

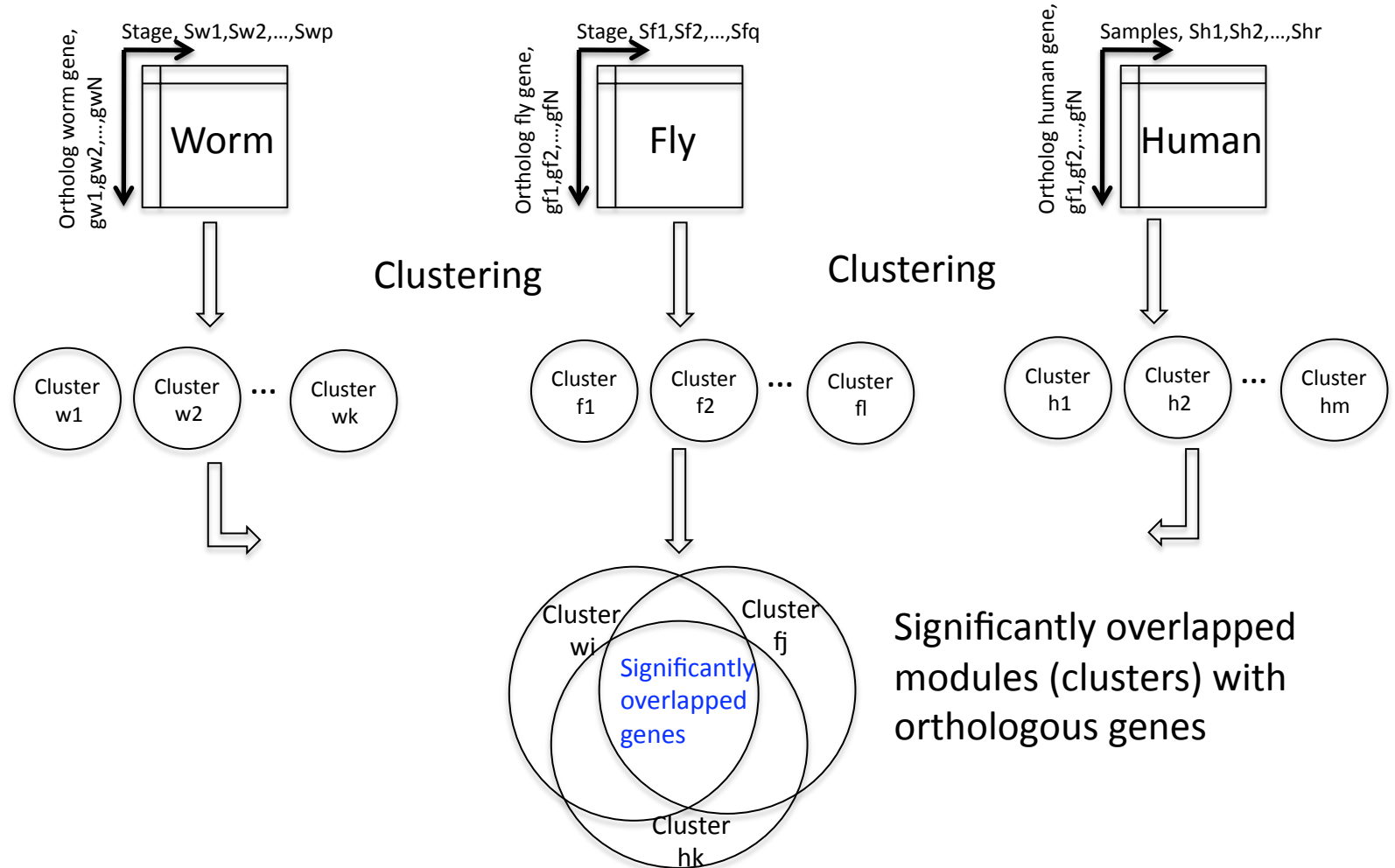
Identification of “orthologous” ncRNAs
among worm, fly and human via
studying their expression correlations
with 1-1-1 orthologous genes

