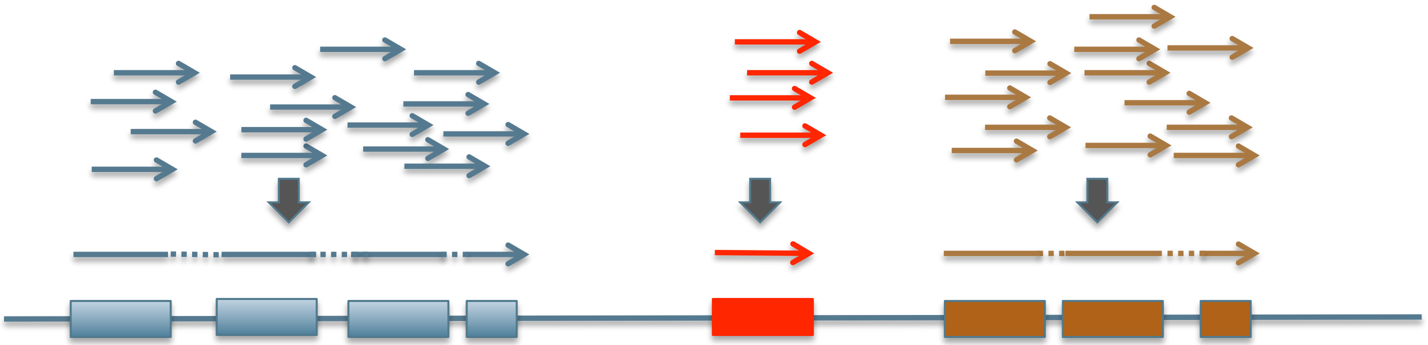
JGI JOINT GENOME INSTITUTE
UNITED STATES DEPARTMENT OF ENERGY

RNA-Seq at JGI

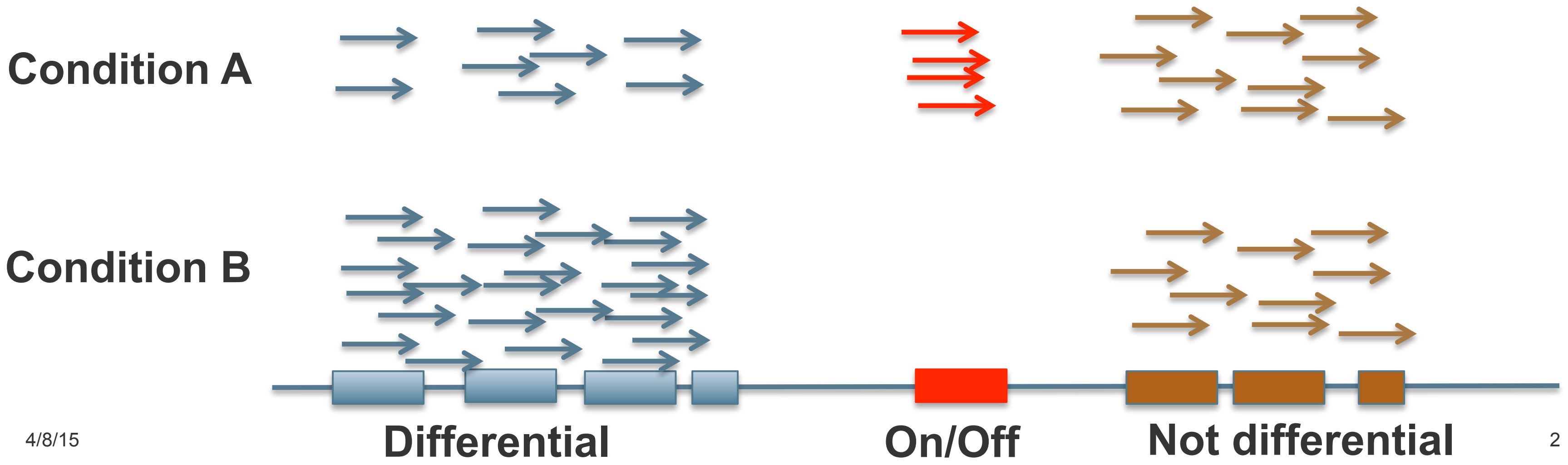
JGI User Meeting 2014
Erika Lindquist
ealindquist@lbl.gov

Why do RNA-Seq?

Genome Annotation



Differential Gene Expression (DGE)



