# Next Generation Sequencing – comparative study results

**Jens Modrof, Andreas Berting, Thomas R. Kreil**
*Global Pathogen Safety*
*Baxalta*

June 10, 2015

# Overview

- Parallel evaluation of broad virus detection methods
  - comparative blinded study with deliberately virus-spiked samples

- Characterization of a human cell substrate using NGS
  - comparative pilot study

# Broad virus detection methods

- New broad virus detection methods should be able to detect
  - all known viruses by multiplexed PCRs or microarray
  - additionally, yet unknown viruses by massively parallel sequencing

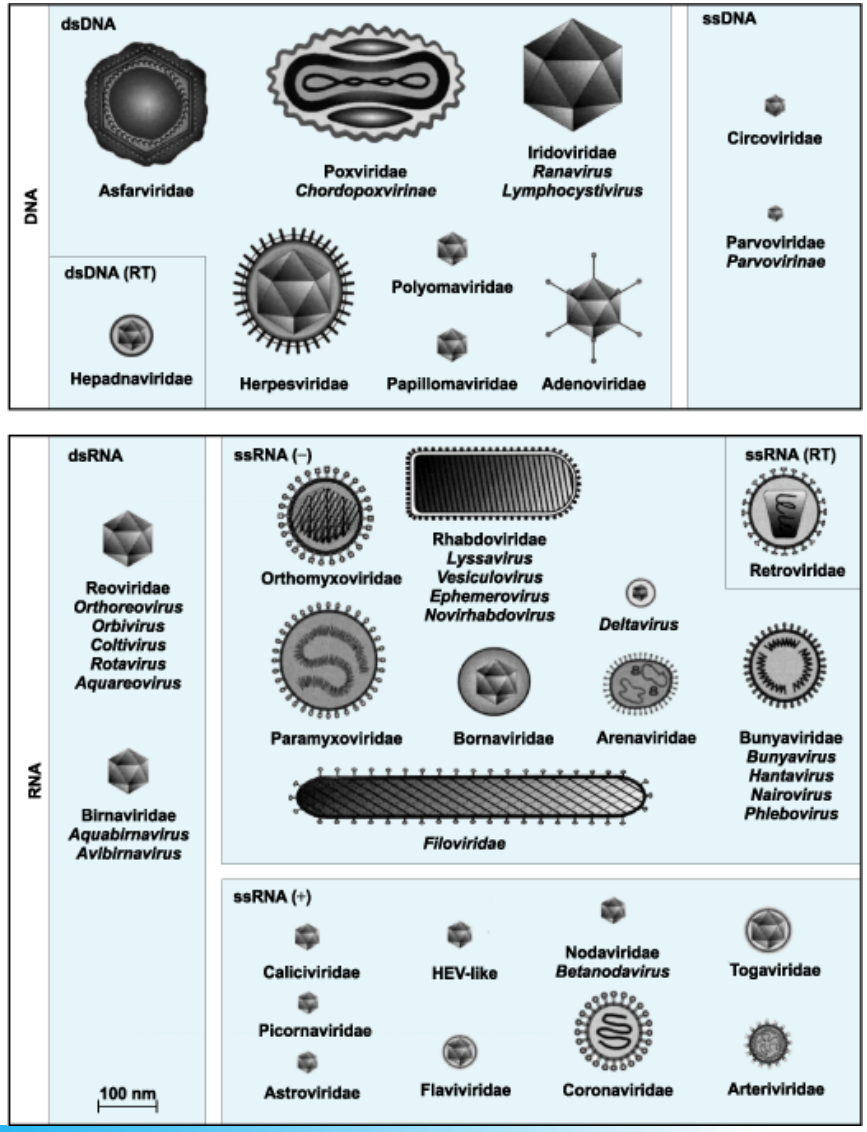
- Scope of application?
  - one time testing of cell banks or contamination
  - routine testing on adventitious agents or process control

# Comparative blinded study using virus spikes[1]

- Selection of model viruses
- Characterization of model viruses
- Sample preparation
- Blinded set of samples was sent to four different investigators
- Results

[1] Modrof J, Berting A, Kreil TR. Parallel evaluation of broad virus detection methods, PDA J Pharm Sci Technol, 2014, Nov-Dec;68(6):572-8.

