SOLiD Sequencing Applications
Sample Prep and Workflow

Dr. Deborah Grove, Director for Genetic Analysis
Dr. Craig Praul, Director for Gene Expression

The Genome Core Facility
The Huck Institutes of the Life Sciences
Penn State University
Services

Sanger Sequencing
SOLiD and 454 Next Gen Sequencing
qPCR by Real-Time
Microarray
Genotyping – Fragment Analysis, SNPs
DNA extraction and Storage of DNA from Buccal Swabs
DNA Synthesis
Nullius in Verba

Don’t Believe a Word of It!
Доверя́й, но проверя́й.

Trust

Then VERIFY!
SOLiD Applications Performed at the PSU Genome Facility

- Whole Genome
- ChIP-Seq
- Transcriptome
- Small RNA
- Degradome
- RIP - RNA Immunoprecipitation
- Future: SAGE
  - Multiplexing
Quality and Quantity

- Qubit
- Agilent Bioanalyzer

NOT the NanoDrop!
Sample Triage

Customer brings in a sample – Nanodrop Concentration

- RNA
  - Nanodrop concentration
    - Amount by Qubit
      - If less than 50%
      - If greater than 50% dsDNA

- dsDNA
  - Poor Quality
    - Bioanalyzer
  - Proceed

- Qubit
  - Poor Quality
    - Return to Customer
  - Proceed
# Agilent 2100 Bioanalyzer Chips

<table>
<thead>
<tr>
<th></th>
<th>RNA 6000 Nano total RNA Kit</th>
<th>RNA 6000 Nano mRNA Kit</th>
<th>RNA 6000 Pico total RNA Kit</th>
<th>RNA 6000 Pico mRNA Kit</th>
<th>Small RNA Kit</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Analytical Specifications</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Quantitative range</strong></td>
<td>25–500 ng/μL</td>
<td>25–250 ng/μL</td>
<td>-</td>
<td>-</td>
<td>50–2000 pg/μL of purified miRNA in water</td>
</tr>
<tr>
<td><strong>Qualitative range</strong></td>
<td>5–500 ng/μL</td>
<td>25–250 ng/μL</td>
<td>50–5000 pg/μL in water</td>
<td>250–5000 pg/μL in water</td>
<td>50–2000 pg/μL of purified miRNA in water</td>
</tr>
<tr>
<td><strong>Sensitivity (S/N&gt;3)</strong></td>
<td>5 ng/μL in water</td>
<td>25 ng/μL in water</td>
<td>50 pg/μL in water 200 pg/μL in TE</td>
<td>250 pg/μL in water 500 pg/μL in TE</td>
<td>50 pg/μL in water**</td>
</tr>
</tbody>
</table>
**Preparation of whole transcriptome RNA**

0.5-1 ug rRNA-depleted total RNA or poly(A) RNA

- Fragment the RNA
- Clean up the RNA

**Preparation of small RNA**

Obtain total RNA then determine the quality

- Purify/enrich small RNA if necessary
- Quantitate small RNA sample and determine input amount

**SOLiD™ amplified library construction**

Hybridize and ligate the RNA adapters over night

- Perform reverse transcription
- Purify the cDNA
- Size select the cDNA
- Amplify the cDNA
- Purify the amplified DNA
- Assess the yield and size distribution of the amplified DNA

**SOLiD™ System templated bead preparation and sequencing**
Small RNA ~18–40 nucleotides in length

- MicroRNA (miRNA) ~20-22
- short interfering RNA (siRNA) ~20-25
- piwi-interacting RNA (piRNA) ~26-31
- siRNA RNA (rasiRNA) ~24-29
Suggested Equipment?? Procedures??

Don’t Argue!
BUY IT!
Small RNA

Before Flash Gel

After Flash Gel

R&D 1 ng
SREK Protocol

Flash Page

Adaptors determine orientation

5’monophosphate intact
mRNA(5’cap) excluded

Caveats for plant and C. elegans
Reverse Transcription and RNase H Digestion (1 hr)

RNASE RNA from cDNA duplexes

Small RNA Library Amplification (1–1.5 hr)

15 to 18 cycles
105 to 150 bases

Sequence 1 x 10E6 to 5 x 10E7 to get from 100 to 200 different microRNA families.
Whole Transcriptome Library Preparation Example

PolyA + RNA – Ribosome Depleted

CAP ________________ ____________________________ AAAA AAAA

RNAse

Fragment to ~50-100 nt

__ __ __ __ __ __ __

Pico-RNA Chip

SREK Protocol

Can retain Strandedness
Clean up and size select

Clean up and amplify
Transcriptome Preparation

cDNA profile after varying temperature
Clean up and proceed to ePCR

Bead enrichment
### Table 2 Comparison of cDNA-tags assigned to major reference genomes using two sequencing methods

