

# **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



Gregory Grove, Candace Price, Craig Praul, Deb Grove, Ashley Price, Sheila Plock, Kelly Crawford

## **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

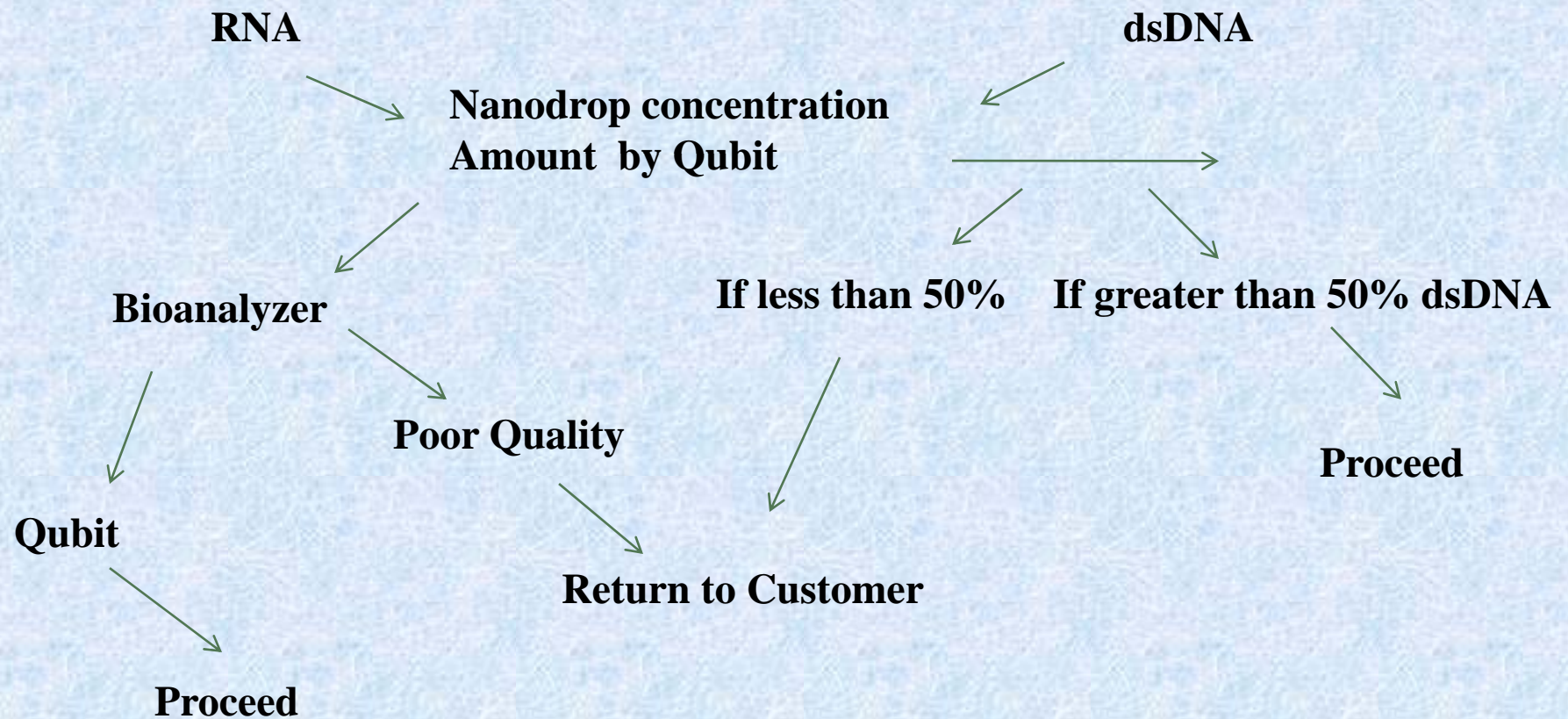
- Qubit
- Agilent Bioanalyzer



NOT the NanoDrop!

# Sample Triage

Customer brings in a sample – Nanodrop Concentration





# Agilent 2100 Bioanalyzer Chips

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



### Preparation of whole transcriptome RNA

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

↓  
Fragment the RNA

↓  
Clean up the RNA

### 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

### Preparation of small RNA

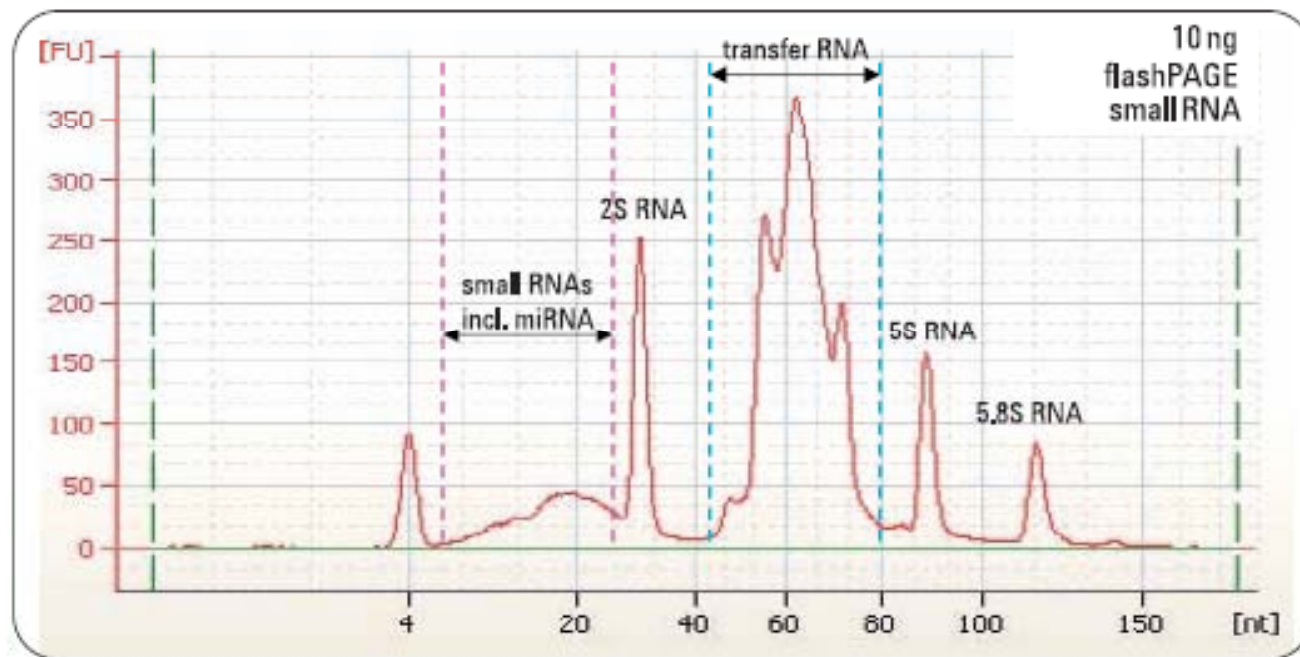
Obtain total RNA then determine the quality

