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:

FIU Center for Aquatic Chemistry and Environment
NSF Center of Research Excellence in Science and Technology

FIU Institute of Water and Environment

NSF

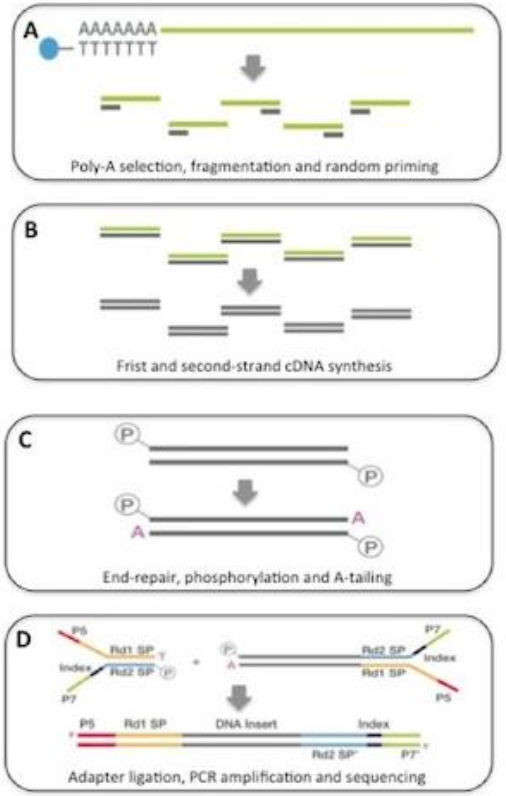
FIU Southeast Environmental Research Center

Environmental Epigenetics Lab

Tag-Seq vs RNAseq: Library Prep

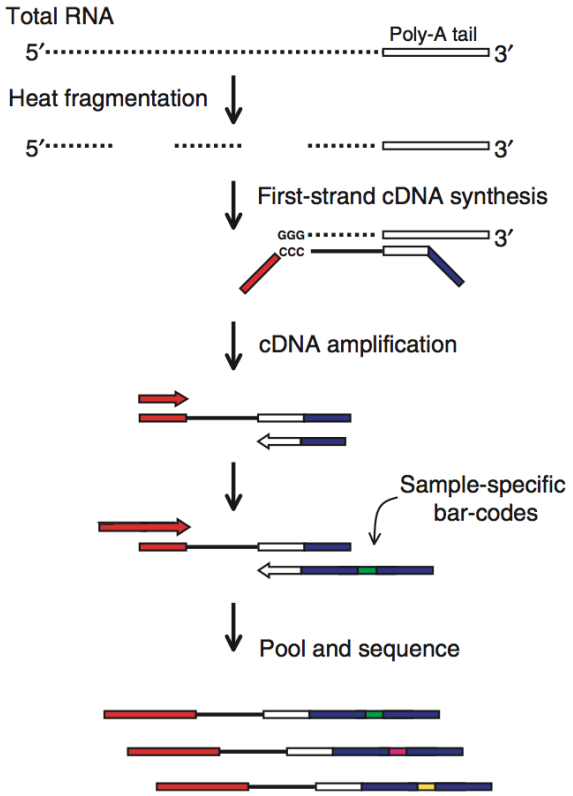
RNAseq:

Isolate random fragments across entire transcript



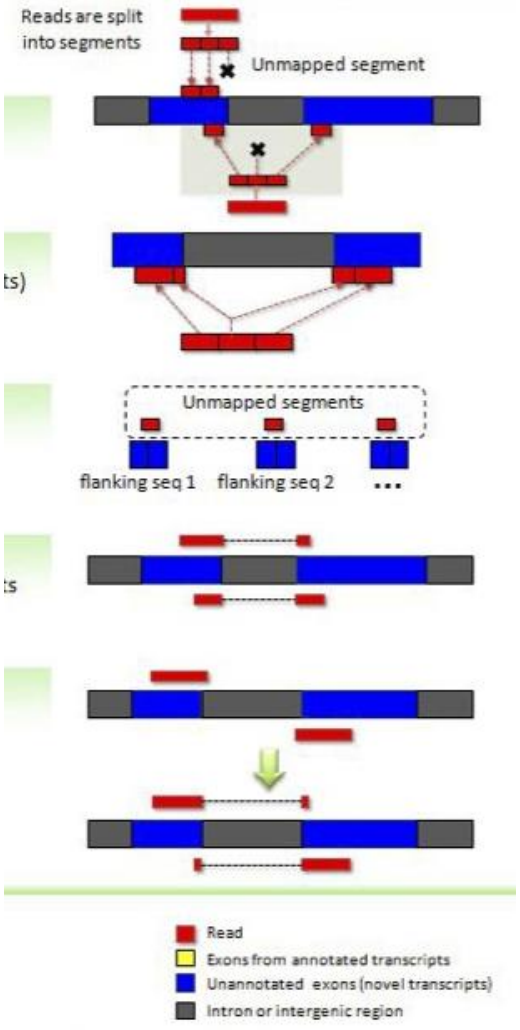
Tag-Seq:

Isolate fragment of 3’ end of transcript

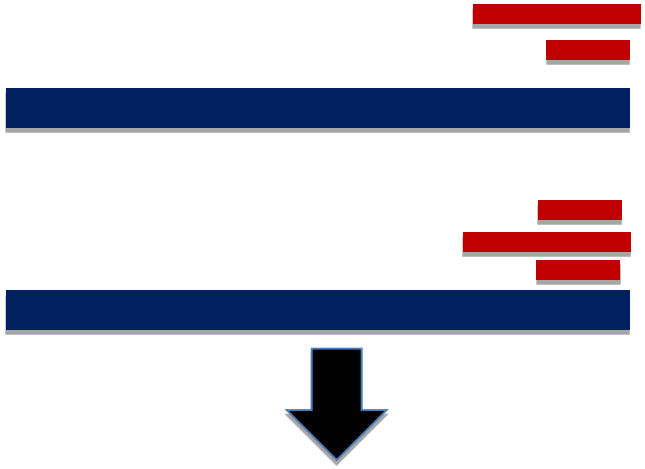


Tag-Seq vs RNAseq : Mapping

RNAseq:
Map to genome



Tag-Seq:
Map to transcriptome



Quantification of
transcript abundance

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

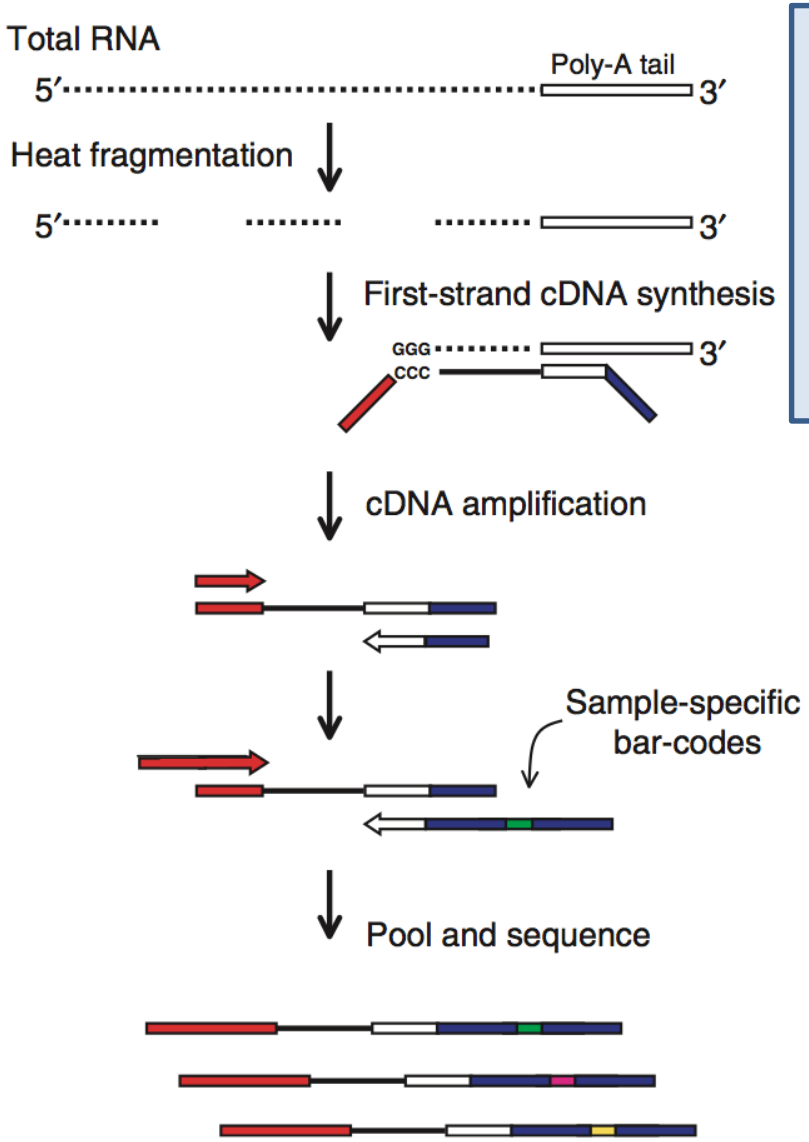
Single Read SR 50 Sequencing

- 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)

Cost per Sample (Lib Prep + HiSeq 4000 SR 50)

- 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)

