

RNA-Seq at JGI

Overview of RNA-Seq products

- Transcriptome assembly

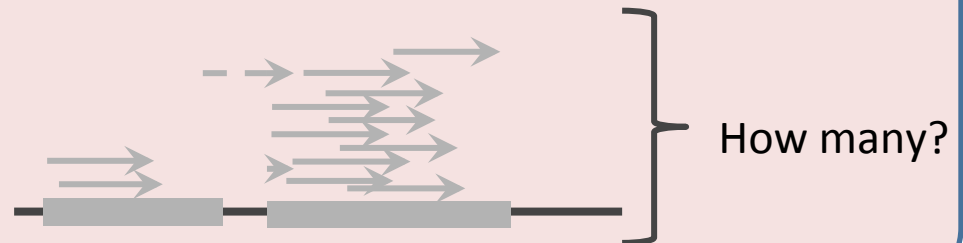
Empirical support in Genome Annotation

How do they splice?



- Differential gene expression

Gene Expression Studies



- small RNA

Gene Regulation Studies



RNA-Seq Science Programs

Fungal Genomics



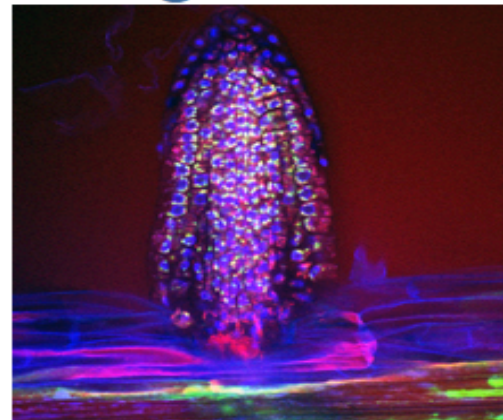
Plant Genomics



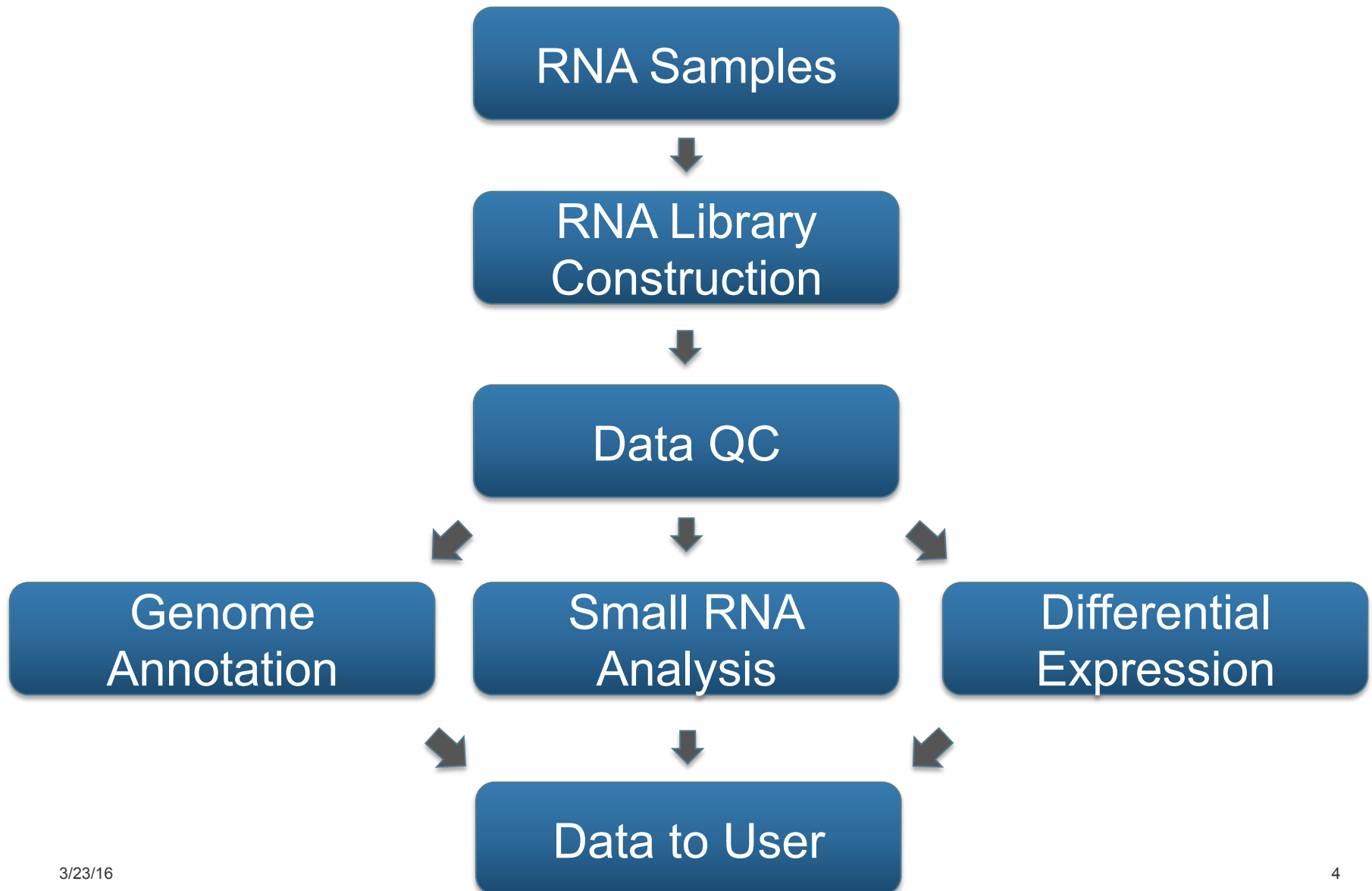
Microbial Genomics



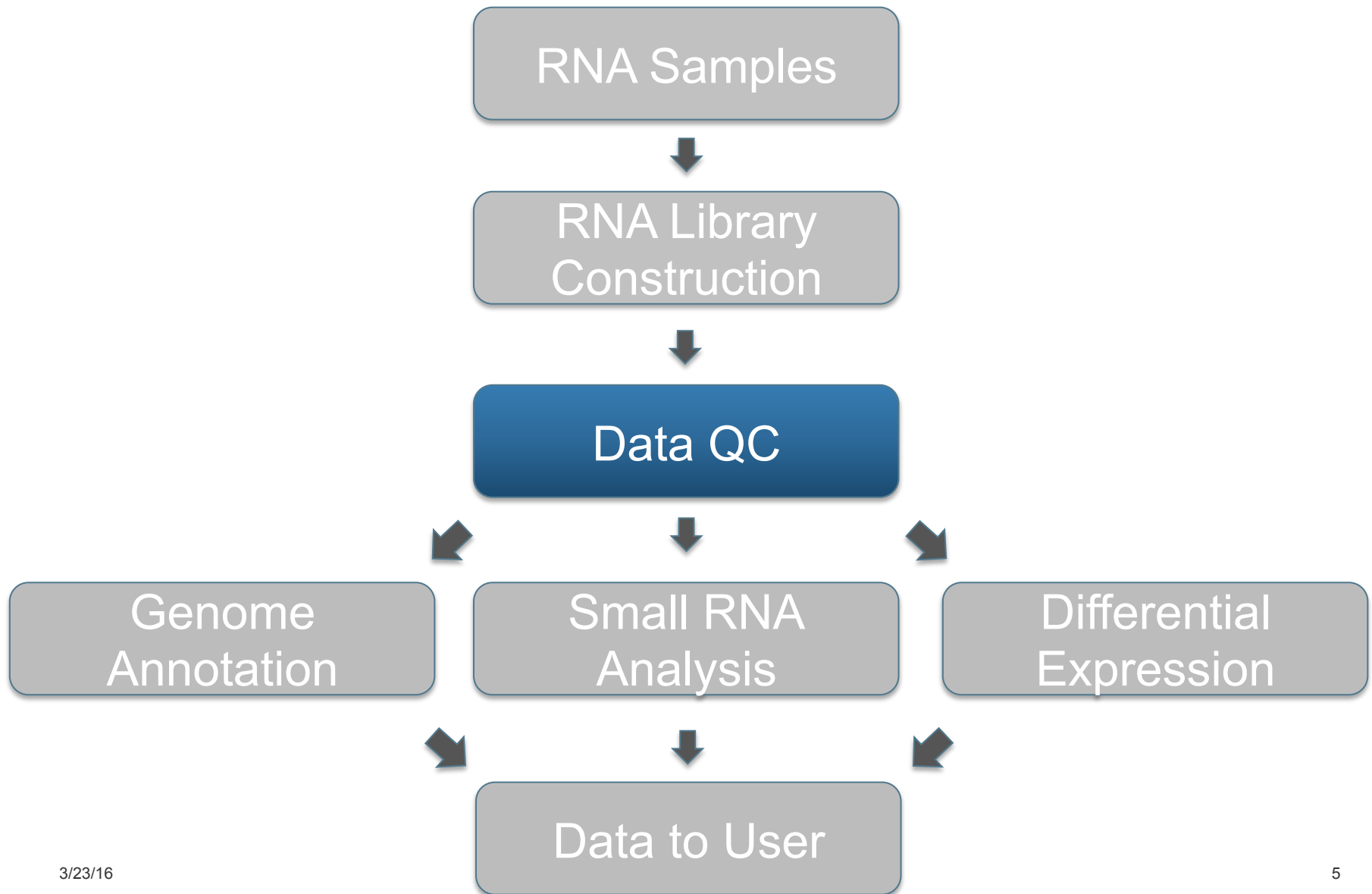
Metagenomics