↓  
Purify/enrich small RNA if necessary

↓  
Quantitate small RNA sample and determine input amount

## **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



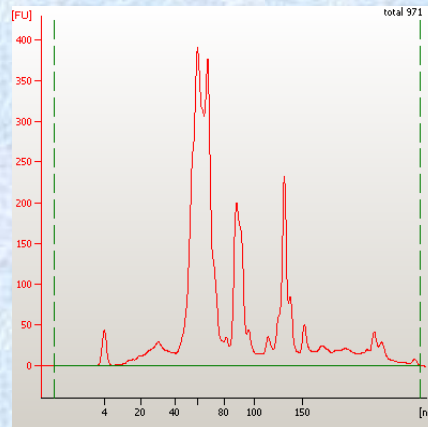


**Suggested  
Equipment??  
Procedures??**

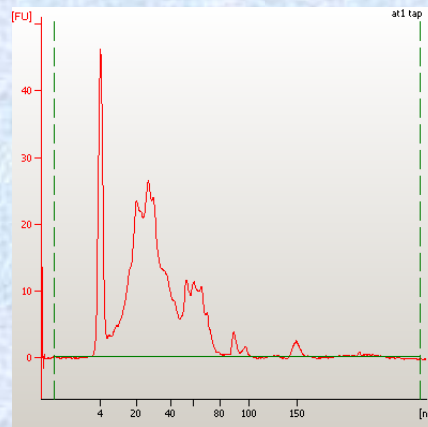
**Don't Argue!  
BUY IT!**

# Small RNA

Before Flash Gel

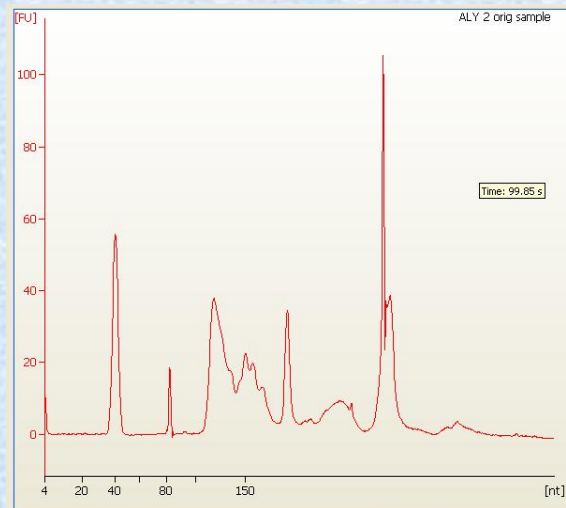


After Flash Gel

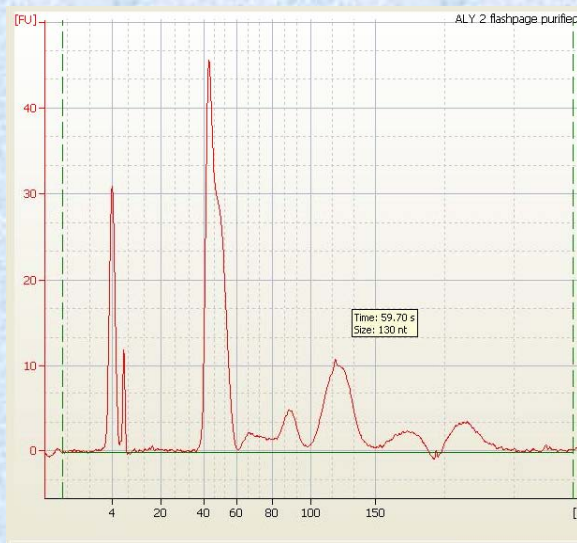


R&D 1 ng

# Before



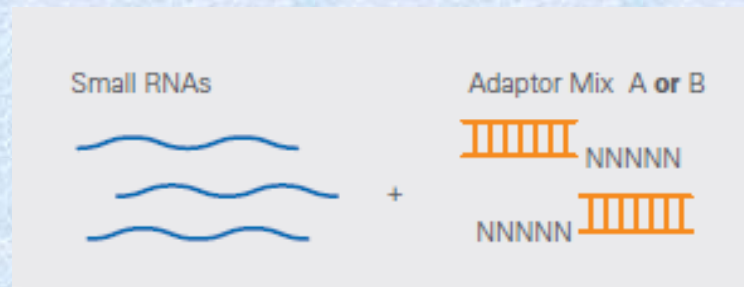
# After





# SREK Protocol

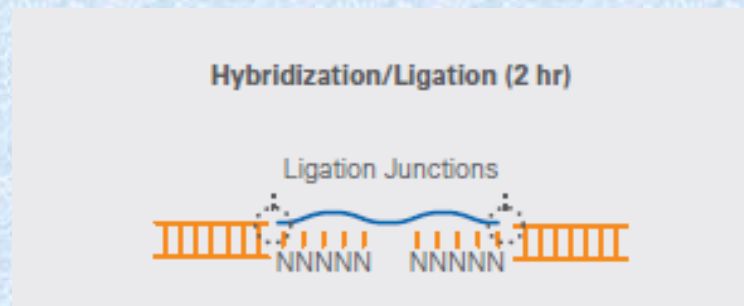
## Flash Page



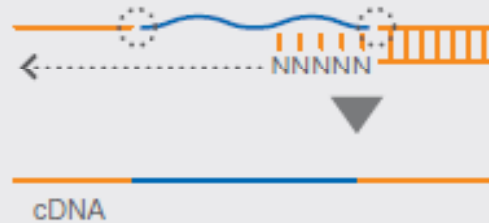
Adaptors determine orientation

5' monophosphate intact  
mRNA(5' cap) excluded

Caveats for plant and *C. elegans*