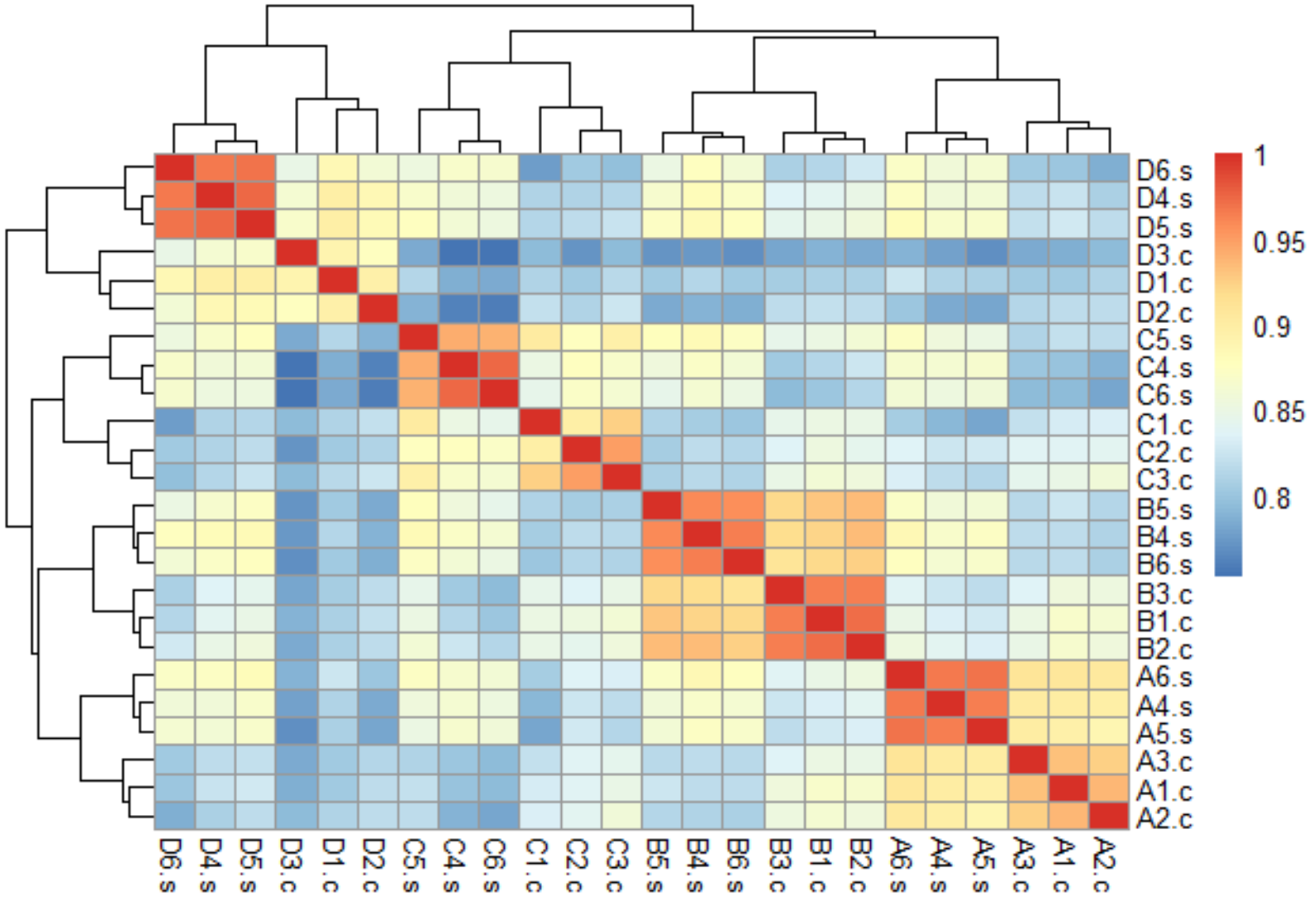
Counts Table to export for analysis

	SampleA	SampleB	SampleC…
Isogroup1	Count	Count	Count
Isogroup2	Count	Count	Count
Isogroup3…	Count	Count	Count

