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

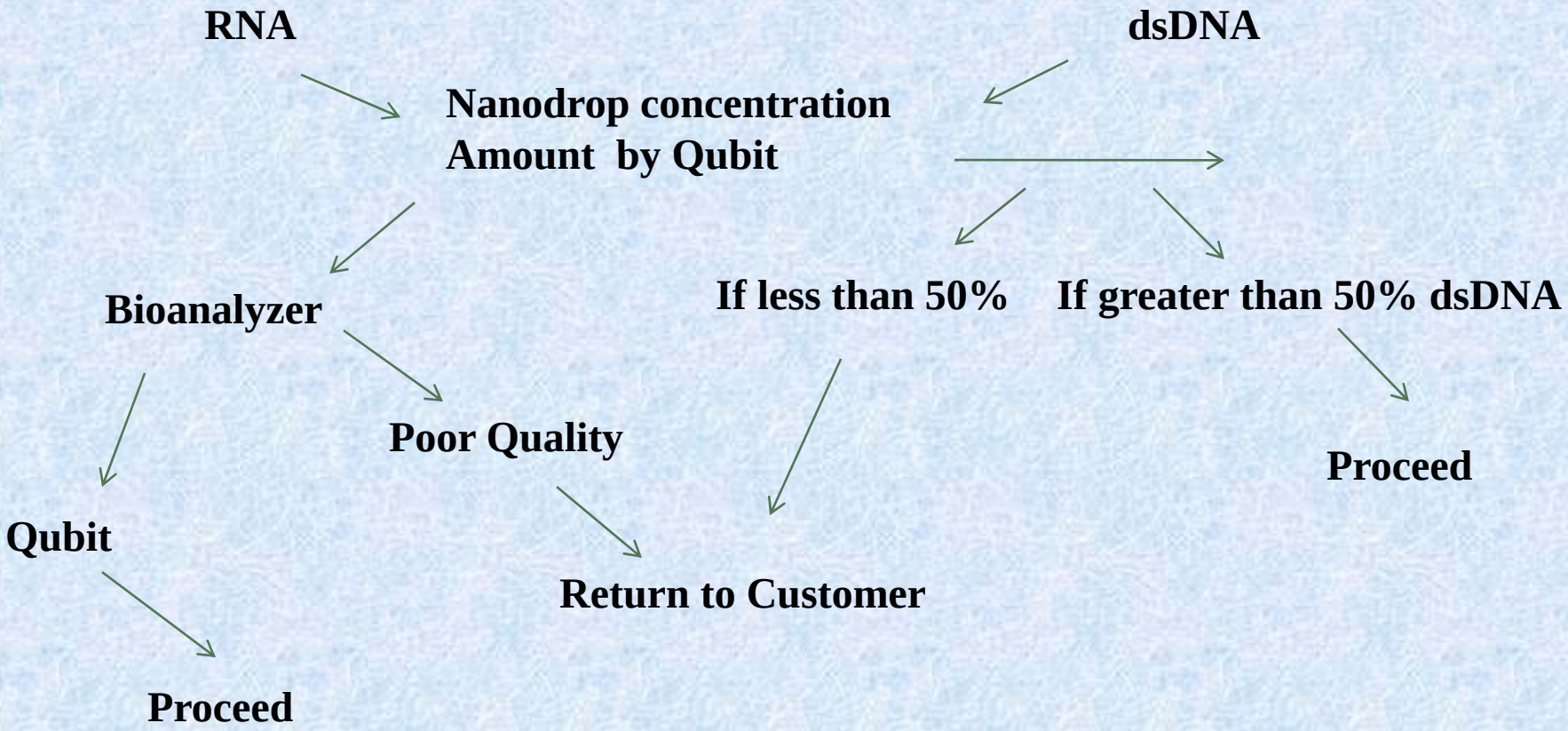
Quality and Quantity

- Qubit
- Agilent Bioanalyzer

NOT the NanoDrop!