RNA-Seq Projects across Programs

Kerrie Barry

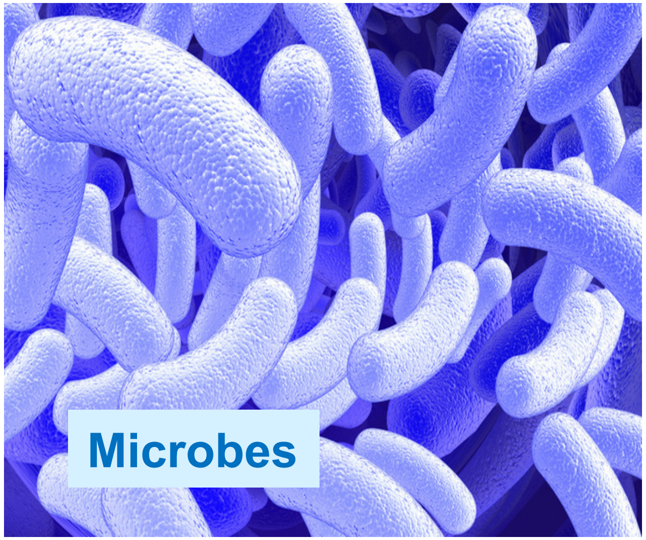
Plants

Fungi

Vivian Ng

Tijana Glavina Del Rio

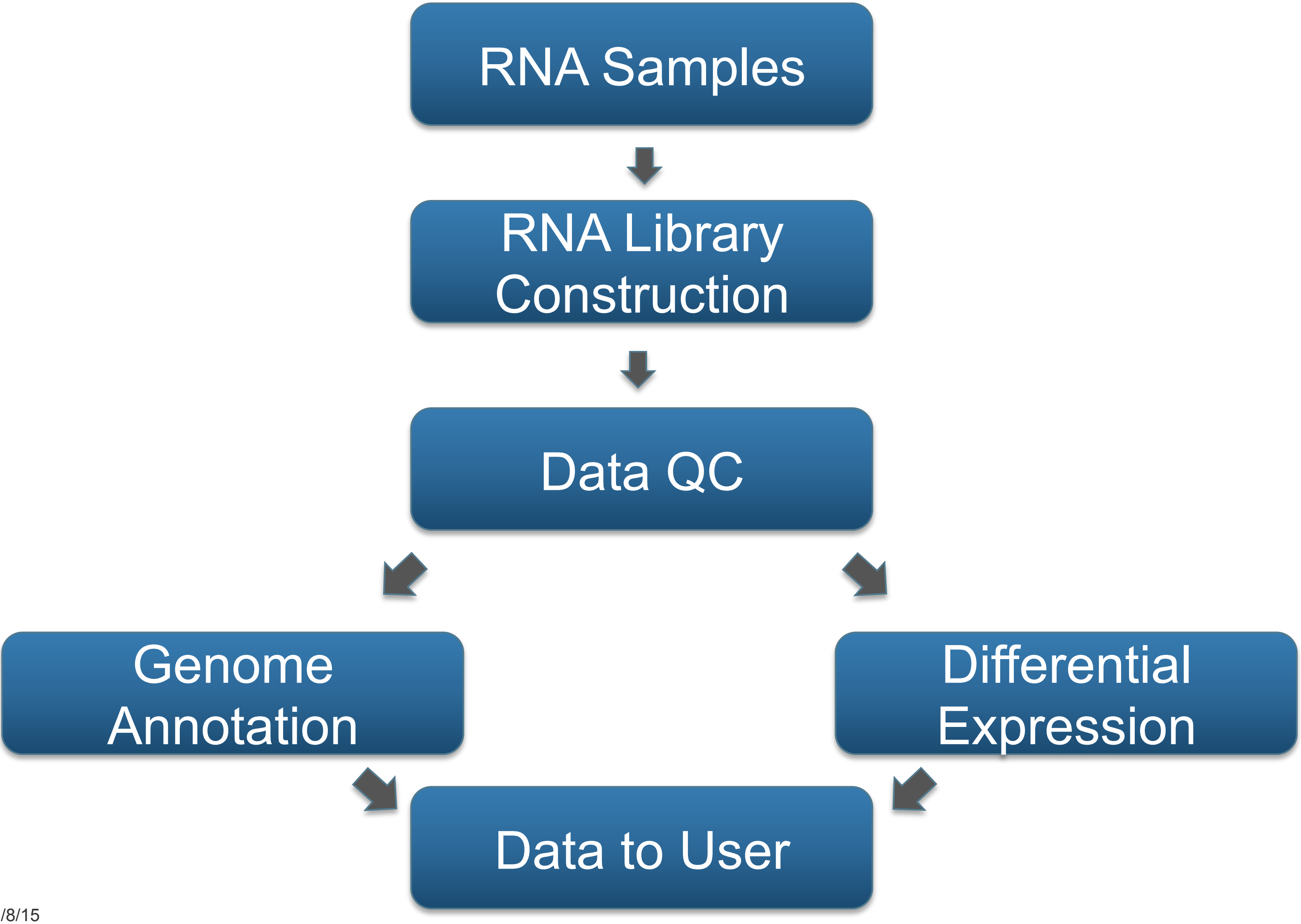
Metagenome



Microbes

Christa Pennacchio

RNA-Seq Outline



RNA Samples

RNA Library Construction

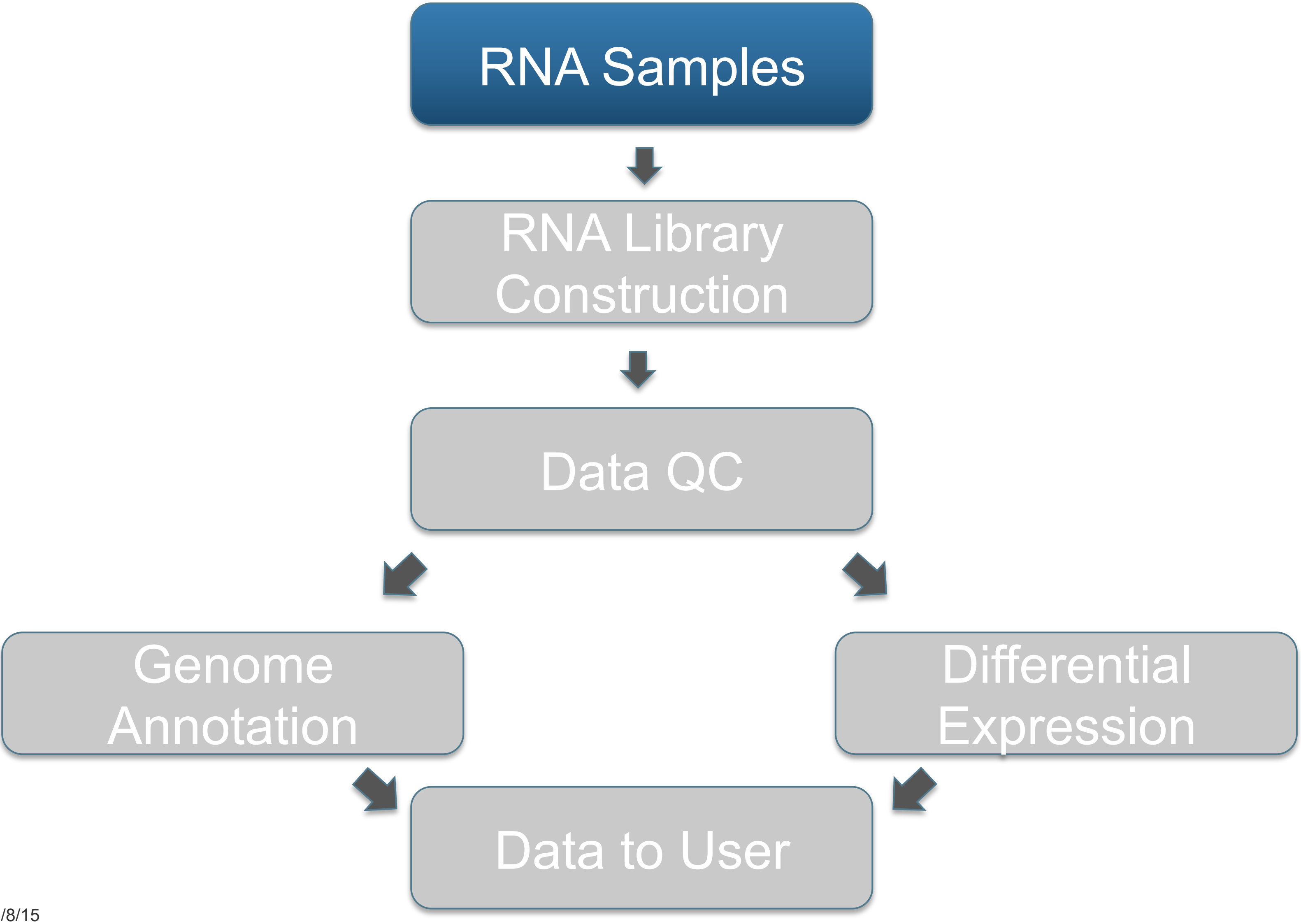
Data QC

Genome Annotation

Differential Expression

Data to User

RNA-Seq Outline



RNA Samples

RNA Library Construction

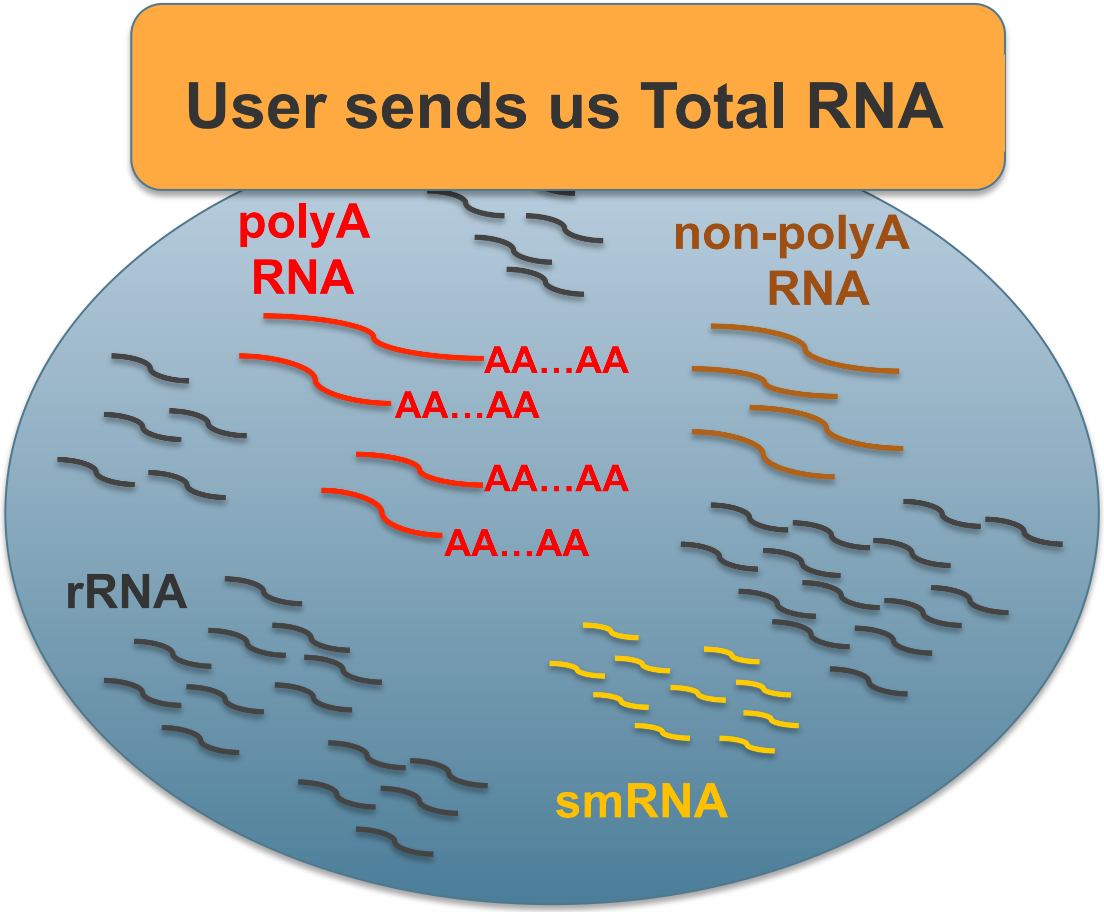
Data QC

Genome Annotation

Differential Expression

Data to User

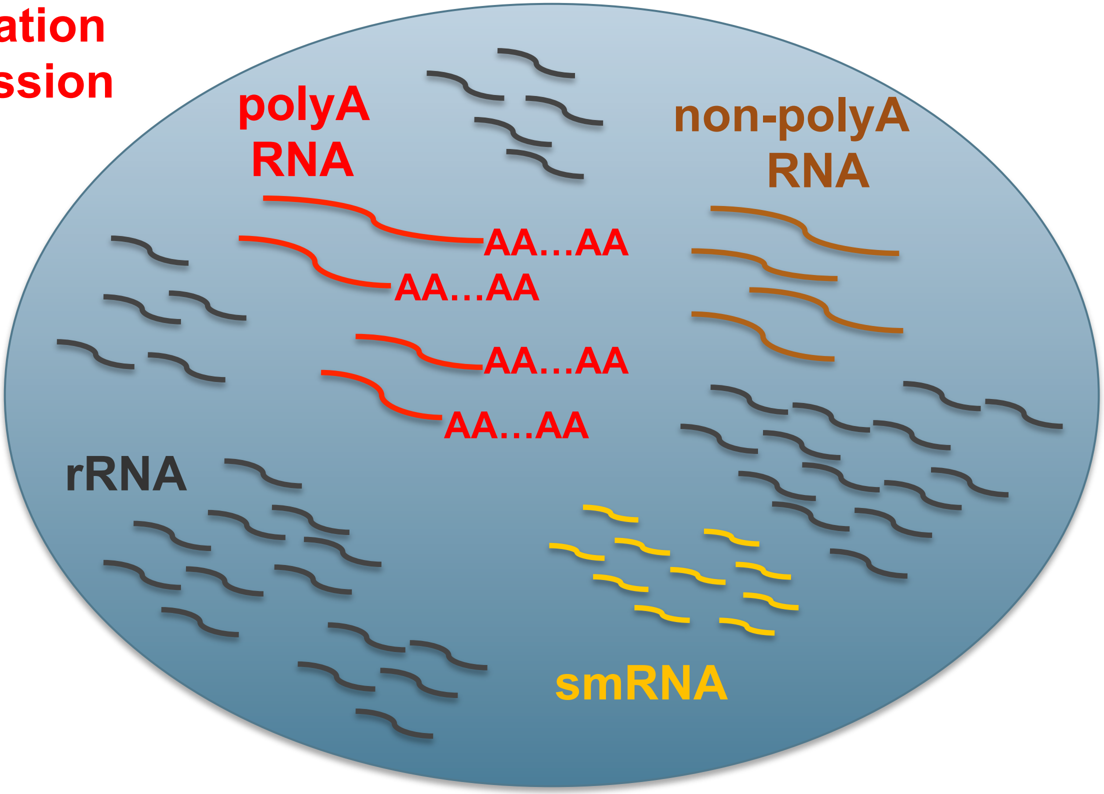
RNA-Seq Types



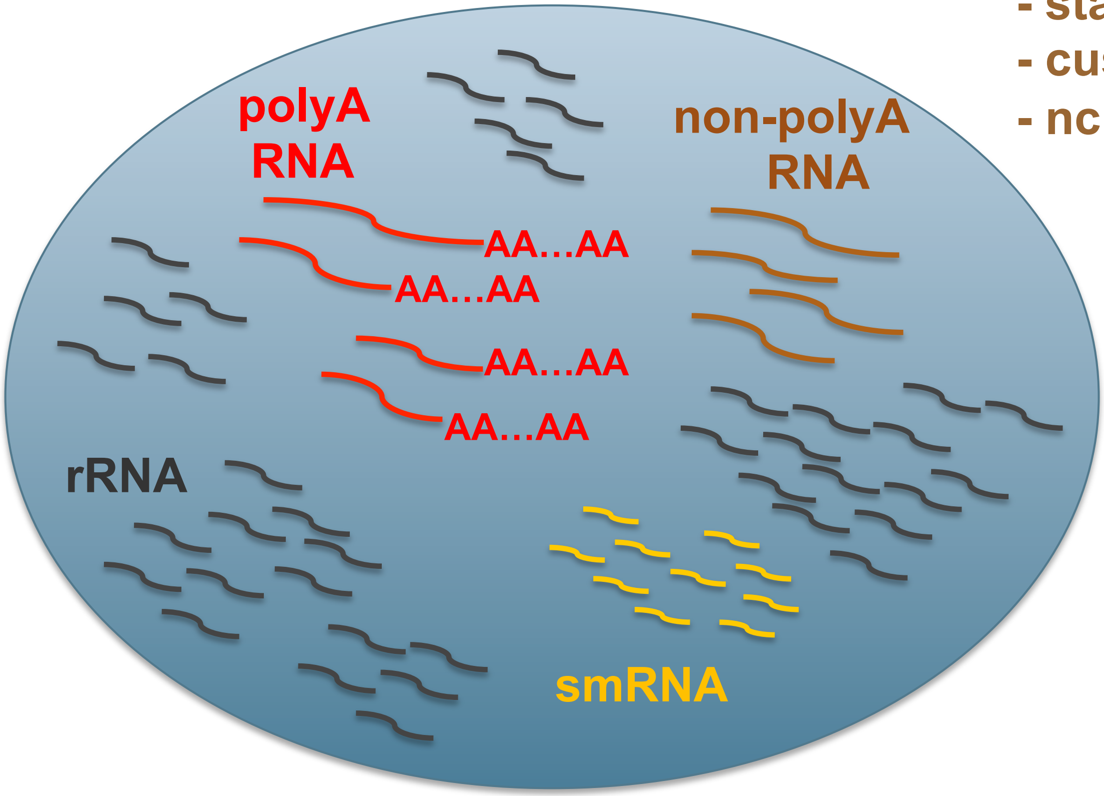
RNA-Seq Types - PolyA

RNA-Seq (polyA):