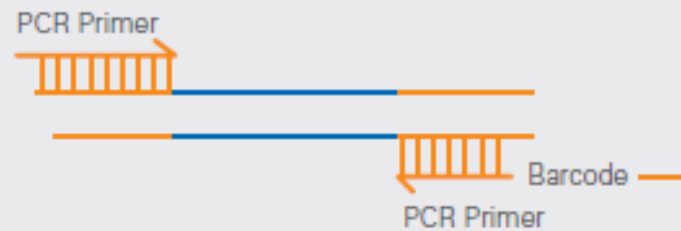


### Reverse Transcription and RNase H Digestion (1 hr)



RNase H from cDNA duplexes

### Small RNA Library Amplification (1–1.5 hr)

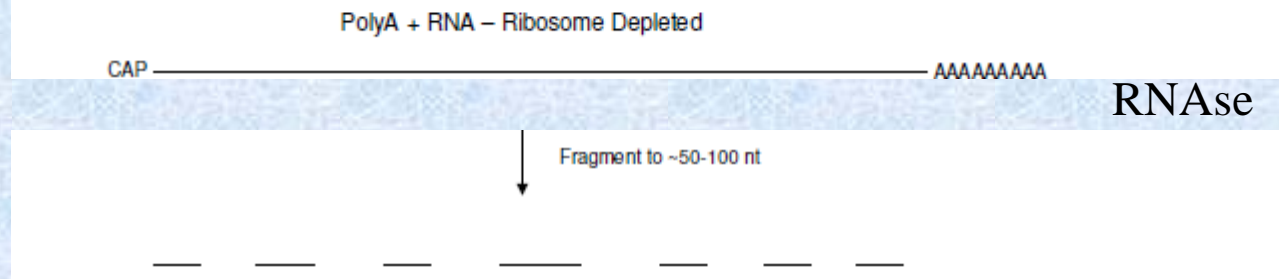


15 to 18 cycles

105 to 150 bases

Sequence  $1 \times 10^6$  to  $5 \times 10^7$  to get from 100 to 200 different microRNA families.

## Whole Transcriptome Library Preparation Example



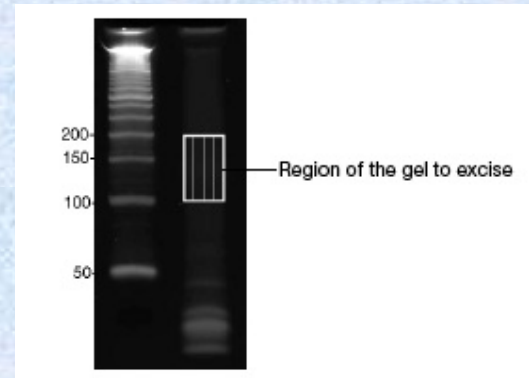
Pico-RNA Chip

SREK Protocol

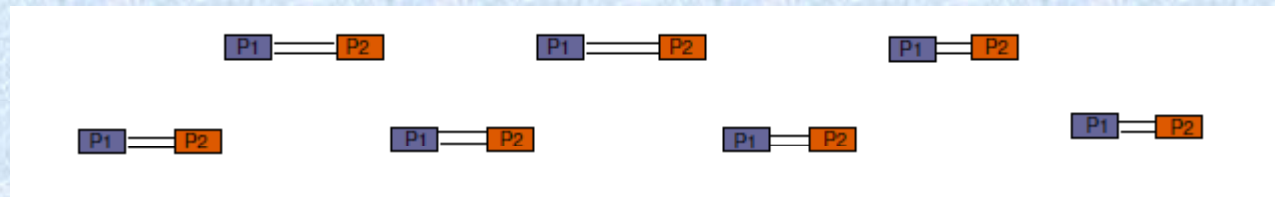
Can retain Strandedness



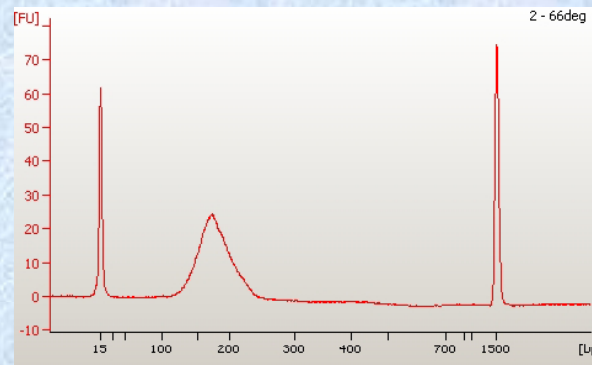
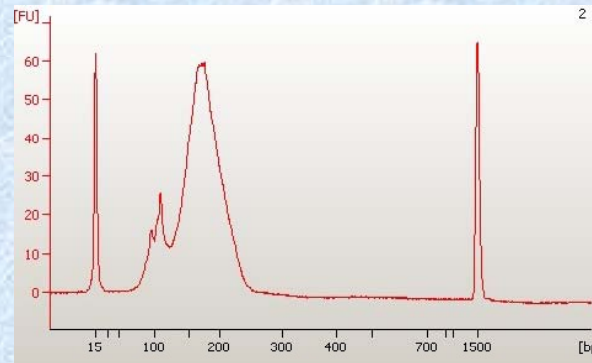
Clean up and size select



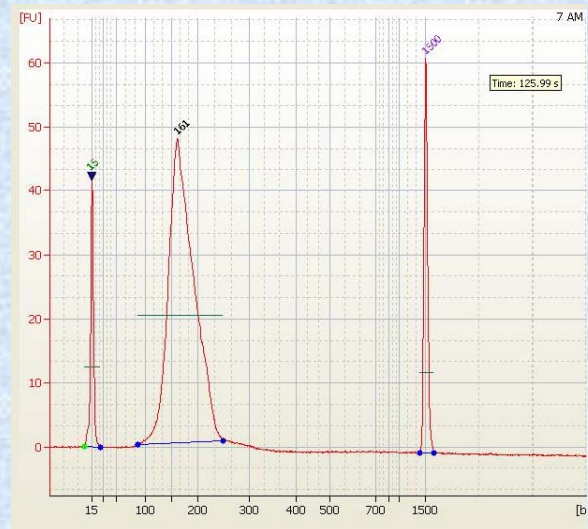
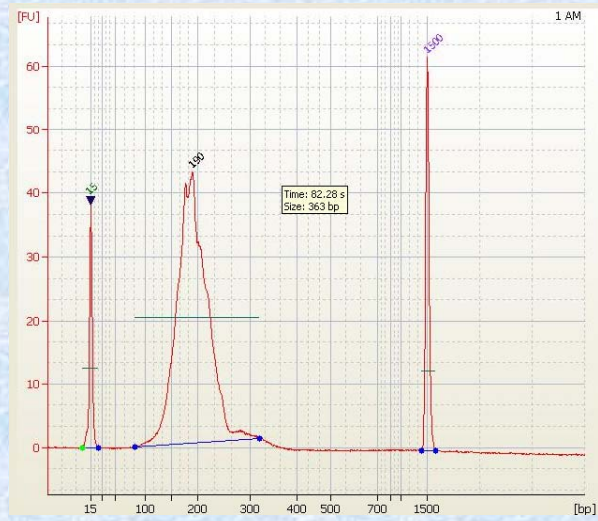
Clean up and amplify



