RNA-Seq library preparation

Brian J. Knaus
USDA Forest Service
Pacific Northwest Research Station
Box 1 | Evolution of the central dogma

The central dogma outlines the flow of information that is stored in a gene, transcribed into RNA and finally translated into protein. The ultimate expression of this information is the phenotype of the organism. Each step of the central dogma is accompanied by recent technological innovations that allow genome-wide analysis. Although the central dogma once presented a view that was essentially descriptive, and limited to gene-by-gene studies, it can now be coupled with technology and viewed as experimental and testable. Hypotheses can be formulated and revised for the purpose of elucidating the detailed connections between genotype and phenotype, therefore unravelling the inner molecular biology of an organism.

Why is library preparation important?

• Yield
  Illumina claims 1-10 μg total RNA

Practicality

Cost effectiveness

• Bias
  Inference of results

• Strandedness
### Supplementary Table 4: Comparison of technical details of library construction methods.

<table>
<thead>
<tr>
<th>Library</th>
<th>Time Required (days)</th>
<th>Total number of steps</th>
<th>Approx. Reagent cost ($)</th>
<th>Starting material used (RNA, ng)</th>
<th>Applicability to small RNA</th>
<th>Kits available?</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA Ligation</td>
<td>8</td>
<td>19</td>
<td>250</td>
<td>1200</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Illumina RNA Ligation</td>
<td>5</td>
<td>16</td>
<td>240</td>
<td>100(^b)</td>
<td>Yes</td>
<td>Partially (Small RNA Library Construction v1.5)</td>
</tr>
<tr>
<td>Illumina RNA Ligation - SPRI</td>
<td>4</td>
<td>12</td>
<td>220</td>
<td>100(^b)</td>
<td>Yes</td>
<td>Partially (Small RNA Library Construction v1.5)</td>
</tr>
<tr>
<td>SMART</td>
<td>5</td>
<td>8</td>
<td>80</td>
<td>100</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Hybrid</td>
<td>5</td>
<td>13</td>
<td>90</td>
<td>500</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>NNSR no actD</td>
<td>4</td>
<td>9</td>
<td>90</td>
<td>250</td>
<td>Unclear</td>
<td>No</td>
</tr>
<tr>
<td>NNSR</td>
<td>4</td>
<td>9</td>
<td>90</td>
<td>250</td>
<td>Unclear</td>
<td>No</td>
</tr>
<tr>
<td>Bisulfite &quot;S&quot;</td>
<td>6</td>
<td>19</td>
<td>540(^a)</td>
<td>1000(^c)</td>
<td>No</td>
<td>Mostly (Bisulfite &amp; Standard Library Construction)</td>
</tr>
<tr>
<td>Bisulfite &quot;H&quot;</td>
<td>6</td>
<td>19</td>
<td>540(^a)</td>
<td>1000(^d)</td>
<td>No</td>
<td>Mostly (Bisulfite &amp; Standard Library Construction)</td>
</tr>
<tr>
<td>dUTP</td>
<td>5</td>
<td>17</td>
<td>430(^a)</td>
<td>200</td>
<td>No</td>
<td>Mostly (Standard Library Construction)</td>
</tr>
<tr>
<td>dUTP oligo(dT)</td>
<td>5</td>
<td>17</td>
<td>440(^a)</td>
<td>200</td>
<td>No</td>
<td>Mostly (Standard Library Construction)</td>
</tr>
<tr>
<td>Control</td>
<td>5</td>
<td>15</td>
<td>430(^a)</td>
<td>200</td>
<td>No</td>
<td>Standard Library Construction</td>
</tr>
<tr>
<td>Control oligo(dT)</td>
<td>5</td>
<td>15</td>
<td>430(^a)</td>
<td>200</td>
<td>No</td>
<td>Standard Library Construction</td>
</tr>
</tbody>
</table>

\(a\): Cost is lower if individual reagents are used instead of Illumina standard library construction kit

\(b\): Starting material was 100 ng -- cDNA was split later for the two variants of this method (see Methods for details).

\(c\): Starting material was 1000 ng -- only 96 ng of 212 ng was used for reverse transcription

\(d\): Starting material was 1000 ng -- only 40 ng of 152 ng was used for reverse transcription

---

Yield from Levin et al. data.

How many reads do you need?
I don’t think there is currently a good answer.
Figure 9: Read bias: Nucleotide frequency vs. mapped read position.

Control: Phi-X  
Fc204, lane 8, OSU CGRB  
5,386 bp, 44% GC
RNA Ligation – SRR059162
_Saccharomyces cerevisiae_
12,156,677 bp, 38% GC

Hybrid - SRR059169

Bisulfite ‘S’ - SRR059174

Random hexamer bias

Random hexamer #2

Oligo dT #2
Reproducible bias within but not among methods.
Strand specificity

RNA-Seq library prep

**Purify RNA**

- mRNA
  - AAA
  - TTTTTT
  - Paramagnetic bead

- mRNA
  - AAA
  - AAA

**Fragment RNA**

- mRNA
  - AAA
  - AAA

- mRNA
  - AAA
  - AAA

**First strand synthesis**

- DNA – 5’
  - AAA
  - DNA-5’
  - DNA-5’

- DNA – 5’
  - AAA
  - DNA-5’
  - DNA-5’

**Second strand synthesis**

- 5’-DNA
  - DNA – 5’
  - DNA-5’
  - DNA-5’

- 5’-DNA
  - DNA – 5’
  - DNA-5’
  - DNA-5’

