Next-Generation Sequencing
Sample Preparation

Benjamin Pinsky, MD, PhD
Assistant Professor of Pathology and Medicine
Medical Director, Clinical Virology Laboratory
Stanford Children’s Health and Stanford Health Care
Disclosures

No relevant disclosures.
Outline

• Next-Generation Sequencing (NGS) definitions
• Nucleic Acid Extraction and Reagents
• Library Preparation Strategies
Definitions

• Next-Generation Sequencing (NGS):

  – Refers to any sequencing method developed to expand upon the capabilities of Sanger sequencing

  – Also referred to as massively parallel, deep, and ultradeep sequencing, etc...

  – The proliferation of sequencing methods has resulted in additional terminology

    • Second-generation – 454, Illumina, Ion Torrent
    • Third-generation – PacBio, Oxford Nanopore
    • Fourth-generation - ?
Sanger Sequencing

- Dideoxy-termination sequencing
- Signal obtained from the population of DNA molecules being sequenced
- Can detect variants at ~10-20% level
- Read lengths up to ~1kb
- Low throughput
- Low cost per run, high cost per base

Schadt et al., Hum Mol Genet 2010
NGS v. Sanger

- NGS - Variety of detection methods
- Signal obtained from DNA molecules spatially separated on a solid support
- Can detect variants at ~1% level or lower

Variable read lengths depending on the NGS method: very short (HiSeq 2x125bp) to very long (PacBio >10kb)
- Moderate to High throughput with multiplexing
- High cost per run/ low cost per base

Schadt et al., Hum Mol Genet 2010  Sahoo and Lefterova et al., J Clin Microbiol 2013
Metagenomic v. Targeted Sequencing

**Metagenomic Sequencing**

- Sequence all of the DNA and/or RNA in a particular specimen
- Includes microbial and human sequence
- Referred to as “Target Agnostic” or “Unbiased” Sequencing

**Targeted Sequencing**

- Specific genomic regions or organisms are interrogated
- Amplification – Amplicon Sequencing
- Hybridization – Target Capture
Remember the Pre-Analytical Steps!

Garbage In = Garbage Out

Photo courtesy of Ellen Jo Baron, PhD
Selecting the Appropriate Specimen Type

- Specimen selection for virologic/microbiologic testing is generally organized by body system.


- Principles of sample selection are the same for NGS.

- Interpretation of Metagenomic sequencing may be particularly complicated by resident microbes in clinical samples.
The Human Virome and Microbiome

Popgeorgiev et al., Intervirology 2013
Nucleic Acid Extraction

- Standard Nucleic Acid extraction methods are suitable for NGS applications.
- One should consider whether total nucleic acid extraction is adequate or whether DNA or RNA-specific extractions are needed.
- Must consider that nucleic extraction kits may contain microbial nucleic acids.
Extraction Kits are Contaminated with Microbes

Salter et al. BMC Biology 2014
# Viral Contamination: Parvovirus-like Hybrid Virus (PHV)

The perils of pathogen discovery: Origin of a novel parvovirus-like hybrid genome traced to nucleic acid extraction spin columns

Samia N. Naccache, Alexander L. Greninger, Deanna Lee, Lark L. Coffey, Tung Phan, Annie Rein-Weston, Andrew Aronsohn, John Hackett, Jr., Eric L. Delwart, Charles Y. Chiu

<table>
<thead>
<tr>
<th>Kit</th>
<th>Spin column</th>
<th>PCR result for: Replicase, nt763-1010 (248 nt)</th>
<th>Bridge, nt1554-2044 (491 nt)</th>
<th>Capsid, nt1922-2044 (121 nt)</th>
<th>Capsid + NCR, nt3288-3448 (161 nt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNeasy MinElute cleanup kit</td>
<td>RNeasy MinElute column</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>RNeasy minikit</td>
<td>RNeasy minicolumn</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>QIAamp UltraSens virus kit</td>
<td>QIAamp minicolumn</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>QIAamp viral RNA minikit</td>
<td>QIAamp minicolumn</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>QIAamp DSP virus kit</td>
<td>QIAamp MinElute column</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>PureLink viral RNA/DNA minikit</td>
<td>PureLink viral column</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TRIzol LS kit</td>
<td>NA</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>F71 viral minikit v2.0</td>
<td>NA</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Water, nuclease-free (Qiagen, Fisher Scientific, and Epicentre)</td>
<td>NA</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

* NCR, noncoding region; C, column elution; F, full extraction; nt, nucleotide; NA, not applicable.
Viral Contamination: Xenotropic Murine Retrovirus (XMRV)

Identification of a Novel Gammaretrovirus in Prostate Tumors of Patients Homozygous for R462Q RNASEL Variant


Detection of an Infectious Retrovirus, XMRV, in Blood Cells of Patients with Chronic Fatigue Syndrome


• Associations with Prostate Cancer and CFS and were disproven.
• Several high profile manuscripts were retracted.
• XMRV was determined to be a contaminant from PCR components, laboratory material, and cell culture.