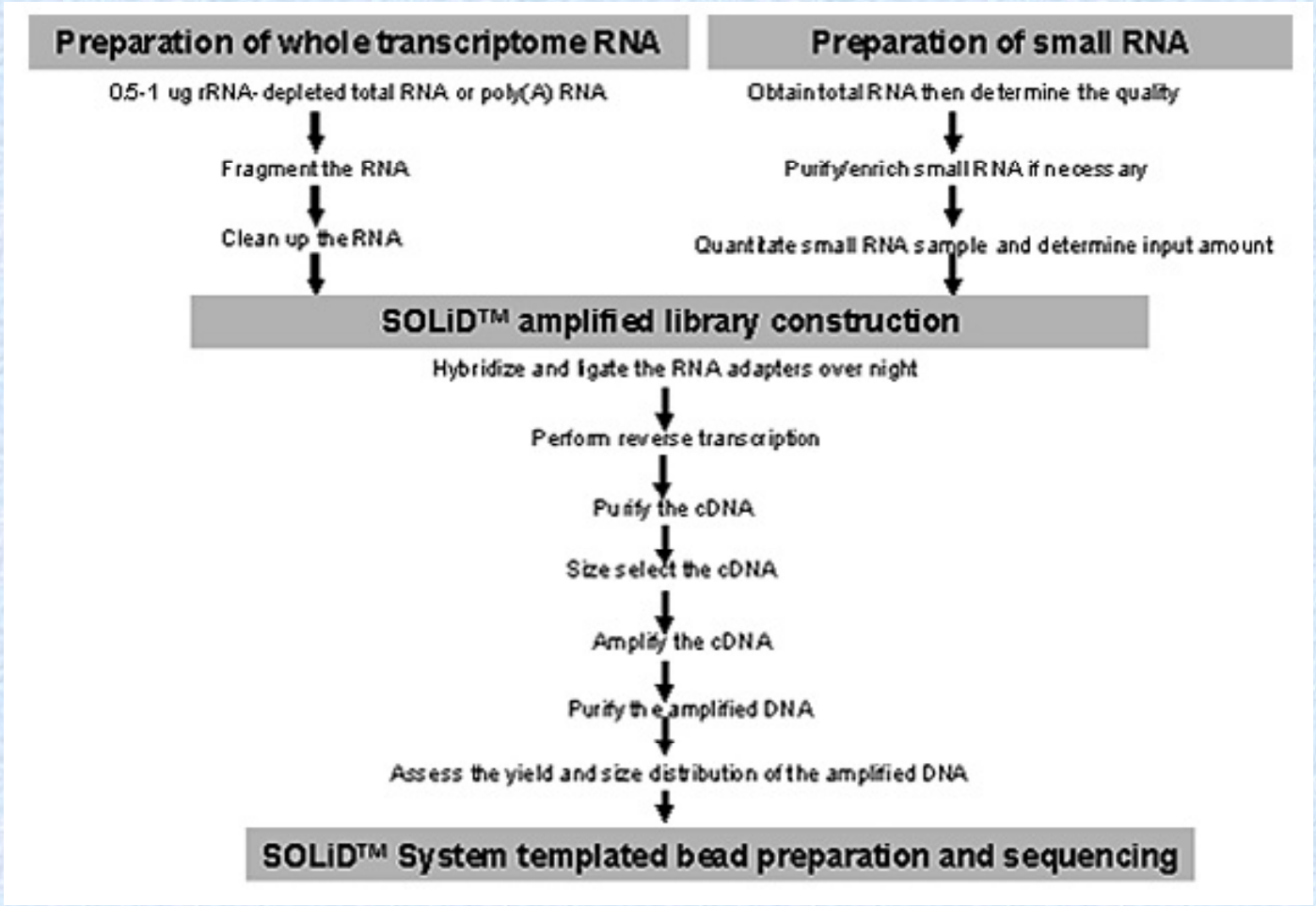
Sample Triage

Customer brings in a sample – Nanodrop Concentration



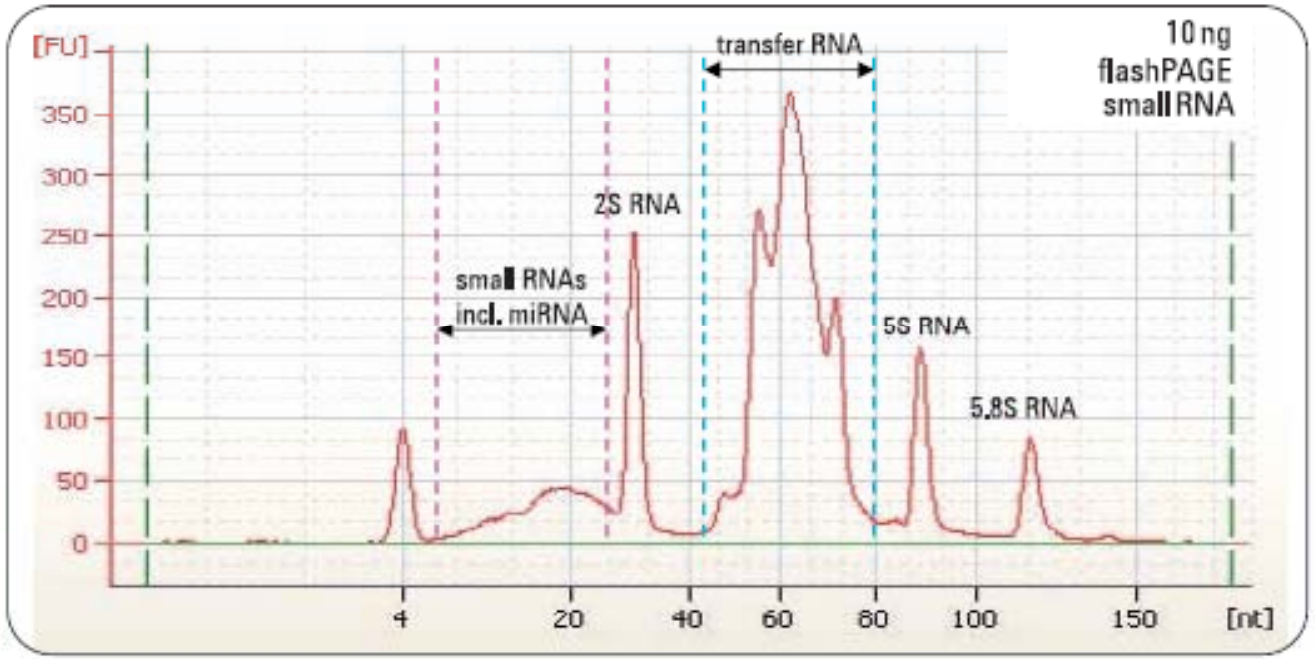
Agilent 2100 Bioanalyzer Chips

	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
Analytical Specifications					
Quantitative range	25–500 ng/µL	25–250 ng/µL	-	-	50-2000 pg/µL of purified miRNA in water
Qualitative range	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
Sensitivity (S/N>3)	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
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


[FU]
350
300
250
200
150
100
50
0
transfer RNA
10 ng
flashPAGE
small RNA
2S RNA
small RNAs
incl. miRNA
5S RNA
5.8S RNA
4
20
40
60
80
100
150
[nt]

Suggested Equipment??

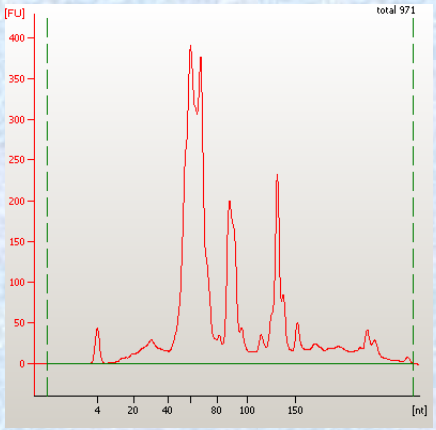
Procedures??

Don’t Argue!

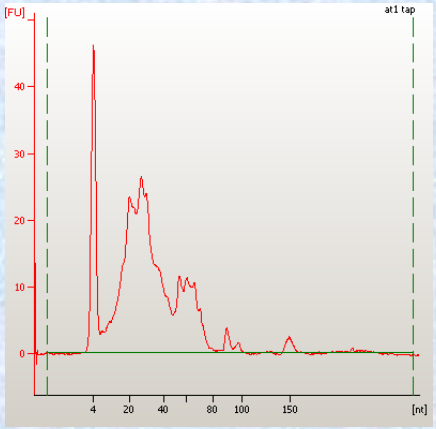
BUY IT!

