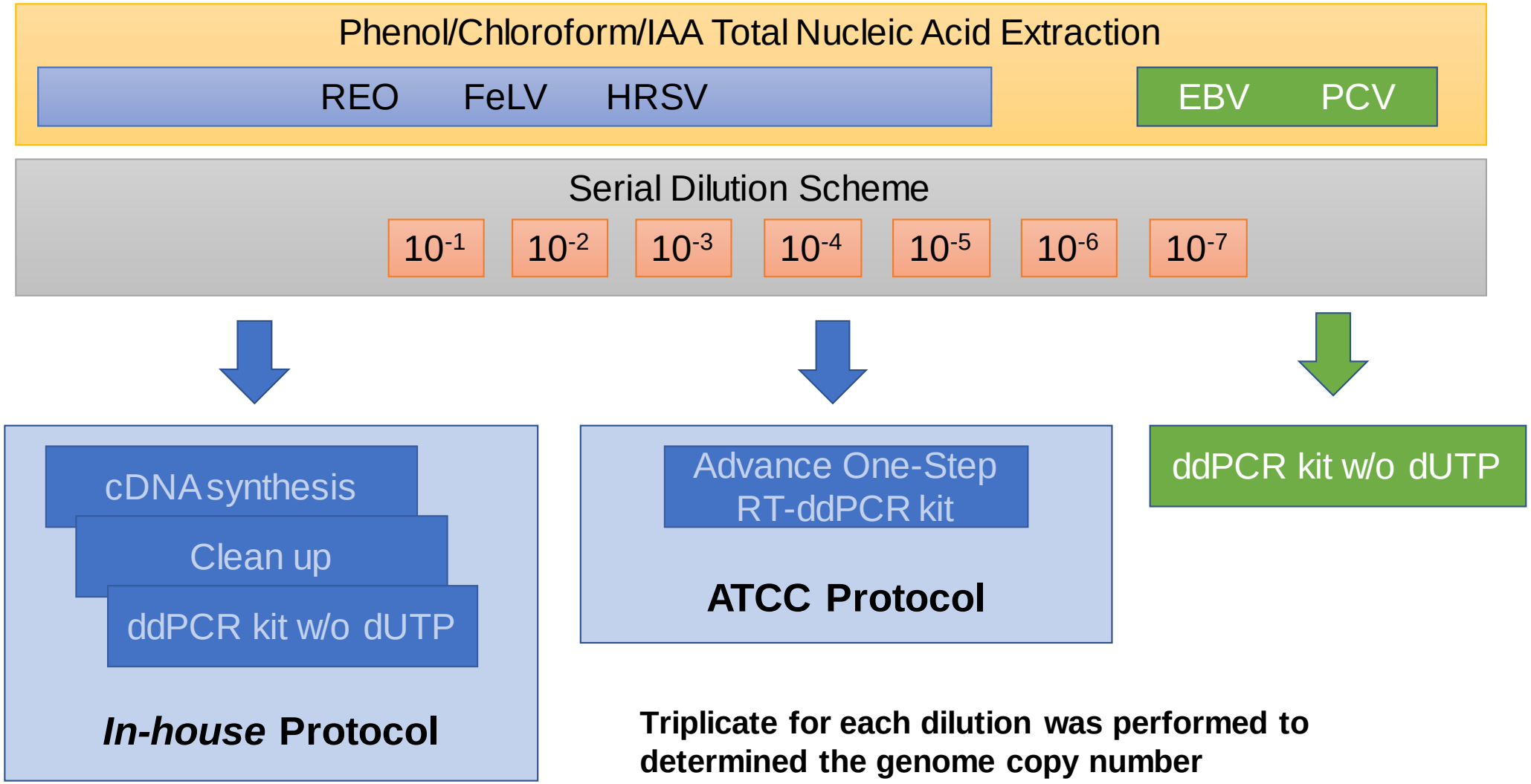
Assay Name	Assay Origin
ddPCR_ATCC_CN_HRSV	ATCC
ddPCR_ATCC_CN_REO1	ATCC
ddPCR_ATCC_CN_EBV	ATCC
ddPCR_ATCC_CN_FeLV	ATCC
ddPCR_ATCC_CN_PCV1	ATCC
ddPCR_CN_SMRV	In-house
ddPCR_CN_PERV	In-house

Evaluation of ddPCR Sensitivity and Accuracy by dsDNA Standards

10^{-1} 10^{-2} 10^{-3} 10^{-4} 10^{-5} 10^{-6} 10^{-7} 10^{-8} 10^{-9} 10^{-10} 10^{-11}



Genomic Copy Number Determination by (RT)-ddPCR



Triplicate for each dilution was performed to determined the genome copy number

Genome Copy Number Determination for RNA Viruses- One-Step versus Two-Step RT-ddPCR

Name	Copy Number (per mL)	
	Two-Step RT-ddPCR	One-Step RT-ddPCR
FeLV	≈10^{11}	≈10^{11}
RSV	≈10^{9}	≈10^{9}

- **One-Step RT-ddPCR is comparable to Two-Step protocol**

Genome Copy Number Determination for DNA Viruses with Large Genome-Restriction Enzyme Digestion

Name	Copy Number (per mL)	
	With BamHI-HF Digestion	Without BamHI-HF Digestion
EBV	$\approx 10^8$	$\approx 10^8$

- **Digestion of virus stocks with large genome is not required**

Conclusions

- Genomic characteristics of virus stocks were determined by NGS
- Linear range of (RT)-ddPCR assay was determined
  - Serial dilutions provides the confidence that the loading titer is within the linear range of ddPCR assay
- Phenol/Chloroform/IAA nucleic acid extraction protocol has at least 2-fold more yield than the column-based protocol (data not shown)
- One-Step RT-ddPCR protocol avoids the bias of random priming issue, as well as the uncertain recovery yield after clean up in Two-Step procedure
- Restriction enzyme digestion for EBV is not required for ddPCR reaction

Acknowledgements

Arifa Khan's Group

- Sandra Fuentes

MerckGroup

- Fabio Le Neve
- Davide Scaglione

ATCC

- Heather Couch
- Sujatha Rashid

Advanced Virus Detection Technologies Interest Group (AVDTIG)

NIST

- Megan Cleveland

❖ ***FDA Funding:*** *Medical Countermeasures Initiative*