Sample and Library Preparation

Siemon Ng

Notch Therapeutics

IABS NGS Training Workshop 03 Dec 2024

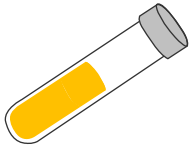
Agenda

- Overview of NGS workflow
- Considerations for upstream sample preparation
  - Nucleic Acid Extraction
  - Viral Enrichment
  - Library Preparation
- Controls for an NGS assay
- Summary and conclusion

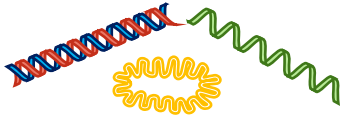
General NGS Workflow

- Each step has specific considerations when NGS is use for adventitious virus detection

Sample



Nucleic Acid Extraction

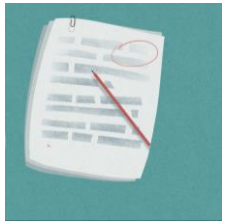


Sequencing Library preparation

Sequencing

Bioinformatics Analysis

Follow-up analysis



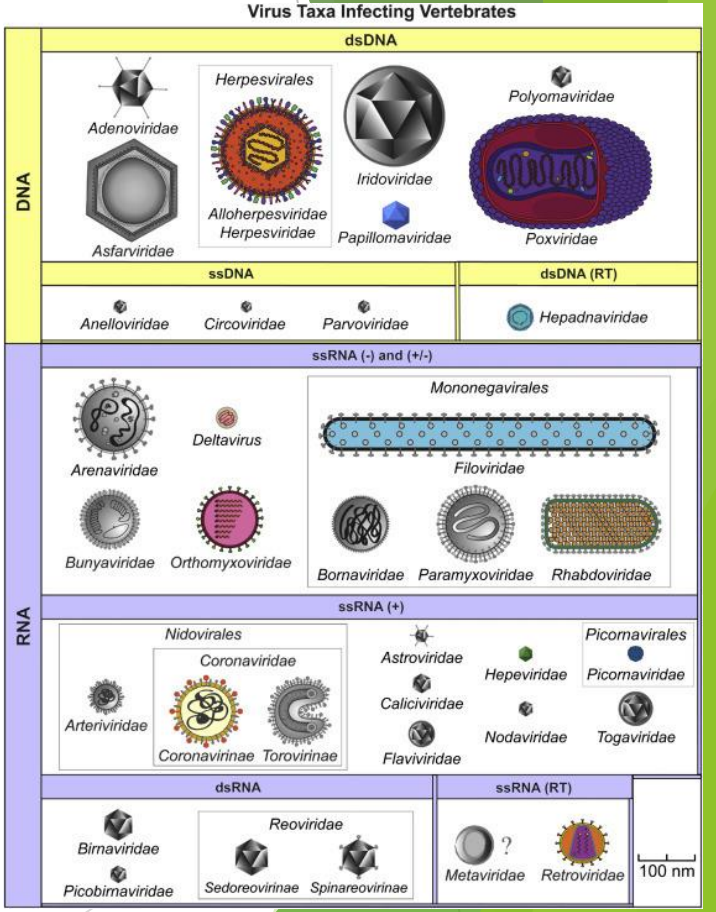
Sample Preparation

- Viruses are very diverse in their genomic and physical properties

Properties	Variability
Genome nucleic acid	ssDNA, dsDNA, ssRNA, dsRNA
Genome structure	Linear, circularized, segmented
Genome size	~3kb to 300kb +
Viral capsid	icosahedral or helical
Viral envelop	Enveloped or non-Enveloped

- Important to have efficient recovery of all nucleic acid types to minimize bias against any viral families.

- Generation of high-quality dsDNA or RNA as input material
  - Results in better quality and more consistent sequencing library and final data.



Sample Preparation

- Different NGS Approaches requires different considerations
  - Viromics – nucleic acids from encapsidated viral particles
  - Genomics – total nucleic acids
  - Transcriptomics – detection of viral transcripts

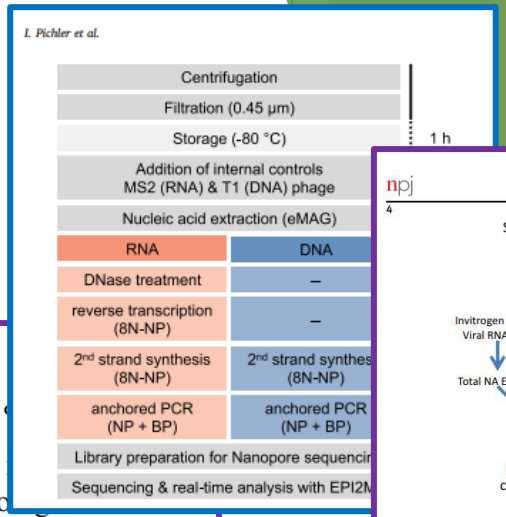
- Potential downstream impact / considerations
  - Extract DNA and RNA together or separately (e.g. library preparation or sequencing)
  - Length of the nucleic acids (e.g. long vs short reads)
  - Bioinformatic analysis (e.g. duplication of reads, accuracy of reads, abundance of host sequences)