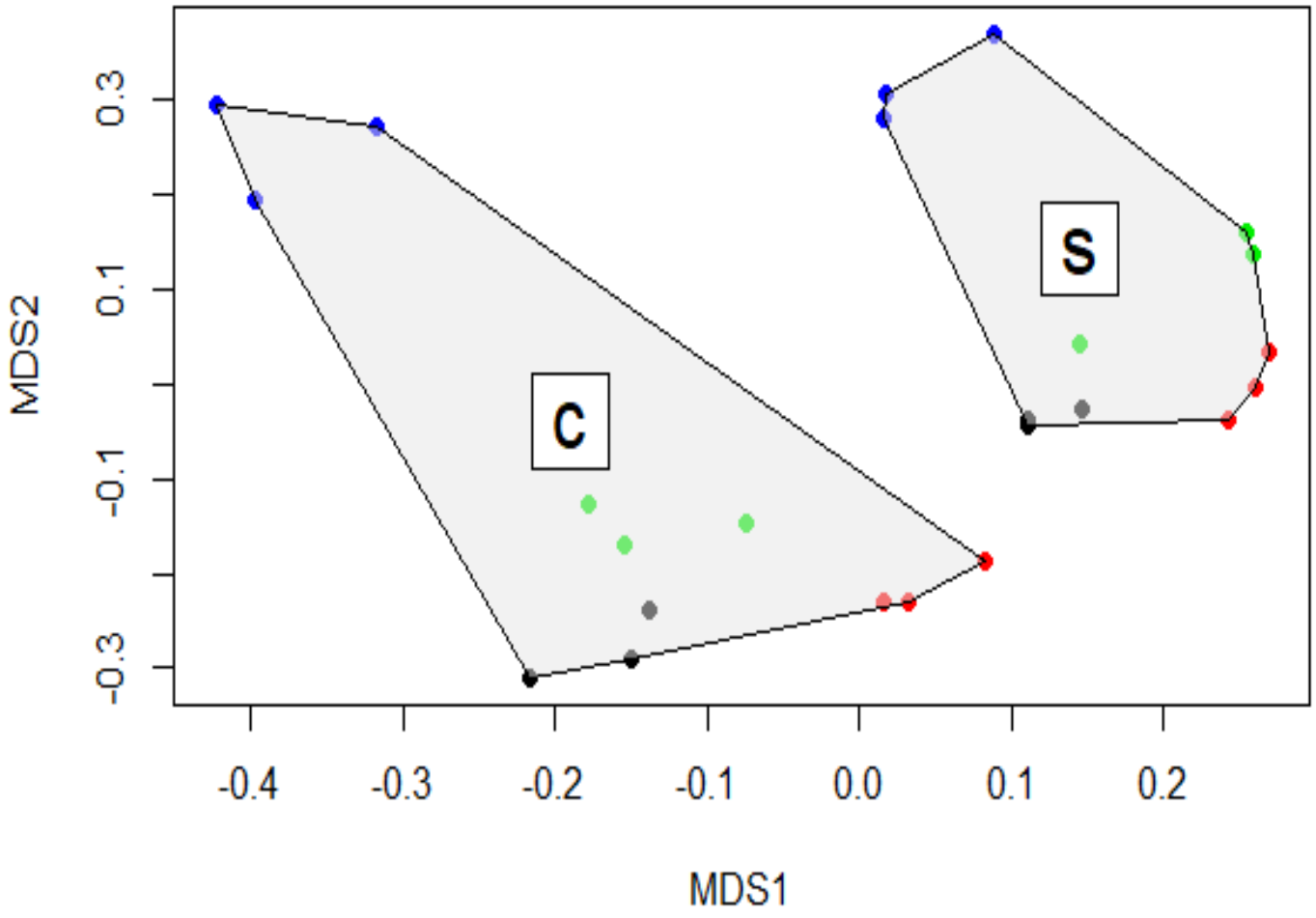
Gene Expression Analysis in R

- 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