Small RNA

Before Flash Gel

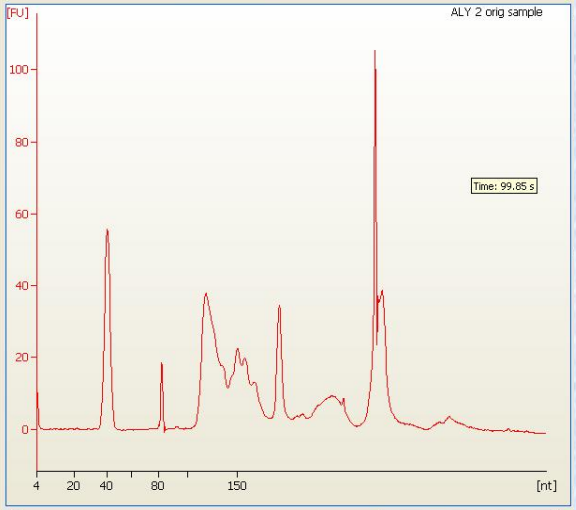


After Flash Gel

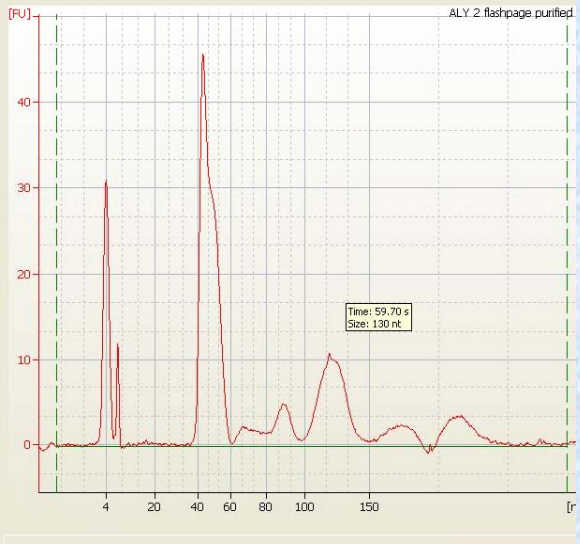


R&D 1 ng

Before

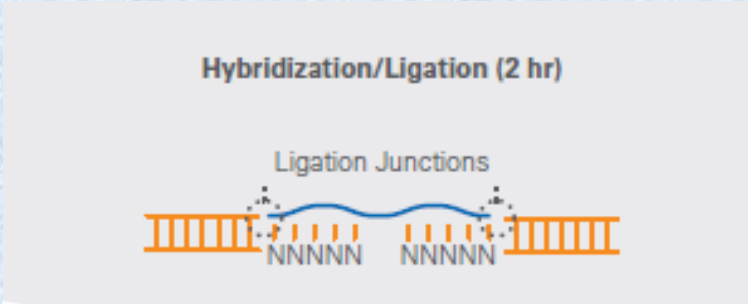


After



SREK Protocol

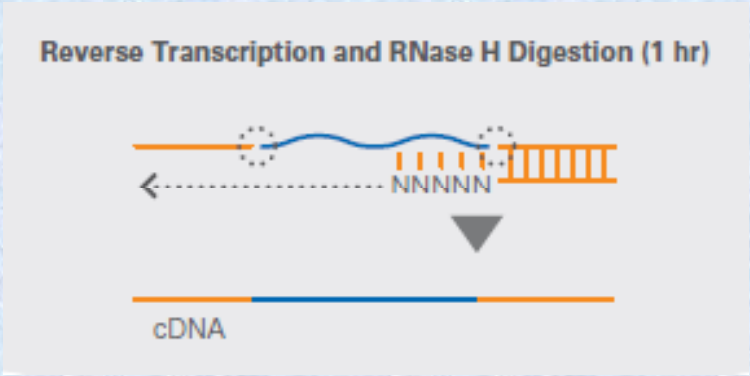
Flash Page



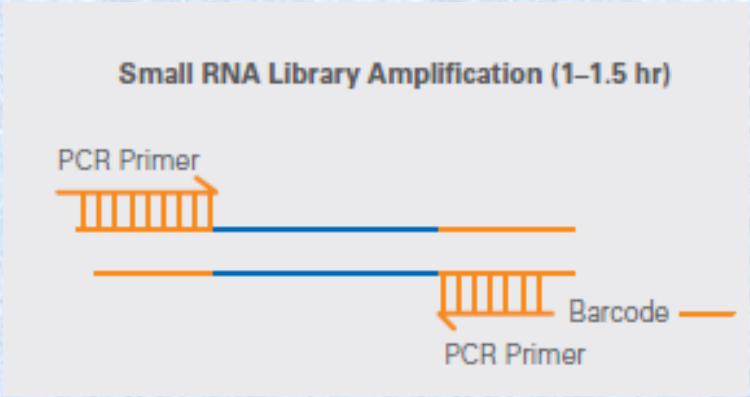
Adaptors determine orientation

5’monophosphate intact mRNA(5’cap) excluded

Caveats for plant and C. elegans



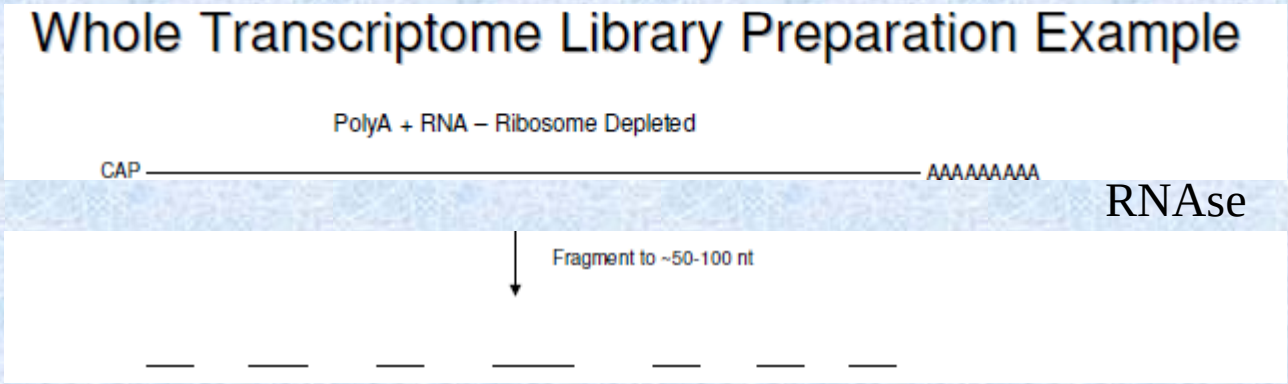
RNASE RNA from cDNA duplexes



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.