# Selection of model viruses: taxonomy




From M. H. V. van Regenmortel et al., eds., Virus Taxonomy: 7th Report of the International Committee on Taxonomy of Viruses, p. 30, copyright 2000.

# Selection of model viruses: virus contaminations & Biotechnology (!)

- **More frequent** than realized …

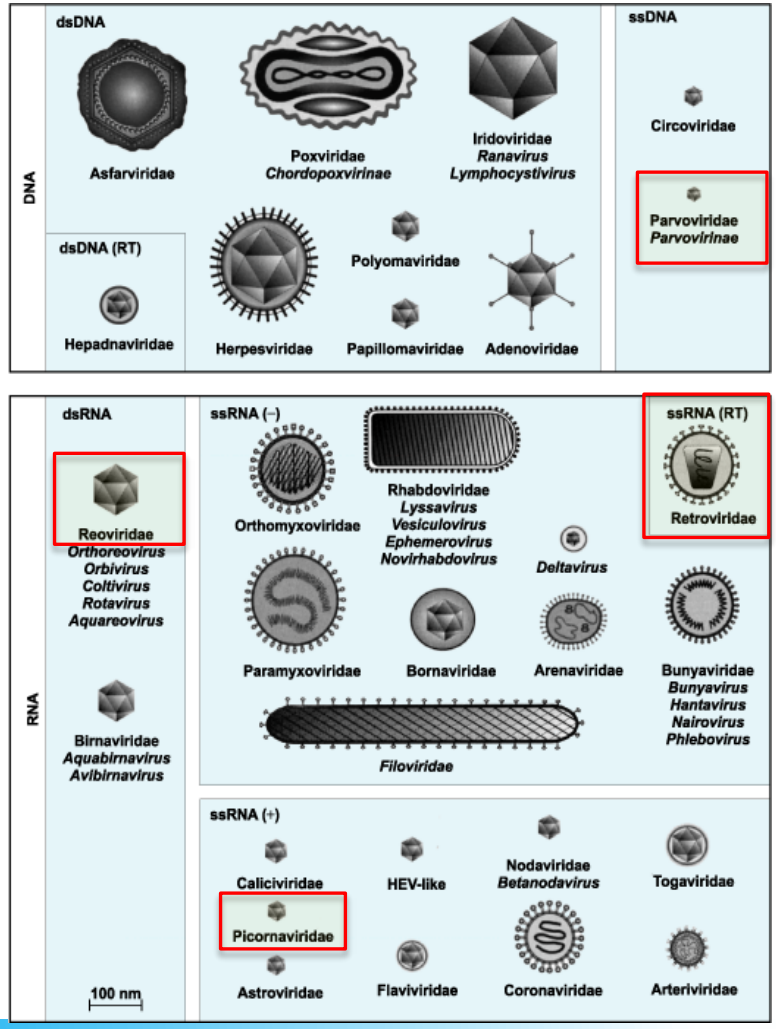
- Sources
  - MIT
  - Others

| | Virus /Host Cell | Events |
|---|---|---|
| 1985-1989 | Orbivirus /CHO<br>EHDV /CHO | 2 |
| 1990-1994 | MMV /CHO<br>MMV /CHO | 2 |
| 1995-1999 | Reovirus /Hu 1° Kidney<br>Vesivirus /CHO<br>CVV /CHO | 3 |
| 2000-2004 | CVV /Unknown<br>CVV /Unknown | 2 |
| 2005-2009 | Vesivirus /CHO<br>Vesivirus /CHO<br>MMV /CHO<br>CVV /CHO<br>MMV /CHO<br>Vesivirus /CHO | 6 |
| 2010+ | MMV /CHO<br>PCV-1 /Vero | 2 |
| Unknown | MMV /BHK-21<br>Human Adenovirus /HEK293<br>Reovirus /Unknown | 3 |
| Total | | 20 |

From Mike Wiebe, MIT **Consortium on Adventitious Agent Contamination in Biomanufacturing**

# Selection of model viruses: taxonomy & contamination events

- **MMV**
- **ReoIII**
- **EMCV**
- **X-MuLV**




From M. H. V. van Regenmortel et al., eds., Virus Taxonomy: 7th Report of the International Committee on Taxonomy of Viruses, p. 30, copyright 2000.