- capture protein coding genes
- for annotation
- for expression

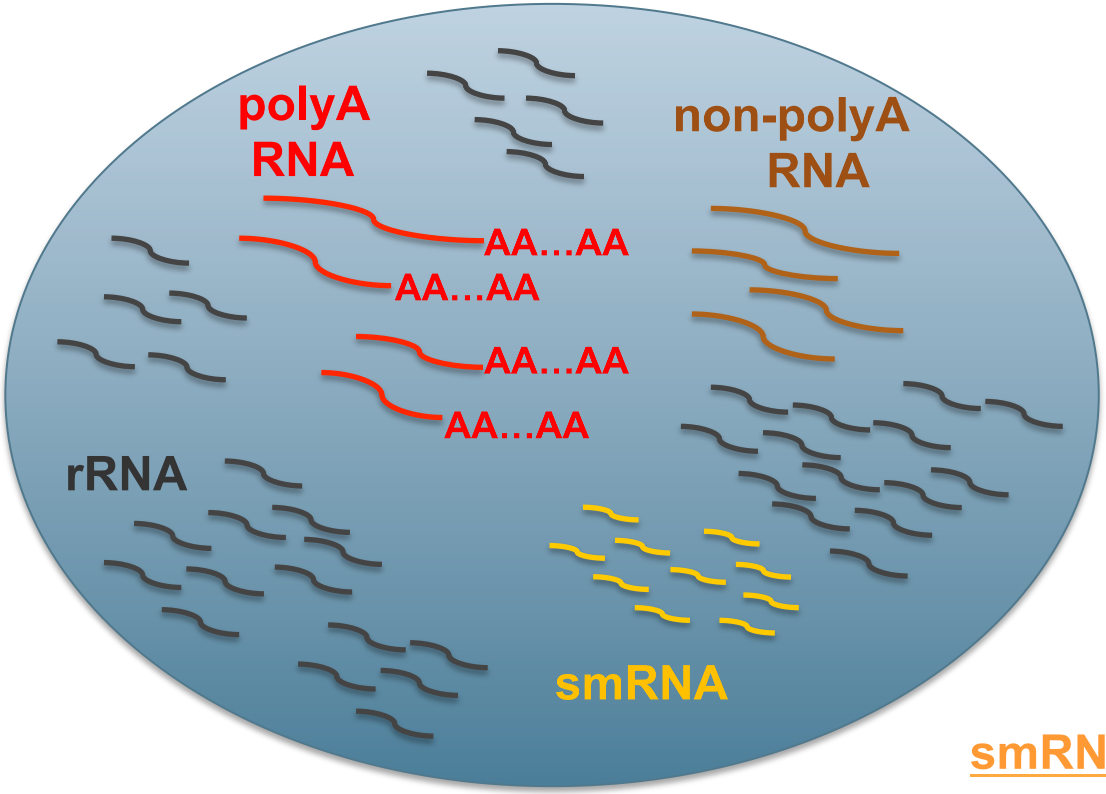


RNA-Seq Types - mRNA



Total RNA-Seq:
(rRNA subtraction)
- standard for Proks
- custom probes for Euks
- ncRNA

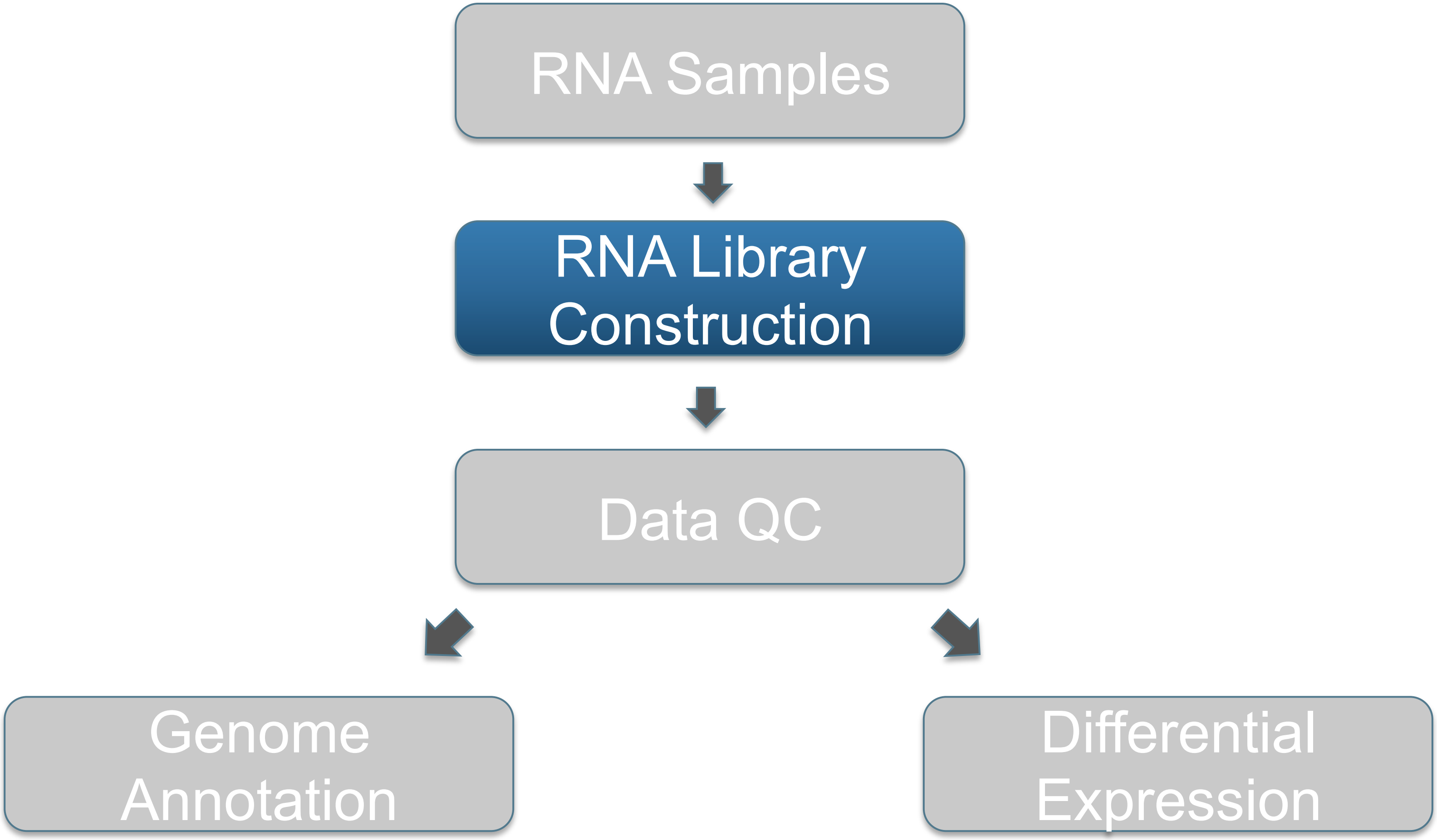
RNA-Seq Types – small RNA



smRNA-Seq:

- small RNA
- gene regulation
- Prok targets 50-150bp
- Euk targets 20-40 bp

RNA-Seq Outline



RNA Samples

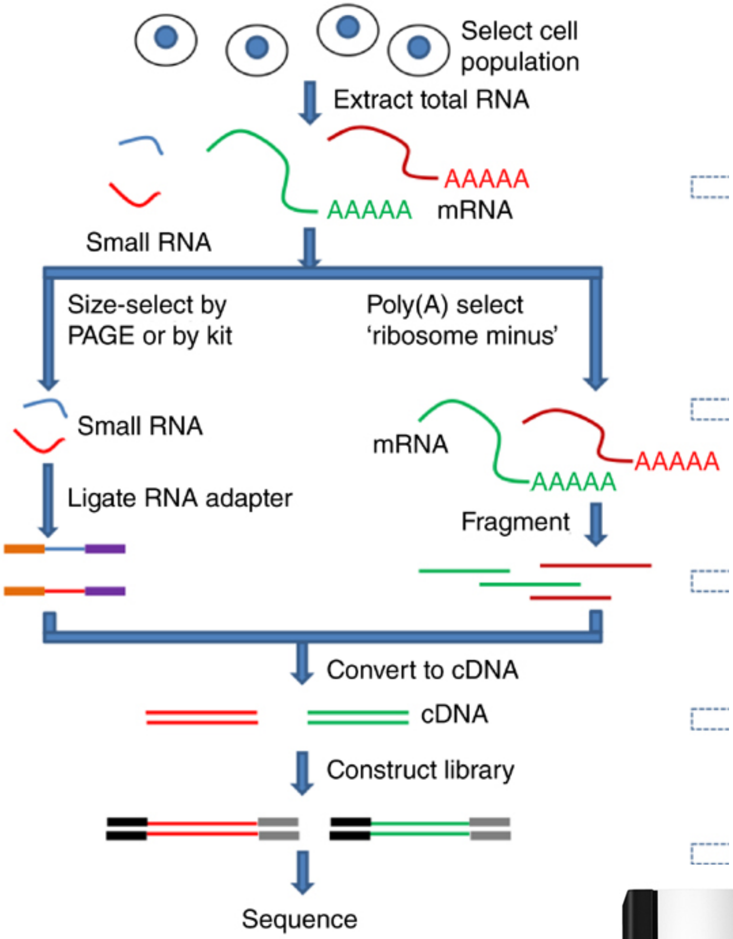
RNA Library Construction

Data QC

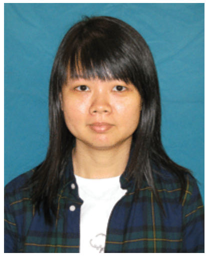
Genome Annotation

Differential Expression

RNA-Seq Library Construction



Quantity of Total RNA	
polyA RNA	1ug
rRNA subtraction	1ug
smRNA	5ug
Low input	100ng