<table>
<thead>
<tr>
<th>Reference Genomes</th>
<th>Number of genes in reference genome</th>
<th>Number of mRNA-tags assigned</th>
<th>Number of genes covered</th>
<th>Number of genes significantly regulated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rosei</td>
<td>454 SOLID</td>
<td>454 SOLID</td>
<td>454 SOLID</td>
<td>454 SOLID</td>
</tr>
<tr>
<td>SyneA + SyneB</td>
<td>5622 4444</td>
<td>4425 309113</td>
<td>1532 4271</td>
<td>22 2274</td>
</tr>
<tr>
<td>Cab</td>
<td>3285 2850</td>
<td>1439 193996</td>
<td>473 2779</td>
<td>3 1782</td>
</tr>
<tr>
<td>GSB</td>
<td>2775 2775</td>
<td>1062 274210</td>
<td>475 2553</td>
<td>6 1708</td>
</tr>
<tr>
<td>Sum of above</td>
<td>16199 14456</td>
<td>15175 1529085</td>
<td>4567 13852</td>
<td>71 7777</td>
</tr>
</tbody>
</table>


### Table 3 Summary of cDNA-tags sequenced using SOLiD technology

<table>
<thead>
<tr>
<th></th>
<th>Sunset 21:00 h</th>
<th>Pre-dawn 05:15 h</th>
<th>Morning, low light 06:40 h</th>
<th>Morning, high light 08:40 h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total cDNA sequences</td>
<td>22,874,860</td>
<td>20,434,931</td>
<td>19,073,009</td>
<td>20,247,518</td>
</tr>
<tr>
<td>Total mapped&lt;sup&gt;a&lt;/sup&gt;</td>
<td>12,520,192</td>
<td>10,840,071</td>
<td>9,974,349</td>
<td>10,574,300</td>
</tr>
<tr>
<td>Mapped to CDS&lt;sup&gt;b&lt;/sup&gt;</td>
<td>743,902</td>
<td>318,205</td>
<td>477,362</td>
<td>1,057,433</td>
</tr>
<tr>
<td>Uniquely mapped to CDS&lt;sup&gt;b&lt;/sup&gt;</td>
<td>623,239</td>
<td>261,838</td>
<td>393,655</td>
<td>883,799</td>
</tr>
</tbody>
</table>

a. Percentage of total cDNA-tags

b. Percentage of total mapped cDNA-tags
Degradome

Procedure for uncapped mRNA

Degraded RNA

miRNA hairpin processing

Accumulations of siRNA precursors

Basically SREK protocol but without FlashPAGE
Size Select 100 to 800 bp
Sliced microRNA targets and precise loop-first processing of MIR319 hairpins revealed by analysis of the Physcomitrella patens degradome

CHARLES ADDO-QUAYE,1,2 JO ANN SNYDER,1,3 YONG BUM PARK,1,3,4 YONG-FANG LI,5 KAMANJU SUINAK,6 AND MICHAEL J. AXTELL1,5

1Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, Pennsylvania 16802, USA
2Department of Computer Science and Engineering, Pennsylvania State University, University Park, Pennsylvania 16802, USA
3Department of Biology, Pennsylvania State University, University Park, Pennsylvania 16802, USA
4Plant Biology Graduate Program, Pennsylvania State University, University Park, Pennsylvania 16802, USA
5Department of Biochemistry and Molecular Biology, Oklahoma State University, Stillwater, Oklahoma 74078, USA

ABSTRACT

Expression profiling of the 5' ends of uncapped mRNAs (“degradome” sequencing) can be used to empirically catalog microRNA (miRNA) targets, to probe patterns of miRNA hairpin processing, to examine miRNA decay, and to analyze accumulation of endogenous short interfering RNA (siRNA) precursors. We sequenced and analyzed the degradome of the moss Physcomitrella patens, an important model system for functional genomic analyses in plant evolution. A total of 52 target miRNAs of 27 different Physcomitrella miRNA families were identified. Many targets of both more conserved and less conserved miRNA families encoded putative regulatory proteins. Remnants of MIRNA hairpin processing also populated the degradome data and indicated an unusual “loop-first” mode of precise processing for the MIR319 gene family. Precise loop-first processing was confirmed for native Physcomitrella, rice, and Arabidopsis MIR319 hairpins, as well as an Arabidopsis artificial MIRNA (aMIRNA) based upon a MIR319 backbone. MIR319 is thus a conserved exception to the general rule of loop-last processing of MIRNA hairpins. Loop-first MIR319 processing may contribute to the high efficacy of a widely used MIR319-based strategy for aMIRNA production in plants.