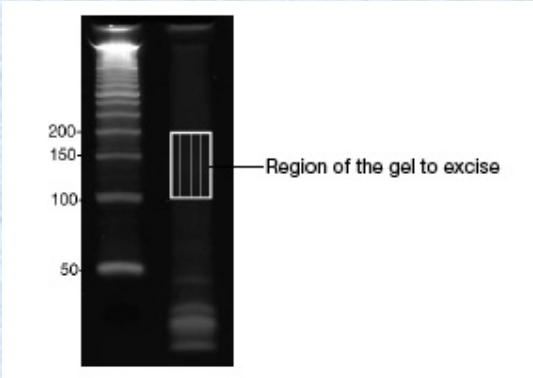


Pico-RNA Chip

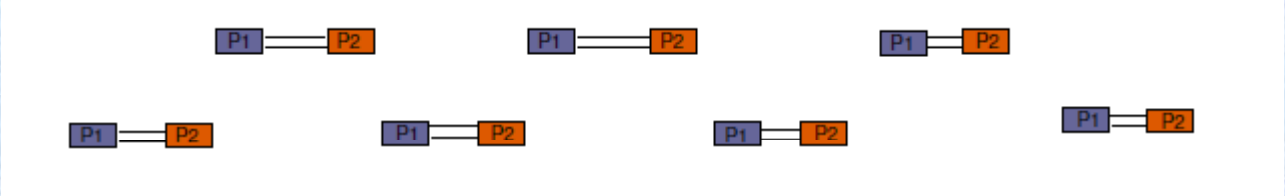
SREK Protocol

Can retain Strandedness

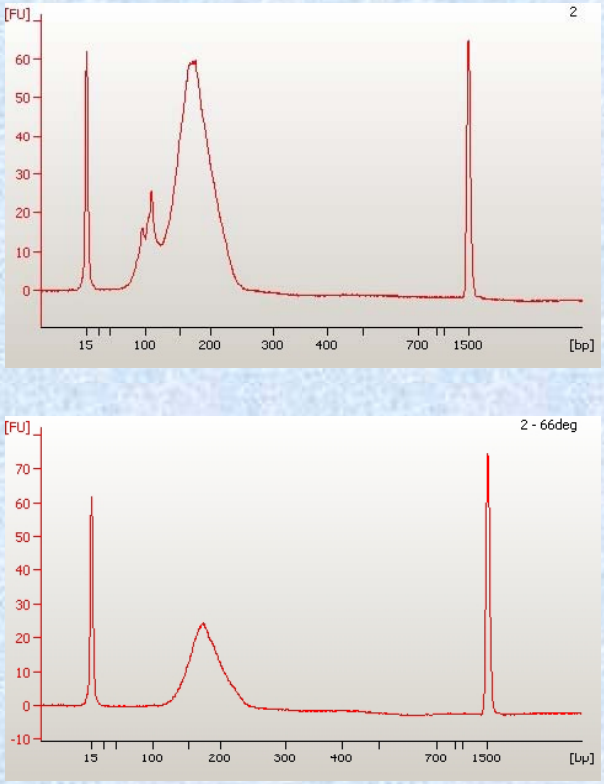
Clean up and size select



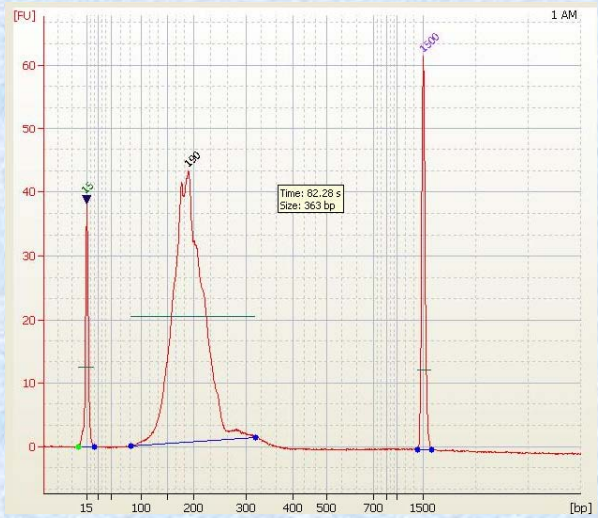
Clean up and amplify

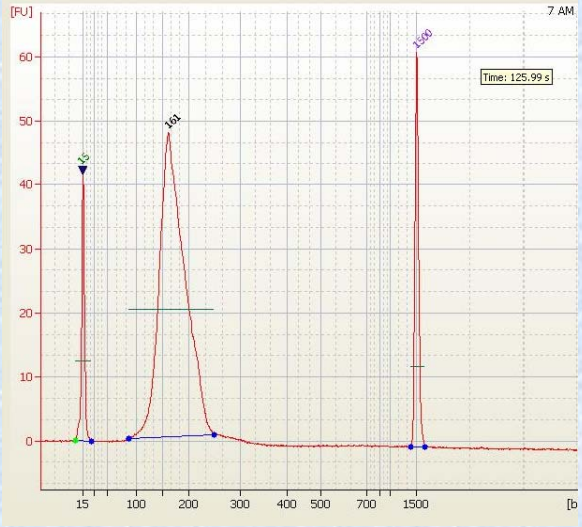


Transcriptome Preparation



cDNA profile after varying temperature


[FU]
1 AM
60
50
40
30
20
10
0
15
190
1500
Time: 82.28 s
Size: 363 bp
15
100
200
300
400
500
700
1500
[bp]


[FU]
7 AM
60
50
40
30
20
10
0
15
191
1500
Time: 125.99 s
15
100
200
300
400
500
700
1500
[b

Clean up and proceed to ePCR

Bead enrichment

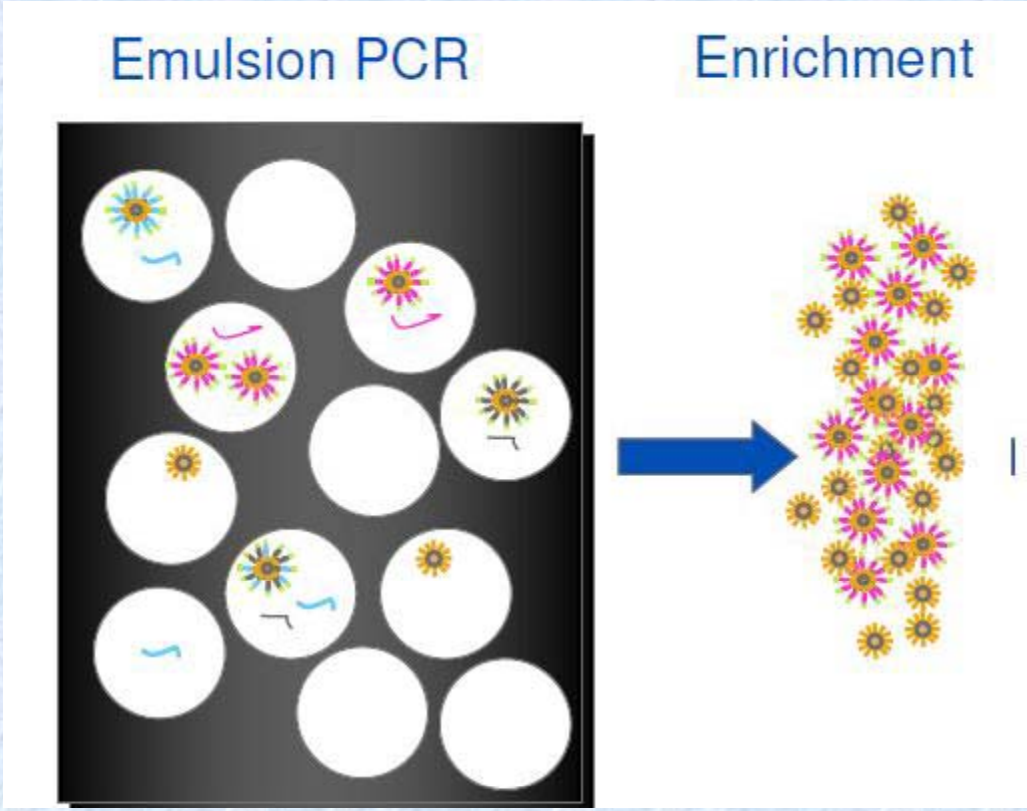


Table 2 Comparison of cDNA-tags assigned to major reference genomes using two sequencing methods

Reference Genomes[a]	Number of genes in reference genome[b]		Number of mRNA-tags assigned[c]		Number of genes covered[d]		Number of genes significantly regulated[e]	
	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'; 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[a]	12,520,192 (54.7%)	10,840,071 (53.0%)	9,974,349 (52.3%)	10,574,300 (52.2%)
Mapped to CDS[b]	743,902 (5.9%)	318,205 (2.9%)	477,362 (4.8%)	1,057,433 (10.0%)
Uniquely mapped to CDS[b]	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