An Endogenous Murine Leukemia Viral Genome Contaminant in a Commercial RT-PCR Kit is Amplified Using Standard Primers for XMRV

Eiji Sato, Rika A Furuta, Takayuki Miyazawa

Sato et al. Retrovirology 2010, 7:110
http://www.retrovirology.com/content/7/1/110

SHORT REPORT

Open Access
Library Preparation

Metagenome → Fragmentation Library

Long-Range PCR Product → fragmentation → end-repair adapter/barcode ligation size selection

Template DNA → Library preparation

target-specific primers adapters, barcodes → PCR enrichment

Targeted Sequencing or Metagenomic Sequencing

Targeted amplicon Library

Targeted Sequencing
Barcodes/Indices Allow Significant Multiplexing
Illumina

- **Bridge PCR**: Physical separation and clonal amplification of individual molecules
- **Cyclic reversible termination**
- **Read length (MiSeq)**: up to 2x300bp
- **Throughput**: ~15 Gb
- **Run time**: ~29-56 hours

**Illumina’s library-preparation work flow**

- DNA fragments
- Blunting by fill-in and exonuclease
- Phosphorylation
- Addition of A-overhang
- Ligation to adapters

**Cyclic reversible termination**

- **First cycle denaturation**
- **First cycle annealing**
- **First cycle extension**
- **Second cycle denaturation**
- **Second cycle annealing**
- **Second cycle extension**

**Read length (MiSeq)**

- **Up to 2x300bp**

**Throughput**

- ~15 Gb

**Run time**

- ~29-56 hours

---

**Anderson and Schrijver. Genes 2010**

**Mardis ER. Annu Rev Annal Chem 2013**
Illumina Library

Fragmented DNA input

Clean Up/Size Selection

PCR Enrichment

End Repair and 5’ Phosphorylation

Cleaning

dA-Tailing

Adaptor Ligation with optional NEBNext Adaptor

U Excision

Clean Up
Life Technologies – Ion Sequencing

- **Emulsion PCR**: Physical separation and clonal amplification of individual molecules
- **Emitted hydrogen ions**: proportional to dNTPs added to primer
- **Read length**: 200-400 bp
- **Throughput**: 100 Mb - 2 Gb
- **Run time**: 2 – 7 hours
Ion Library

AMPure Beads

Agilent BioAnalyzer

Ion Chef
Add adaptors to fragmented DNA: generating closed circular DNA molecules

Create DNA-polymerase complex in zero mode waveguide (ZMW)

Single molecule real-time sequencing

Read length: 4,200-8,500 bp

Throughput: 275-375 Mb

Run time: 30-180 minutes
SMRT Library - PacBio
Oxford Library

DNA sample

constriction

+ high molecular weight gDNA

fragmentation

end-repair and dA tailing

adapter ligation

tether attachment

30 mins (optional)

20 mins

12 mins

12 mins

25 mins
Oxford Nanopore

- Add adaptors/enzymes to DNA
- Measures changes in current as DNA passes through protein nanopore
- Read length: ~6,500-8,000 bp
- Throughput: 100 Mb - 200 Mb
- Run time: 48 hours

Ip et al., F1000Research 2015
Oxford Rapid Library

VolTrax Automated Sample Prep
Metagenomic NGS
Pre-Analytical Enrichment Strategies

Specialized strategies are needed to enrich viral RNA or DNA from the predominantly human nucleic acids.

Laboratory Strategies

- Physical separation: ultracentrifugation, density gradient centrifugation
- Nuclease pretreatment: remove host nucleic acids, while preserving capsid-protected viral particles
- Oligonucleotides for host nucleic acid depletion.
- Oligonucleotides for pathogen enrichment
DASH: Depletion of Abundant Sequences by Hybridization

- Utilizes the Cas9 (CRISPR associated) nuclease
- Single Guide RNAs (sgRNAs) used to program dsDNA cleavage
- sgRNAs designed against ribosomal RNA cDNAs
**ViroCap**: solution-based capture designed to enrich nucleic acid from DNA and RNA viruses from 34 families that infect vertebrate hosts.

**VirCapSeq-VERT**: solution-based capture designed to enrich nucleic acid from the 207 viral species known to infect vertebrates.
Library Preparation

Metagenome → Long-Range PCR Product → Fragmentation Library

- fragmentation
- end-repair
- adapter/barcode ligation size selection

Template DNA

Library preparation

target-specific primers

Adaptors, barcodes → PCR enrichment

Targeted amplicon Library

Targeted Sequencing or Metagenomic Sequencing

Targeted Sequencing
PCR Error – Variant Calling

• Basis for Error
  – Nested PCR is frequently required (CMV, HIV)
  – High fidelity enzymes have the lowest processivity

• Types of Error
  – Overestimation of variants present before PCR
  – Biased distribution of amplified variants
  – Substitution errors
  – Recombination
Evaluating for Error: Sequencing Clones

![Graph showing observed error vs. read depth and genome position for Sabin 1, Sabin 2, and Sabin 3.](attachment:image.png)
Primer ID: Create a Random Tag for Each Template

- Eliminates resampling
- Makes it possible to identify and ignore most PCR and sequencing errors


Jabara, CB et al. PNAS 2011
The Future Is Now!
Automation of Library Preparation
Thank You!