Daifeng Wang

Gerstein Lab

Aug 23 2012

Before KKY's model is ready, the tentative workflow is as follows:



- Correlate ncRNAs with modular eigengenes (1st PCs) in their corresponding species
 - “Orthologous” ncRNAs if they all highly correlate with significantly overlapped modules

1-1-1 orthologous genes

- MERGED ORTHOLOG list (2012/06/11)
<File:Modencode.merged.orth.txt.gz> - MIT orthologs (except for human-mouse) + merged MIT-Ensembl human-mouse orthologs
 - 1935 1-1-1 orthologous genes (triplets) in total
- The freeze RNA-seq gene expression data in June 2012 in mod/ENCODE
 - 1925 worm orthologous genes across 33 stages
 - 1887 fly orthologous genes across 30 stages
 - 1907 human orthologous genes across 21 Tier1 samples (Thanks to Arif)
- ncRNAs
 - Non-protein coding genes from Ensembl
 - Non-zero gene expressions > 50% across samples

Statistics on orthologs & ncRNAs in freeze data

Number	Worm	Fly	Human
Sample	33	30	21
1-1-1 Orthologous genes	1925	1887	1907
pseudogene	493	2	1181
miRNA	46	77	
ncRNA	1988	175	
snRNA	25		
snoRNA	99	201	
tRNA	13	25	
rRNA		55	
processed_transcript			217
lincRNA			305
polymorphic_pseudogene			16
antisense			199
sense_intronic			30
3prime_overlapping_ncrna; IG_C_pseudogene; IG_V_gene; TR_C_gene; sense_overlapping			6

Clustering via WGCNA

Species	module	min size	max size	Un-clustered genes
Worm	20	59	195	91
Fly	15	55	349	120
Human	15	51	244	73

* min size=50, detect cut height=0.99 in WGCNA

Significantly overlapped modules

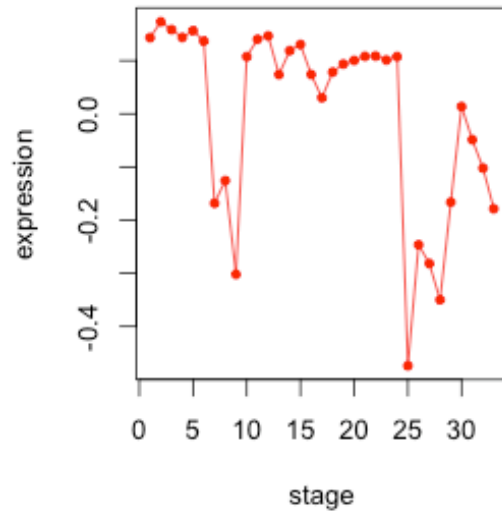
Overlapped orthologs	TOP 1	TOP2	TOP3
Worm-Fly	50	35	31
W-F-H	21	11	10

Results

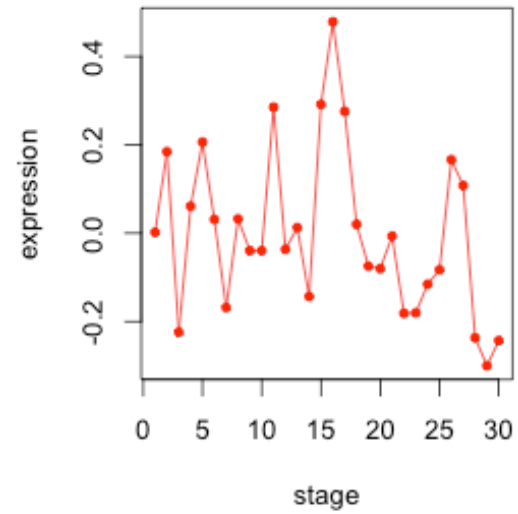
- Eigengenes (1st PCs) in top overlapped modules
 - Highly correlated ncRNA expression profiles (spearman correlation > 0.6)
- Spearman correlation vs. Length
 - for all types of ncRNAs
- Show Worm-Fly top 2 module pairs, and then W-F-H top 2 module triplets