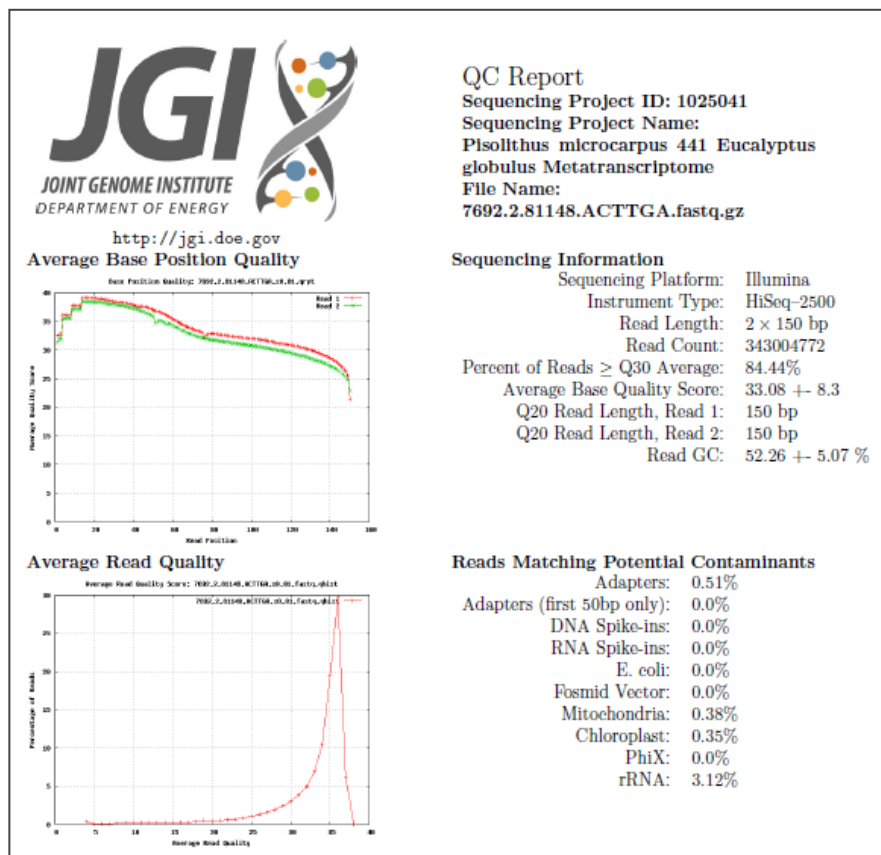
RNA-Seq Workflow



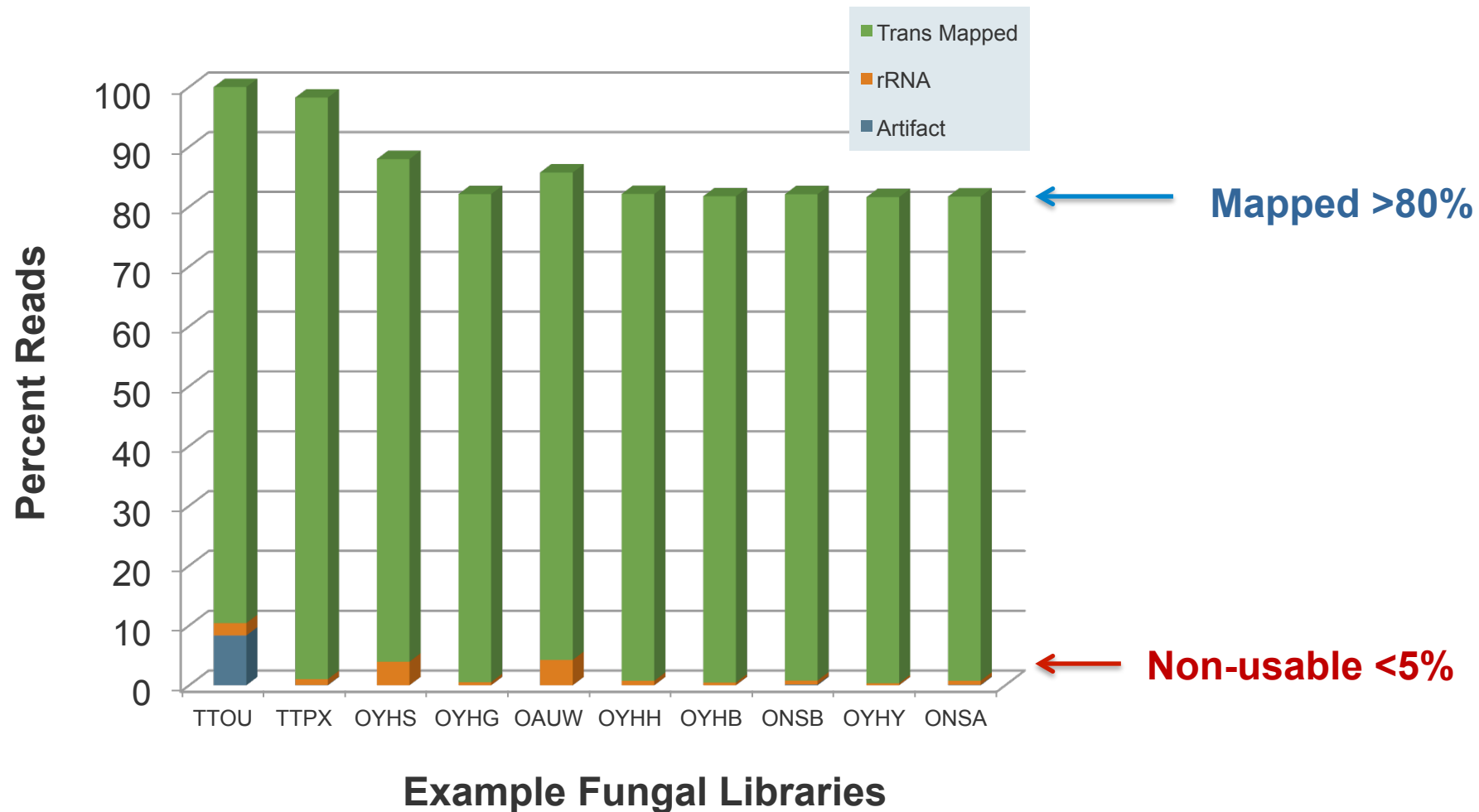
RNA-Seq Workflow



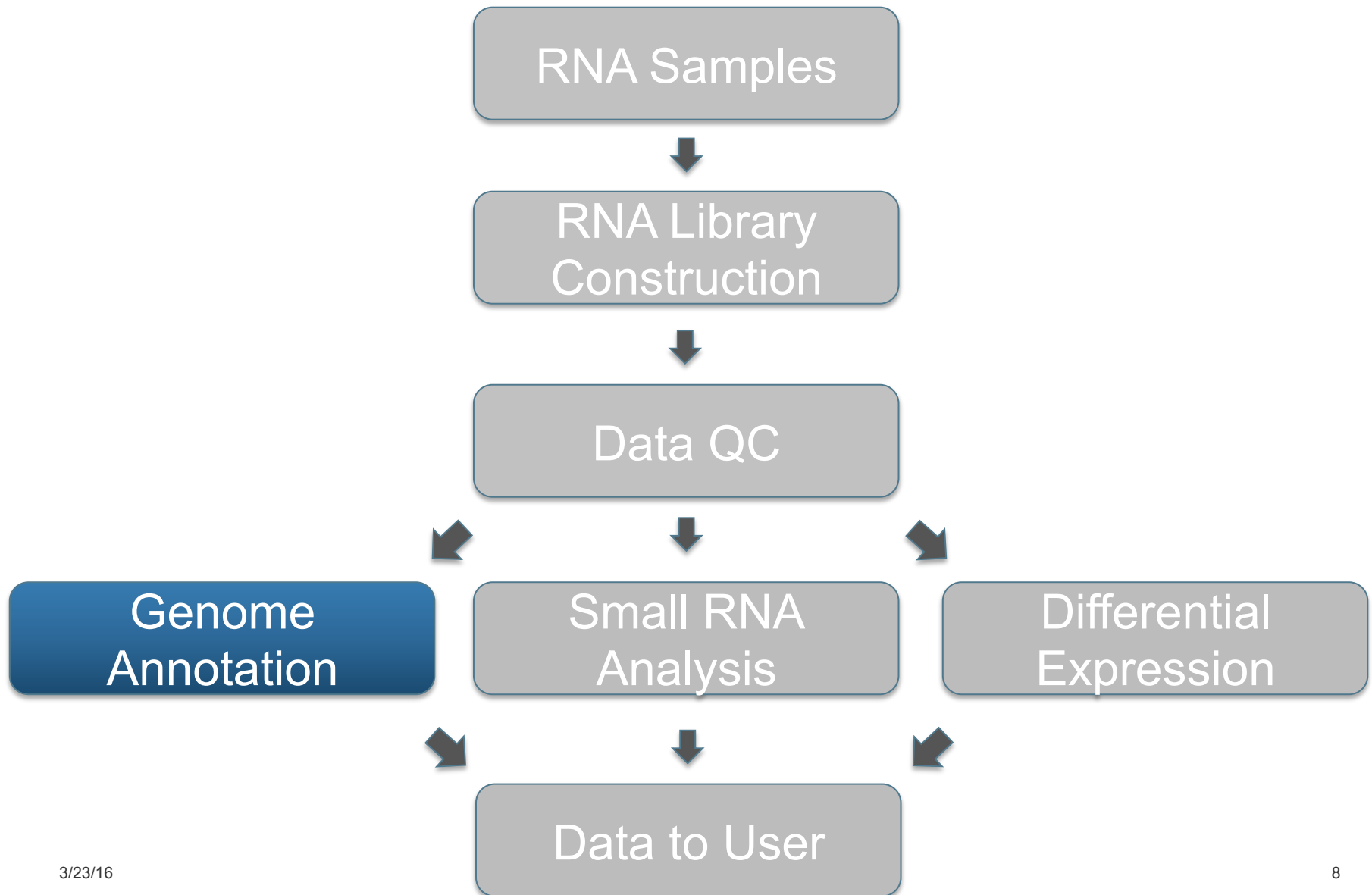
QC Report



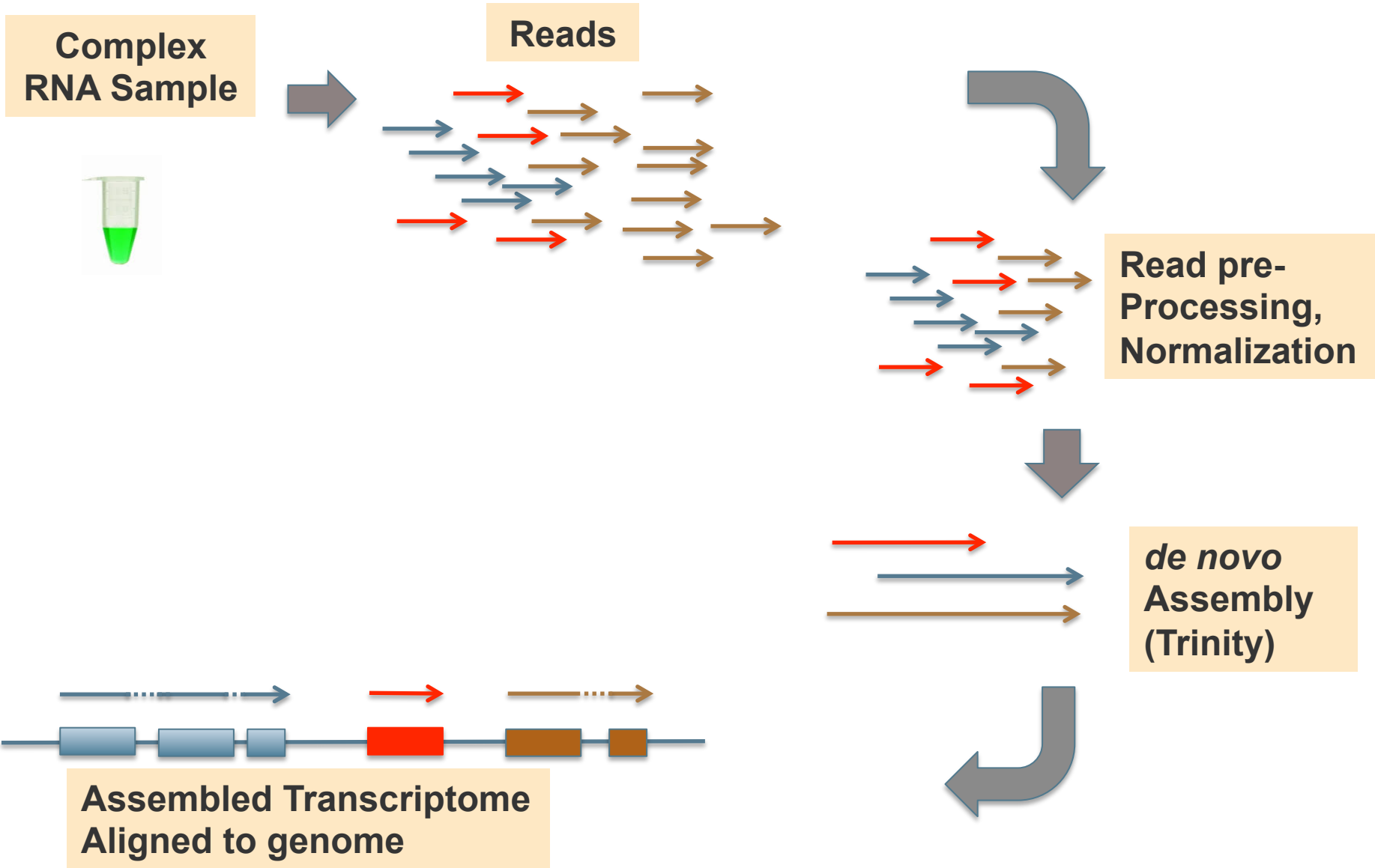
RNA-Seq Library QC – Usable reads