# Characterization of model viruses

- **MMV** (ATCC VR-1346), grown on A9 cells
- **ReoIII** (ATCC VR-824), grown on Vero cells
- **EMCV** (ATCC VR-129B), grown on Vero cells
- **X-MuLV** (ATCC VR-1447), grown on M. dunni cells

- All virus stocks were tested for
  - identity
  - possible virus cross-contamination
  - mycoplasma
  - protein concentration
  - aggregation

# Sample preparation

- Two virus conc. spiked in two different media
- Purification of total NA with DNeasy blood and tissue kit
  - in 50µl extract: between 2.8 and 6.2 log copies
- Blinded set of 20 samples was sent to 4 investigators

# Methods investigated

- TIGER (Abbott's Plex-ID / Carlsbad, CA)
- Lawrence Livermore Microbial Detection Array (Livermore, CA)
- Massively Parallel Sequencing (MPS)
  - TRON (Mainz, Germany)
  - Eurofins (Munich, Germany)

# Analysis using Plex-ID

- Broad range PCR
- Mass spectometry and signal processing
- Virus identification by base count ratio analysis

- Detection of known and related virus families

# Analysis on Lawrence Livermore Microbial Detection Array (LLMDA) – methods

- Whole transcriptome amplification
- Microarray hybridization
  - Hybridization to the 388K format
  - Scan on the Roche MS-200 2 µm scanner
- Microarray data analysis
  - Algorithm according to Gardner et al. 2010[1]
  - Signals of >99% or >95% above controls

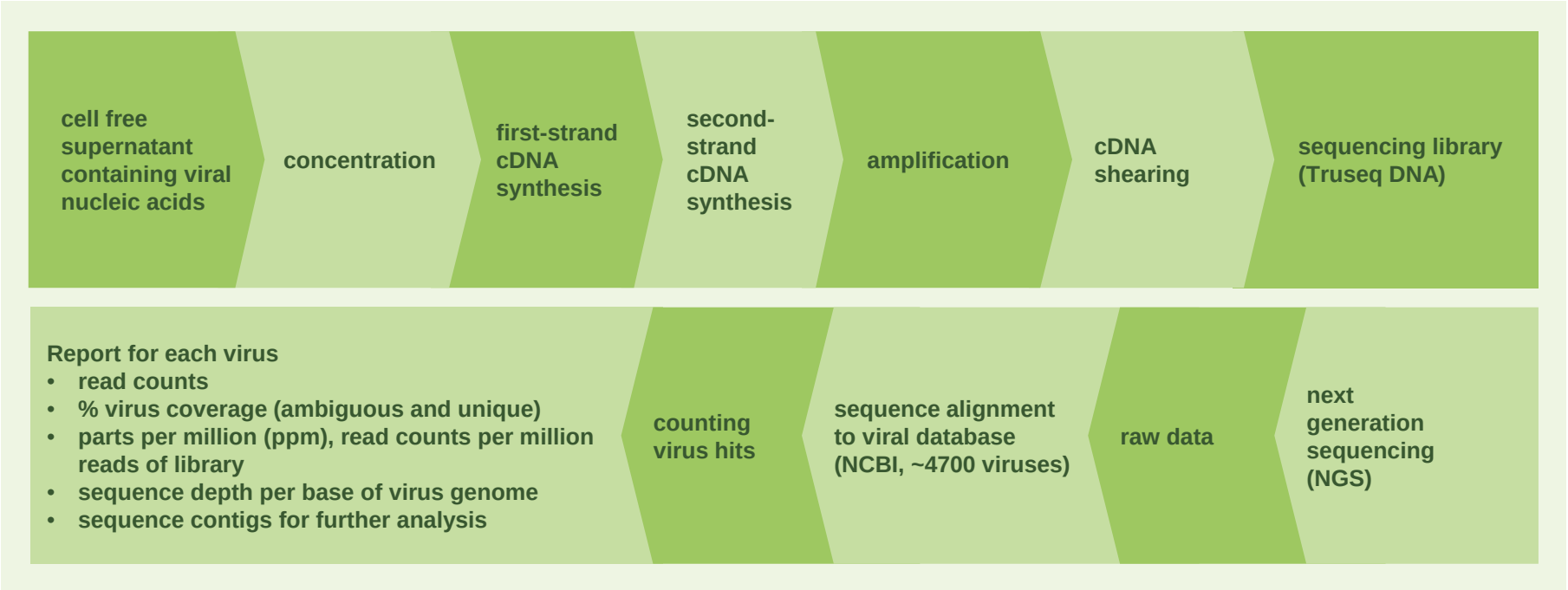
[1] Gardner et al. A microbial detection array (MDA) for viral and bacterial detection. *BMC Genomics,* 2010, 11:668

