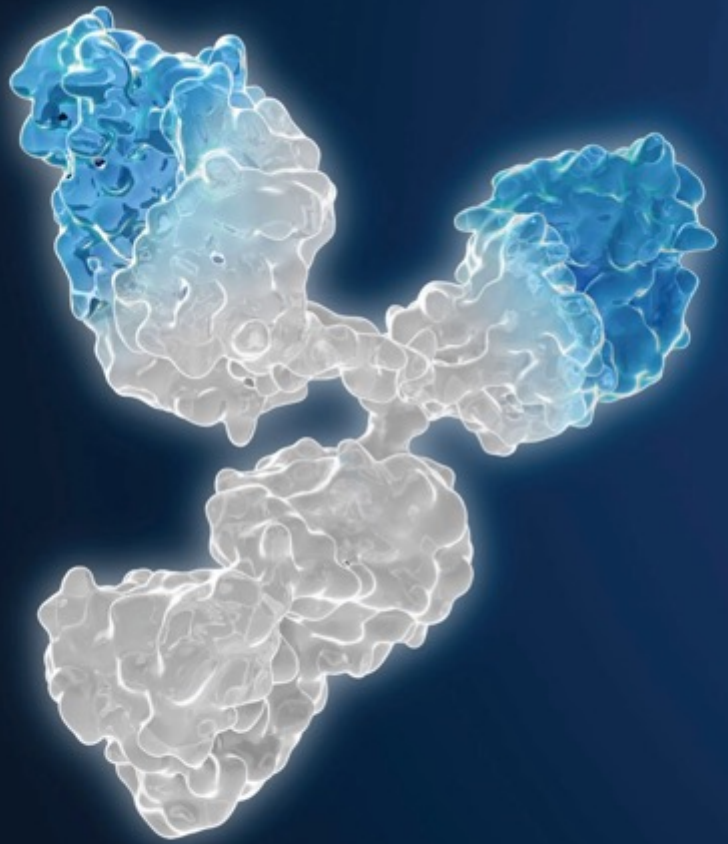


BACKGROUND NGS DATA FOR SAMPLE MATRIX AND REFERENCE CONTROL STANDARDS USING NANOPORE SEQUENCING

NIST CONFERENCE SEPTEMBER 18TH AND 19TH 2019

SERGE MONPOEHO, PHD
SENIOR DIRECTOR QUALITY CONTROL, CHIEF VIROLOGIST
REGENERON PHARMACEUTICALS, QC VIROLOGY

REGENERON
SCIENCE TO MEDICINE®

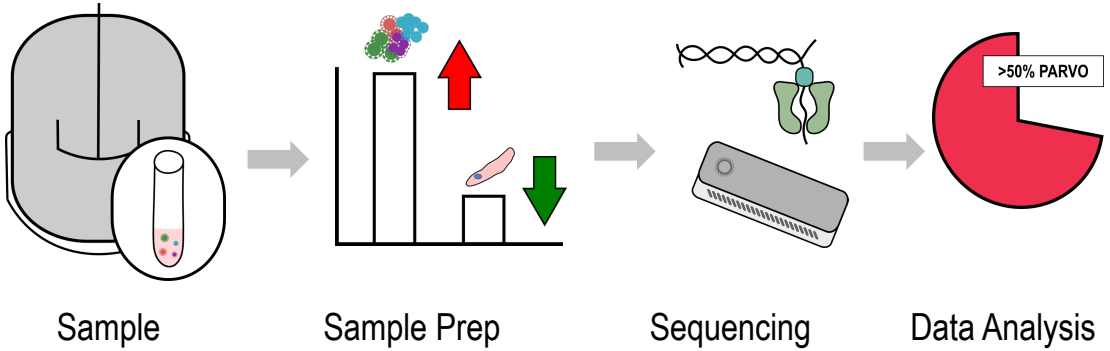
CURRENT USE OF NGS

- **Testing Requirement for MAb Lot Release**
  - In Vitro Virus Screening assay (14 Day)
  - Mycoplasma Screening (28 Day)
  - PCR Rodent Parvovirus (for known contaminant not optimally detected during In Vitro Virus Assay)