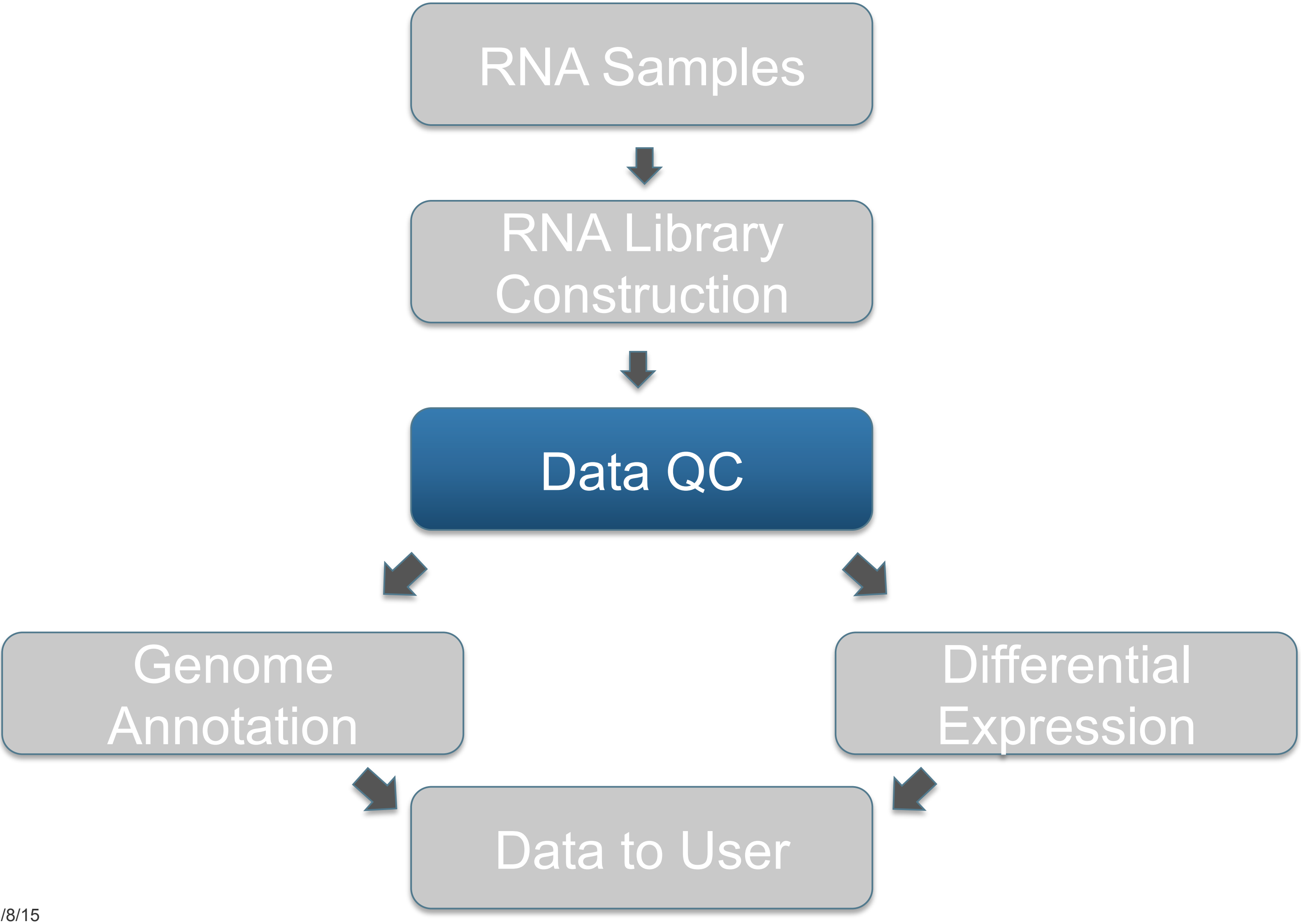
Chew-Yee Ngan

100M PE reads per Annotation Sample
20M PE or SE reads per DGE sample

Chris Daum

Zeng W, Mortazavi A. (2012) **Technical considerations for functional sequencing assays.** *Nat Immunol* 13(9), 802-7.

RNA-Seq Outline



RNA Samples

RNA Library Construction

Data QC

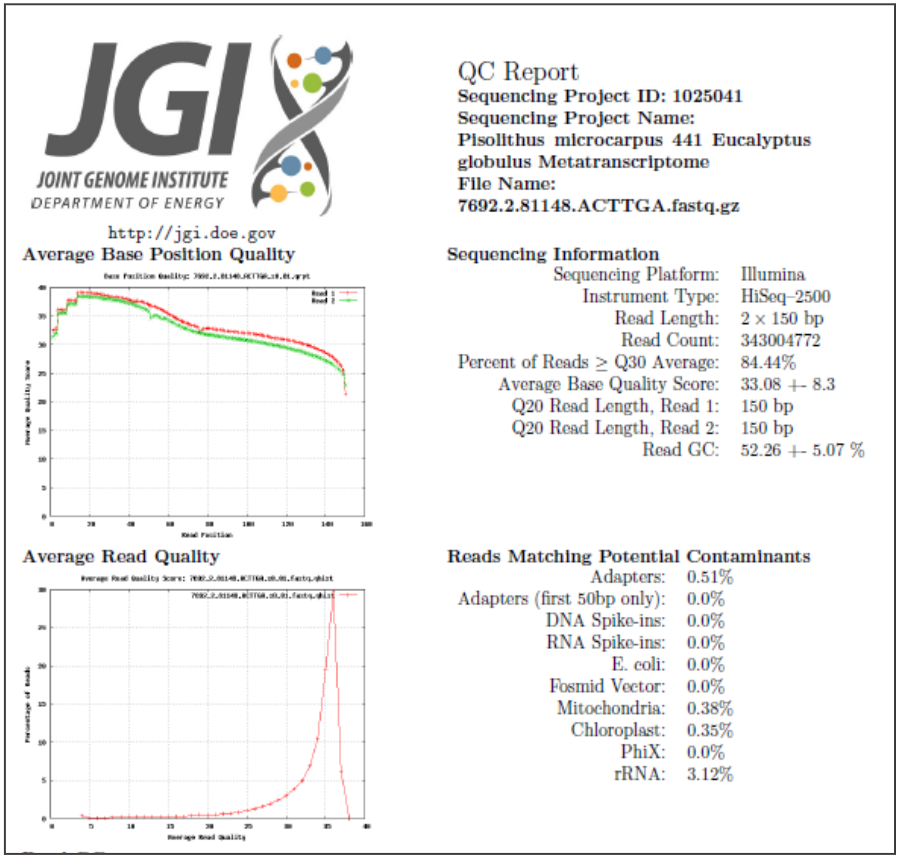
Genome Annotation

Differential Expression

Data to User

RNA-Seq Library QC

QC Report



JOINT GENOME INSTITUTE
DEPARTMENT OF ENERGY

http://jgi.doe.gov

Average Base Position Quality

Average Read Quality

QC Report
Sequencing Project ID: 1025041
Sequencing Project Name:
Pisolithus microcarpus 441 Eucalyptus globulus Metatranscriptome
File Name:
7692.2.81148.ACTTGA.fastq.gz