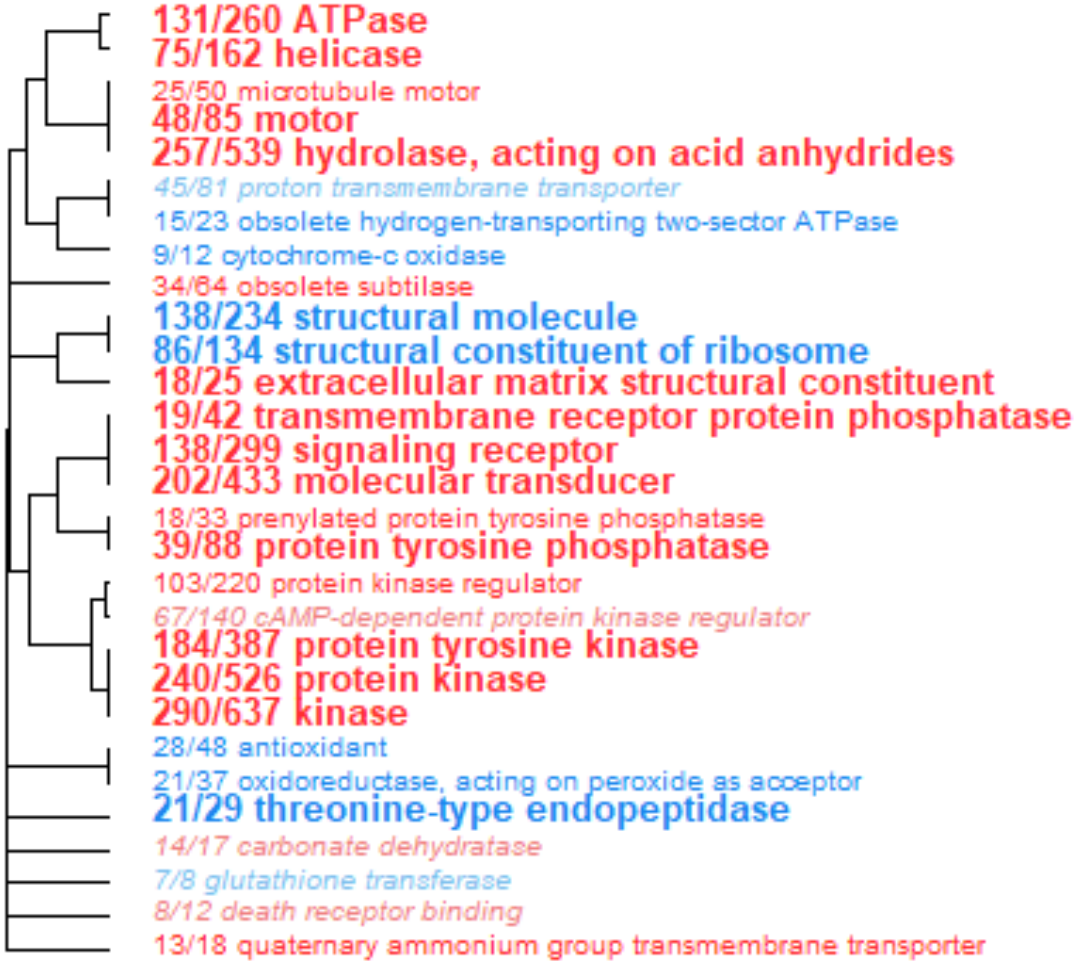


Principle Coordinate Analysis



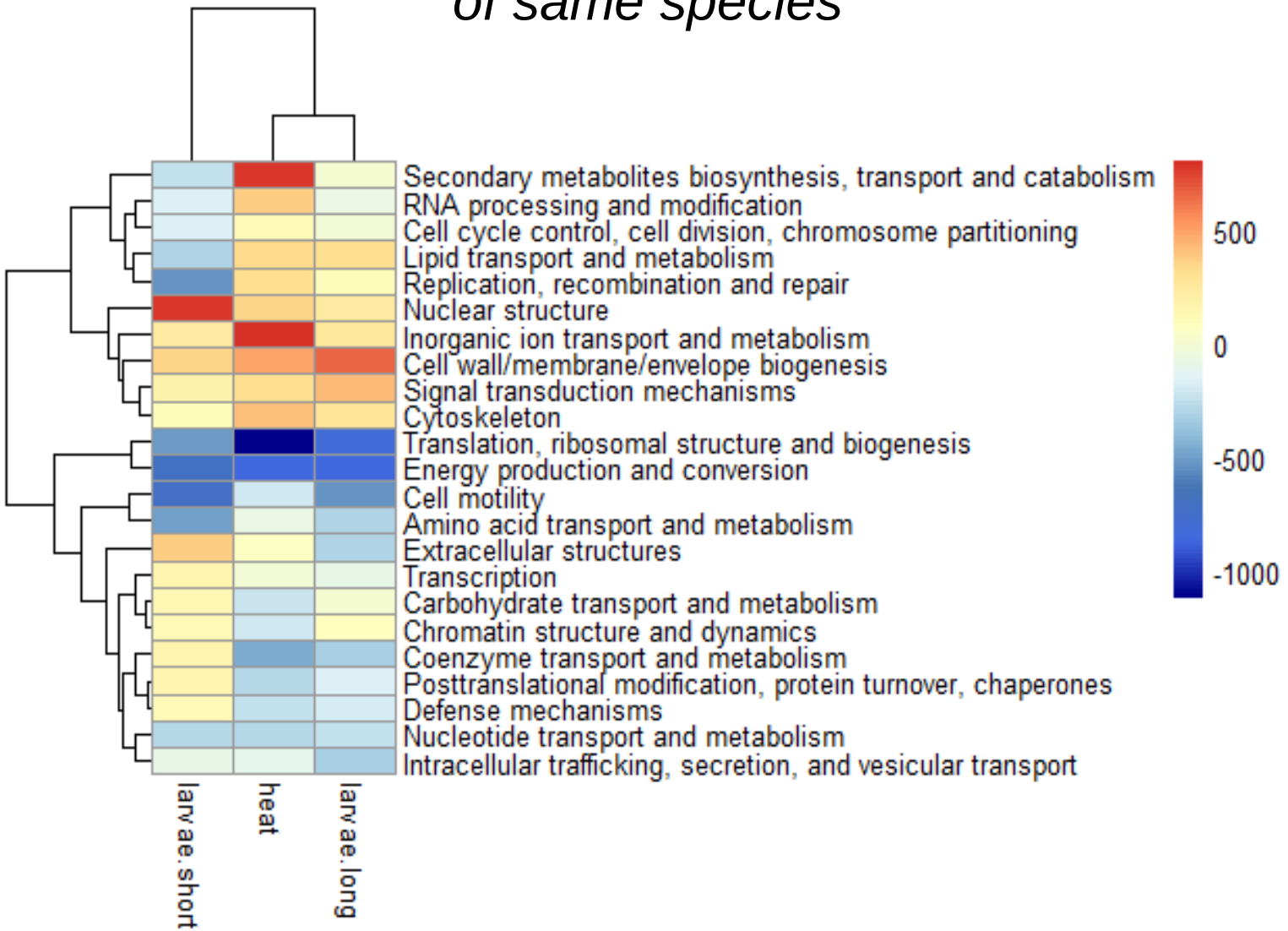
GO MWU (Molecular Function)