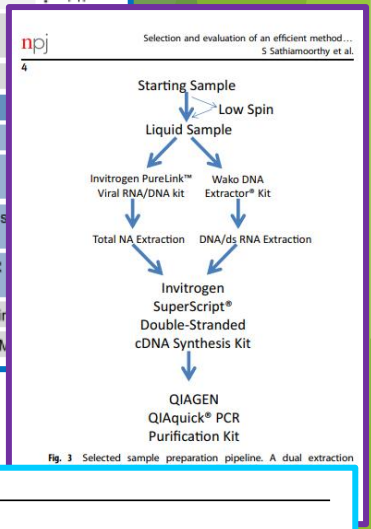
Disclaimer

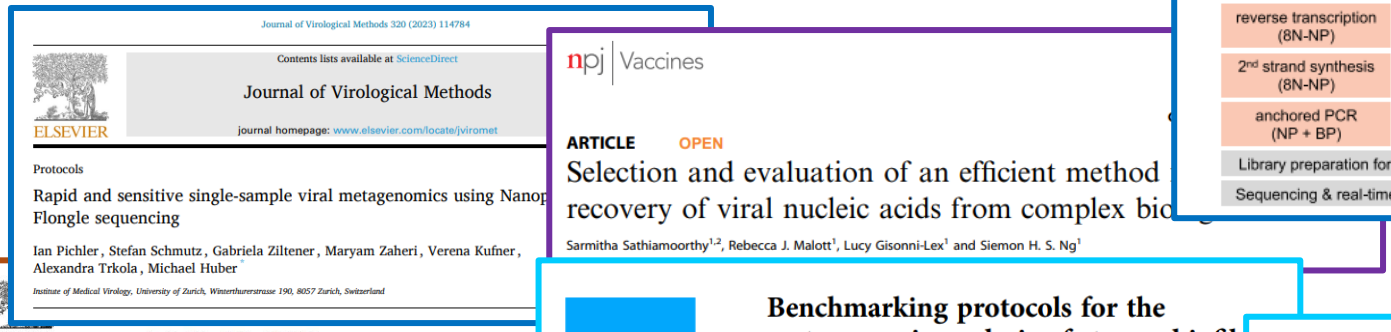
- The methods and kits in the following slides are not a recommendation or endorsement of any specific technique. They are only examples to showcase existing techniques for the different sample preparation considerations. In many cases, multiple commercial kits are available with similar performances.

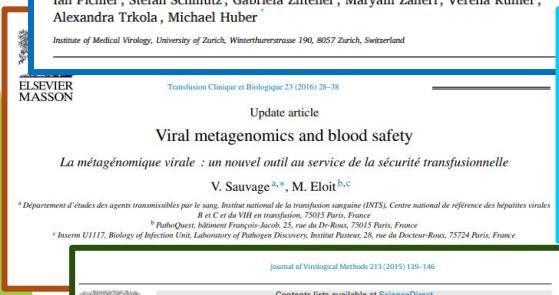
Nucleic Acid Extraction

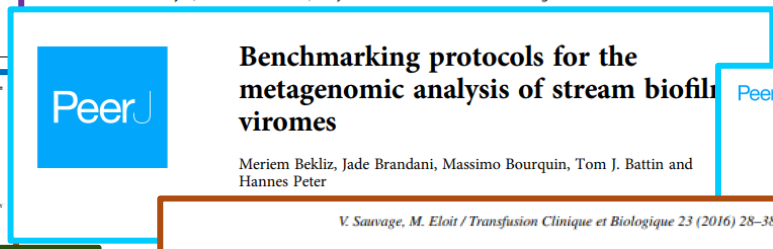
- Multiple methods for nucleic acid extraction