No.1 Worm-Fly module pair

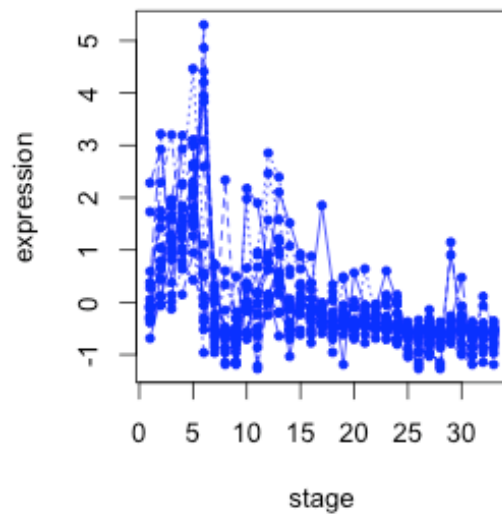
1st PC in worm magenta



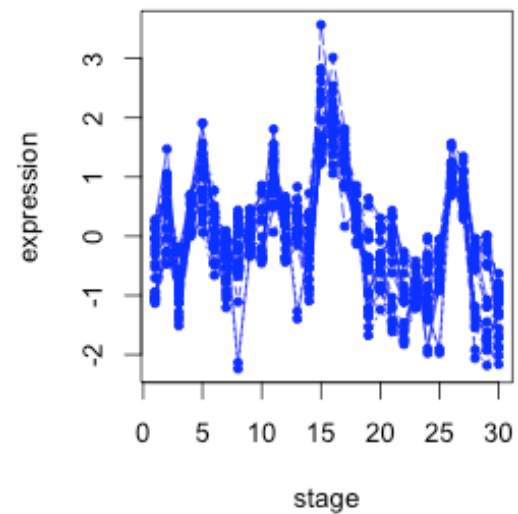
1st PC in fly salmon



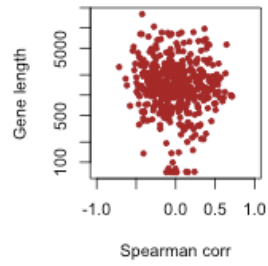
high cor worm ncRNAs(18)



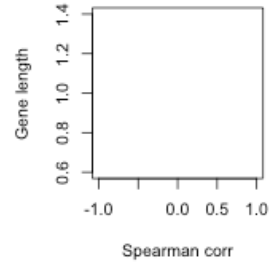
high cor fly ncRNAs(21)