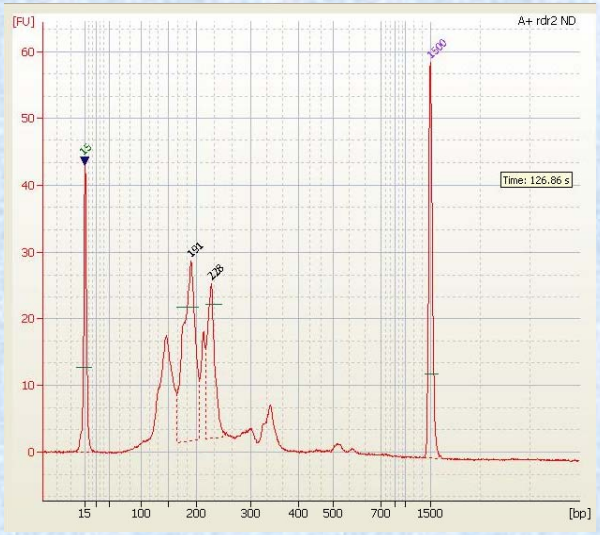
Procedure for uncapped mRNA

Degraded RNA

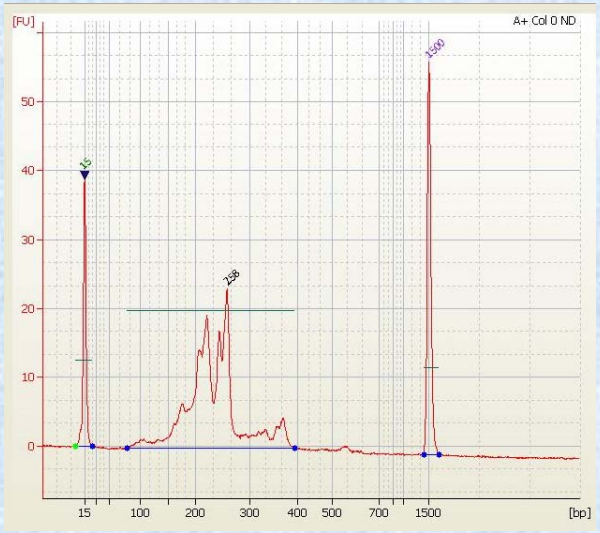
miRNA hairpin processing

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

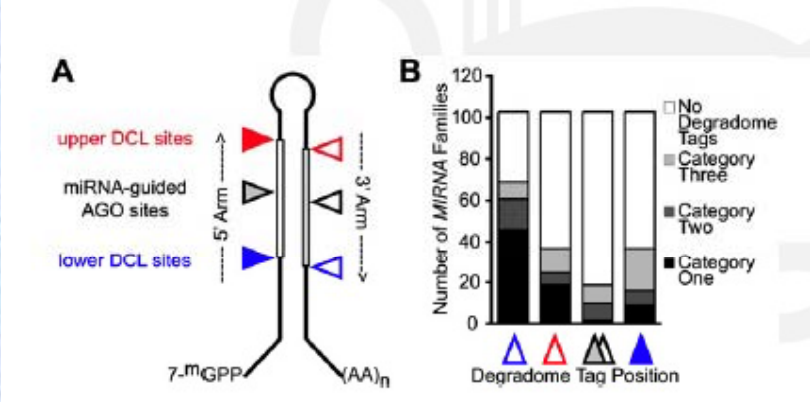
CHARLES ADDO-QUAYE,[1,2] JO ANN SNYDER,[1,3] YONG BUM PARK,[1,3,4] YONG-FANG LI,[5] RAMANJULU SUNKAR,[5] and MICHAEL J. AXTELL[1,3,4]

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

ABSTRACT

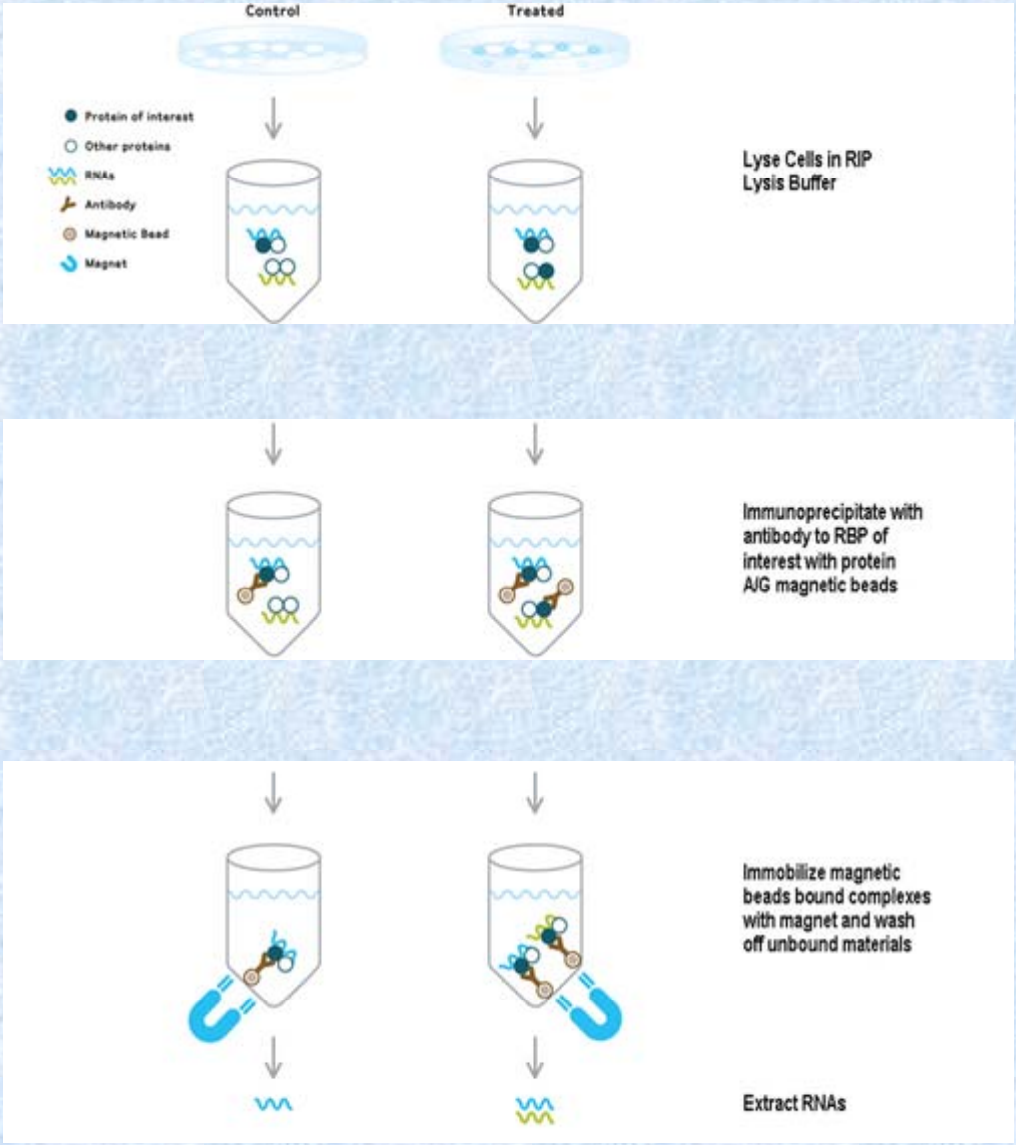
AU1 Expression profiling of the 5′ ends of uncapped mRNAs ("degradome" sequencing) can be used to empirically catalog microRNA (miRNA) targets, to probe patterns of miRNA hairpin processing, to examine mRNA decay, and to analyze accumulation of endogenous short interfering RNA (siRNA) precursors. We 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* (*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