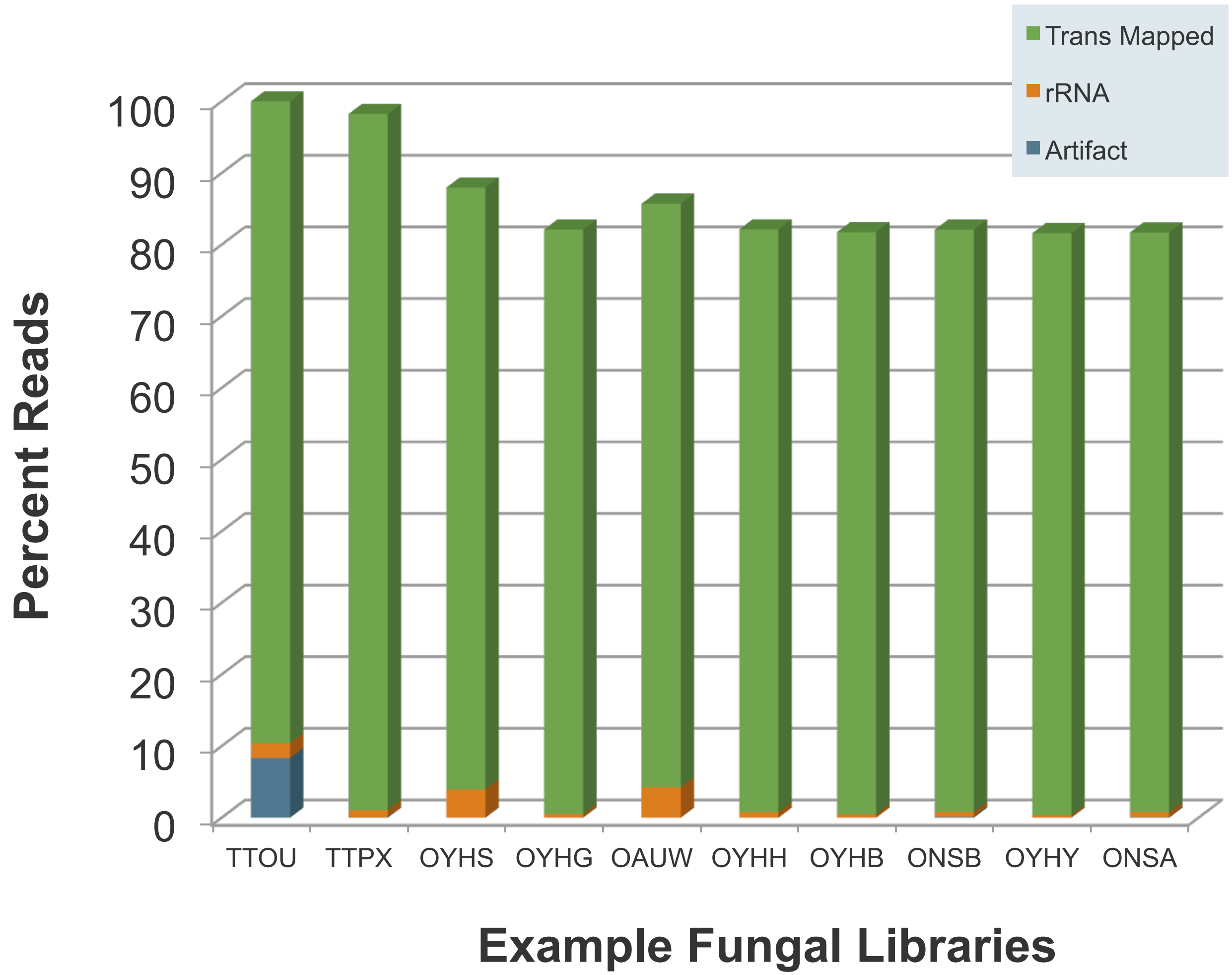
Sequencing Information

Sequencing Platform:	Illumina
Instrument Type:	HiSeq–2500
Read Length:	2 × 150 bp
Read Count:	343004772
Percent of Reads ≥ Q30 Average:	84.44%
Average Base Quality Score:	33.08 +- 8.3
Q20 Read Length, Read 1:	150 bp
Q20 Read Length, Read 2:	150 bp
Read GC:	52.26 +- 5.07 %

Reads Matching Potential Contaminants

Adapters:	0.51%
Adapters (first 50bp only):	0.0%
DNA Spike-ins:	0.0%
RNA Spike-ins:	0.0%
E. coli:	0.0%
Fosmid Vector:	0.0%
Mitochondria:	0.38%
Chloroplast:	0.35%
PhiX:	0.0%
rRNA:	3.12%

RNA-Seq Library QC – Usable reads

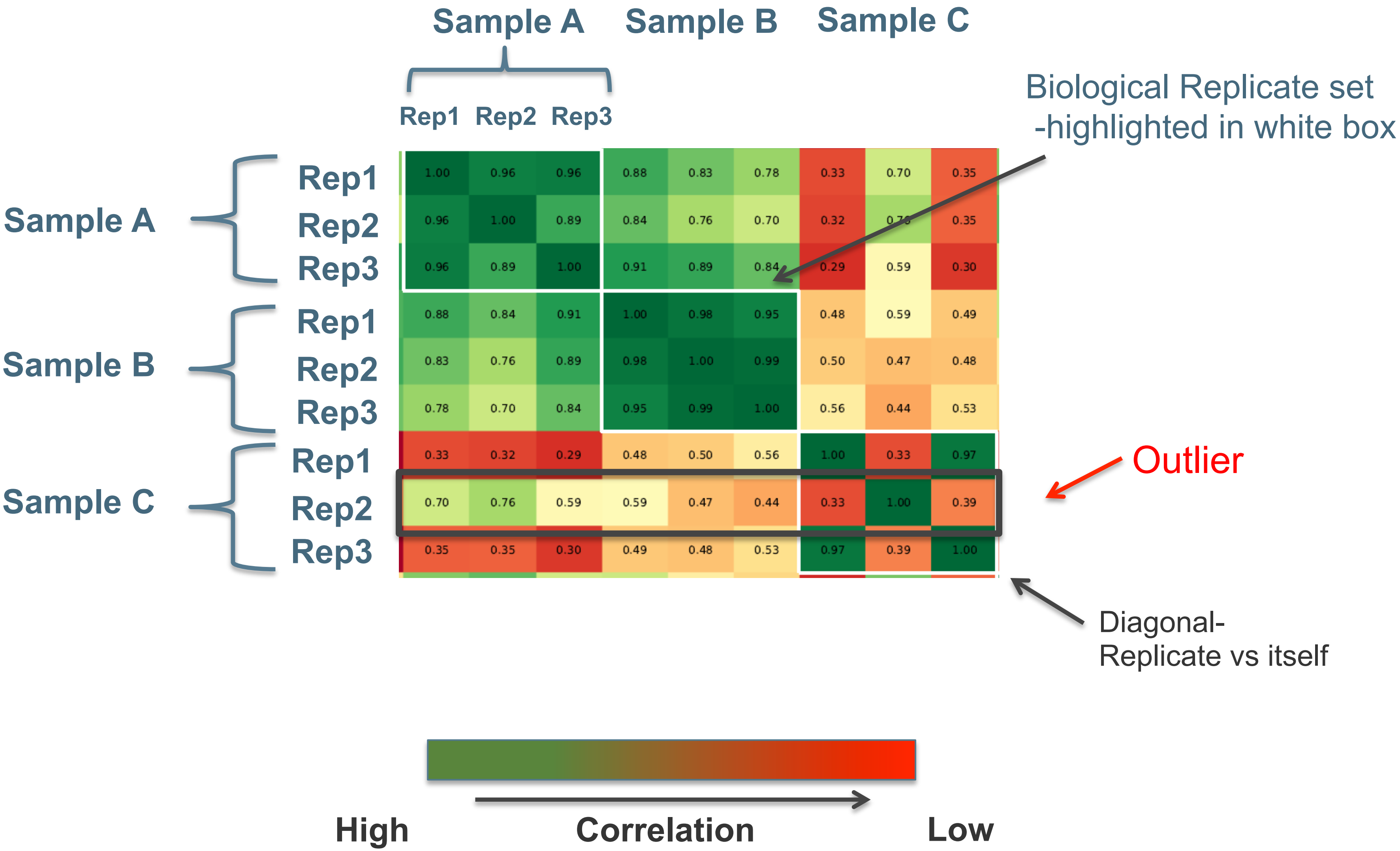


Mapped >80%

Non-usable <5%

Anna
Lipzen

Are replicates correlated?



		Sample A Rep1	Sample A Rep2	Sample A Rep3	Sample B Rep1	Sample B Rep2	Sample B Rep3	Sample C Rep1	Sample C Rep2	Sample C Rep3
Sample A	Rep1	1.00	0.96	0.96	0.88	0.83	0.78	0.33	0.70	0.35
	Rep2	0.96	1.00	0.89	0.84	0.76	0.70	0.32	0.76	0.35
	Rep3	0.96	0.89	1.00	0.91	0.89	0.84	0.29	0.59	0.30
Sample B	Rep1	0.88	0.84	0.91	1.00	0.98	0.95	0.48	0.59	0.49
	Rep2	0.83	0.76	0.89	0.98	1.00	0.99	0.50	0.47	0.48
	Rep3	0.78	0.70	0.84	0.95	0.99	1.00	0.56	0.44	0.53
Sample C	Rep1	0.33	0.32	0.29	0.48	0.50	0.56	1.00	0.33	0.97
	Rep2	0.70	0.76	0.59	0.59	0.47	0.44	0.33	1.00	0.39
	Rep3	0.35	0.35	0.30	0.49	0.48	0.53	0.97	0.39	1.00

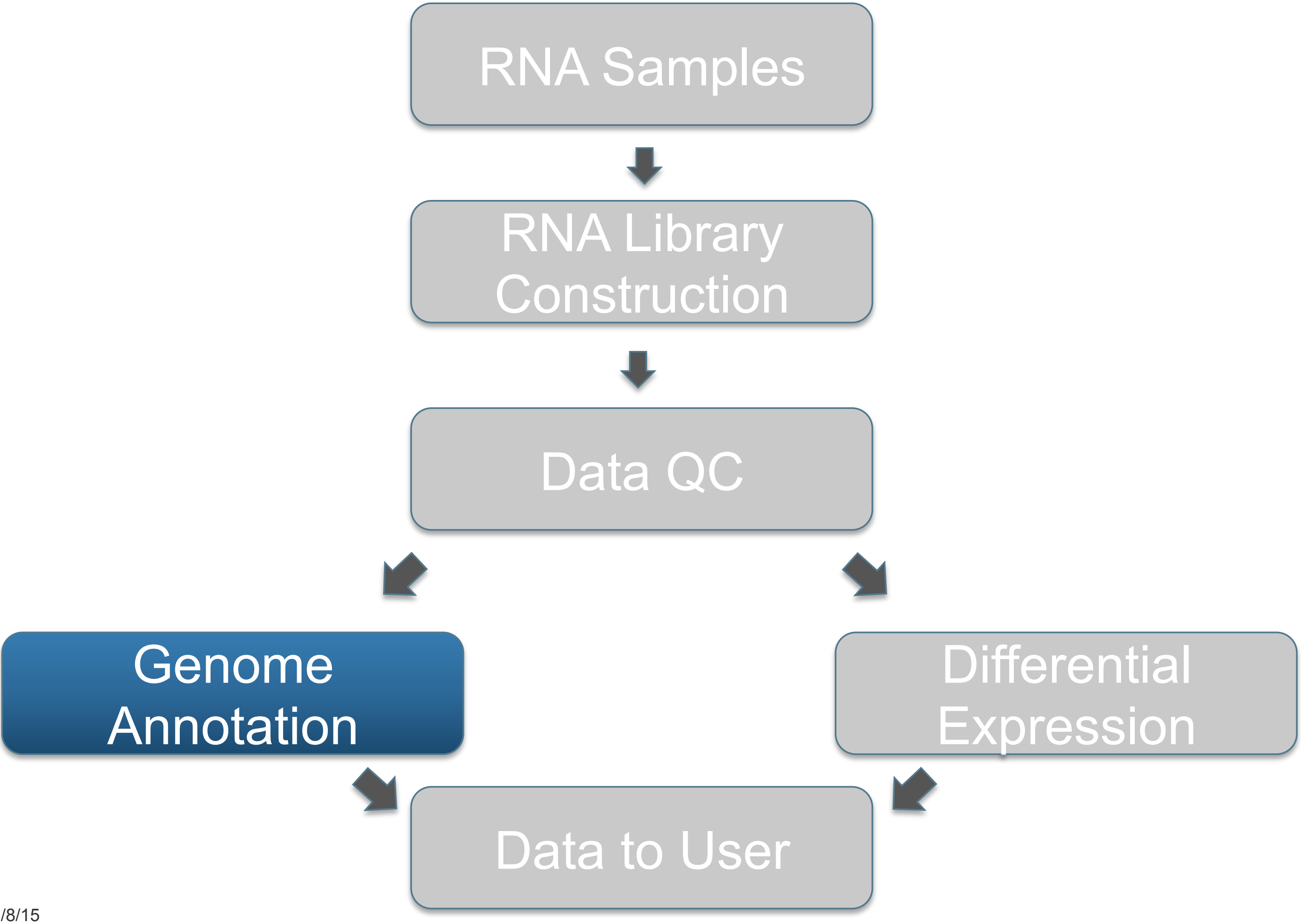
Biological Replicate set
-highlighted in white box