# Transcriptome Preparation



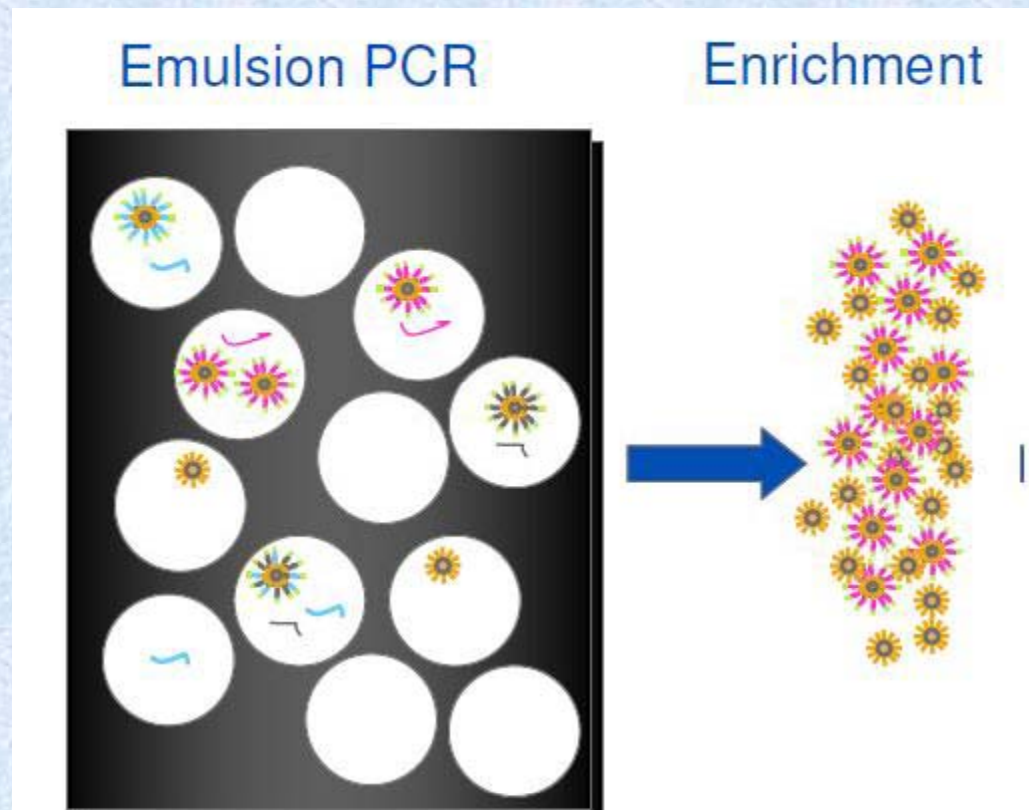
cDNA profile after varying temperature





Clean up and proceed to ePCR

Bead enrichment



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

Reference Genomes <sup>a</sup>	Number of genes in reference genome <sup>b</sup>		Number of mRNA-tags assigned <sup>c</sup>		Number of genes covered <sup>d</sup>		Number of genes significantly regulated <sup>e</sup>	
	454	SOLiD	454	SOLiD	454	SOLiD	454	SOLiD
Rosei	4517	4387	8249	751766	2087	4249	40	2013
SyneA + SyneB	5622	4444	4425	309113	1532	4271	22	2274
Cab	3285	2850	1439	193996	473	2779	3	1782
GSB	2775	2775	1062	274210	475	2553	6	1708
Sum of above	16199	14456	15175	1529085	4567	13852	71	7777

a. Genome abbreviations. Rosei, *Roseiflexus* sp. RS-1; SyneA, *Synechococcus* sp. strain A; SyneB, *Synechococcus* sp. strain B<sup>+</sup>; Cab, *Candidatus* *Chloracidobacterium thermophilum*; GSB, a *Chlorobiales* species.

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

	Sunset 21:00 h	Pre-dawn 05:15 h	Morning, low light 06:40 h	Morning, high light 08:40 h
Total cDNA sequences	22,874,860	20,434,931	19,073,009	20,247,518
Total mapped <sup>a</sup>	12,520,192 (54.7%)	10,840,071 (53.0%)	9,974,349 (52.3%)	10,574,300 (52.2%)
Mapped to CDS <sup>b</sup>	743,902 (5.9%)	318,205 (2.9%)	477,362 (4.8%)	1,057,433 (10.0%)
Uniquely mapped to CDS <sup>b</sup>	623,239 (5.0%)	261,838 (2.4%)	393,655 (3.9%)	883,799 (8.4%)

a. Percentage of total cDNA-tags  
b. Percentage of total mapped cDNA-tags

# Degradome

Procedure for uncapped mRNA

Degraded RNA

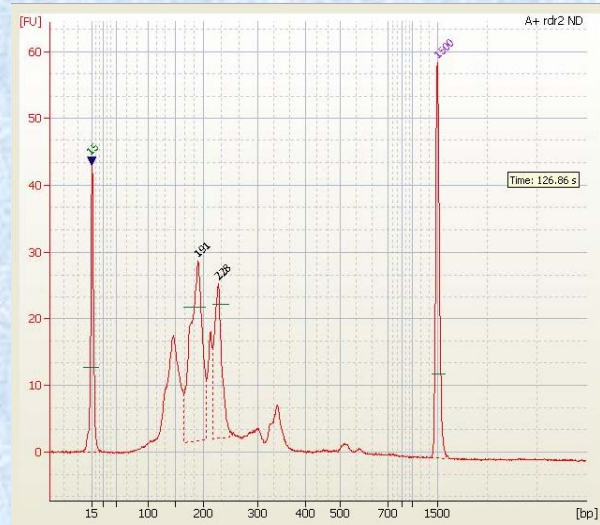
miRNA hairpin processing

Accumulations of siRNA precursors

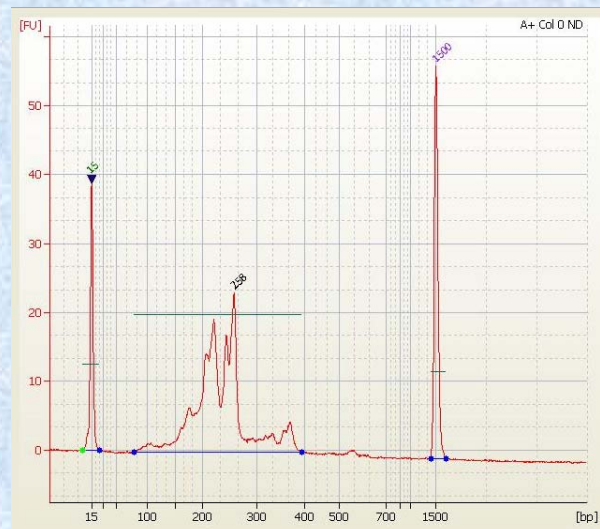
Basically SREK protocol but without  
FlashPAGE