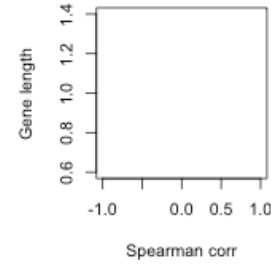
w-magenta-pseudogene-493



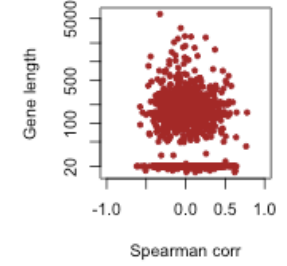
w-magenta-rRNA-0



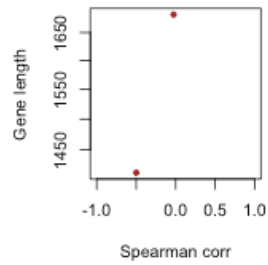
w-magenta-snrRNA-0



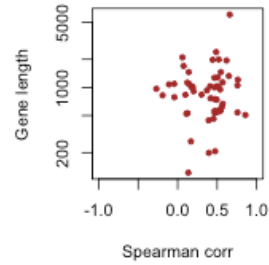
w-magenta-ncRNA-1988



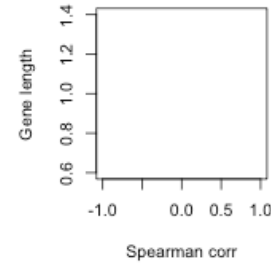
f-salmon-pseudogene-2



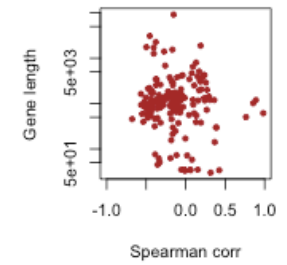
f-salmon-rRNA-55



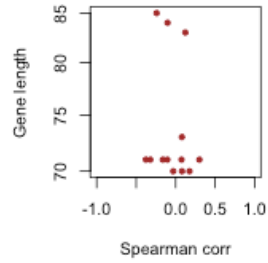
f-salmon-snrRNA-0



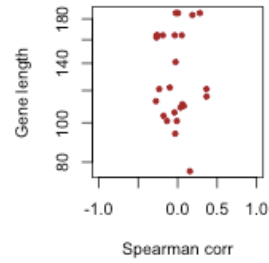
f-salmon-ncRNA-175



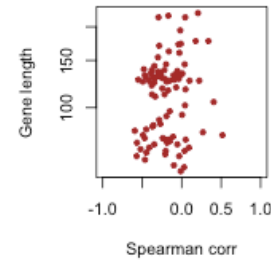
w-magenta-tRNA-13



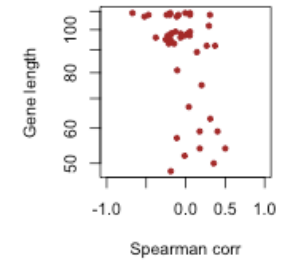
w-magenta-snrRNA-25



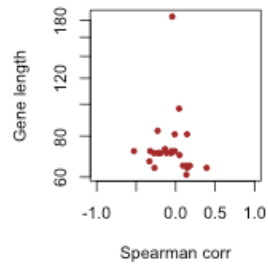