Outlier

Diagonal-
Replicate vs itself

High **Correlation** **Low**

RNA-Seq Outline



RNA Samples

RNA Library Construction

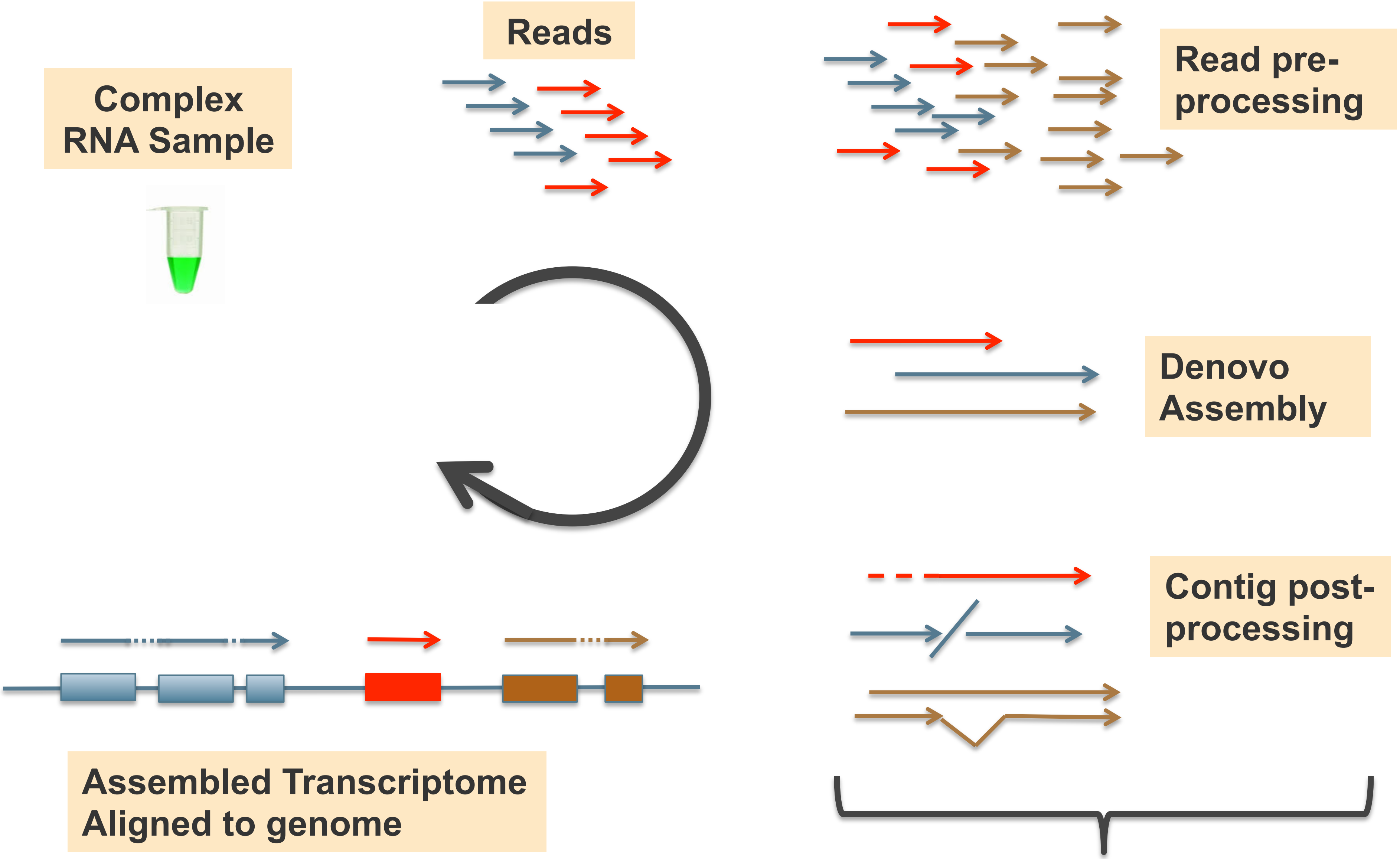
Data QC

Genome Annotation

Differential Expression

Data to User

RNA-Seq for Genome Annotation



Complex RNA Sample

Reads

Read pre-processing

Denovo Assembly

Contig post-processing

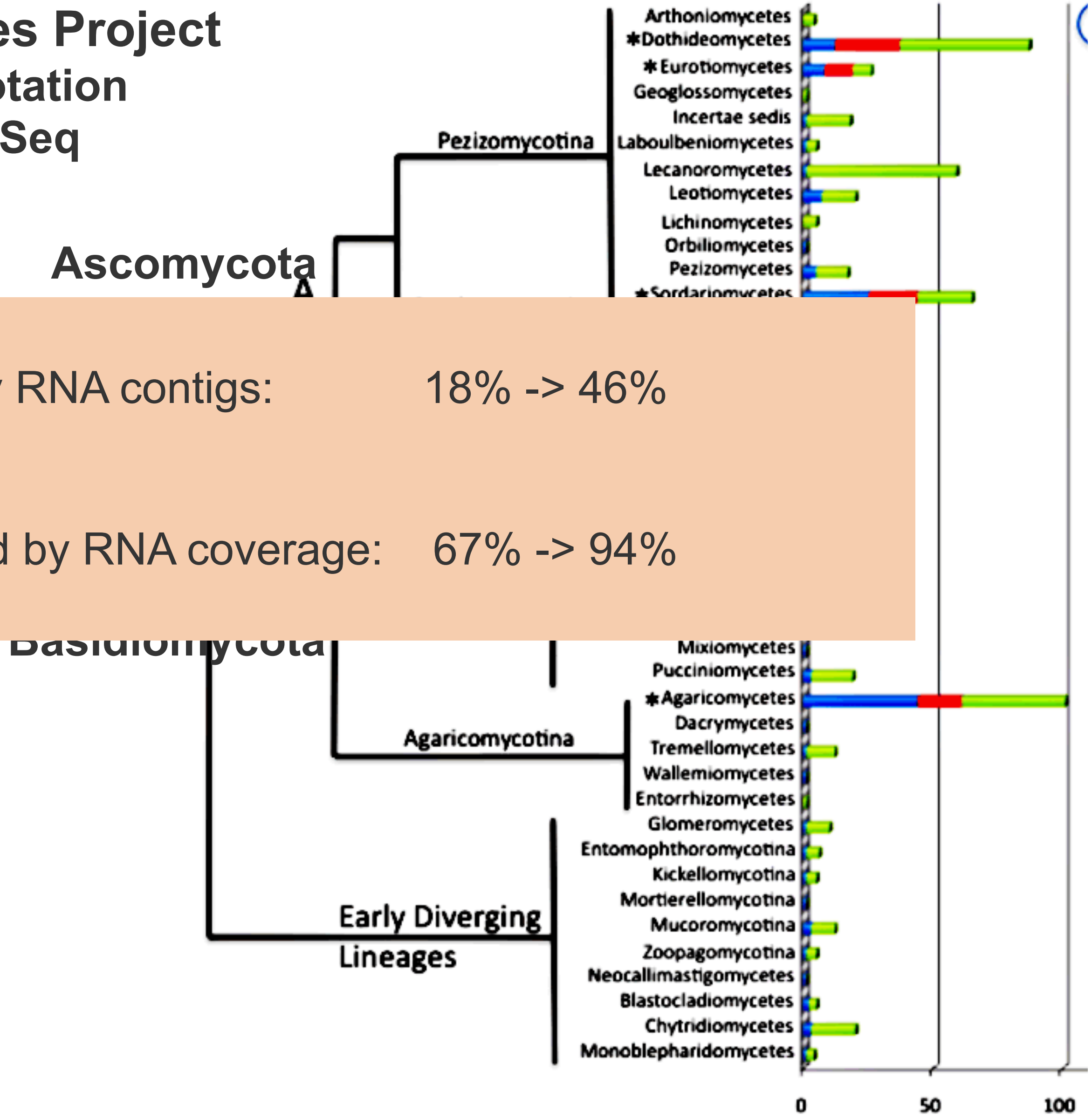
Assembled Transcriptome Aligned to genome

Rnnotator; Martin etal. BMC Genomics 2010,11:663

How has RNA-Seq helped Annotation?