Also Cellular Component and Biological Process



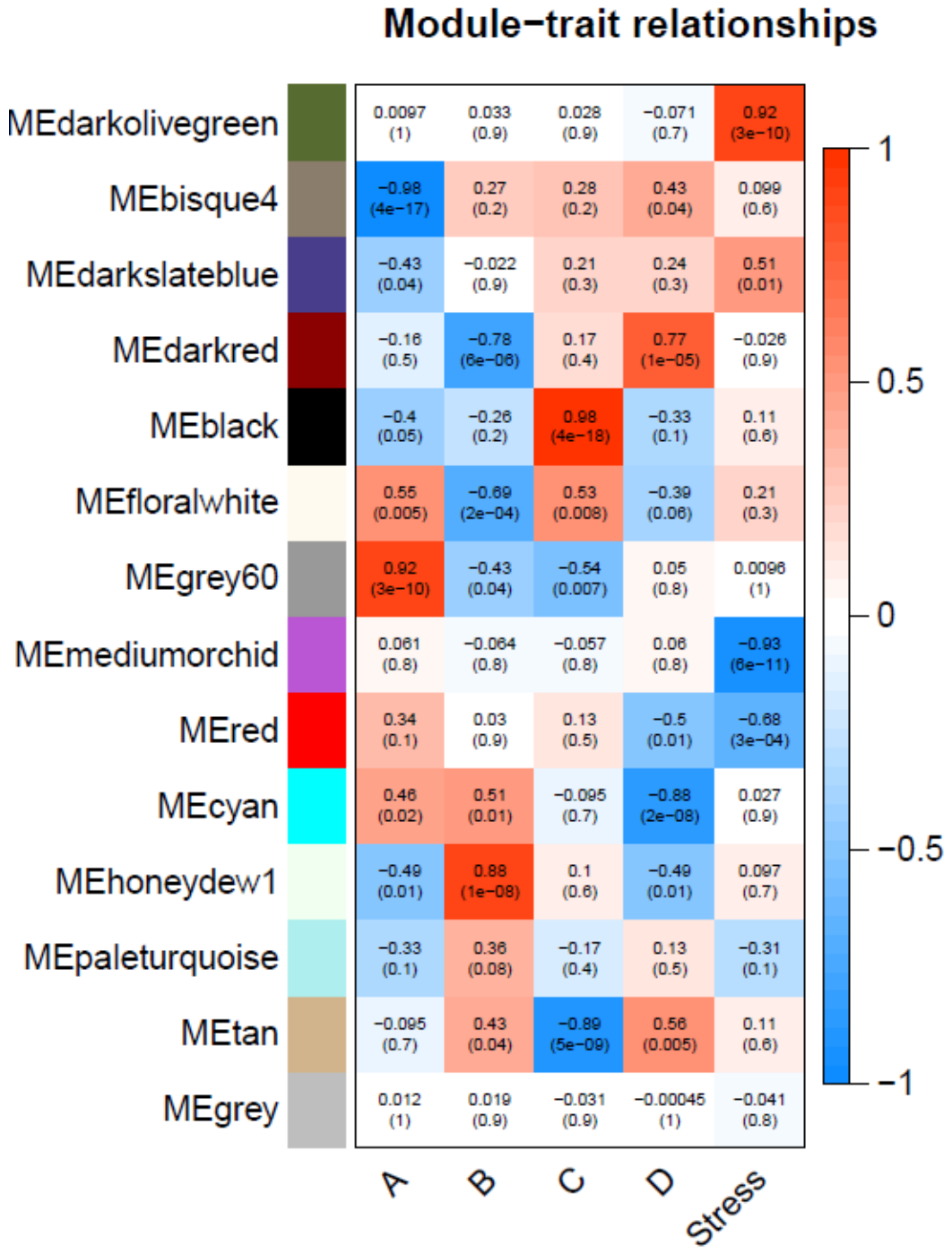
KOG

Including comparison between different studies of same species