w-magenta-snoRNA-99



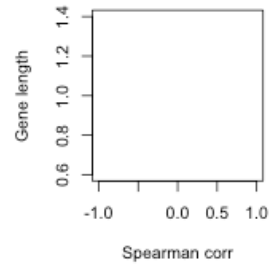
w-magenta-miRNA-46



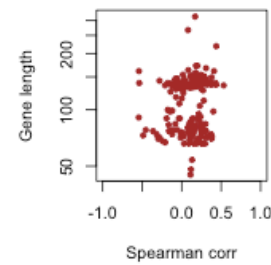
f-salmon-tRNA-25



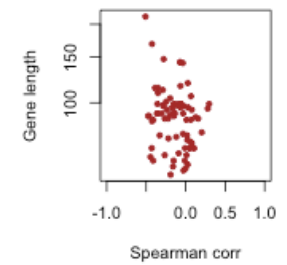
f-salmon-snrRNA-0



f-salmon-snoRNA-201

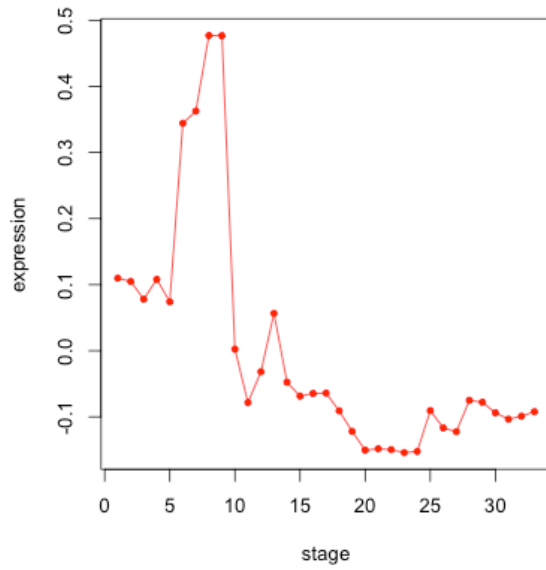


f-salmon-miRNA-77

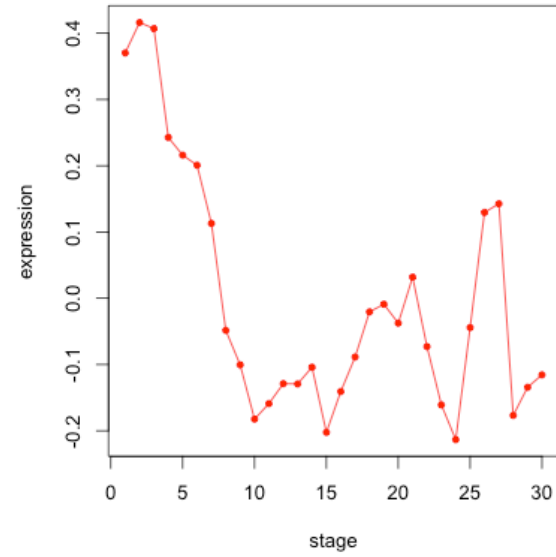


No.2 Worm-Fly module pair

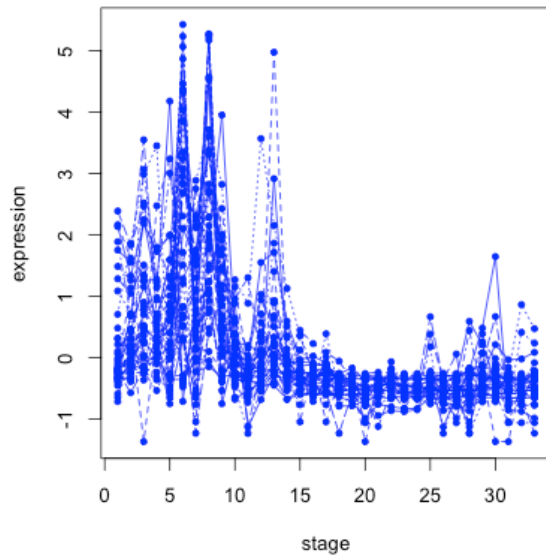
1st PC in worm turquoise



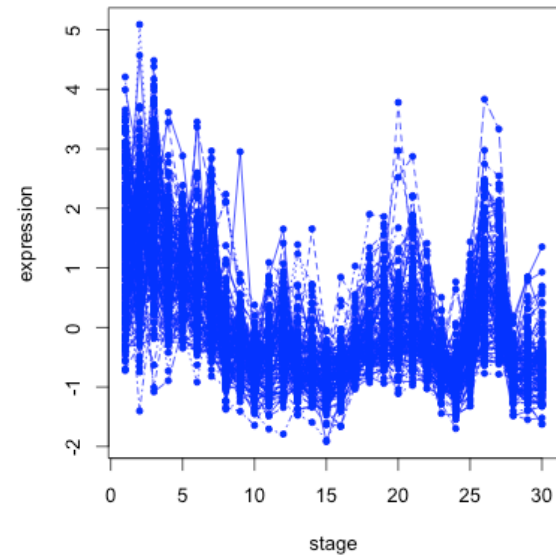
1st PC in fly turquoise



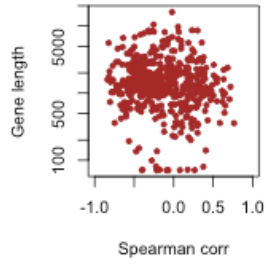
high cor worm ncRNAs(38)



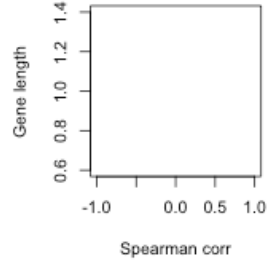
high cor fly ncRNAs(160)



