TagSeq MEGA Workshop July 2019

Serena Hackerott August 13, 2019

Background material adapted from M. Matz and E. Abbott

Funding for Workshop Attendance Provided by:



Tag-Seq vs RNAseq: Library Prep

RNAseq:

Isolate random fragments across entire transcript



Tag-Seq:

Isolate fragment of 3' end of transcript



Tag-Seq vs RNAseq : Mapping

RNAseq:

Map to genome



Tag-Seq: Map to transcriptome



Identification of novel transcripts, splice sites and quantification of transcript abundance

Tag-Seq vs RNAseq

	Tag-based	RNAseq			
Reference	uses transcriptome	uses genome			
Reads per transcript	One (after duplicate removal)	multiple across entire length			
Regions of transcript sequenced	3' end only	entire transcript			
Normalization for gene length for expression analysis	not required	required			
Required Sequencing Depth for expression analysis	less	more			
Cost of library preparation	less	more			
Can detect splice variants	no	yes			
Can detect variation in paralogous gene expression	no	yes			

Taq-Seq Sequencing and Costs

Single Read SR 50 Sequencing

- Goal: ~1 million reads per sample after removing PCR duplicates (~2 million before dup. removal)
- Min ~500,000 reads per sample after dup. removal
- 50-70 samples per lane (max 100 samples/lane)

<u>Cost per Sample (Lib Prep + HiSeq 4000 SR 50)</u>

- Low Coverage (5M reads): \$40 / sample
- High Coverage (10M reads): \$60 / sample

Tag-Seq library prep



- RNA Extraction and Clean
- Heat Fragmentation
- cDNA Synthesis
- cDNA Amplification
 - PicoGreen Quantification
- Barcoding and Size Selection

Tag-Seq linux pipeline

Concatenating and renaming raw read files

Trimming, removing PCR duplicates, quality filtering

Mapping to transcriptome

Deriving gene counts (per isogroup)

Counts Table to export for analysis

	SampleA	SampleB	SampleC
lsogroup1	Count	Count	Count
lsogroup2	Count	Count	Count
lsogroup3	Count	Count	Count

Gene Expression Analysis in R

- Differential Gene Expression
 - Heatmap and hierarchical clustering
 - Principle Coordinate Analysis
 - Model fitting and significance testing
- GO: Gene Ontology
 - GO Enrichment Test (MWU)
- KOG: euKaryotic Orthologous Groups
 - KOG Enrichment Test (MWU)
- WGCNA: Weighted Gene Coexpression Network Analysis

Heatmap and Hierarchical Clustering



Principle Coordinate Analysis



GO MWU (Molecular Function)

Also Cellular Component and Biological Process



p < 0.001 p < 0.005 p < 0.01

KOG

Including comparison between different studies of same species Secondary metabolites biosynthesis, transport and catabolism RNA processing and modification Cell cycle control, cell division, chromosome partitioning Lipid transport and metabolism 500 Replication, recombination and repair Nuclear structure Inorganic ion transport and metabolism Cell wall/membrane/envelope biogenesis Signal transduction mechanisms 0 Cvtoskeleton Translation, ribosomal structure and biogenesis Energy production and conversion -500 Cell motility Amino acid transport and metabolism Extracellular structures -1000 Transcription Carbohydrate transport and metabolism Chromatin structure and dynamics Coenzyme transport and metabolism Posttranslational modification, protein turnover, chaperones Defense mechanisms Nucleotide transport and metabolism Intracellular trafficking, secretion, and vesicular transport heat larv ae. short larv ae. long

Module-trait relationships

WGCNA Modules

Correlation and Significance with Genotype and Stress

MEdarkolivegreen	0.0097 (1)	0.033 (0.9)	0.028 (0.9)	-0.071 (0.7)	0.92 (3e-10)	-1
MEbisque4	-0.98 (4e-17)	0.27 (0.2)	0.28 (0.2)	0.43 (0.04)	0.099 (0.6)	· ·
MEdarkslateblue	-0.43 (0.04)	-0.022 (0.9)	0.21 (0.3)	0.24 (0.3)	0.51 (0.01)	
MEdarkred	-0.16 (0.5)	-0.78 (6e-06)	0.17 (0.4)	0.77 (1e-05)	-0.026 (0.9)	_05
MEblack	-0.4 (0.05)	-0.26 (0.2)	0.98 (4e-18)	-0.33 (0.1)	0.11 (0.6)	0.0
MEfloralwhite	0.55 (0.005)	-0.69 (2e-04)	0.53 (0.008)	-0.39 (0.06)	0.21 (0.3)	
MEgrey60	0.92 (3e-10)	-0.43 (0.04)	-0.54 (0.007)	0.05 (0.8)	0.0096 (1)	0
MEmediumorchid	0.061 (0.8)	-0.064 (0.8)	-0.057 (0.8)	0.06 (0.8)	-0.93 (6e-11)	-0
MEred	0.34 (0.1)	0.03 (0.9)	0.13 (0.5)	-0.5 (0.01)	-0.68 (3e-04)	
MEcyan	0.46 (0.02)	0.51 (0.01)	-0.095 (0.7)	-0.88 (2e-08)	0.027 (0.9)	0.5
MEhoneydew1	-0.49 (0.01)	0.88 (1e-08)	0.1 (0.6)	-0.49 (0.01)	0.097 (0.7)	0.5
MEpaleturquoise	-0.33 (0.1)	0.36 (0.08)	-0.17 (0.4)	0.13 (0.5)	-0.31 (0.1)	
MEtan	-0.095 (0.7)	0.43 (0.04)	-0.89 (5e-09)	0.56 (0.005)	0.11 (0.6)	
MEgrey	0.012 (1)	0.019 (0.9)	-0.031 (0.9)	-0.00045 (1)	-0.041 (0.8)	- −1
	8	\$	C	۲ م	1055	

GO MWU of Dark Olive Green Module (Molecular Function)



Suggestions for Gene Expression Studies

- Design experiments so that the question can be answered without knowing the identity of the genes
 - Avoid "tea leaf reading"
- Ideal designs have factors with two levels
 - Rather than BACI designs, Control Impact designs are easier to analyze
- Can compare gene expression data from different experiments if species is the same