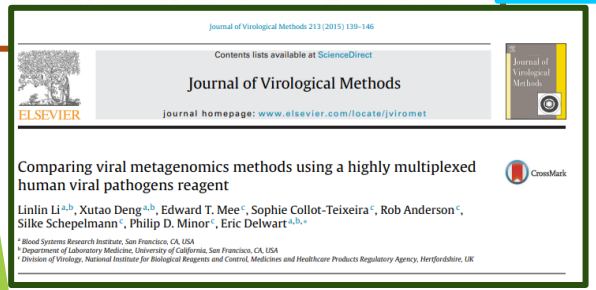


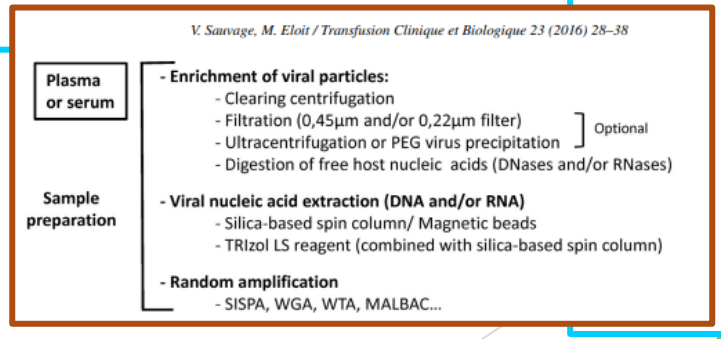


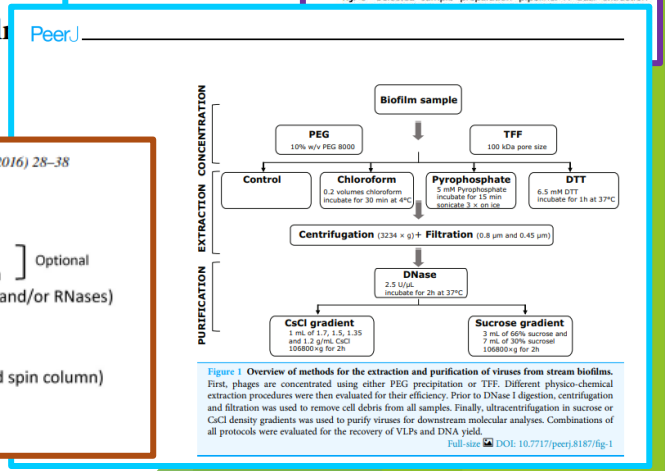












Nucleic Acid Extraction

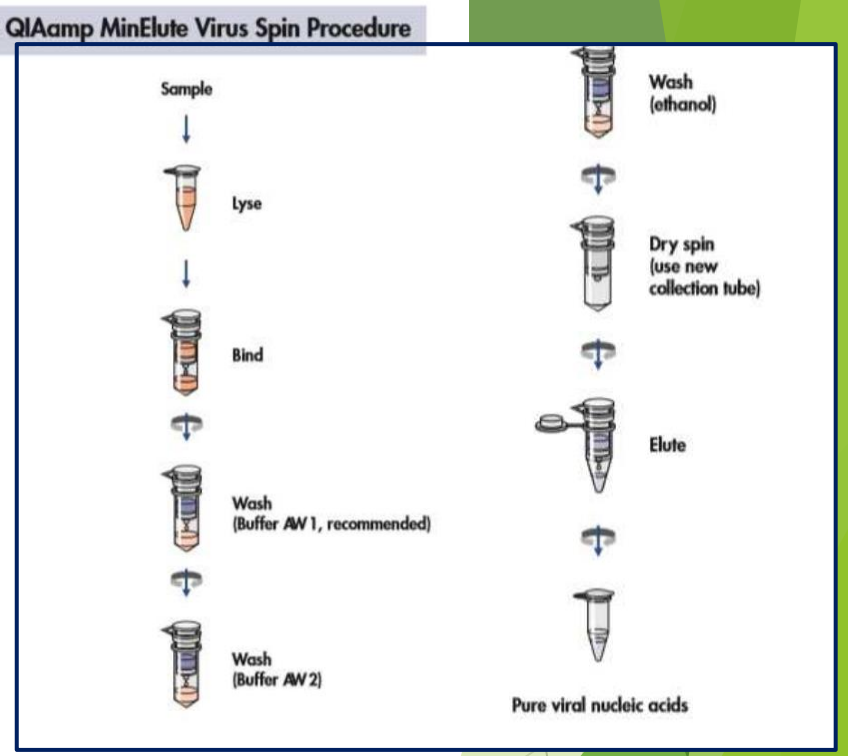
- Multiple methods for nucleic acid extraction
  - Silica columns
  - Bead-based
  - Precipitation-based

Silica Membrane

- Recovery of very high purity nucleic acids
- Commercially available for DNA, RNA or Total nucleic acids
  - Some extraction kits are further specialized for viruses

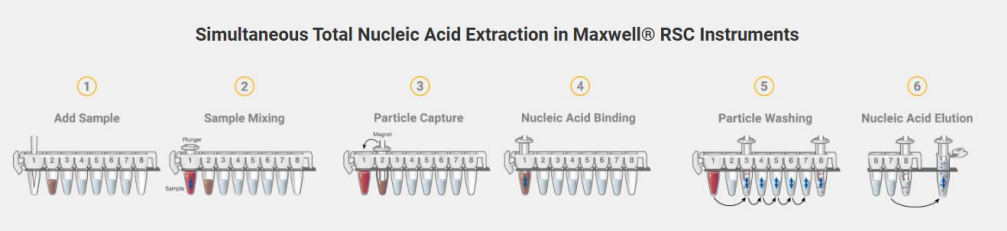
1. Start with Lysis buffer containing a chaotropic salt to open cell membrane and denature proteins
2. Lysate is applied to the silica membrane column which binds the nucleic acids
3. The column is washed with various buffers to remove cellular debris
4. Elution of nucleic acids under low salt condition

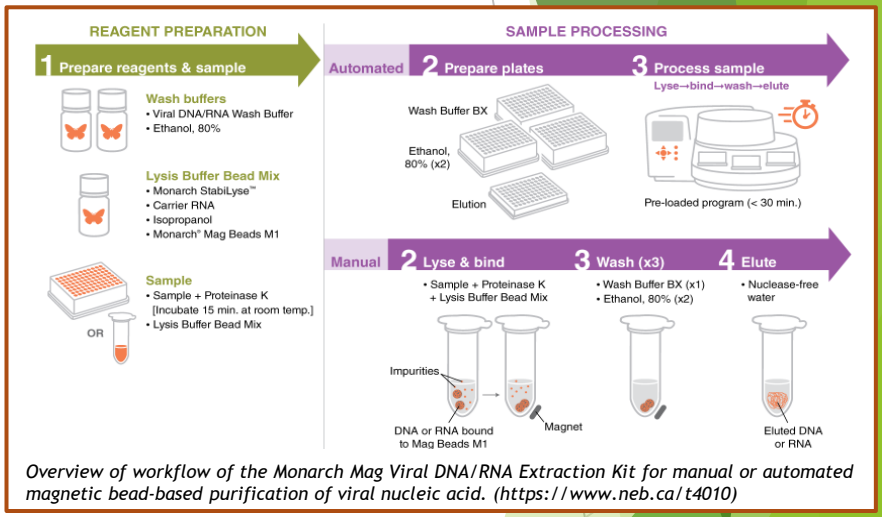
- Optional Steps:
  - Proteinase K to help digest away proteins
  - Carrier RNA to help with recovery of low amount of nucleic acids
  - RNase / DNase treatment while nucleic acid is bound to the column



Bead-based extraction

- Silica coated magnetic beads have similar workflow
- Washing of beads can be easier and more consistent with magnetic beads
- More adaptable to a liquid handler for high throughput extraction





Overview of workflow of the Monarch Mag Viral DNA/RNA Extraction Kit for manual or automated magnetic bead-based purification of viral nucleic acid. (https://www.neb.ca/t4010)

ThermoFisher KingFisher

Hamilton Liquid Handler

Precipitation-based extraction

- Phenol-Chloroform extraction
  - Uses guanidinium thiocyanate and phenol (e.g. Trizol) to lyse cells
  - Chloroform is added to separate the solution into three phases: aqueous phase for RNA, interphase for DNA and organic phase for protein
  - Isopropanol or Ethanol is added to the aqueous phase or interphase, respectively, for extracting RNA or DNA.

- Ethanol Precipitation for DNA
  - Decreases the solubility of the nucleic acid using salt and ethanol
    1. Lysis cells with a lysis buffer (e.g. SDS)
    2. Add salt to neutralize the charges on nucleic acids (e.g. 0.1X volume of Sodium Acetate)
    3. Add 100% ethanol (e.g. 2X volume)
    4. Incubate on ice for 10-15 minutes
    5. Centrifuge at high speed to principate the DNA
    6. Wash the pellet with 70% Ethanol
    7. Air-dry the pellet and resuspend the DNA in water or buffer

Enrichment of viral signal

- Generally an important part of an NGS viral adventitious agent detection pipeline.
- Without enrichment of viral signals, host sequences can make up the majority of the sequence data.