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,<sup>1,2</sup> JO ANN SNYDER,<sup>1,3</sup> YONG BUM PARK,<sup>1,3,4</sup> YONG-FANG LI,<sup>5</sup> RAMANJULU SUNKAR,<sup>5</sup> and MICHAEL J. AXTELL<sup>1,3,4</sup>

<sup>1</sup>Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, Pennsylvania 16802, USA

<sup>2</sup>Department of Computer Science and Engineering, Pennsylvania State University, University Park, Pennsylvania 16802, USA

<sup>3</sup>Department of Biology, Pennsylvania State University, University Park, Pennsylvania 16802, USA

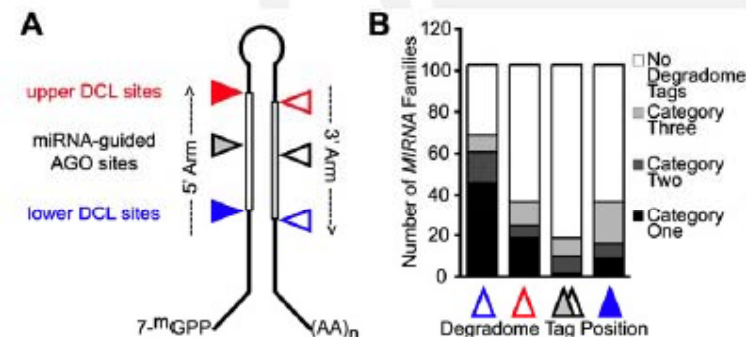
<sup>4</sup>Plant Biology Graduate Program, Pennsylvania State University, University Park, Pennsylvania 16802, USA

<sup>5</sup>Department of Biochemistry and Molecular Biology, Oklahoma State University, Stillwater, Oklahoma 74078, USA

## ABSTRACT

**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 mRNA 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 mRNAs 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* (a*MIRNA*) 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 a*MIRNA* production in plants.

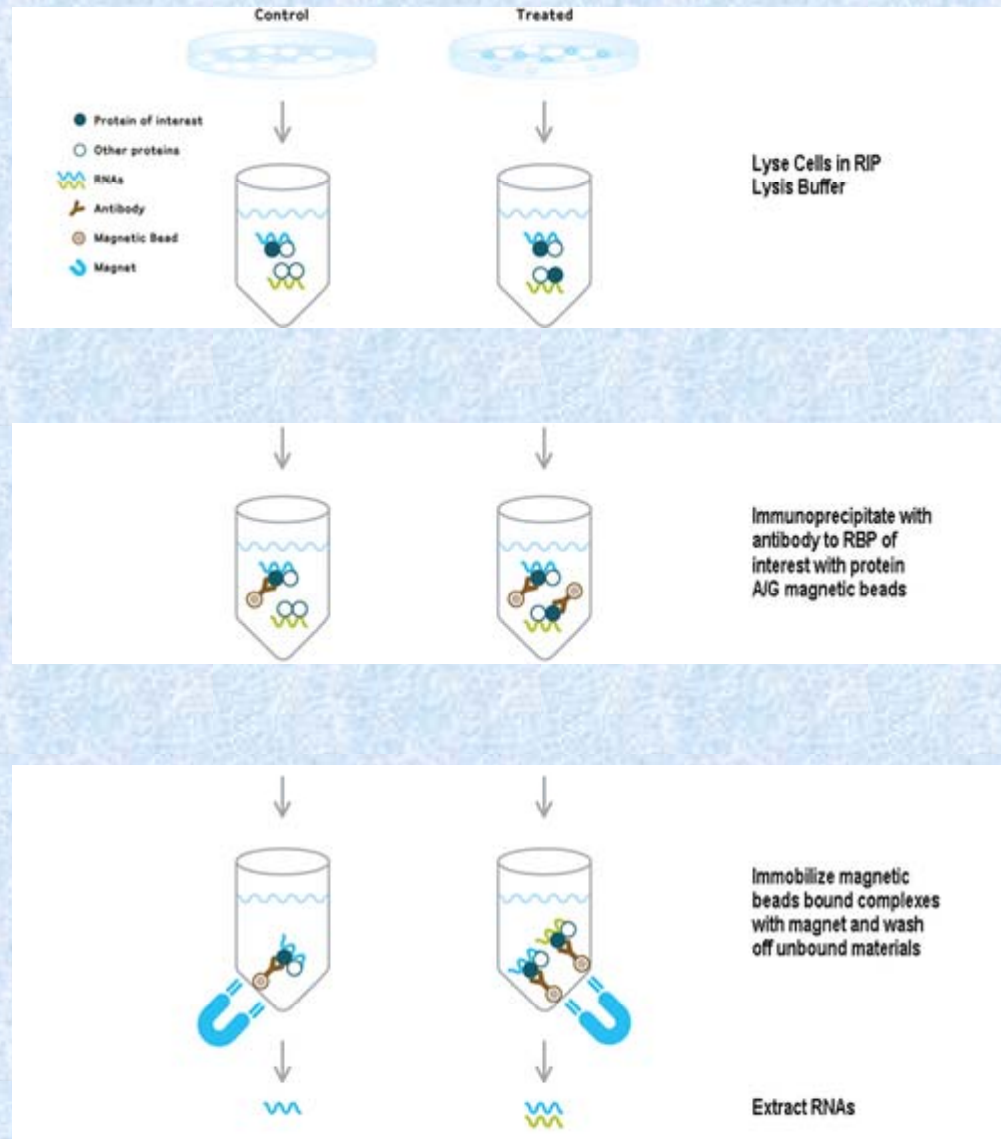
**Keywords:** microRNA; uncapped mRNAs; plants; *Physcomitrella*; genomics; second-generation sequencing