- 5’-DNA
  - DNA – 5’
  - DNA-5’
  - DNA-5’
End repair

A’ overhang

Ligate adapters

Size select electrophoresis

PCR Enrichment

<table>
<thead>
<tr>
<th>5’-DNA</th>
<th>5’-DNA</th>
<th>5’-DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA – 5’</td>
<td>DNA-5’</td>
<td>DNA-5’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5’-DNA</th>
<th>5’-DNA</th>
<th>5’-DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>A DNA – 5’</td>
<td>A DNA-5’</td>
<td>A DNA-5’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Adapter</th>
<th>5’-DNA</th>
<th>Adapter</th>
</tr>
</thead>
<tbody>
<tr>
<td>T 5’-DNA A</td>
<td>A DNA-5’</td>
<td>T</td>
</tr>
</tbody>
</table>

A' overhang

DNA – 5’ DNA-5’ DNA-5’
RNA ligation method

RNA extracted

Poly(A)$^+$ isolation via oligo dT selection
Structure of messenger ribonucleic acid (mRNA) in eukaryotes.
Illumina mRNA method – not strand specific

- **Fragment**
  - Metal hydrolysis (e.g., Mg, NaOAC)

- **First strand synthesis**
  - random hexamer (oligo dT)

- **Second strand synthesis**
  - E.coli pol I, RNase H

- **Clean**
  - QIAquick

- **End blunting**
  - T4 DNA polymerase, DNA pol I, Klenow, T4 polynucleotide kinase

- **Clean**
  - QIAquick

- **dA Tailing**
  - Klenow fragment (3’->5’ exo )

- **Clean**
  - MinElute

- **Adapter ligation**
  - T4 DNA ligase

- **Clean**
  - MinElute

- **Size selection**
  - Gel-based

- **PCR enrichment**
  - Phusion, 15 cycles

- **Clean**
  - QIAquick
dUTP mRNA method – strand specific

Fragment
Metal hydrolysis (e.g., Mg, NaOAC)

First strand synthesis
random hexamer (oligo dT)

Second strand synthesis
+ dUTP
E.coli pol I, RNase H

Clean
QIAquick

End blunting
T4 DNA polymerase, DNA pol I, Klenow, T4 polynucleotide kinase

Clean
QIAquick

dA Tailing
Klenow fragment (3’->5’ exo )

Clean
MinElute

Adapter ligation
T4 DNA ligase

Clean
MinElute

Size selection
Gel-based

Create gaps at Uracils
USER enzyme

PCR enrichment
Phusion, 15 cycles

Clean
QIAquick
Illumina RNA ligation method

**Poly(A)$^+$ RNA**
100 ng

**De-cap (5’ m7G)**
Tobacco acid pyrophosphatase

**Clean**
PCIA & EtOH

**Fragment**
Na citrate

**Clean**
EtOH precipitation

**3’ dephosphorylate**
Antarctic phosphatase

**5’ phosphorylate**
T4 polynucleotide kinase

**Ligate 3’ adapter**
T4 RNA ligase 2, truncated
Pre-adenylated adapter

**Ligate 5’ adapter**
T4 RNA ligase 1

**Reverse transcription**
SuperScript III

**Clean**
Rnase H, PCIA, EtOH

**Size selection**
Gel (175-225 nt)

**Clean**
Gel extract, EtOH

**PCR**
Phusion, 14 cycles

**Clean**
AMPure beads

**Clean**
Rneasy MinElute
## Illumina RNA ligation method SPRI

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poly(A)^+ RNA</td>
<td>100 ng</td>
</tr>
<tr>
<td>De-cap (5’ m7G)</td>
<td>Tobacco acid pyrophosphatase</td>
</tr>
<tr>
<td>Clean</td>
<td>PCIA &amp; EtOH</td>
</tr>
<tr>
<td>Fragment</td>
<td>Na citrate</td>
</tr>
<tr>
<td>Clean</td>
<td>EtOH precipitation</td>
</tr>
<tr>
<td>3’ dephosphorylate</td>
<td>Antarctic phosphatase</td>
</tr>
<tr>
<td>5’ phosphorylate</td>
<td>T4 polynucleotide kinase</td>
</tr>
<tr>
<td>Clean</td>
<td>Rneasy MinElute</td>
</tr>
<tr>
<td>Ligate 3’ adapter</td>
<td>T4 RNA ligase 2, truncated Pre-adenylated adapter</td>
</tr>
<tr>
<td>Ligate 5’ adapter</td>
<td>T4 RNA ligase 1</td>
</tr>
<tr>
<td>Reverse transcription</td>
<td>SuperScript III</td>
</tr>
<tr>
<td>Clean</td>
<td>Rnase H, PCIA, EtOH</td>
</tr>
<tr>
<td>PCR</td>
<td>Phusion, 14 cycles</td>
</tr>
<tr>
<td>Clean</td>
<td>AMPure beads</td>
</tr>
</tbody>
</table>
dUTP method

- **Poly(A)^+ RNA**
  - 200 ng

- **Fragment**
  - Na citrate

- **First strand synthesis**
  - Random hexamers, Superscript III

- **Clean**
  - 2X PCIA, EtOH

- **Second strand synthesis**
  - E.coli ligase, DNA pol I

- **Clean**
  - QIAquick

- **End blunting**
  - T4 DNA polymerase, DNA pol I, Klenow, T4 polynucleotide kinase

- **dA Tailing**
  - Klenow fragment (3’- > 5’ exo’)

- **Clean**
  - QIAquick

- **Adapter ligation**
  - T4 DNA ligase

- **Clean**
  - QIAquick

- **Size selection**
  - Gel-based

- **PCR enrichment**
  - Phusion, 14 cycles

- **Clean**
  - QIAquick
RNA-Seq library preparation

Brian J. Knaus
USDA Forest Service
Pacific Northwest Research Station