# Sample analysis using MPS

- MPS was used by two companies
  - **TRON**
  - **Eurofins**

    - Illumina® HiSeq2000
    - No host substraction method was used
    - All sequence reads were aligned to the NCBI databases

# MPS analysis by TRON – workflow




- Sequence reads aligned against NCBI-nt viral/phage database (July12, 2012)

# MPS analysis by TRON – sample 12/5

- Sample 12: medium 1, Reo III 4.1 log Copies / 50µl

| Virus ID | Description | Genome length | Read count | % covered | % covered with uniquely mapped reads | PPM |
|---|---|---|---|---|---|---|
| NC_004274.1 | Mammalian orthoreovirus 3 segment L3, complete sequence | 3901 | 19633 | 85,72 | 85,72 | 780,75 |
| NC_004282.4 | Mammalian orthoreovirus 3 segment L1, complete sequence | 3860 | 9027 | 79,22 | 79,22 | 359,24 |
| NC_004276.1 | Mammalian orthoreovirus 3 segment S4, complete sequence | 1196 | 591 | 69,57 | 69,57 | 23,44 |
| NC_004281.1 | Mammalian orthoreovirus 3 segment M3, complete sequence | 2235 | 11908 | 69,62 | 69,53 | 469,01 |
| NC_004278.1 | Mammalian orthoreovirus 3 segment M2, complete sequence | 2207 | 4894 | 65,20 | 65,20 | 194,35 |
| NC_004277.1 | Mammalian orthoreovirus 3 segment S1, complete sequence | 1416 | 7156 | 58,97 | 58,97 | 284,39 |
| NC_004279.1 | Mammalian orthoreovirus 3 segment S2, complete sequence | 1331 | 2898 | 51,54 | 50,11 | 115,35 |
| NC_004280.1 | Mammalian orthoreovirus 3 segment M1, complete sequence | 2304 | 1908 | 48,79 | 48,79 | 75,93 |
| NC_002019.1 | Influenza A virus (A/Puerto Rico/8/34(H1N1)) segment 5, complete sequence | 1565 | 8186 | 31,25 | 30,16 | 325,44 |
| NC_008912.1 | Glypta fumiferanae ichnovirus segment C9, complete sequence | 3141 | 955722 | 29,10 | 22,32 | 37834,63 |

# Eurofins – bioinformatic analysis I

- Removal of artificial and interfering sequences
  - In-house pipeline based on Convey tools

| | Raw Data | Cleaned Data | | | After Bos taurus Mapping | | | | After Bos taurus and Phix Mapping | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sample | #pairs raw | #pairs cleaned | % | #reads SE cleaned | #pairs unmapped | % | #reads SE unmapped | % | #pairs unmapped | % | #reads SE unmapped | % |
| 1D | 93,233,948 | 84,402,954 | 90.5 | 6,792,031 | 10,486,441 | 12.4 | 2,666,525 | 39.3 | **10,452,506** | **12.4** | **2,660,284** | **39.2** |
| 2C | 93,378,587 | 85,516,156 | 91.6 | 6,041,403 | 16,710,275 | 19.5 | 2,684,534 | 44.4 | **16,672,755** | **19.5** | **2,678,164** | **44.3** |
| 3B | 96,721,334 | 88,808,202 | 91.8 | 5,804,451 | 17,882,353 | 20.1 | 2,559,025 | 44.1 | **17,828,284** | **20.1** | **2,551,095** | **44.0** |
| 4B | 100,026,772 | 92,593,145 | 92.6 | 5,345,696 | 14,680,154 | 15.9 | 2,120,019 | 39.7 | **14,629,014** | **15.8** | **2,111,838** | **39.5** |
| 5B | 101,399,740 | 93,455,589 | 92.2 | 6,621,407 | 11,446,611 | 12.2 | 2,202,965 | 33.3 | **11,398,881** | **12.2** | **2,197,166** | **33.2** |
| 6C | 78,450,544 | 74,475,334 | 94.4 | 4,890,022 | 24,444,594 | 29.6 | 2,376,414 | 48.6 | **24,400,993** | **29.5** | **2,369,999** | **48.5** |