1000 Fungal Genomes Project

- All JGI Fungal annotation projects have RNA-Seq support

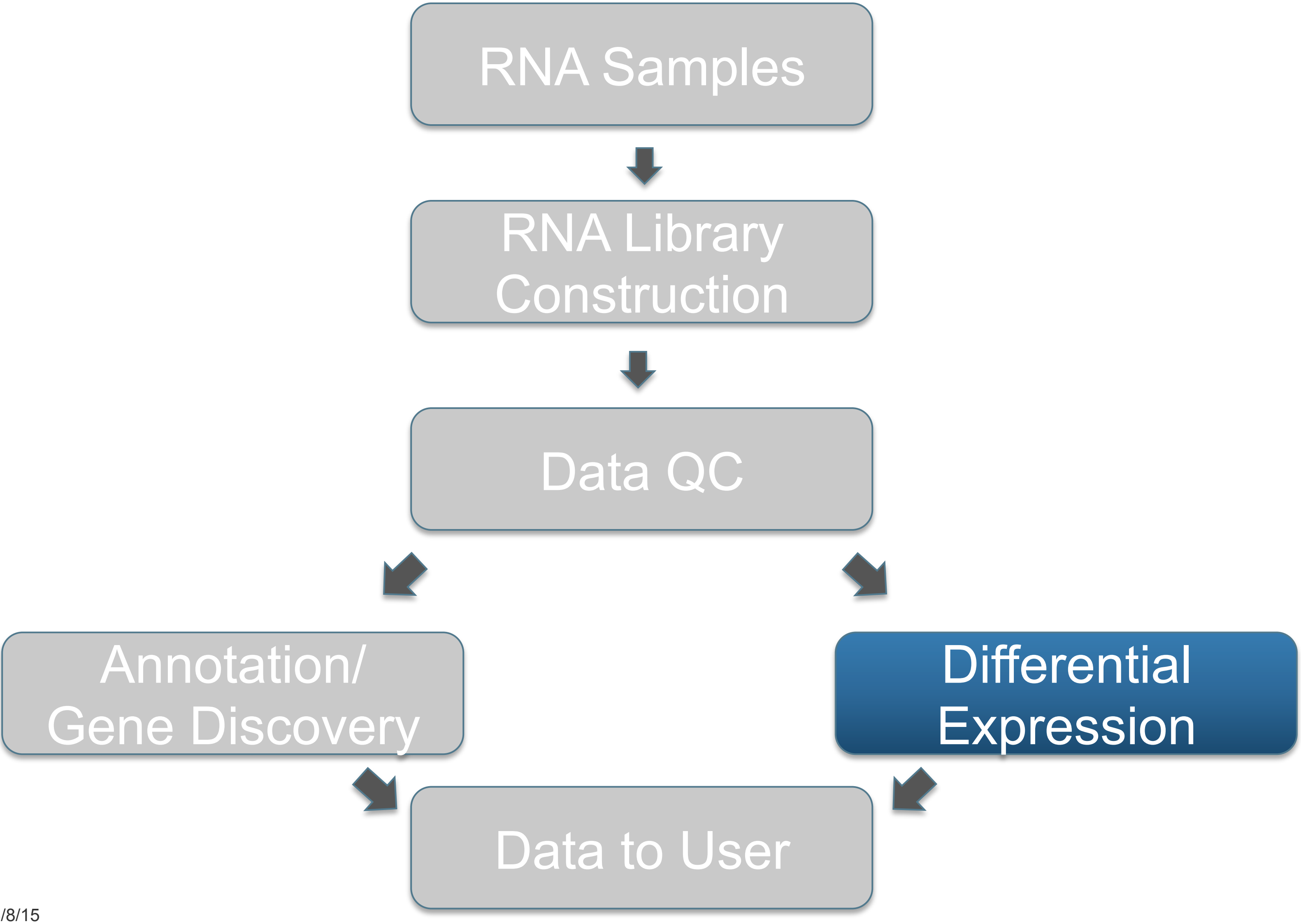


Genes found by RNA contigs: 18% -> 46%

Genes validated by RNA coverage: 67% -> 94%

🟦 Complete/in progress
🟥 Proposed
🟩 Unsampled families

RNA-Seq Outline



RNA Samples

RNA Library Construction

Data QC

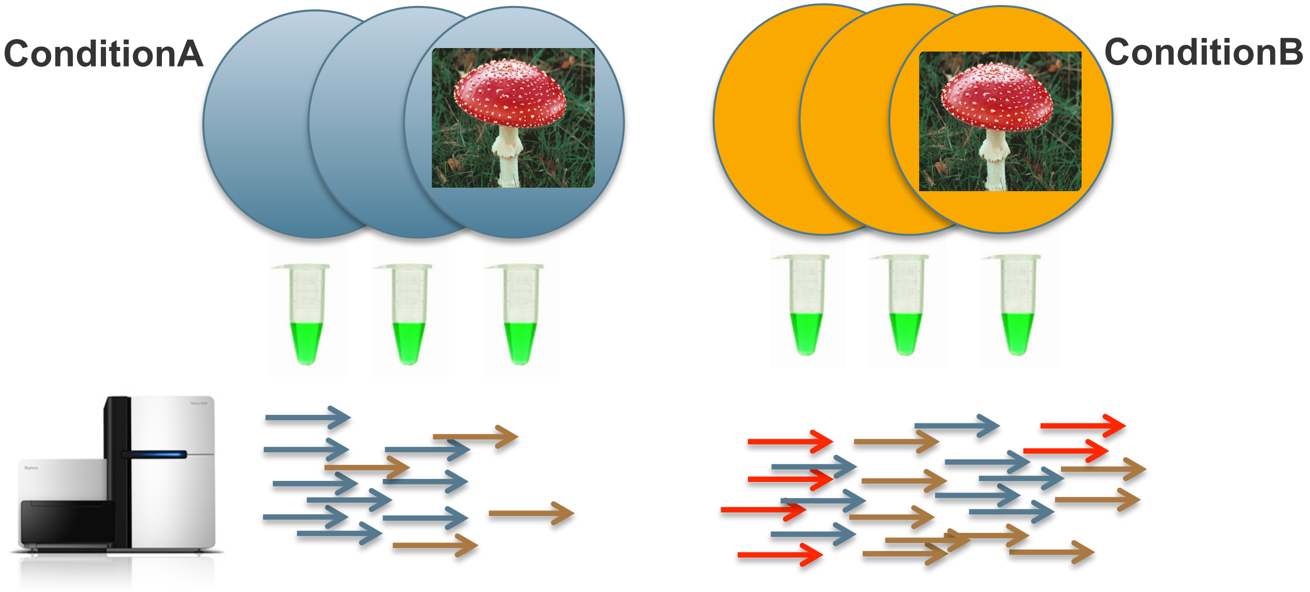
Annotation/ Gene Discovery

Differential Expression

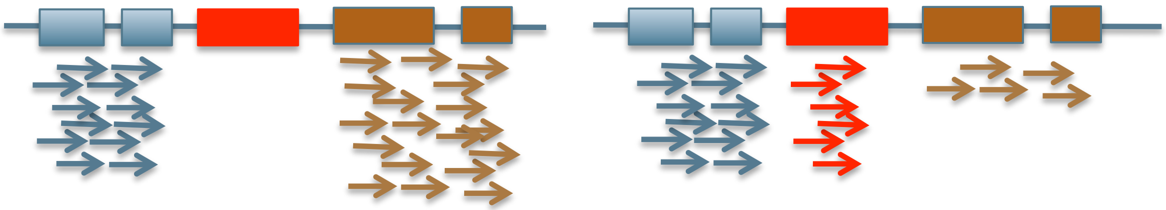
Data to User

RNA – Differential Gene Expression

BIOLOGICAL REPLICATES!