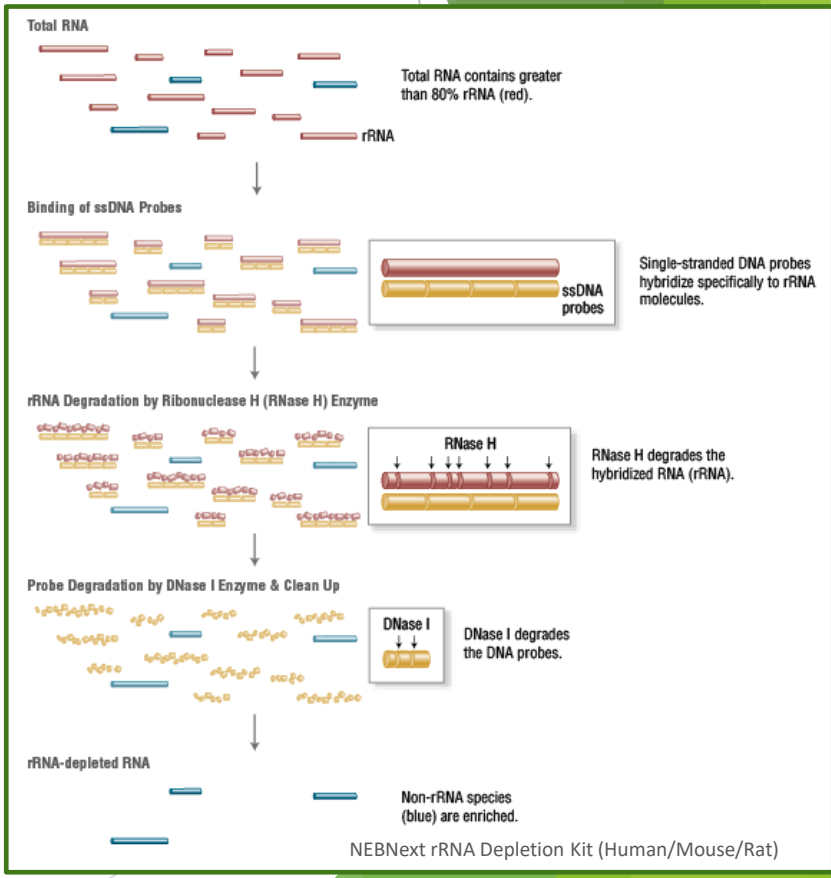
- Examples of potential enrichment techniques:
  - Nuclease treatment
  - Ribosomal depletion
  - Centrifugation
  - Filtration

Nuclease Treatment

- Removal of ‘free’ nucleic acids from host cells or viral particles
  - DNase and RNase (e.g. DNase I, RNase A)
  - Universal nuclease (e.g. Benzonase, Dr Nuclease)
- Nuclease can also be used to remove DNA or RNA if they are being extracted independently
- Potential optimization consideration: buffer, concentration, time of incubation, inactivation
- Assess the recovery of viruses as some might be sensitive to nuclease treatment
- Ensure inactivation or remove of nuclease to prevent downstream impact of extract nucleic acid and library generation.
  - Important to ensure no degradation of extracted nucleic acids
  - Chemical used for inactivation should be removed to minimize any inhibition of downstream steps

Host Ribosomal RNA Removal

- Ribosomal RNA can make up a large proportion of the extraction nucleic acid (RNA or Total nucleic acids)
- Generally a probe-based technique where the complementary oligos binds to host ribosomal RNA and is removed by magnetic beads or enzymatic digestion
- Numerous commercial kits are available for different species



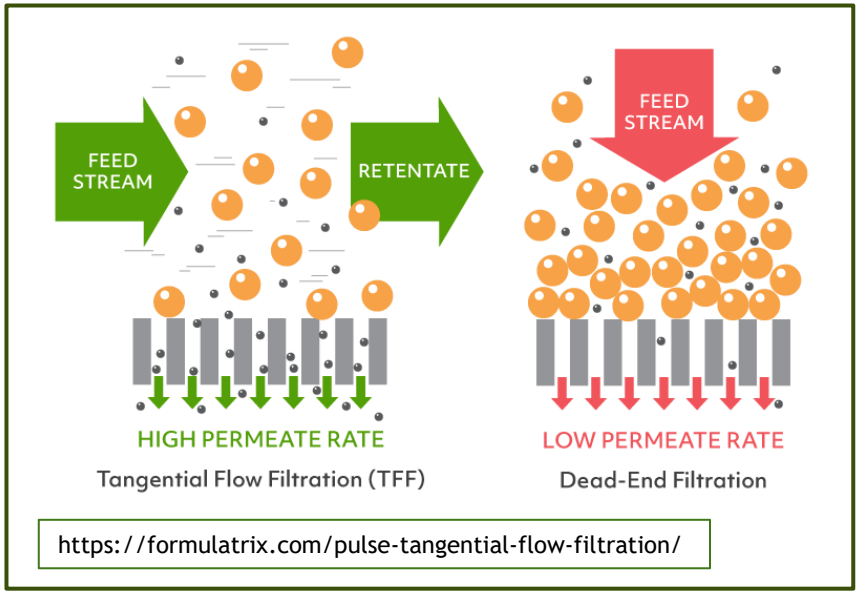
NEBNext rRNA Depletion Kit (Human/Mouse/Rat)

Centrifugation

- Low speed centrifugation (e.g. <700 x g) can be used to remove cellular debris
- Ultracentrifugation can further concentrate viral particles.
- Centrifugation speed and time should be assessed to ensure unbiased recovery of different virus and the different viral properties (e.g. size, density)