# Eurofins – bioinformatic analysis II

- De-novo assembly
- Contigs/scaffolds blasted twice using in-house pipelines based on
  - NCBI viral/phage database (April 12, 2013)
  - NCBI nt database
- Filter: minimum 75bp and 90% identity

# Spiking study results, summary

| Virus | Concentration [log copies / mL] | Medium | Plex-ID | LLMDA | MPS - TRON | MPS - Eurofins |
|---|---|---|---|---|---|---|
| MMV | 4.7 / 5.0 | 1 | + | +* | + | + |
| | | 2 | + | + | - | + |
| | 5.5 / 5.6 | 1 | + | + | + | - - |
| | | 2 | + | + | + | - - |
| Reo III | 4.1 / 4.6 | 1 | +* | +* | - - | + |
| | | 2 | +* | +* | + | + |
| | 5.4 / 6.0 | 1 | +* | + | + | + |
| | | 2 | +* | +* | + | + |
| EMCV | 5.8 / 6.2 | 1 | +* | + | + | + |
| | | 2 | +* | + | + | + |
| | 7.3 / 7.5 | 1 | +* | + | + | + |
| | | 2 | +* | + | + | + |
| X-MuLV | 4.8 / 5.3 | 1 | - | - | - | + |
| | | 2 | - | + | - | + |
| | 6.5 / 6.8 | 1 | - | + | + | + |
| | | 2 | - | + | + | + |
| Negative controls | n.a. | 1 | + | + | + | + |
| | | 2 | - - | + | + | + |

*after unblinding using an alternate amplification method

Modrof J, Berting A, Kreil TR. Parallel evaluation of broad virus detection methods,PDA J Pharm Sci Technol, Nov-Dec;68(6):572-8.

# Unexpected assignments, summary

| Virus | Medium | Plex-ID | LLMDA | MPS - TRON | MPS - Eurofins |
|---|---|---|---|---|---|
| MMV | 1 | | | **EMCV** | **MuLV/** MSV |
| | 2 | | | | |
| | 1 | | | **EMCV** | **MuLV** |
| | 2 | | | **EMCV** | **EMCV/MuLV** |
| Reo III | 1 | B19V | HIV-1 | **EMCV** | |
| | 2 | HAV | Rubella (BRD-II) | **EMCV, MMV** | |
| | 1 | | RBSDV, Porcine Rotavirus | **EMCV** | |
| | 2 | | | | |
| EMCV | 1 | | | | |
| | 2 | B19V | | | |
| | 1 | | | | **MMV** |
| | 2 | B19V | | | **Reovirus** |
| X-MuLV | 1 | | | | MSV |
| | 2 | | | | **Reovirus** |
| | 1 | | MSV | | |
| | 2 | | | | **MMV,** MSV |
| Negative controls | 1 | | | | |
| | 2 | AAV | | | **MMV** |

# Conclusion – spiking study

- All investigated methods detected most samples correctly (72% - 95%)
- All approaches had difficulties with particular viruses
- Unexpeced assignments were determined with all methods (20 - 45%)

- Methods are already useful for e.g. rapid virus detection during a contamination event

# Characterization of a cell substrate - a comparative pilot study

- Considerations:
  - Contaminations of fully tested cell lines were revealed by use of NGS
  - NGS testing is not yet established as routine analysis