Keywords: microRNA; uncapped mRNAs; plants; Physcomitrella; genomics; second-generation sequencing
RIP- Equivalent of ChIP-Seq
RNAs and nc RNAs
Remove crosslinks and denature

Follow Transcriptome Protocol

RNAse - ~ 100 bp

Hybridize and Ligate adaptors

RT to get cDNA

Size-Select 100 - 150 bp

Amplify and quantify
3’ SAGE Library Preparation Workflow Example

Region of 3’ UTR

<table>
<thead>
<tr>
<th>exon</th>
</tr>
</thead>
</table>

AAAAAAA d cDNA
TTTTTTTTT-EcoP15 site

NlaIII

AAAAAAA
TTTTTTTTT-EcoP15 site

Ligate

AAAAAAA
TTTTTTTTT-EcoP15 site

EcoP15 Digest, SA Capture

Advantage – No concatenation or bacterial cloning bias

© 2008 Applied Biosystems
Multiplexing -- Barcodes

Barcode 1
'0032'

<table>
<thead>
<tr>
<th>Barcode 1</th>
<th>Sequence</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’GCCCTGGGCGGTACCAGGCCCCTGCTTGGAGAAATGAGGAACCCGGGCACTT</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>3’GCGGACGCGGCTAGTGGGTCCCGGAATCTCTACACTCCCTGGGCCCCTG</td>
<td>48</td>
<td></td>
</tr>
</tbody>
</table>

Library Amplification Design
Multiplex Analysis Enables Simpler Workflow: Save Labor, Time, $
Barcoded Transcriptome Samples

<table>
<thead>
<tr>
<th>Sample</th>
<th>Total reads</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>34,893,856</td>
</tr>
<tr>
<td>Dark ANaerobic</td>
<td>34,204,786</td>
</tr>
<tr>
<td>Dark O₂</td>
<td>15,512,068</td>
</tr>
<tr>
<td>Air (low CO₂)</td>
<td>26,949,795</td>
</tr>
<tr>
<td>Glycerol</td>
<td>22,194,244</td>
</tr>
<tr>
<td>High light</td>
<td>16,842,005</td>
</tr>
<tr>
<td>OD 0.4</td>
<td>26,488,336</td>
</tr>
<tr>
<td>OD 1.0</td>
<td>39,063,718</td>
</tr>
<tr>
<td>OD 3.0</td>
<td>26,691,541</td>
</tr>
<tr>
<td>OD 5.0</td>
<td>20,220,020</td>
</tr>
<tr>
<td>22°C</td>
<td>29,868,989</td>
</tr>
<tr>
<td>30°C</td>
<td>25,094,301</td>
</tr>
<tr>
<td>Glycerol dark</td>
<td>25,788,734</td>
</tr>
<tr>
<td>N-deprived</td>
<td>26,430,653</td>
</tr>
<tr>
<td>Barcode</td>
<td>0 Mismatches</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>bcB20_02</td>
<td>29623690</td>
</tr>
<tr>
<td>Subtotals</td>
<td>29623690</td>
</tr>
<tr>
<td>bcB20_08</td>
<td>32073199</td>
</tr>
<tr>
<td>Subtotals</td>
<td>32073199</td>
</tr>
<tr>
<td>bcB20_01</td>
<td>31190566</td>
</tr>
<tr>
<td>Subtotals</td>
<td>31190566</td>
</tr>
<tr>
<td>bcB20_06</td>
<td>22166492</td>
</tr>
<tr>
<td>Subtotals</td>
<td>22166492</td>
</tr>
<tr>
<td>bcB20_03</td>
<td>26252079</td>
</tr>
<tr>
<td>Subtotals</td>
<td>26252079</td>
</tr>
<tr>
<td>bcB20_05</td>
<td>34222321</td>
</tr>
<tr>
<td>Subtotals</td>
<td>34222321</td>
</tr>
<tr>
<td>bcB20_07</td>
<td>37031380</td>
</tr>
<tr>
<td>Subtotals</td>
<td>37031380</td>
</tr>
<tr>
<td>bcB20_04</td>
<td>28681842</td>
</tr>
<tr>
<td>Subtotals</td>
<td>28681842</td>
</tr>
</tbody>
</table>
**SOliD Metamorphoses**

<table>
<thead>
<tr>
<th>Version</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>v2.0</td>
<td>150 million mappable reads</td>
</tr>
<tr>
<td>v3.0</td>
<td>280 million</td>
</tr>
<tr>
<td>v3.5</td>
<td>375 million</td>
</tr>
<tr>
<td>v4.0</td>
<td>700 million</td>
</tr>
<tr>
<td>HQ</td>
<td>3 x v4.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Version</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>V2.0</td>
<td>35 bases</td>
</tr>
<tr>
<td>V3.0</td>
<td>50 bases</td>
</tr>
<tr>
<td>HQ</td>
<td>75 bases</td>
</tr>
</tbody>
</table>
qPCR

Used to normalize multiplexed samples before ePCR

Taqman Kit only amplifies dsDNA with primer sequences.
Analysis

• High Level Labs
• Galaxy
• “Plug and Play”
• AB Software Bioskope
PSU’s Anton Nekrutenko’s Galaxy site at [http://usegalaxy.org](http://usegalaxy.org)
Mapping SOLiD reads
A single-end example
Galaxy Sample Tracking System

Anton Nekurtenko
Ramkrishna Chakrabarty
Dept of Biochemistry and Molecular Biology
Penn State University
Create a Sequencing Request

1. Admin tab
2. Manage Requests link
3. Create new request button

4. Fill request form & Add samples

5. Fill samples info & Save
Thanks to:

The Huck Institutes of the Life Sciences
Lloyd and Dorothy Huck

And the others in the lab…