Filtration

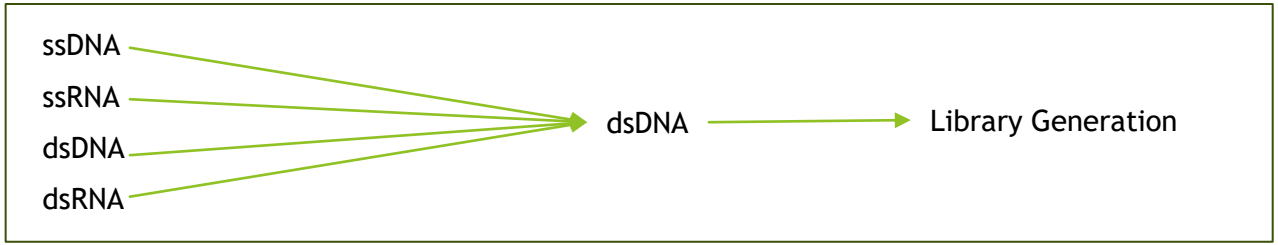
- Filtration (membrane or TFF) can also be done to remove cellular debris with care to again not to be bias to the recovery of different types of viruses.



Additional Sample Preparation

Conversion to dsDNA

- dsDNA are generally the input to sequencing library generation.



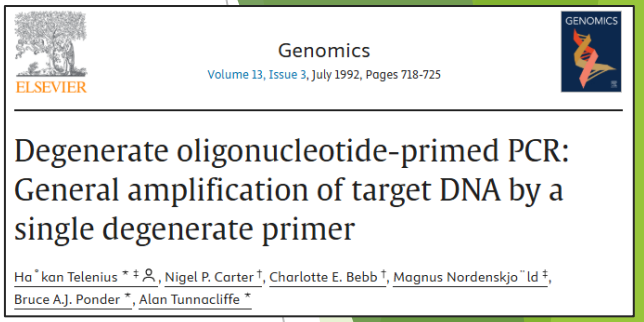
  - Many kits available to cover RNA to ssDNA to dsDNA

- dsRNA genomes are more stable and more challenging to denature. Some group has shown successes with higher temperature during denaturation (e.g. Reoviridae is a family of dsRNA viruses)

- NOTE: Use of Oligo dT to convert mRNA to DNA is not recommended as not all viral transcript is poly-adenylated (e.g. some flaviviruses). Using Oligo dT could potentially miss the detection of non-Poly(A) viral transcripts

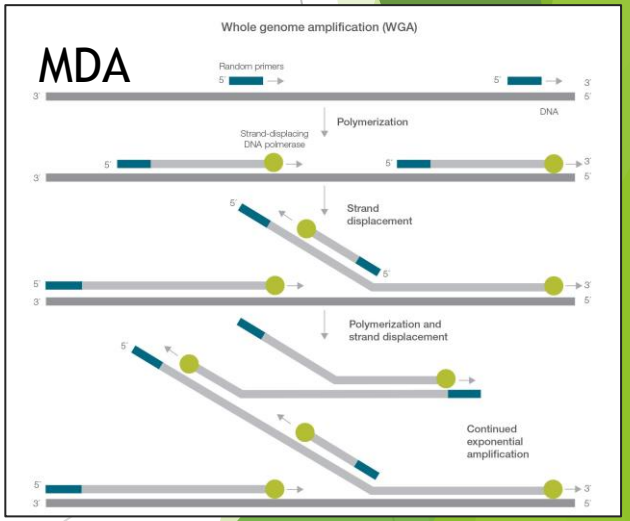
Whole Genome Amplification

- Increase the amount of nucleic acids for input into library generation and can be useful when the recovered nucleic acid is very low
- PCR based amplification
  - For example: DOP-PCR (Degenerate Oligonucleotide-Primed PCR)
    - Uses degenerate primers to anneal to the DNA and amplify the nucleic acid by PCR.
    - Can potentially be biased against regions where there is no binding site
- Isothermal amplification
  - For Example: Multiple Displacement Amplification (MDA)/ Rolling Circle Amplification
    - Uses Phi29 DNA polymerase to amplify DNA under isothermal condition
    - DNA is primed using random oligos with continuous amplification without denaturation
- There are also other potential techniques
- These methods can increase the amount of input DNA but can also results in potential bias on the input nucleic acids (E.g. bias of primer sites, shorter fragments is less efficient using MDA)


ELSEVIER
Genomics
Volume 13, Issue 3, July 1992, Pages 718-725
GENOMICS

Degenerate oligonucleotide-primed PCR: General amplification of target DNA by a single degenerate primer

Ha˚kan Telenius * ‡, Nigel P. Carter †, Charlotte E. Bebb †, Magnus Nordenskjo¨ld ‡, Bruce A.J. Ponder *, Alan Tunnacliffe *



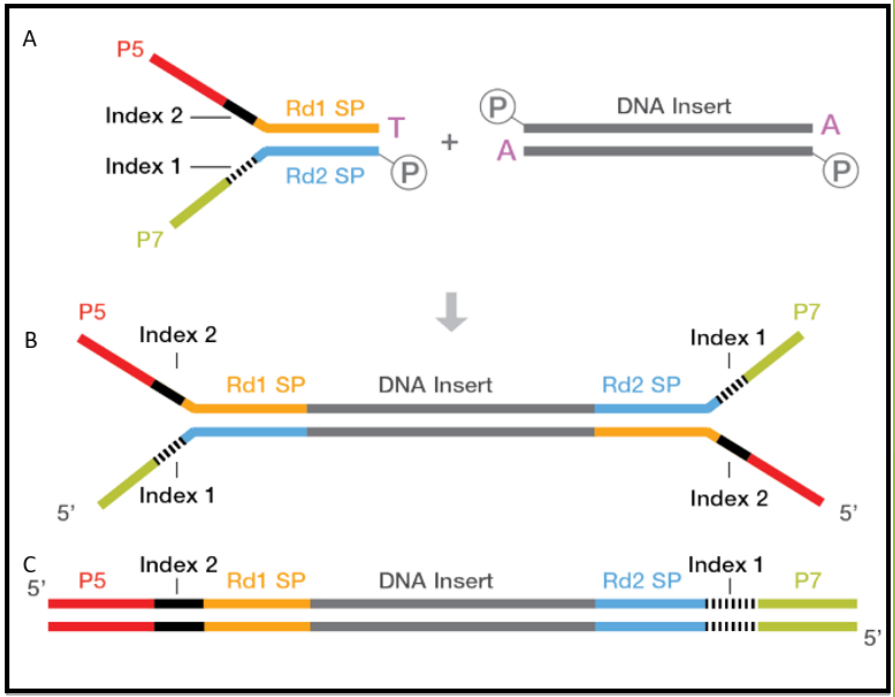
ThermoFisher: https://www.thermofisher.com/ie/en/home/life-science/pcr/isothermal-nucleic-acid-amplification/multiple-displacement-rolling-circle-amplification.html