- **NGS is currently used:**

As Investigational Tool	For Adventitious Virus Screening Method Development
- OOS investigation during adventitious virus testing - Abnormal Bioreactor Events (cell crash) investigation	- Direct NGS Method development - Hybrid Method (Culture + NGS) development

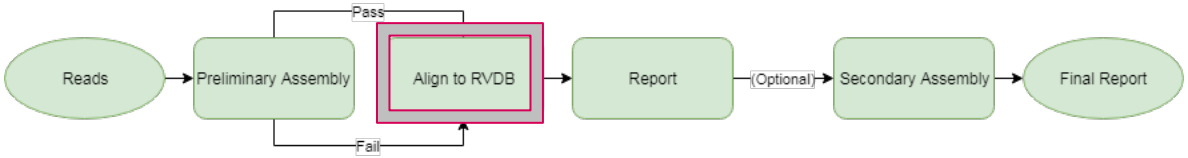
SEQUENCING TECHNOLOGY USED



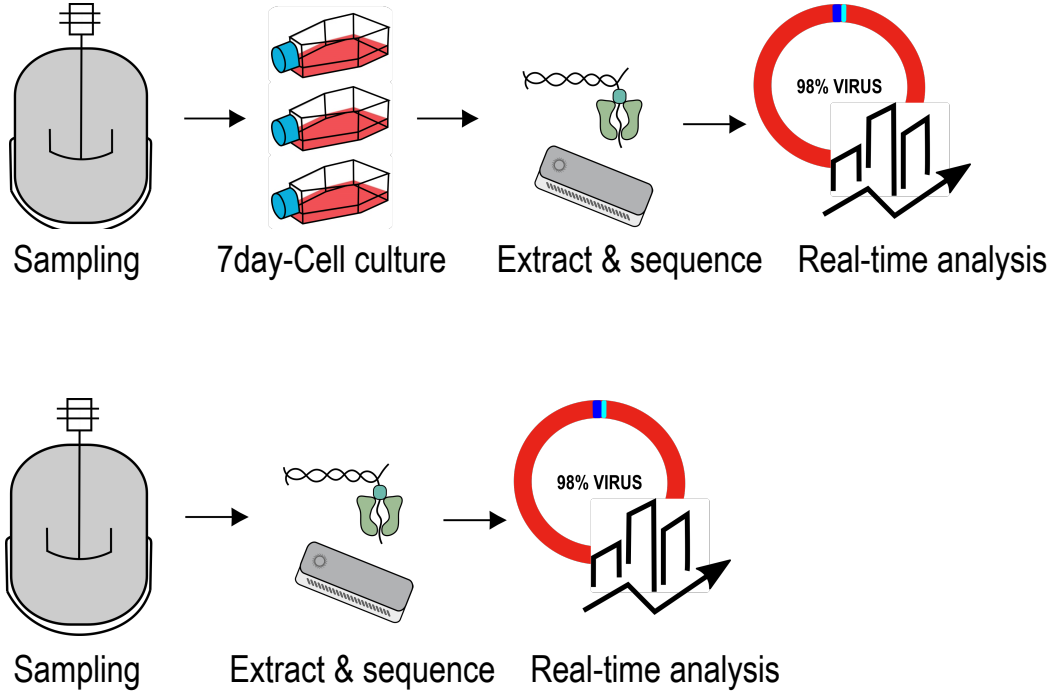
Oxford Nanopore MinION

- Direct DNA sequencing (no amplification)
- Long reads technology
- Real time streaming of sequence data

Data Analysis Workflow



DATA ANALYSIS CHALLENGES



Challenges:

- Nucleic acids in the bioreactor
- Nucleic acid or dead virus in Cell culture reagents e.g. FBS (even gamma irradiated)
- Nucleic acids or virus from indicator cells
- Nucleic acids from nucleic acid purification reagents
- Nucleic acids from sequencing reagents
- Nucleic acids from the internal controls