# Testing of cells by MPS: procedures

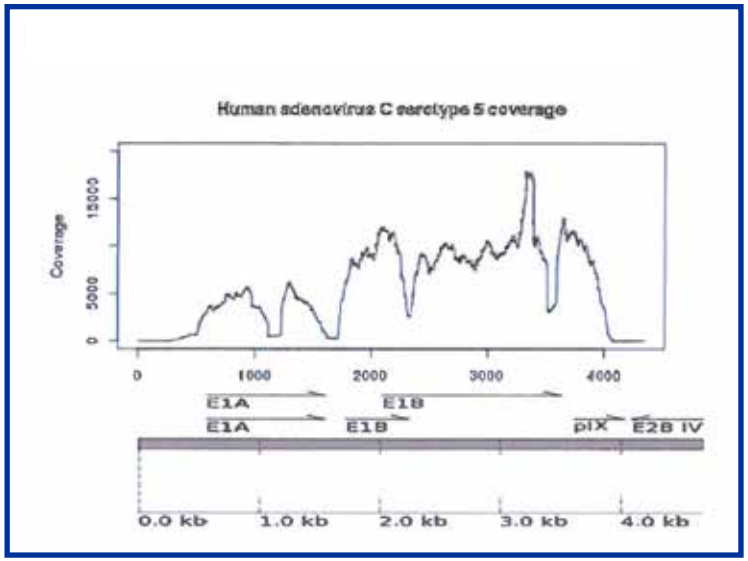
- Appr. $1x10^7$ cells sent to two different investigators
- Sequencing method: MPS using Illumina®
  - **LAB 1**
    - RNA
    - poly A purification step
  - **LAB 2**
    - RNA and DNA
    - rRNA depletion step

# Sequencing / Analysis

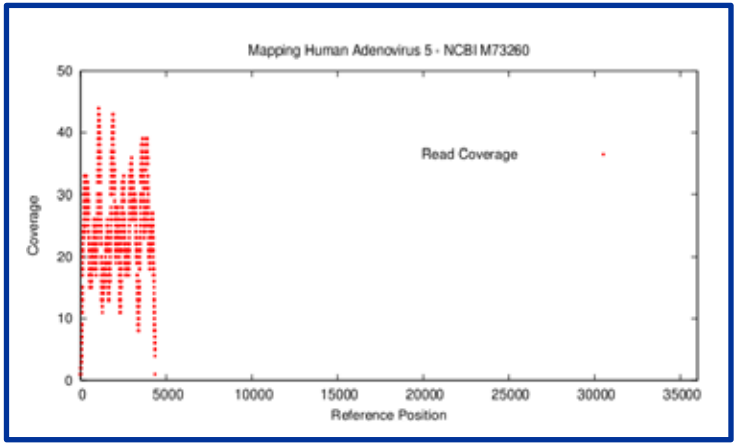
| | LAB 1 | LAB 2 |
|---|---|---|
| Reads (raw data) | RNA 476,000,000 | RNA 357,000,000 / DNA 287,000,000 |
| Reads after Q filtering | RNA 3,798,000 | RNA 5,027,000 / DNA 8,694,000 |
| Av length | RNA 104 bp | RNA 89 bp / DNA 130 bp |
| Contigs | RNA 61,910 | RNA 92,866 / DNA 7,358,588 |
| Alignment | - EMBL-EBI nt database (Sept 2014)<br>- Curated version of UniProt protein seq db UniRef100 (Sept 2014) | - NCBI Viral Database August 2014<br>- NCBI NT database April 2014 (hits only) |
| Decision making | **RNA** identity of at least 90%, match length of at least 100 nt. | **RNA** 90% seq identity, match length of at least 75 nt / **DNA** 90% seq identity, match length of at least 200 nt |