Library Preparation

- All sequencing platforms generally require a NGS library construction step
  - Shearing of templates
  - Addition of platform specific adaptors for amplification or sequencing

Illumina

- ng to ug of input DNA
- Dual indexing of Illumina library
  1. Fragmentation
  2. End-repair
  3. Adaptor ligation
  4. Size selection
  5. Limited cycle PCR amplification
- Tagmentation, where fragmentation of the DNA and the ligation of the adaptors can be done in a single step, can potentially streamline some of the steps above
  - Less labor intensive
  - Potentially more consistent once established

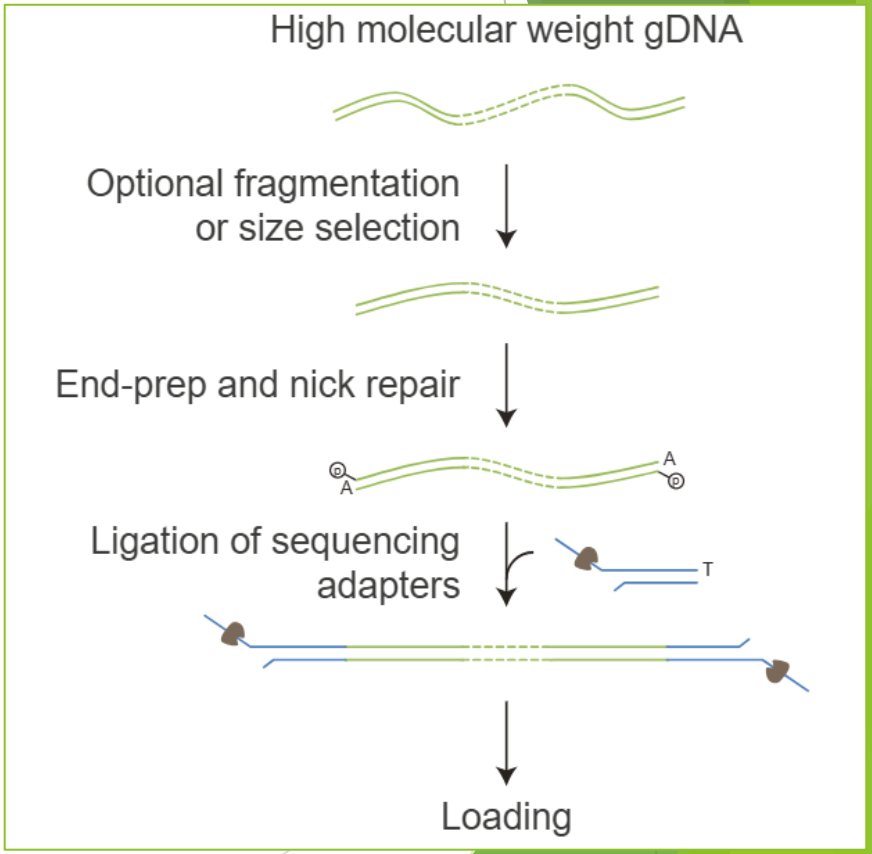


https://knowledge.illumina.com/library-preparation/general/library-preparation-general-reference_material-list/000003874

Oxford Nanopore

- ~100ng of input DNA

1. Tagmentation your DNA
2. Ligation of sequencing adapters
3. Clean up the sample

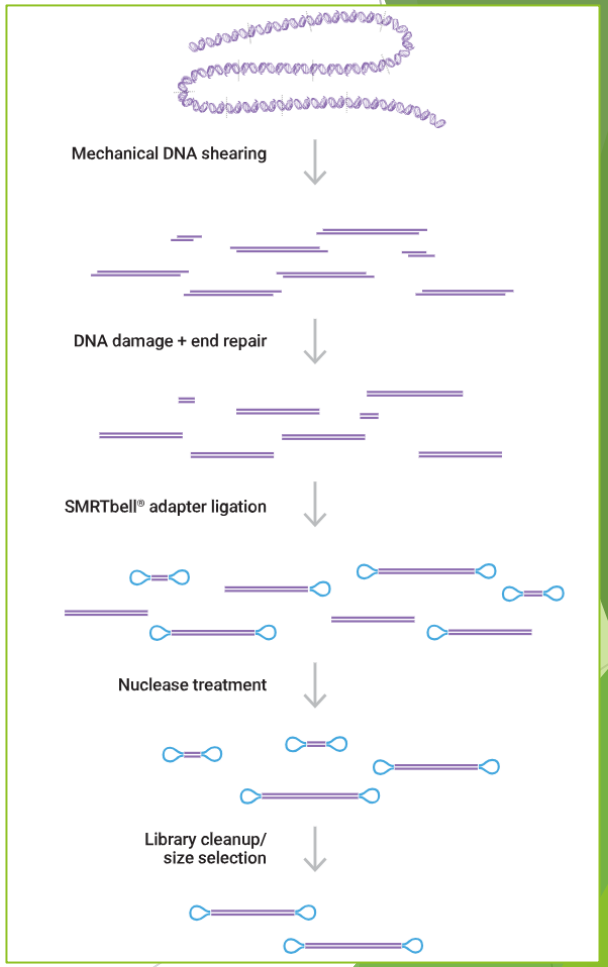


https://nanoporetech.com/document/chemistry-technical-document

PacBio

- 1 – 2 ug of DNA (depending on the platform)