# RIP- Equivalent of ChIP-Seq

## RNAs and nc RNAs





Remove crosslinks and denature



Follow Transcriptome Protocol



RNAse - ~ 100 bp



Hybridize and Ligate adaptors



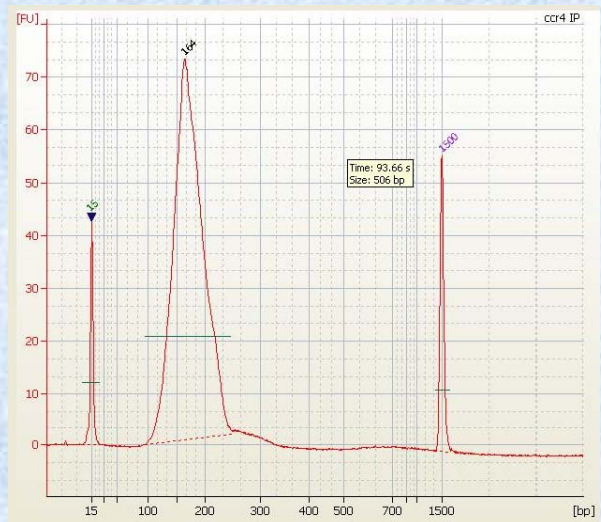
RT to get cDNA



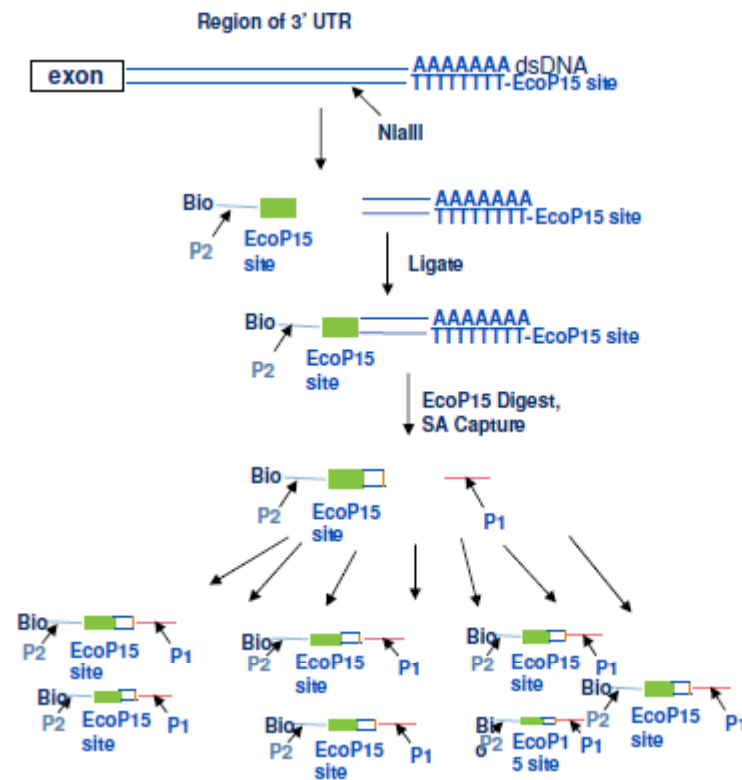
Size-Select 100 - 150 bp



Amplify and quantify



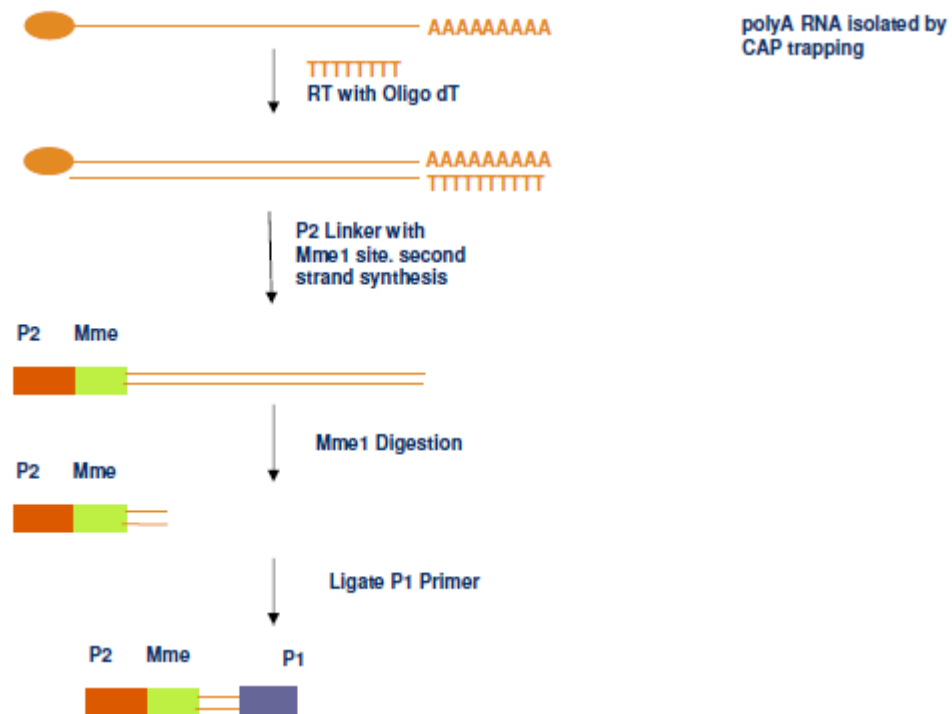
## 3' SAGE Library Preparation Workflow Example



**Advantage – No  
concatenation or bacterial  
cloning bias**



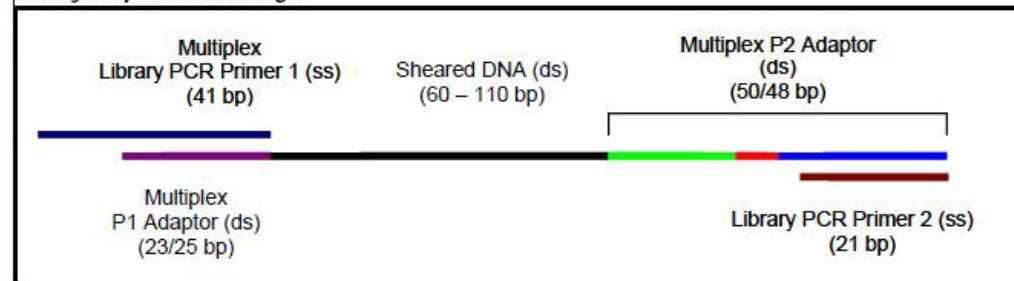
## CAGE Library Preparation Workflow Example