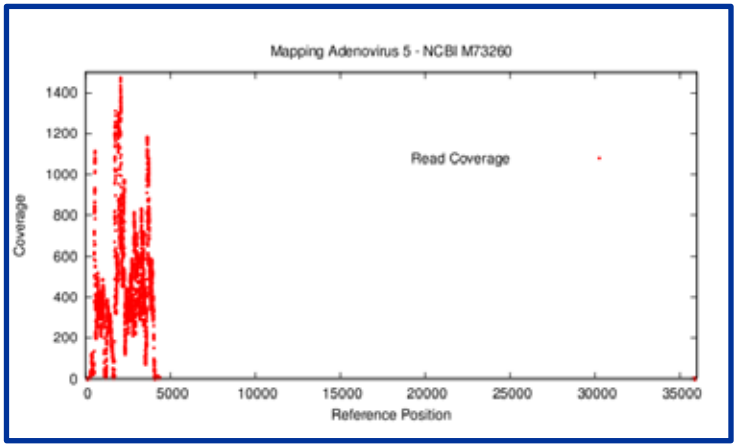
# Results: Ad5 Genome Sequences

- LAB 1




- LAB 2



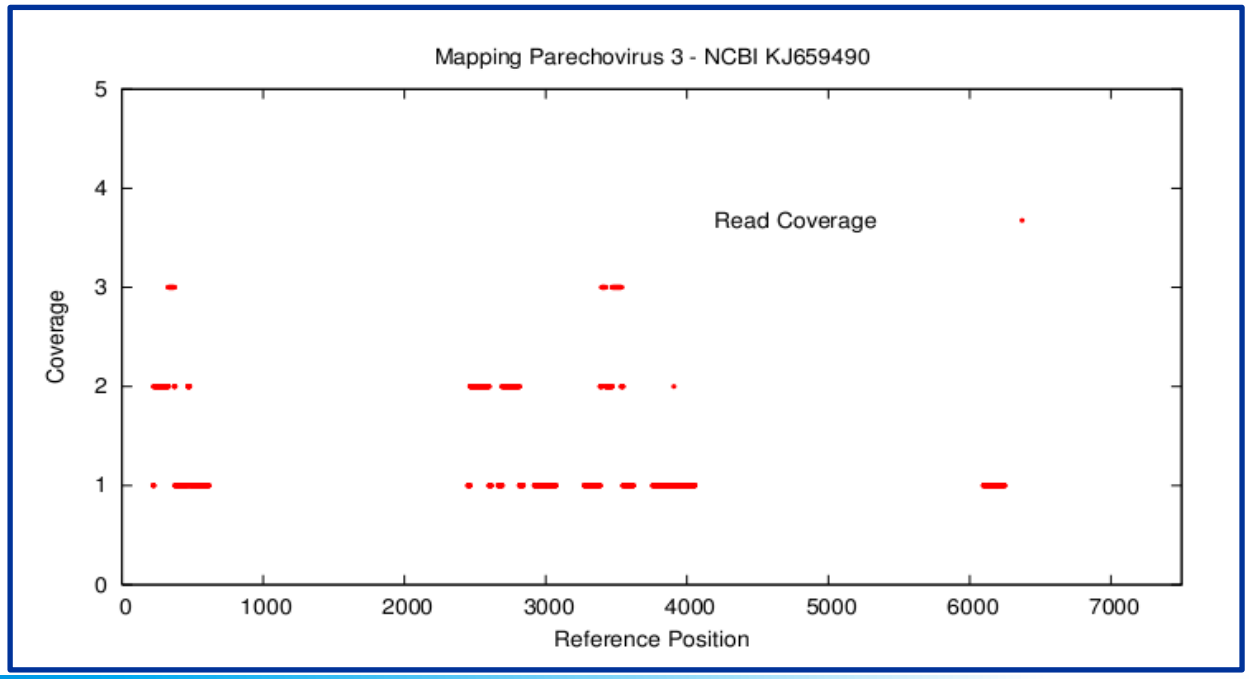
# Other Results I

- LAB 1
  No viral transcripts close to known vertebrate viruses were found (except from Ad5 genome known to be present in these cells). No long contigs with a taxonomy distant to known viruses, which would have suggested a viral infection were identified.

# Other Results II

- LAB 2

"Human parechovirus 3" could be detected <u>only</u> in the DNA sample. Due to the fact that this result is not plausible...further experiments may be applied...".



# Next steps: Confirmation

- In addition to cells, investigation of cell supernatant and medium

| Cells | Supernatant | Medium | Possible interpretation |
|---|---|---|---|
| + | - | - | Latent infection, evolution left-over |
| - | + | - | Irrelevant signal |
| - | - | + | Raw material contamination |
| + | + | - | Positive signal |

# Cell substrate characterization: Conclusion

- Adenovirus type C (Ad-5) genome sequences were detected, as expected (transformation of cells)
- Both levels RNA and DNA should be investigated
- Inclusion of culture supernatant as well as medium aids in result interpretation

- Unexpected signals require (thorough) follow-up analysis

Your Life. Our Inspiration.