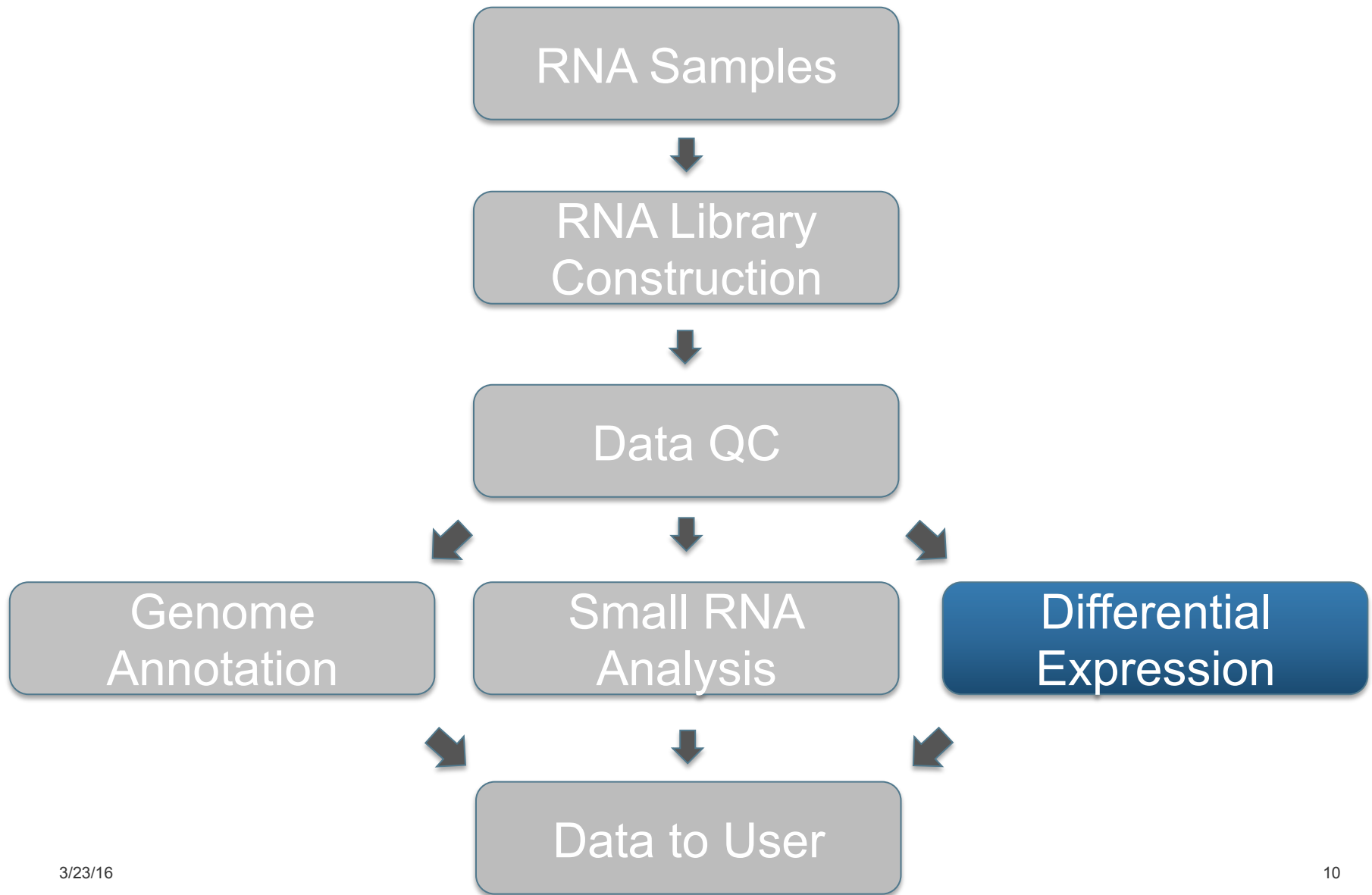
RNA-Seq Workflow



de novo Assembly - Trinity

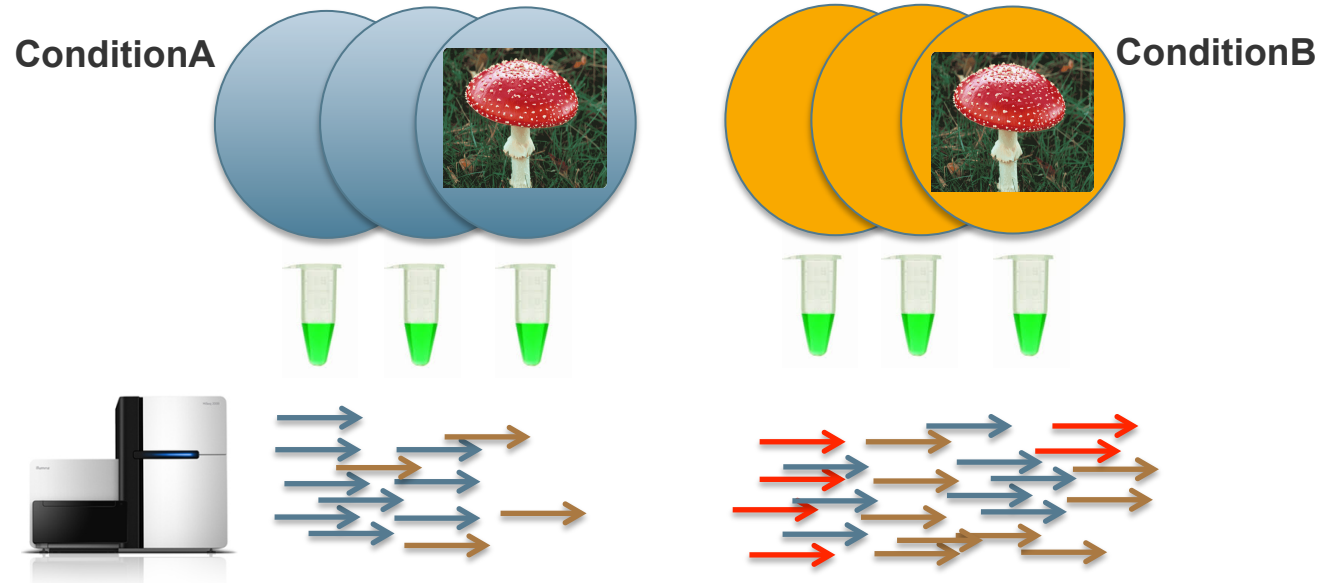


RNA-Seq Workflow



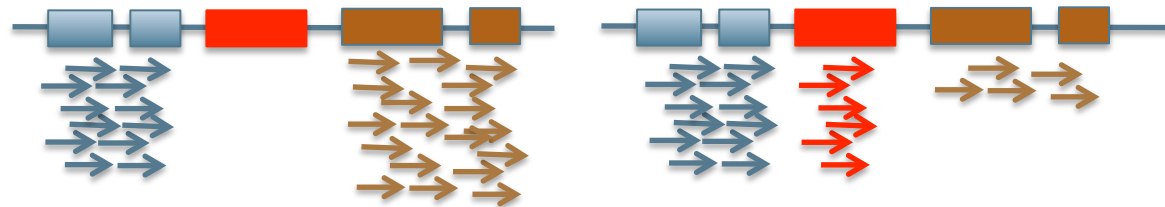
RNA – Differential Gene Expression

**BIOLOGICAL
REPLICATES!**