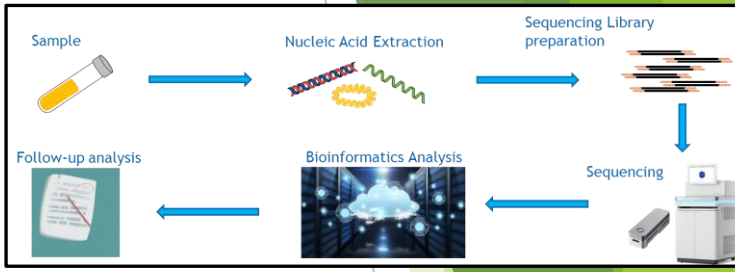
1. Removal of reads less than 10kb (ensure long read length in the sequencing data)
2. DNA shearing
3. End repairs and A-tailing
4. Adapter ligations
5. Removal of unligated adaptor (nuclease treatment)
6. Final cleanup and size selection



https://www.pacb.com/products-and-services/applications/whole-genome-sequencing/

Overall combination of approaches

- Different combination of kits and approaches will influence the recovery efficiency for viral genome
- Important to understand how your specific matrix can be optimized



Biologicals 85 (2024) 101741

Contents lists available at ScienceDirect

Biologicals

journal homepage: www.elsevier.com/locate/biologicals

Efficient extraction of adventitious virus nucleic acid using commercially available methods

William G. Valiant [a], Jon Borman [b], Kang Cai [b], Peter M. Vallone [a,*]

[a] Biomolecular Measurement Division, Applied Genetics Group, National Institute of Standards and Technology, Gaithersburg, MD, USA
[b] Biopharmaceutical Development, Biopharmaceuticals R&D, AstraZeneca, Gaithersburg, MD, USA

npj | Vaccines

www.nature.com/npjvaccines

Corrected: Author correction

ARTICLE OPEN

Selection and evaluation of an efficient method for the recovery of viral nucleic acids from complex biologicals

Sarmitha Sathiamoorthy[1,2], Rebecca J. Malott[1], Lucy Gisonni-Lex[1] and Siemon H. S. Ng[1]

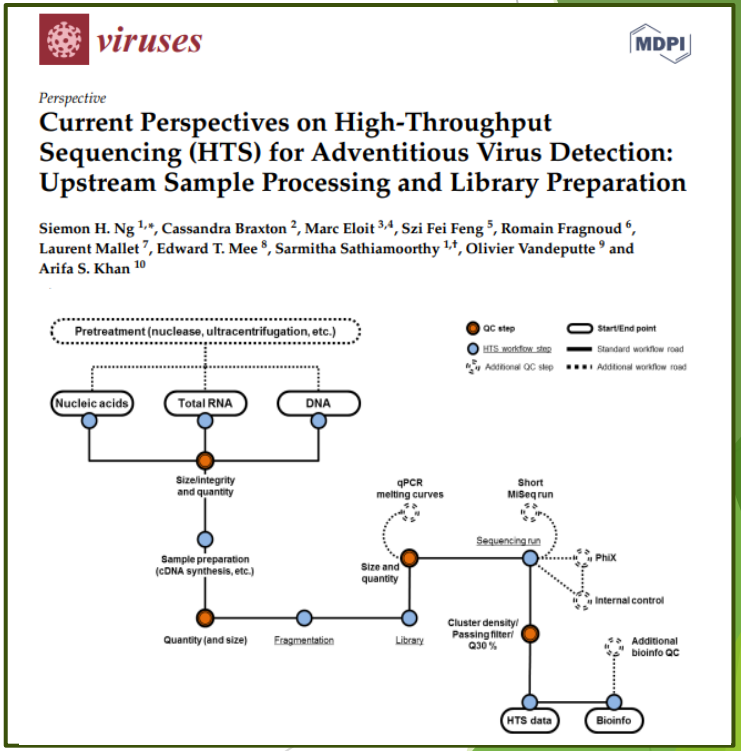
Comparing Viral Metagenomic Extraction Methods

Jeanette Klenner[1,2], Claudia Kohl[1], Piotr W. Dabrowski[1,3] and Andreas Nitsche[1*]

Curr. Issues Mol. Biol. Vol. 24

Quality Control of Sample Preparation

- Extracted nucleic acids
  - Quantity
  - Quality (Size and purity)
- Library preparation
  - Quantity
  - Verification of final library size
  - Ligation of adaptors



viruses

MDPI

Perspective

Current Perspectives on High-Throughput Sequencing (HTS) for Adventitious Virus Detection: Upstream Sample Processing and Library Preparation

Siemon H. Ng [1,*], Cassandra Braxton [2], Marc Eloit [3,4], Szi Fei Feng [5], Romain Fragnoud [6], Laurent Mallet [7], Edward T. Mee [8], Sarmitha Sathiamoorthy [1,†], Olivier Vandeputte [9] and Arifa S. Khan [10]

Controls used in a NGS assay for virus detection

Controls

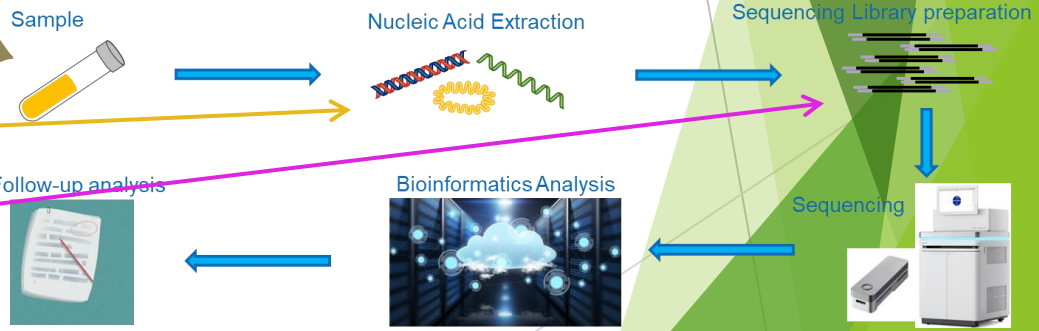
- Provide important check-points in ensuring performance of the assay
- Assess the integrity of the starting material and subsequent steps
- Verify the integrity, size, quantity of sequencing library
- System in place to detect cross-contamination or minimize potential contaminations from the environment
- Evaluating sensitivity and useful for assay qualification / validation