Generate **background signal** to be distinguished from the real contaminant

NGS BACKGROUND EXAMINATION THROUGH 3 STUDIES

1. CELL LINES & PBS 
   - CHO-K1
   - MRC-5
   - VERO
   - PBS
   - Water

2. FBS 
   - Nuclease-treated, enriched
   - Enriched
   - Unenriched

3. SPIKED ASSAY CONTROLS
   - M13
   - MS2 
   - In host background at 1E7 pfu/µl

→ **Sample Prep** → **Nanopore Sequencing** → **Align to RVDB**

STUDY-1: NGS BACKGROUND IN PBS, WATER, AND CELL LINES CHO, VERO, MRC-5

1. CELL LINES & PBS 
   - CHO-K1
   - MRC-5
   - VERO
   - PBS
   - Water

2. FBS 
   - Nuclease-treated, enriched
   - Enriched
   - Unenriched

3. SPIKED ASSAY CONTROLS 
   - M13
   - MS2
   - In host background at 1E7 pfu/µl

Sample Prep → **Nanopore Sequencing** → **Align to RVDB**

STUDY-1: NGS BACKGROUND IN PBS AND WATER

PBS			
Findings	Reads	Average Length	Coverage %
Total Reads	7,131		
Unaligned	5,379 (75%)		
CHO	1,190 (17%)		
Human	407 (6%)		
SV40	129 (2%)	379	81%
Semliki_forest_virus	1 (~0%)	424	2%
Guanarito_mammarenavirus	3 (~0%)	391	6%

PBS			
Findings	Reads	Average Length	Coverage %
Hum Picobirnavirus	1 (~0%)	485	11%

Water			
Findings	Reads	Average Length	Coverage %
Total Reads	1,474		
Unaligned	1,473 (~100%)		
Human	1 (~0%)	245	

STUDY-1: NGS BACKGROUND IN INDICATOR CELL LINE CHO-K1

CHO-K1			
Findings	Reads	Average Length	Coverage %
Total Reads	3,173,164		
Unaligned	497,674 (17%)		
CHO	2,675,490 (84%)	524	
Human	30,766 (1%)	508	
Guanarito_mammarenavirus	18 (~0%)	887	11%
White_spot_syndrome_virus	6 (~0%)	637	<1%
RVLP	708 (~0%)	870	95%

CHO-K1			
Findings	Reads	Average Length	Coverage %
BVDV	1 (~0%)	1943	15%
Bosavirus	97 (~0%)	786	92%

STUDY-1: NGS BACKGROUND IN INDICATOR CELL LINE VERO

VERO			
Findings	Reads	Average Length	Coverage %
Total Reads	3,130,681		
Unaligned	1,151,406 (37%)		
CHO	1,170 (~0%)	634	
Human	1,193,298 (38%)	1,090	
Guanarito_mammarenavirus	206 (~0%)	504	13%
White_spot_syndrome_virus	4 (~0%)	250	<1%
Endogenous retrovirus	2,598 (~0%)		

VERO			
Findings	Reads	Average Length	Coverage %
BVDV	3 (~0%)	1453	54%
Bosavirus	98 (~0%)	832	92%
SV40	725,093	671	90%
Shamonda	1 (~0%)	273	6%

STUDY-1: NGS BACKGROUND IN INDICATOR CELL LINE MRC-5

MRC-5			
Findings	Reads	Average Length	Coverage %
Total Reads	527,568		
Unaligned	130,000 (25%)		
CHO	218 (~0%)	264	
Human	396,643 (75%)	405	
Guanarito_mammarenavirus	177 (~0%)	470	11%
White_spot_syndrome_virus	6 (~0%)	194	<1%

MRC-5			
Findings	Reads	Average Length	Coverage %
BVDV	1 (~0%)	666	6%
Bosavirus	457 (~0%)	492	93%
Hum Hep B	6 (~0%)	281	66%

STUDY-1: SUMMARY

Findings	Reads	Coverage
Unaligned	Expect 10-40% of reads unalligned	
Indicator Cells (host)	Expect 40-85% of reads	
Guanarito_mammarenavirus	[3-206]	[6%-13%]
White_spot_syndrome_virus	[4-6]	<1%
SV40	[129-725,093]	[80%-90%]
Hum Picobirnavirus	1 read	11%
RVLPs	708	95%
Semliki_forest_virus	1	2%

Findings	Reads	Coverage
BVDV	[1-3]	[6%-54%]
Bosavirus	[97-457]	[92%-93%]
Hum Hep B	6	66%
Endogenous retrovirus	2,598	
Shamonda	1	6%