Align reads to genome = HISAT

Read count = featureCounts



Normalize/Diff Exp = DESeq2

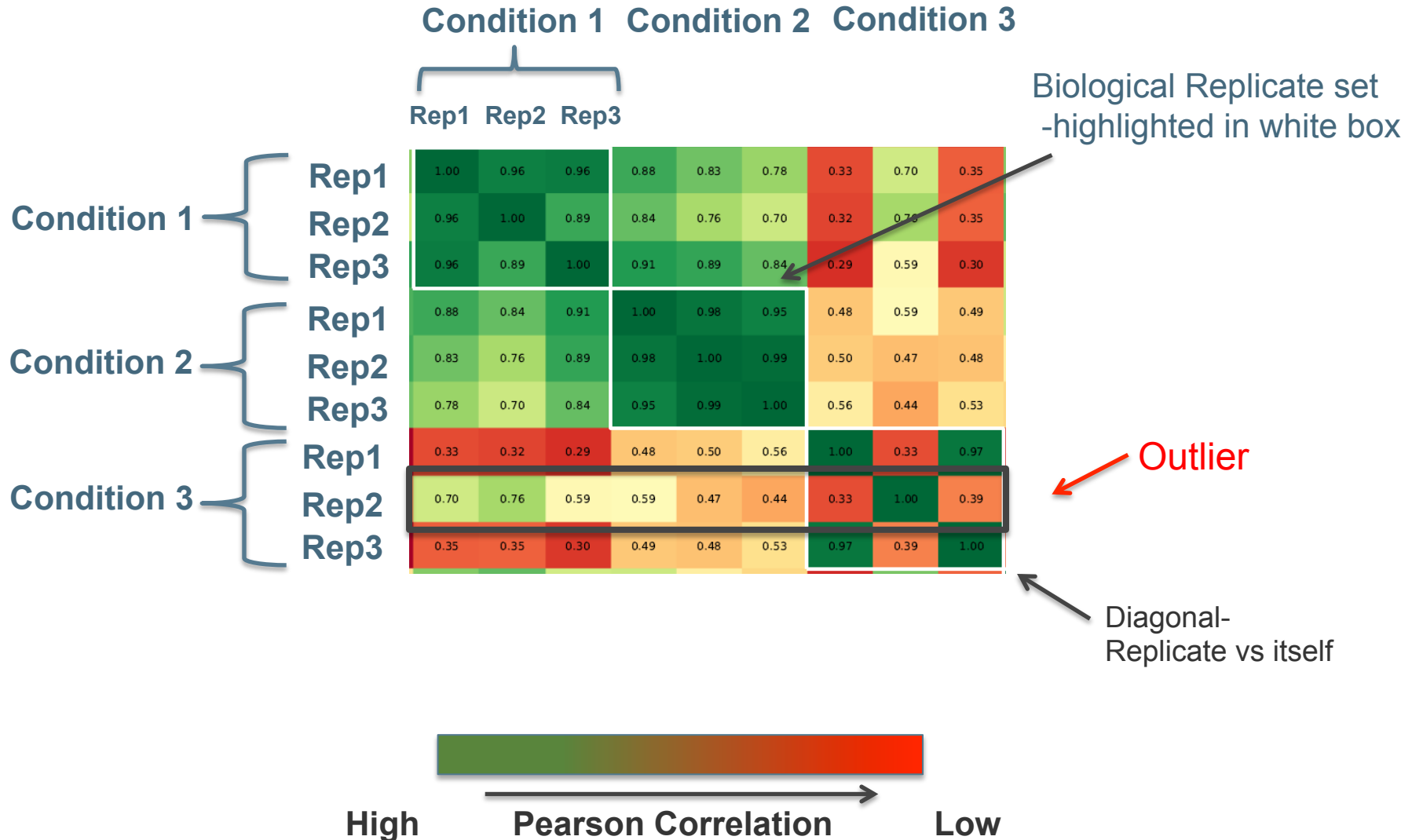
Differential Gene Expression

- Is the difference in gene expression statistically significant ?

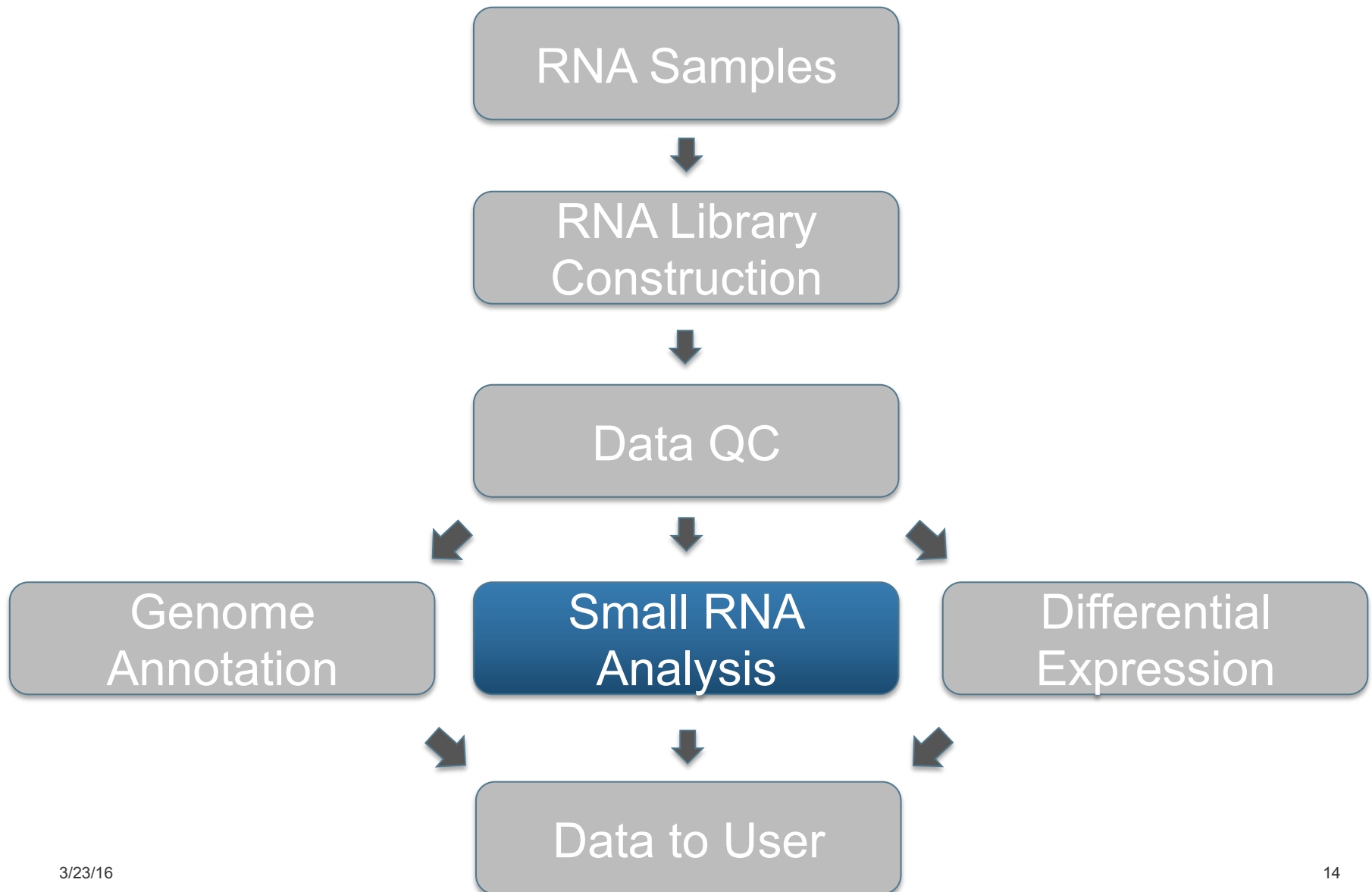
Gene ID	CONDITION A			CONDITION B			FOLD CHANGE	P-VAL
	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3		
001	132	151	98	1239	849	1563	-3.3	1.9E-46
002	2063	1825	1911	2107	2046	2031	-0.12	0.51
003	12585	12158	12858	320	362	316	5.2	1E-224