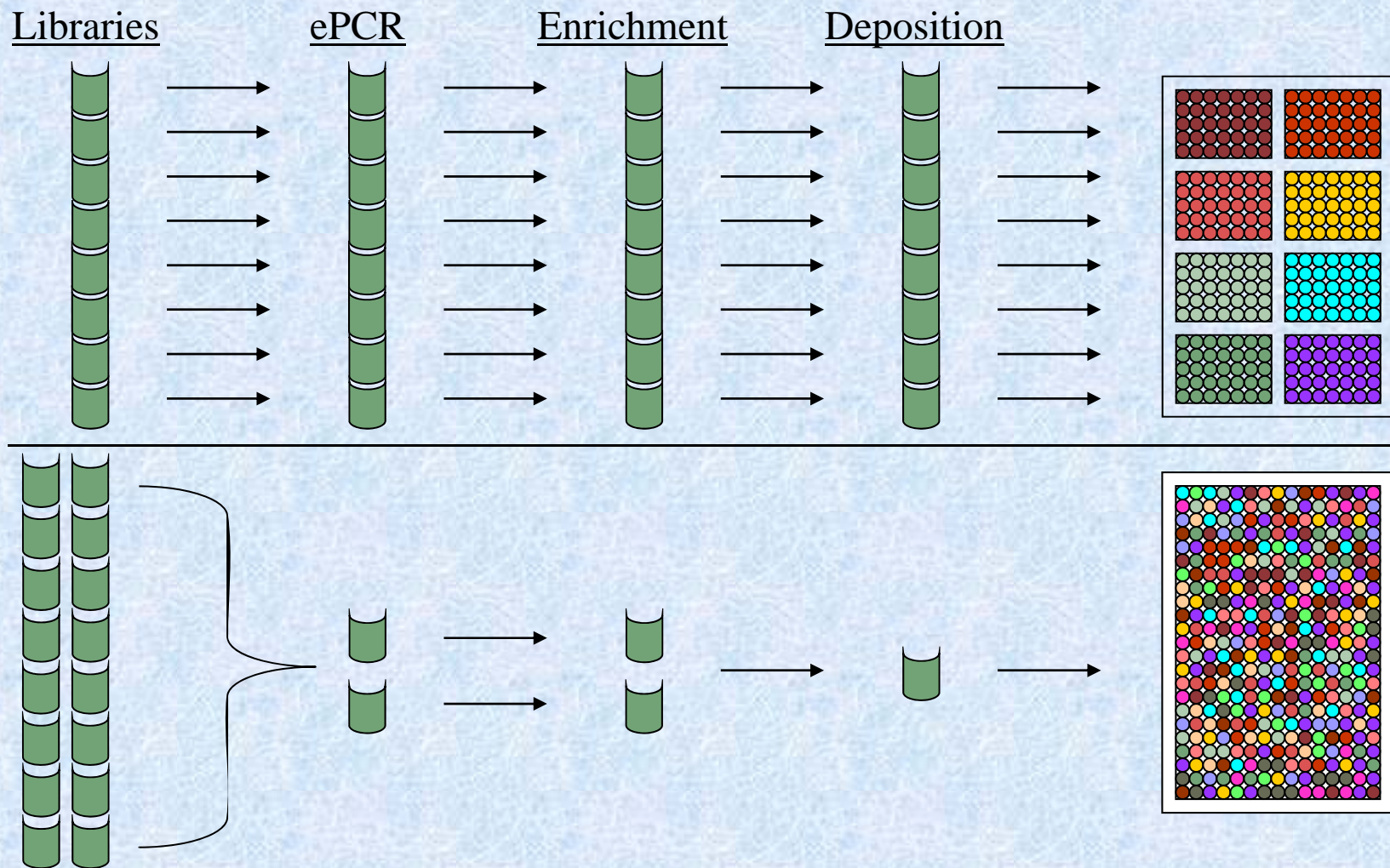
# Multiplexing -- Barcodes

Barcode 1 "0032"	5'	CGCCTTGGCCGTACAGCAGGGGCTTAGAGAATGAGGAACCCGGGGCAGTT	3'	Length: 50
	3'	GCGGAACCGGCATGTCGTC	5'	Length: 48

## Library Amplification Design



# Multiplex Analysis Enables Simpler Workflow: Save Labor, Time, \$





# Barcoded Transcriptome Samples

Sample	Total reads
Standard	34,893,856
Dark ANaerobic	34,204,786
Dark O <sub>2</sub>	15,512,068
Air (low CO <sub>2</sub> )	26,949,795
Glycerol	22,194,244
High light	16,842,005
OD 0.4	26,488,336
OD 1.0	39,063,718
OD 3.0	26,691,541
OD 5.0	20,220,020
22°C	29,868,989
30°C	25,094,301
Glycerol dark	25,788,734
N-deprived	26,430,653

Barcode	0 Mismatches	1 Mismatch	
bcB20_02	29623690	0	29623690
Subtotals	29623690	0	29623690
bcB20_08	32073199	0	32073199
Subtotals	32073199	0	32073199
bcB20_01	31190566	0	31190566
Subtotals	31190566	0	31190566
bcB20_06	22166492	0	22166492
Subtotals	22166492	0	22166492
bcB20_03	26252079	0	26252079
Subtotals	26252079	0	26252079
bcB20_05	34222321	0	34222321
Subtotals	34222321	0	34222321
bcB20_07	37031380	0	37031380
Subtotals	37031380	0	37031380
bcB20_04	28681842	0	28681842
Subtotals	28681842	0	28681842

# **SOLiD Metamorphoses**

v2.0	150 million mappable reads
v3.0	280 million
v3.5	375 million
v4.0	700 million
HQ	3 x v4.0