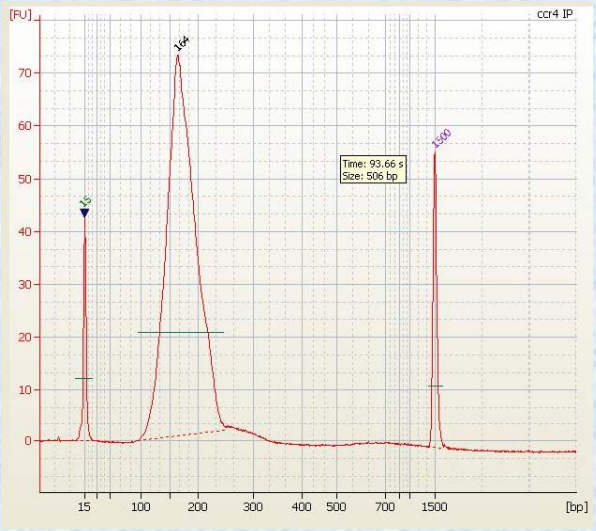
RNAse - ~ 100 bp

Hybridize and Ligate adaptors

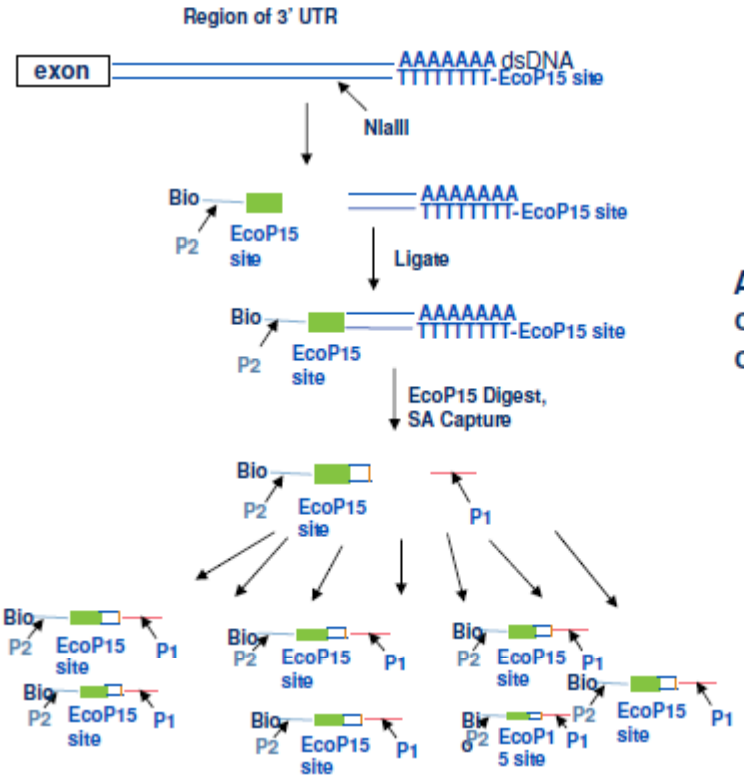
RT to get cDNA

Size-Select 100 - 150 bp

Amplify and quantify

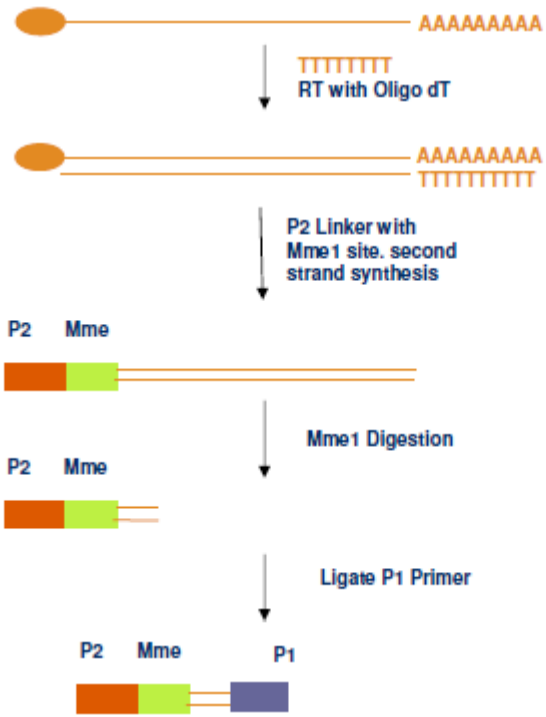

[FU]
ccr4 IP
70
60
50
40
30
20
10
0
15
164
1500
Time: 93.66 s
Size: 506 bp
15
100
200
300
400
500
700
1500
[bp]

3' SAGE Library Preparation Workflow Example



Advantage – No concatenation or bacterial cloning bias

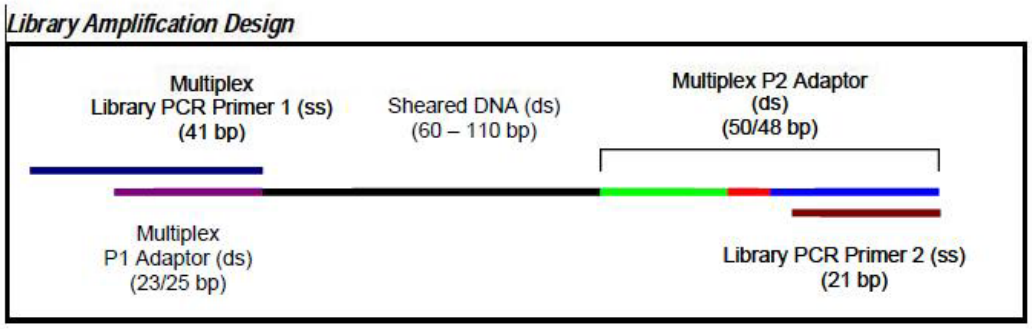
CAGE Library Preparation Workflow Example