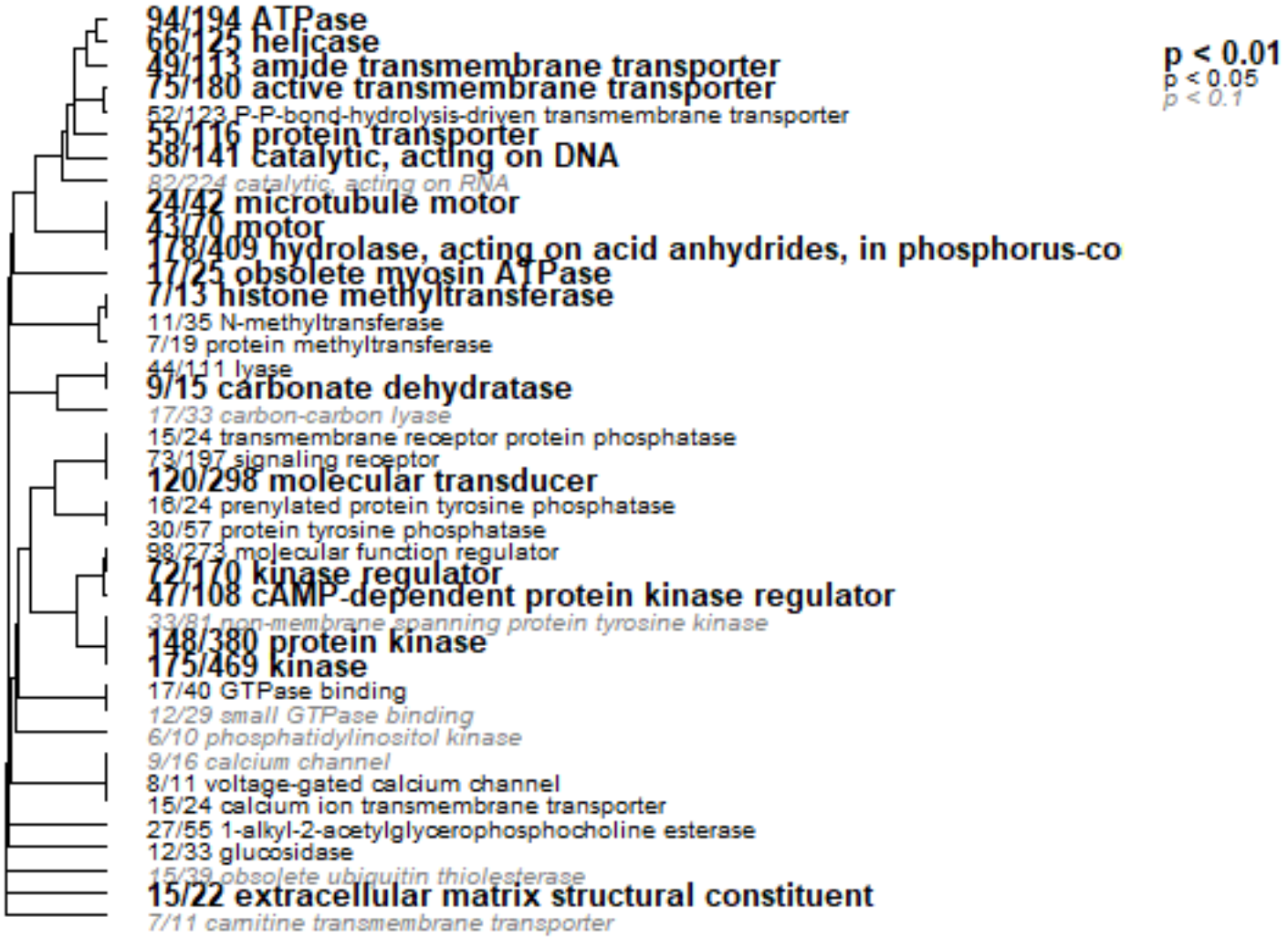


WGCNA Modules

Correlation and Significance with Genotype and Stress



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