Example of acceptance criteria:
>10Reads and >60%Coverage

- Real Contaminants

-Artifacts

STUDY-2

1. CELL LINES & PBS 
   - CHO-K1
   - MRC-5
   - VERO
   - PBS
   - Water

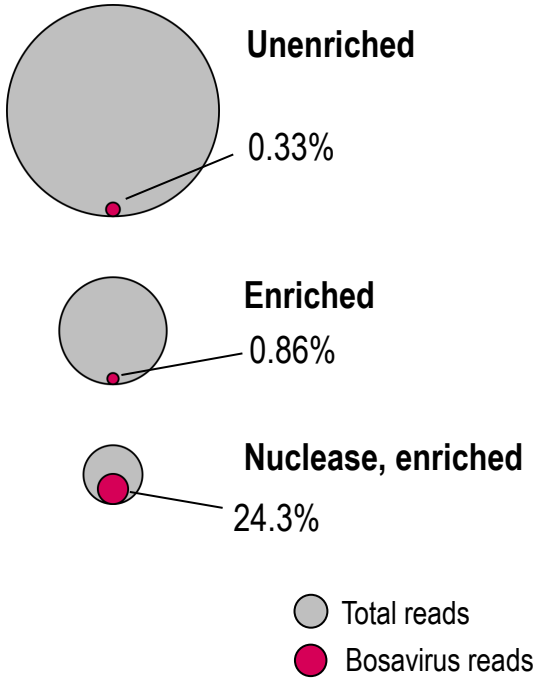
2. FBS 
   - Nuclease-treated, enriched
   - Enriched
   - Unenriched

3. SPIKED ASSAY CONTROLS
   - M13 
   - MS2
   - In host background at 1E7 pfu/µl

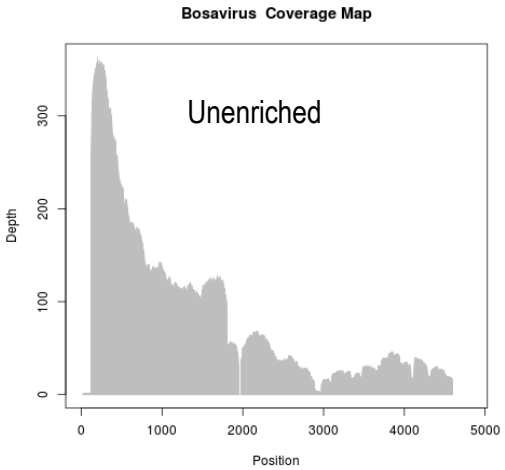
→ **Sample Prep** → **Nanopore Sequencing** → **Align to RVDB**

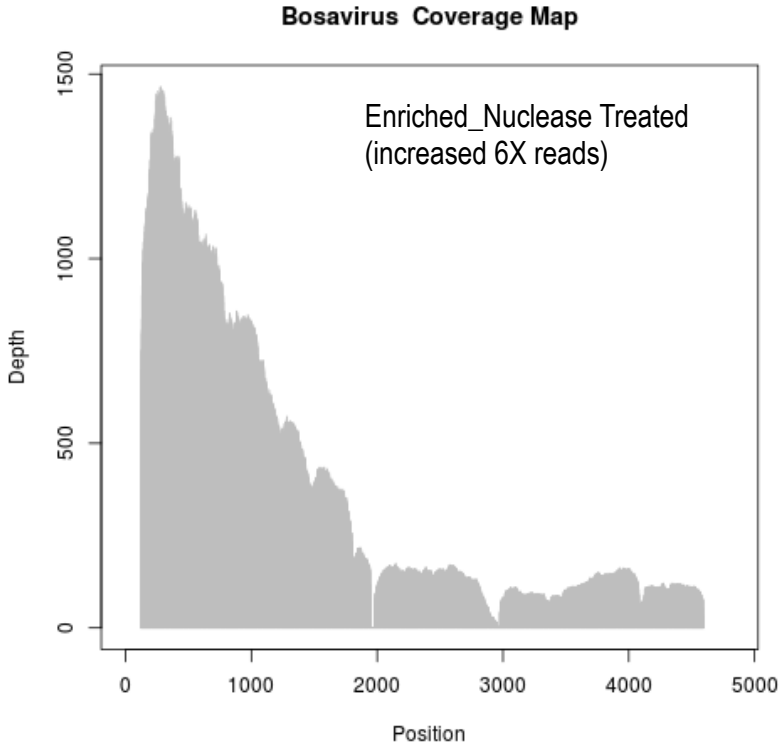
STUDY-2: FBS (GAMMA IRRADIATED)