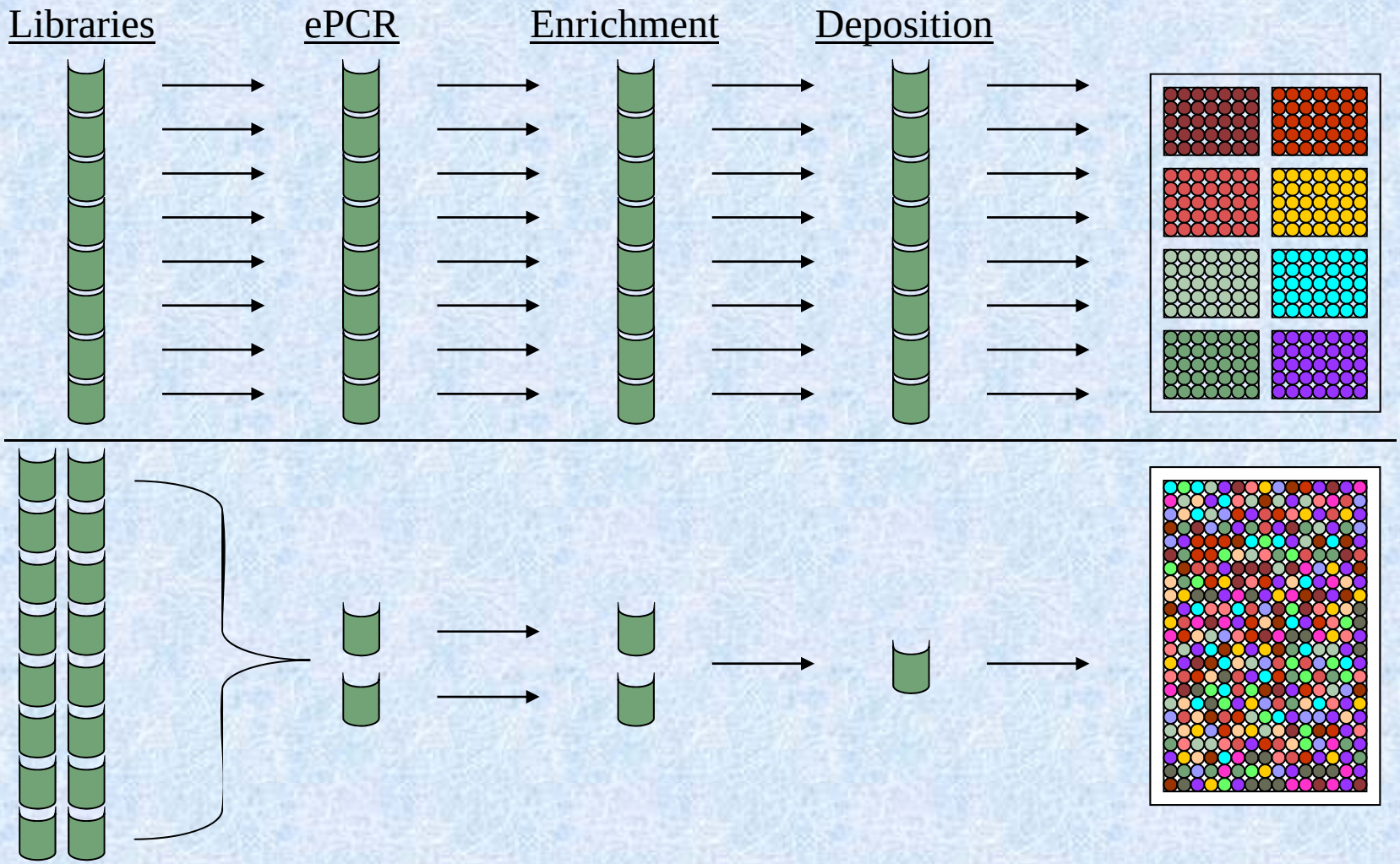
polyA RNA isolated by
CAP trapping

Multiplexing -- Barcodes

Barcode 1 "0032"	5′ CGCCTTGGCCGTACAGCAGGGGCTTAGAGAATGAGGAACCCGGGGCAGTT 3′ Length: 50 3′ GCGGAACCGGCATGTCGTCCCCGAATCTCTTACTCCTTGGGCCCCGTC 5′ Length: 48



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_2	15,512,068
Air (low CO_2)	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
22ºC	29,868,989
30ºC	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

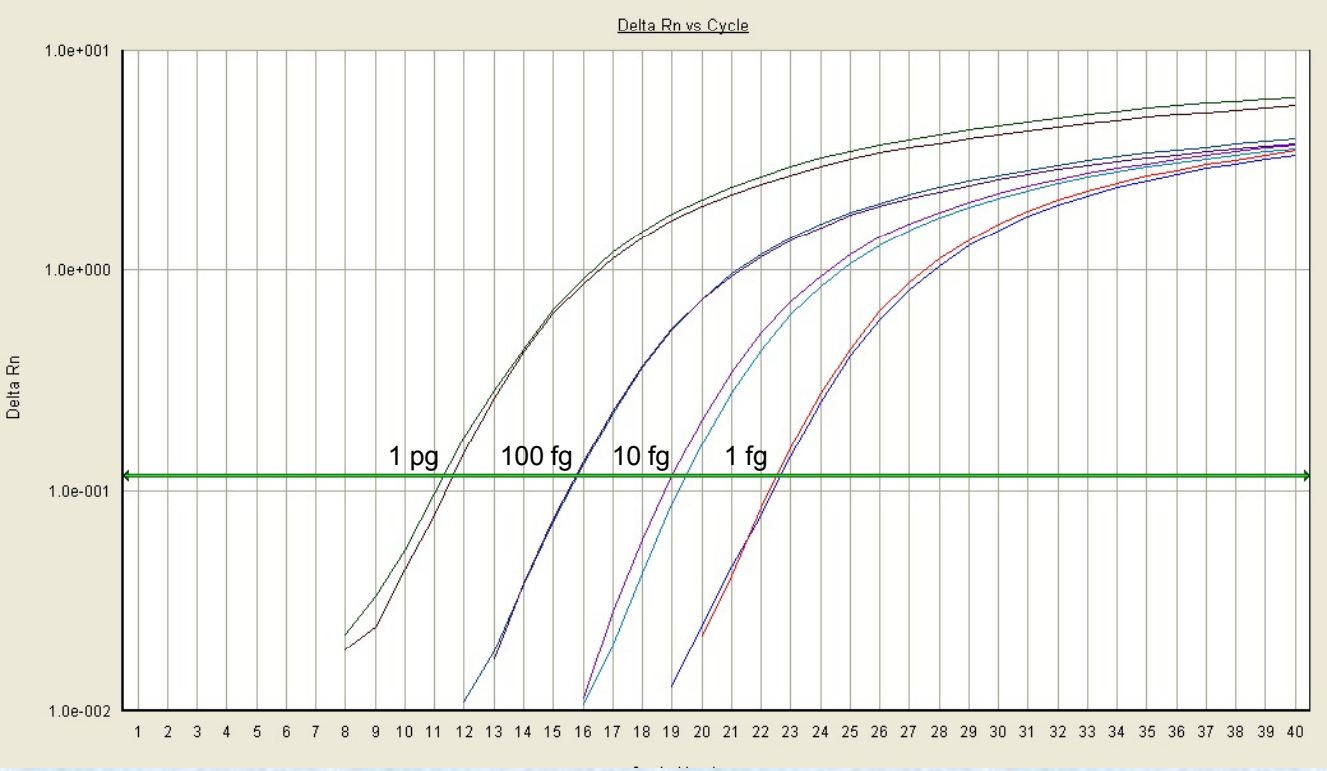
v2.0	150 million mappable reads
v3.0	280 million
v3.5	375 million
v4.0	700 million
HQ	3 x v4.0