Table of raw counts

Are replicates correlated?



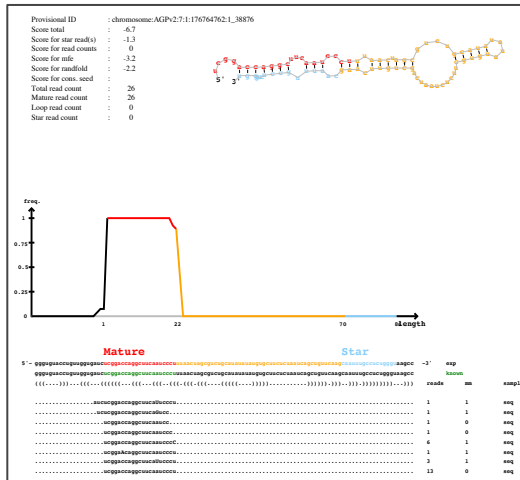
RNA-Seq Workflow



Select a range of sizes

PCR amplification

The diagram shows a document feeder on the left, with a document being scanned. To the right, a series of arrows represent the flow of data or information. A large downward arrow points from the scanning process to the data flow, indicating the transition from physical document to digital data.

[illegible]

The graph displays the distribution of read lengths. The x-axis represents the 'Read Length' from 0 to 50, and the y-axis represents the number of reads ('# Reads') from 0 to 1,800,000. The distribution is highly skewed, with a major peak at a read length of approximately 24, where the number of reads exceeds 1,600,000. There are smaller peaks around 21 and 27, and a gradual increase in reads for lengths above 40.

Read Length	# Reads (approx.)
10	0
20	50,000
21	400,000
22	350,000
23	1,600,000
24	200,000
25	250,000
26	200,000
27	250,000
28	100,000
30	100,000
35	150,000
40	50,000
45	200,000
50	350,000

Example Symbiont Project

Goal: Identify symbiotic gene expression effects

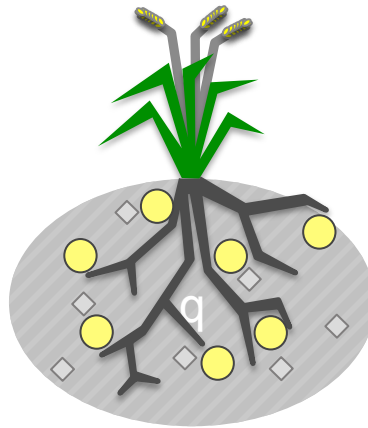
Known: Fungal infection increase plant growth

Design: Plant / Fungi grown in isolation and in contact

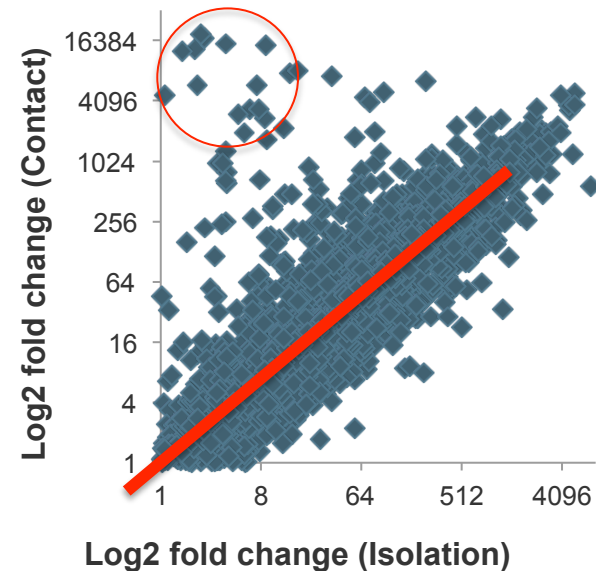
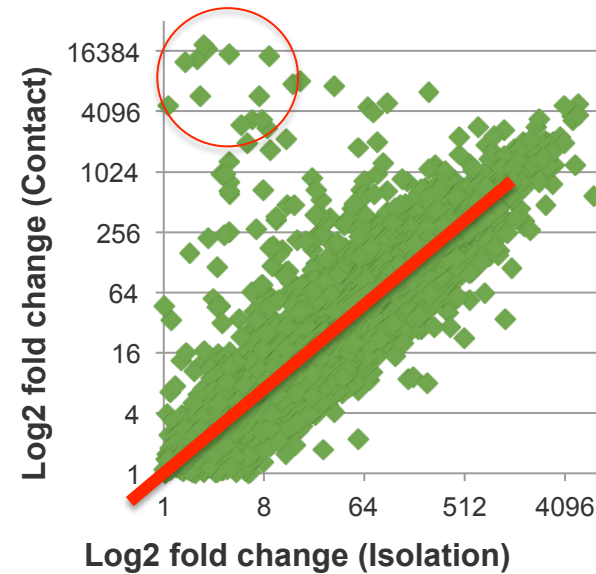
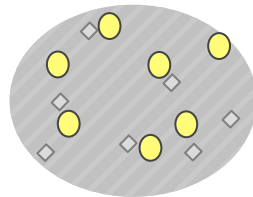
Plant in Isolation