	Unenriched			Enriched			Enriched_Nuclease Treated		
	Reads	Average Length	Coverage %	Reads	Average Length	Coverage %	Reads	Average Length	Coverage %
Host	82,877			19,471			4,868		
Endogenous elements/provirus	4	419	87.74	1	755	73.57	6	304	5.36
Bosavirus	1087	528	94.15	707	894	92.62	6090	389	92.62
BVDV	4	379	9.68	1	1568	13.01	7	439	15.32
Stealth Virus 1	34	707	67.82	7	1008	47.07	29	580	60.33
% (Bosavirus/Total Reads)	0.33			0.87			24.39		
% (Stealth Virus 1/Total Reads)	0.01			0.01			0.12		



STUDY-2 BOSAVIRUS COVERAGE MAP





STUDY-2 SUMMARY

Considering our criteria of :
>10Reads and >60%Coverage = Real Contaminant

Gamma Irradiated FBS				
	Findings	Reads	Average Length	Coverage
Real Contaminants	Bosavirus	6090	389	93%
	Stealth Virus 1	29	580	60%
Artifacts	BVDV	7	439	15%
	Endogenous retrovirus elements	6	304	5%

STUDY-3

1. CELL LINES & PBS 
   - CHO-K1
   - MRC-5
   - VERO
   - PBS
   - Water

2. FBS 
   - Nuclease-treated, enriched
   - Enriched
   - Unenriched

3. SPIKED ASSAY CONTROLS
   - M13 
   - MS2
   - In host background at 1E7 pfu/µl

→ **Sample Prep** → **Nanopore Sequencing** → **Align to RVDB**

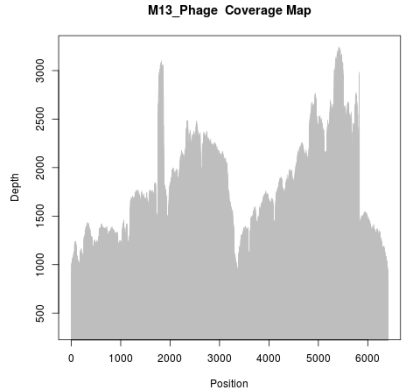
STUDY-3: INTERNAL CONTROL PHAGES ADDED TO CHO-K1 SAMPLES AND PBS

	Sample (n=3)			PBS Control		
	Reads	AVG length	% coverage	Reads	AVG length	% coverage
M13	22,879	394	100	9,926	397	100
MS2	3,533	449	97	9,128	514	99.66
CHO-K1	1,207	337		1	378	
Human	315	376		83	370	
Semliki Forest Virus	2,203	498	8	980	537	7.79
Hamster Retroviral sequence	3	637	77	0		
CHO RVLP	5	621	28	0		
Macacine betaherpesbvirus 3	2	440	<1	1	375	0.13
Stealth_virus_1_clone_3B43	10	684	45	5	623	43.95
Stealth_virus_1_clone_3B43_T3.	4	366	42	5	623	43.95
Guanarito mammarenavirus isolate	1	520	<1	0		
Herpes simplex virus (type 1 /strain RH2)	5	363	<1	4	363	0.71
Total Sequenced Reads	52,332			33,146		
Unaligned reads	22,144			13,015		

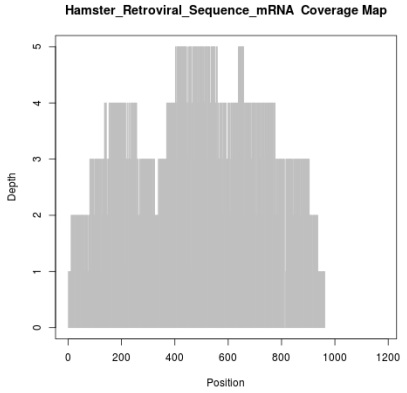
Reads as a percentage of total reads	**Spiked Sample**	**PBS**
M13	43.72%	29.95%
MS2	6.75%	27.54%
CHO-K1	2.31%	0.00%
Human	0.60%	0.25%
Semliki Forest Virus	4.21%	2.96%