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

qPCR

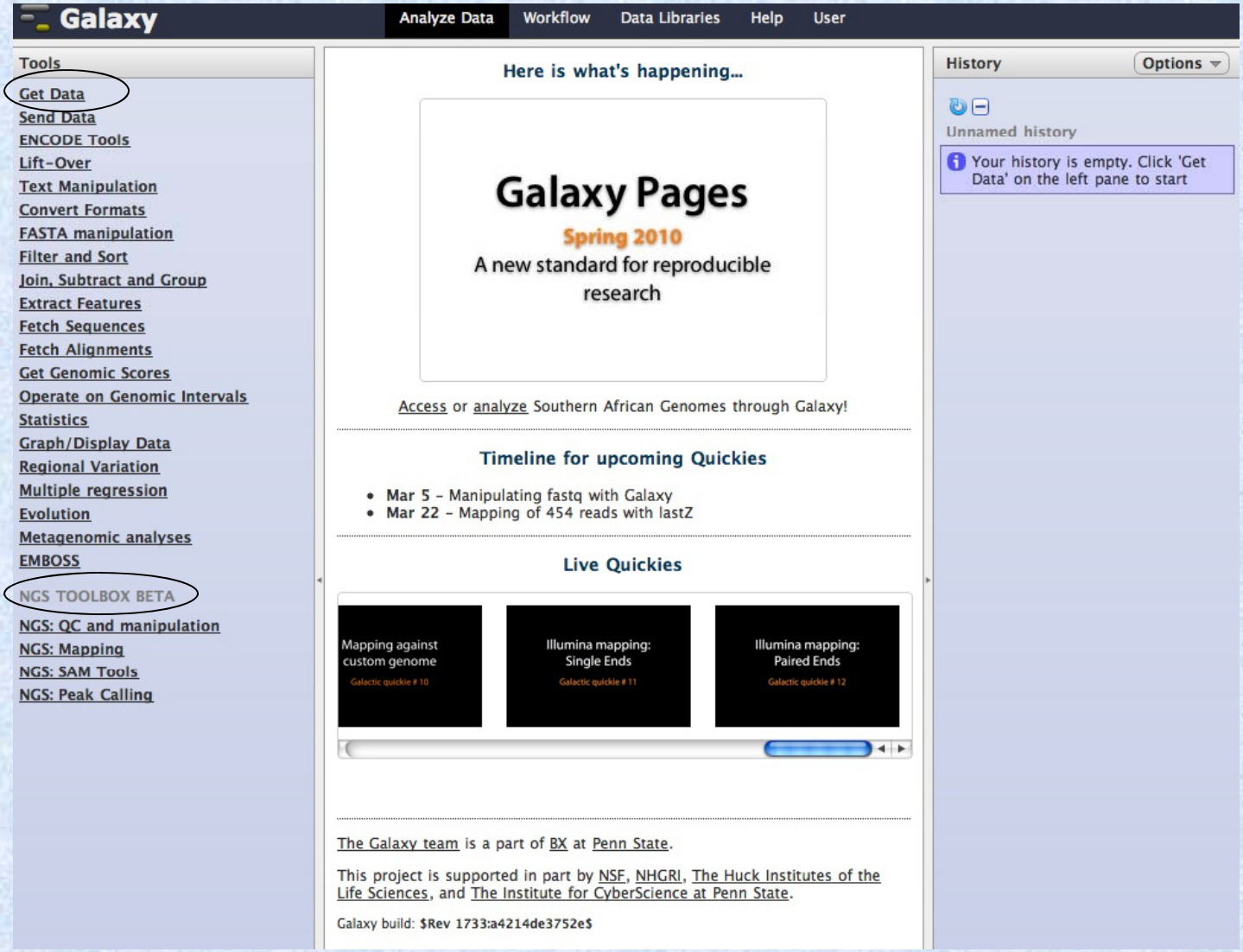
Used to normalize multiplexed samples before ePCR

Taqman Kit only amplifies dsDNA with primer sequences.

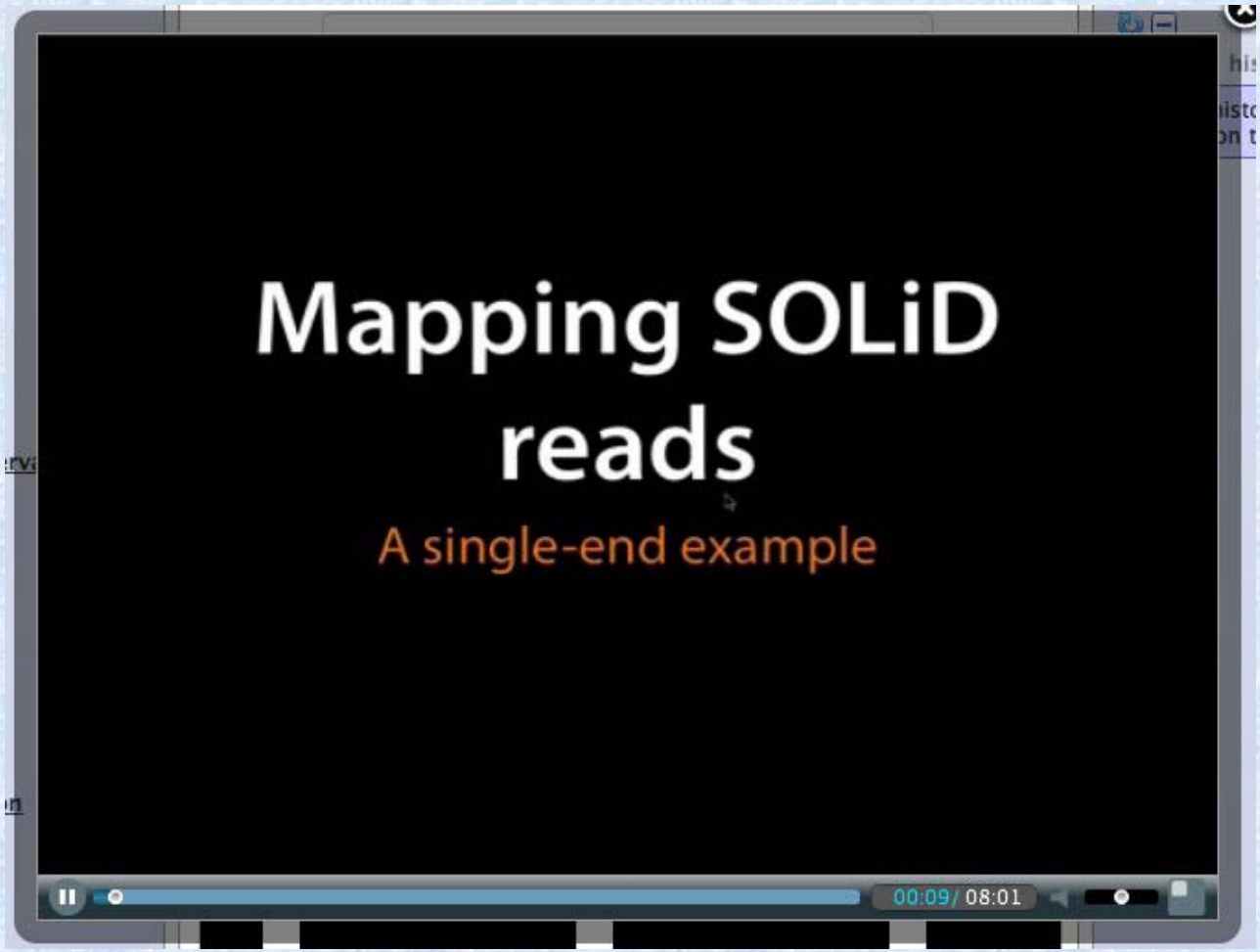


Analysis

- High Level Labs
- Galaxy
- “Plug and Play”
- AB Software Bioscope



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**

5. Fill samples info & **Save**

Thanks to:

The Huck Institutes of the Life Sciences
Lloyd and Dorothy Huck

And the others in the lab…