Environment / Area Controls

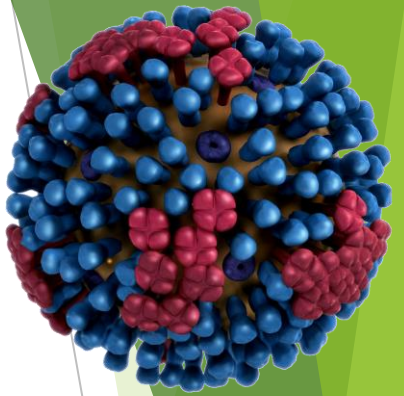
- Separation of activities can be similar to PCR / qPCR assays and rigorous control is needed to minimize any contamination
  - Separate sample extraction, pre/post amplifications
  - Differential airflow to minimize contamination
  - All work are done in dedicated BSC if possible
  - Limit the use of shared reagents as much as possible (i.e., single use aliquots)
  - Each sample should be dual-indexed or dual barcoded so that each sample can be easily distinguished from other samples
  - Multiplexing might result in some barcode hopping between samples due to excess bar-code in the reaction.

Different types of control material for monitoring the steps in NGS workflow

- Live viruses
- Virus-like-particle
- Nucleic acids
  - Premade libraries



Live Virus



- Directly assess different viruses and to detect differences in nucleic acid extraction.
- Select viruses that are not expected to be in any samples or the environment
- Working with live, potentially infectious viruses
- Live viruses should be characterized for copy number, viral titre, viral sequence, host background sequence
  - Copy number and viral titre should match to most viral particles contains a genome

WHO International Reference Panel

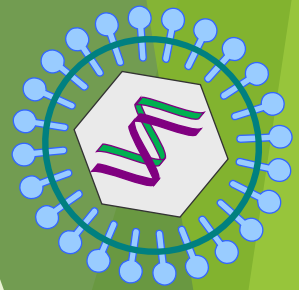
- Seven viruses established as **"First World Health Organization International Reference Panel for Adventitious Virus Detection in Biological Products by High-throughput Sequencing"**
  - Established based on a CBER collaborative study with participants from the AVDTWG
  - Characterization: Infectious titer, particle count, Genome copy number (ddPCR) and NGS.
  - Made specifically for qualification and validation studies (expects to be used as a panel of 7 viruses)

- BEI Resource (Catalog #NR-59630)
  - https://www.beiresources.org/Catalog/animalviruses/NR-59630.aspx
  - 1000 vials of the new virus stocks; limit to 1 order per year.

Virus	Genome type	Genome size	Particle size	Envelope	Chemical resistance
Reovirus (Lang)*	RNA, double-strand; Linear (segmented)	23.6kb(1,196–3,915nt	80nm	No	Medium-high
FeLV (KT)*	RNA; single-strand; Linear (dimeric)	8.5kb	80-100nm	Yes	Low
RSV (A2)*	RNA; single-strand; Linear	15kb	150-200nm	Yes	Low-medium
PCV-1*	DNA, single-strand; circular	1.8kb	16-18nm	No	High
EBV (B95-8)*	DNA, double-strand; Linear	172kb	122-180nm	Yes	Low-medium
Minute Virus of Mice (MVM)	DNA, single-strand; Linear	5.1kb	26 nm	No	HIgh
Human coronavirus (HCoV-OC43)	RNA; single-strand; Linear	30.7kb	80-120nm	Yes	Low

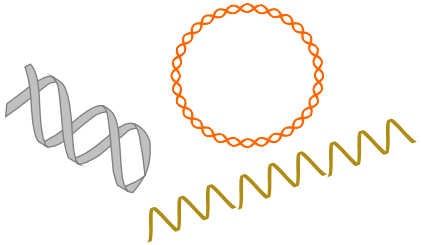
* Additionally 5 of the viruses are available as CBER NGS Virus Reagents (BEI NR-59622) for NGS development

Virus-Like-Particle



- VLPs are particles can that mimic structure of some viruses with the opportunity to replace the viral genome with a custom DNA or RNA sequence.
- Can mimic the physical properties of a whole virus particles during nucleic acid extraction
- VLP are non-infectious and generally suitable for work in Biosafety Level 1
- The length of the packaged nucleic acid can be limited to a few kilobases
- Potential commercial sources are available

Nucleic Acid



- Monitoring steps after nucleic acid extraction
  - Generally simple to generate and nucleic acid sequence can be specifically designed as needed.
    - dsDNA, ssDNA, ssRNA and dsRNA are all possible.
    - Length of the nucleic acids can be a constrain
  - NIST'S External RNA Controls Consortium (ERCC); available via different vendors
  - ATCC and NIBSC also have nucleic acid controls
- Pre-made libraries
  - Suitable for monitoring that the sequencing reaction / equipment work as expected.
  - For Illumina technology, there is a PhiX control. It could also be generated in-house.
  - Important that the sequence of the pre-made library to not resemble potential viral contaminations.

Negative Controls

- Allow to identify viral signals that might be from the assay reagents (i.e. extraction kits) and testing environment.
- Consider if and which type of negative control is most appropriate.
  - Negative control such as water or media might yield insufficient amount of nucleic acids for library generation.
- Possible to retain a control sample and not test unless there is a putative positive hit.

Summary

- Highlighted specific considerations for the upstream steps of sample preparation such as nucleic acid extraction, enrichment and library preparation

- Controls are important to monitor performance for each step during development and to ensure of a NGS assay

Thank You