STUDY-3: READS COVERAGE MAPS

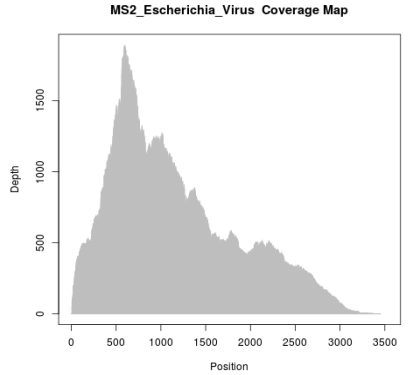
M13 6,407 bp



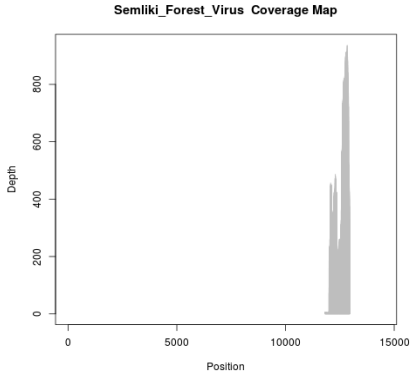
Retrovirus 1,184bp



MS2 3,534 bp



Semliki Forest Virus 14,529bp



STUDY-3 SUMMARY

Considering our criteria of :
>10Reads and >60%Coverage = Real Contaminant

- Recovery of the internal control phages
- No real contaminant was noted
- New artifacts such as Macacine betaherpesbvirus 3 and Herpes simplex virus (type 1 /strain RH2)
- MS2 recovery seems impacted by the sample matrix
- M13 is not impacted
- High reads very low coverage for Semliki Forest Virus
- Unaligned reads is between 39% to 42% of total reads (suspecting host of the phages)

CONCLUSIONS AND FUTURE DIRECTIONS

- Identified real contaminants of our reagents
  - SV40
  - Bosavirus
  - RVLPs
  - Simian Endogenous retrovirus elements
  - Stealth Virus 1
- Identified several artifacts and the most prominent being
  - BVDV
  - Semliki_forest_virus
- These findings are first steps in determining criteria for real contaminants vs. artifacts

Future Directions:

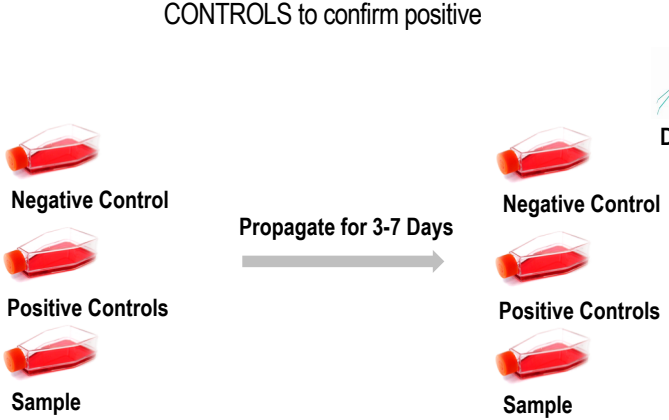
- Continue to investigate and gather sequencing data from all reagents and materials used in NGS assays.
- Optimize the data analysis with machine Learning to characterize background and artifacts
- Based on current LOD of MinION, virus amplification on cell culture will be considered

FUTURE DIRECTION: HYBRID ASSAY (CULTURE + MINION)

- **Assay system suitability**
  - Analyze NEGATIVE CONTROL to determine if different from expected background.
  - NEGATIVE CONTROL reads SUBTRACTED from POSITIVE CONTROLS to confirm positive

- **Sample Suitability**
  - Analyze INTERNAL CONTROLS spiked into all samples to determine if recovery acceptable

- **Sample results**
- Analysis includes:
  - Review of historical background (AI potentially)
  - NEGATIVE CONTROL read analysis subtracted
  - Internal Controls SUBTRACTED from SAMPLE reads
  - Remaining reads are ALIGNED to Viral database



QUESTIONS?