Align reads to genome = HISAT



Read count = featureCounts

Normalize/Diff Exp = DESeq2

See Vasanth Singan's Poster #36

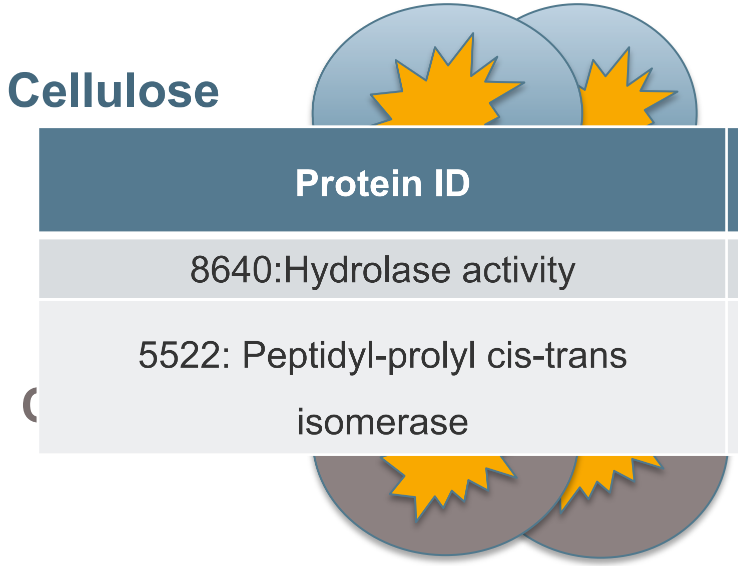
Example Expression Project

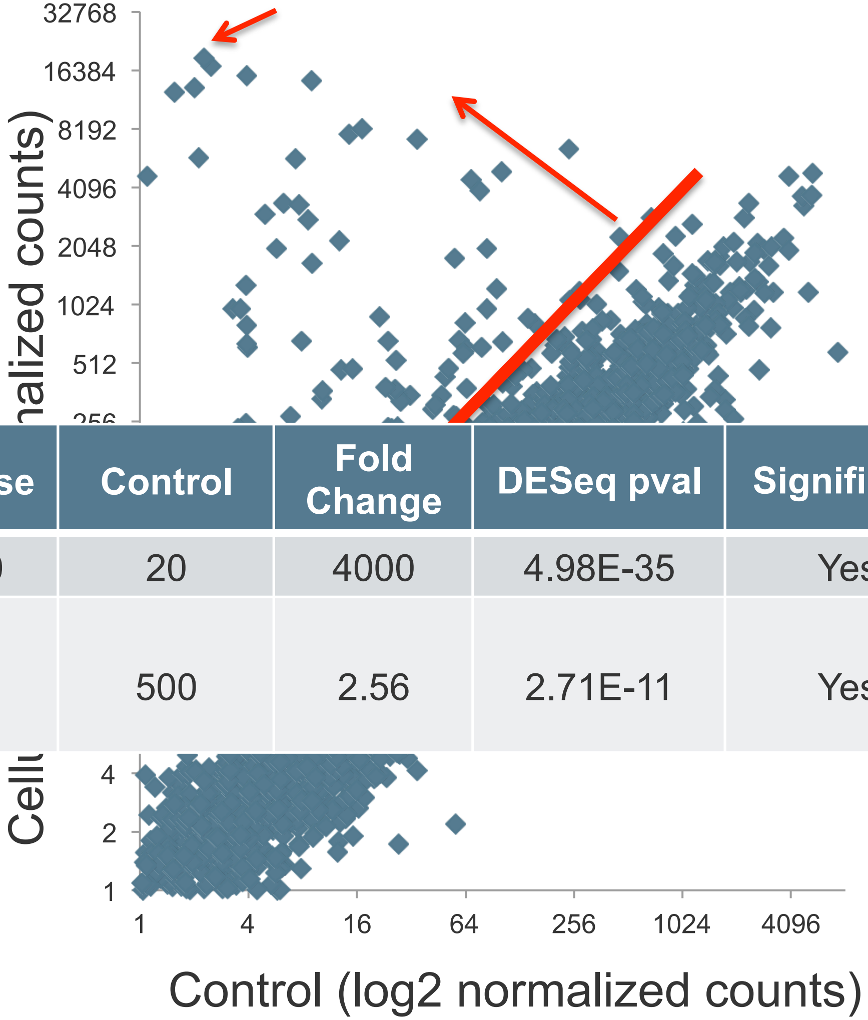
Goal: Identify CaZY enzymes

Known: Fungal growth on Cellulose induces these enzymes

Design: replicates grown as control and on cellulose

Scope: PolyA RNA-Seq libs





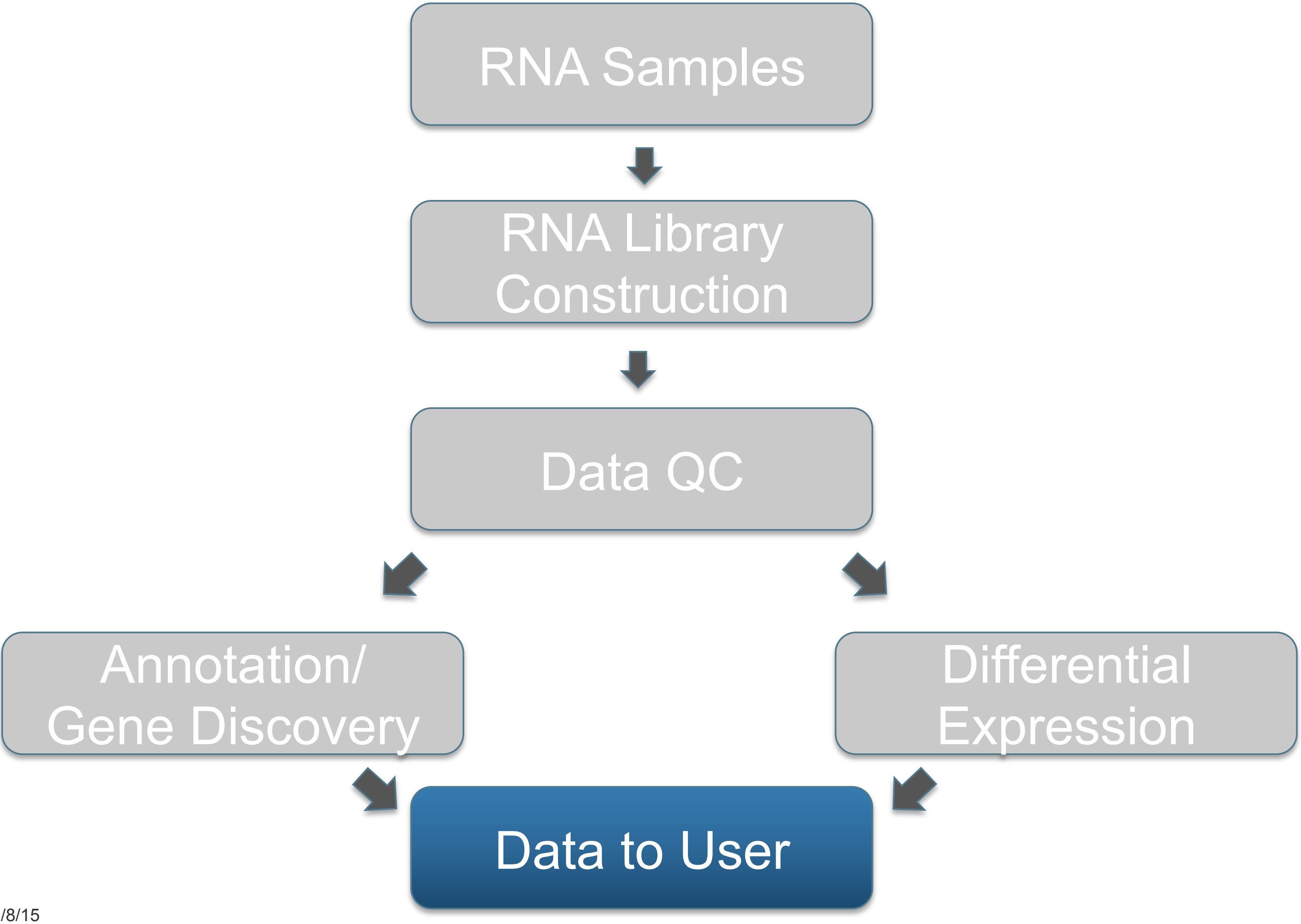
Checked these genes

Prot ID 8640

Protein ID	Cellulose	Control	Fold Change	DESeq pval	Significant
8640:Hydrolase activity	80000	20	4000	4.98E-35	Yes
5522: Peptidyl-prolyl cis-trans isomerase	750	500	2.56	2.71E-11	Yes

* Substrate names for illustration only

RNA-Seq Outline



RNA Samples

RNA Library Construction

Data QC

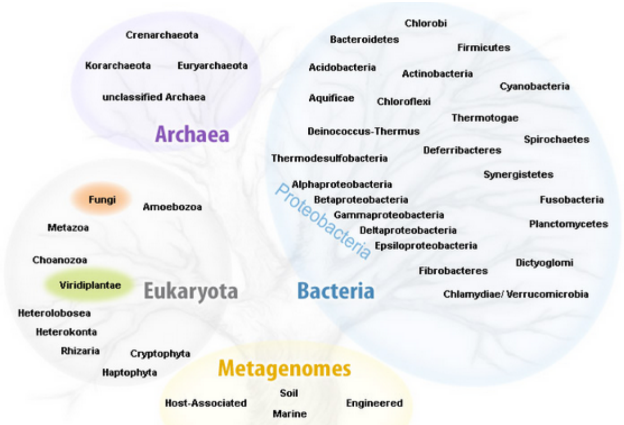
Annotation/ Gene Discovery

Differential Expression

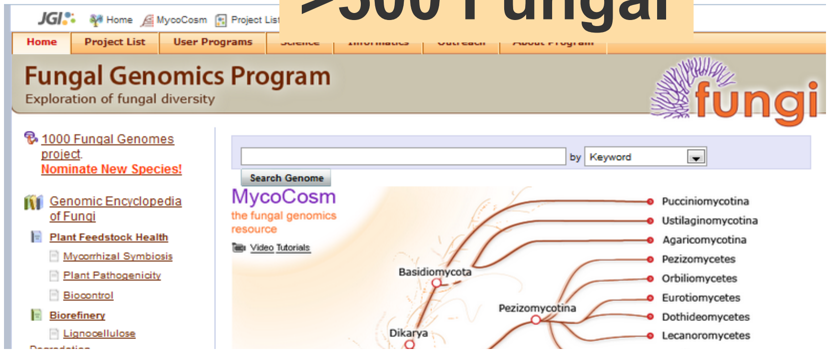
Data to User

RNA-Seq Deliverables thru Portals

2014 RNA Projects
(complete/in progress)



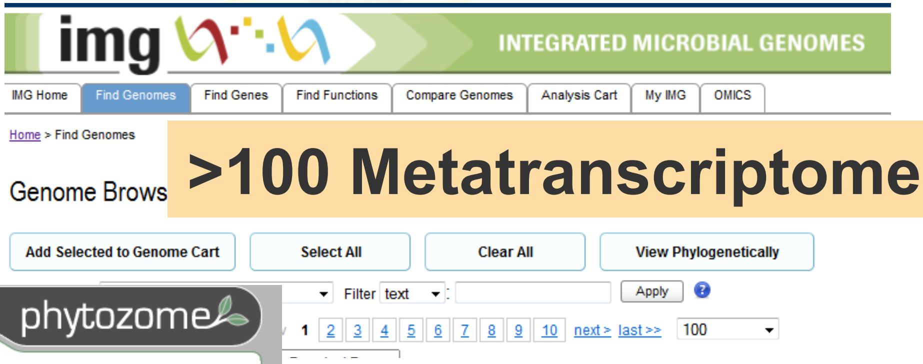
>500 Fungal



Fungal Genomics Program
Exploration of fungal diversity

- 1000 Fungal Genomes project
- Nominate New Species!
- Genomic Encyclopedia of Fungi
- Plant Feedstock Health
  - Mycorrhizal Symbiosis
  - Plant Pathogenicity
  - Biocontrol
- Biorefinery
  - Lignocellulose

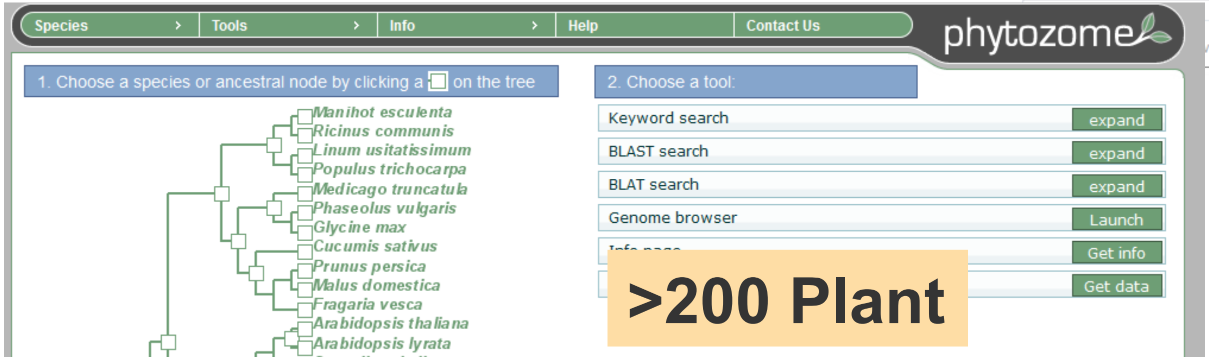
MycoCosm
the fungal genomics resource



Home > Find Genomes

Genome Brows

>100 Metatranscriptome



1. Choose a species or ancestral node by clicking a ☐ on the tree

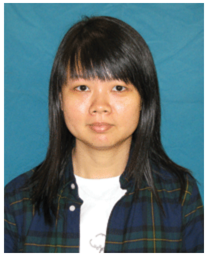
2. Choose a tool:

- Keyword search
- BLAST search
- BLAT search
- Genome browser
- Info page

>200 Plant

Who's who

Anna Lipzen
RNAQC, Assembly
Differential Expression



Chew Yee Ngan
RNA lab

Vasanth Singan
RNASeq Development
DGE Poster**

William Andreopoulos
Prok QC, Analysis

Andrew Tritt
Metatranscriptomes

Chris Daum
Sequencing