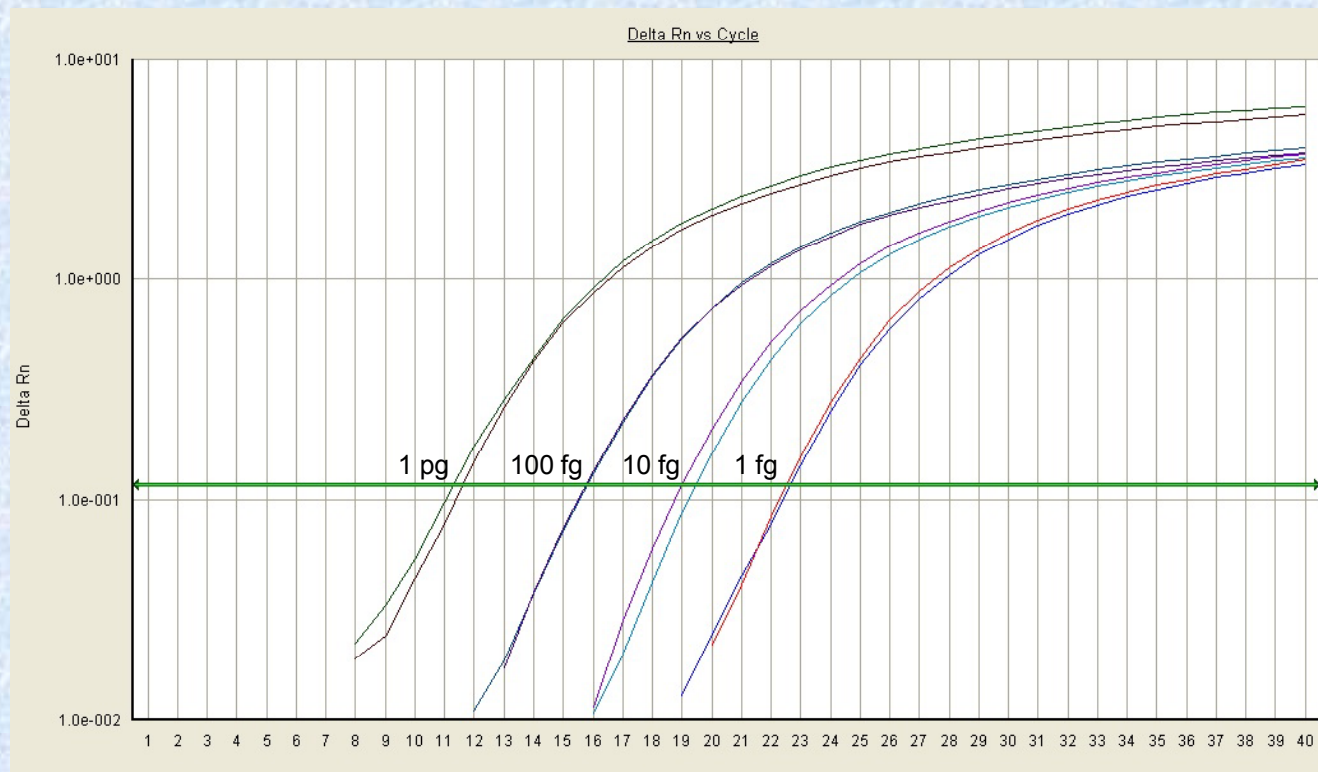
V2.0	35 bases
V3.0	50 bases
HQ	75 bases



# 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 Bioscope

The screenshot displays the Galaxy web interface. The top navigation bar includes links for 'Analyze Data', 'Workflow', 'Data Libraries', 'Help', and 'User'. The left sidebar, titled 'Tools', lists various bioinformatics tools, with 'Get Data' and 'NGS TOOLBOX BETA' circled. The central panel features a large announcement for 'Galaxy Pages Spring 2010', described as 'A new standard for reproducible research'. Below this, it lists 'Timeline for upcoming Quickies' and 'Live Quickies' with specific dates and tasks. The right sidebar, titled 'History', shows an 'Unnamed history' and a message indicating that the history is empty and suggesting to click 'Get Data' to start. The bottom of the central panel contains information about the Galaxy team at Penn State and project funding sources.

**Galaxy**

Analyze Data Workflow Data Libraries Help User

Tools

- Get Data**
- Send Data
- ENCODE Tools
- Lift-Over
- Text Manipulation
- Convert Formats
- FASTA manipulation
- Filter and Sort
- Join, Subtract and Group
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Regional Variation
- Multiple regression
- Evolution
- Metagenomic analyses
- EMBOSS
- NGS TOOLBOX BETA**
- NGS: QC and manipulation
- NGS: Mapping
- NGS: SAM Tools
- NGS: Peak Calling

Here is what's happening...

## Galaxy Pages

Spring 2010

A new standard for reproducible research

Access or analyze Southern African Genomes through Galaxy!

### Timeline for upcoming Quickies

- Mar 5 - Manipulating fastq with Galaxy
- Mar 22 - Mapping of 454 reads with lastZ

### Live Quickies

- Mapping against custom genome  
Galactic quickie # 10
- Illumina mapping: Single Ends  
Galactic quickie # 11
- Illumina mapping: Paired Ends  
Galactic quickie # 12

The Galaxy team is a part of BX at Penn State.

This project is supported in part by NSF, NHGRI, The Huck Institutes of the Life Sciences, and The Institute for CyberScience at Penn State.

Galaxy build: \$Rev 1733:a4214de3752e\$

History Options

Unnamed history

Your history is empty. Click 'Get Data' on the left pane to start

PSU's Anton Nekrutenko's Galaxy site at  
<http://usegalaxy.org>



# Mapping SOLiD reads

A single-end example



00:09 / 08:01



# **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

**Add a new request**

Select Request Type:  
[HD] [v]

Select user:  
[swigun.pawank@...] [v]  
The request would be submitted on behalf of this user (Required)

Name of the Experiment:  
[Experiment001] [v]  
(Required)

Description:  
[Demo experiment] [v]  
(Optional)

Name:  
[Ramkrishna Chakrabarty] [v]

Name of the person submitting the sequencing request (Optional):  
[Dr. Anton Nekrutenko] [v]

Principal Investigator:  
[Dr. Anton Nekrutenko] [v]

Department:  
[Biochemistry and Molecular Biology] [v]

Contact Information:  
[select one] [v]

[Save] [Add samples]

5. Fill samples info & Save

**Sample Information**

Name	Barcode	Index	Data Library	Folder
Sample_1 <small>required</small>		[Fastflow] [v]	[Select one] [v]	
Sample_2 <small>required</small>		[Fastflow] [v]	[Select one] [v]	

**Run details**

Name	Sample type	Library	Workflow
Sample_1 <small>required</small>	[DNA] [v]	[Illumina] [v]	[vul] [v]
Sample_2 <small>required</small>	[DNA] [v]	[Illumina] [v]	[vul] [v]

[No samples] Import from csv file (Upload file) No file chosen [Copy from sample] [Add New]

[Save]



Thanks to:

The Huck Institutes of the Life Sciences  
Lloyd and Dorothy Huck

And the others in the lab...