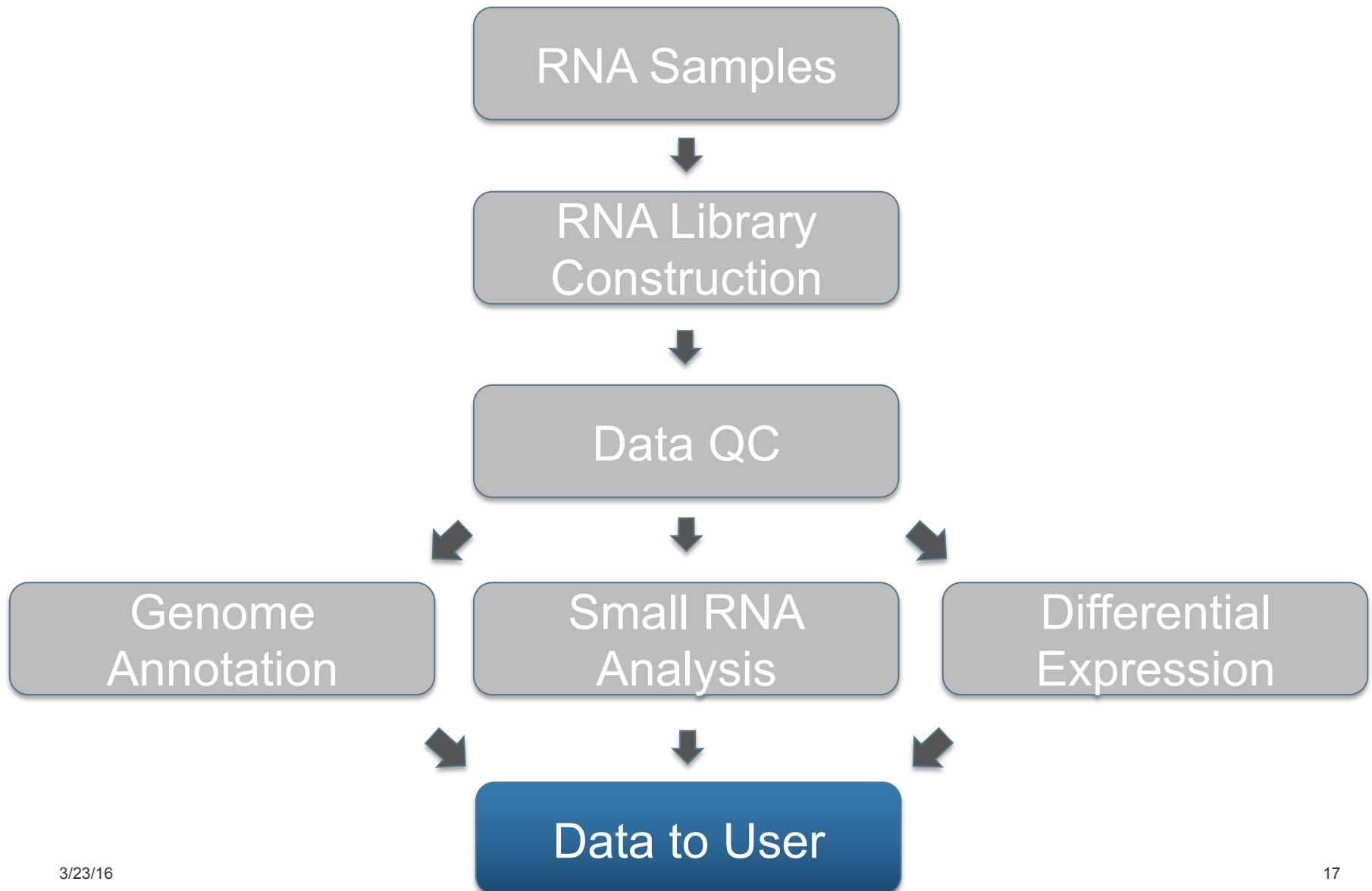
Plant in Contact



Fungi in Isolation



RNA-Seq Workflow

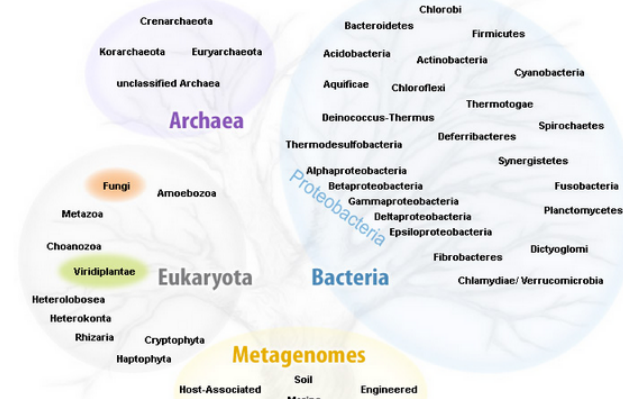


RNA-SEQ Data Provided Through JGI Genome Portals

<http://genome.jgi.doe.gov>

FY 2015 RNA Projects

734 Fungal Samples



Gen **146 Microbial Samples**
Add **466 Metatranscriptome Projects**

Who's Who ?

Project Management



Kerrie Barry



Tijana Galvina del Rio



Christa Pennacchio



Vivian Ng

Sequencing Technologies



Chris Daum



Rita Kuo



Yuko Yoshinaga

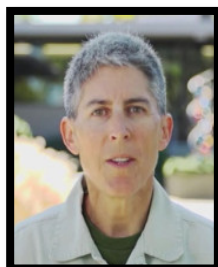
QC



Bryce Foster

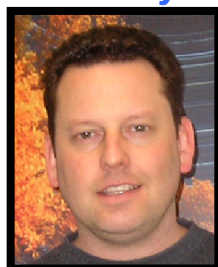
Analysis

Lead



Erika Lindquist

Community MT



Brian Foster

Fungal



Anna Lipzen

Microbial



Bill Andreopoulos