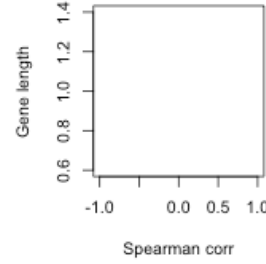
w-turquoise-pseudogene-493



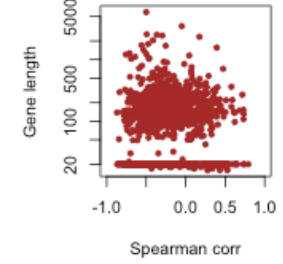
w-turquoise-rRNA-0



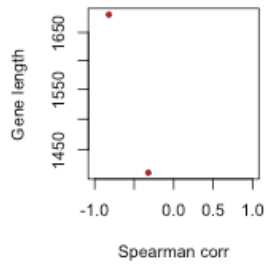
w-turquoise-snrRNA-0



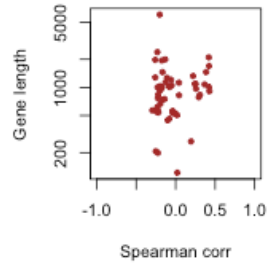
w-turquoise-ncRNA-1988



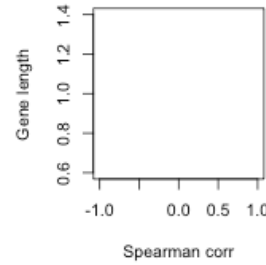
f-turquoise-pseudogene-2



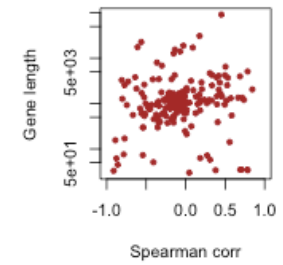
f-turquoise-rRNA-55



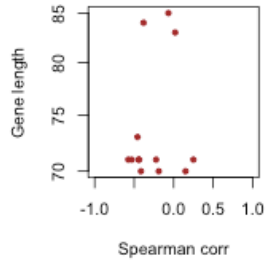
f-turquoise-snrRNA-0



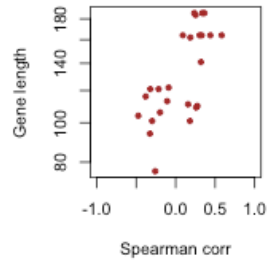
f-turquoise-ncRNA-175



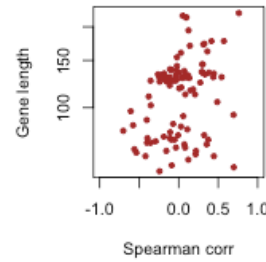
w-turquoise-tRNA-13



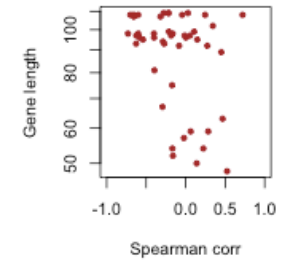
w-turquoise-snrRNA-25



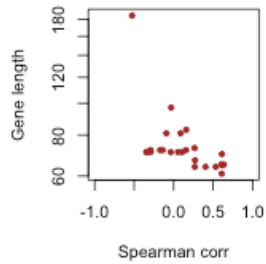
w-turquoise-snoRNA-99



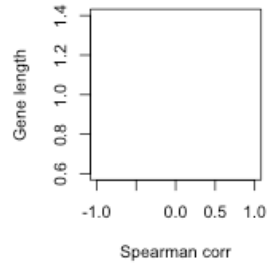
w-turquoise-miRNA-46



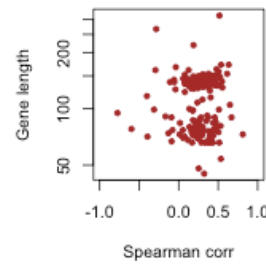
f-turquoise-tRNA-25



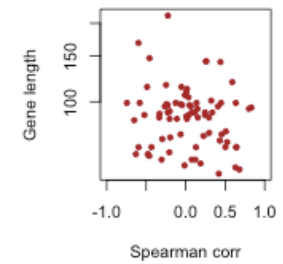
f-turquoise-snrRNA-0



f-turquoise-snoRNA-201

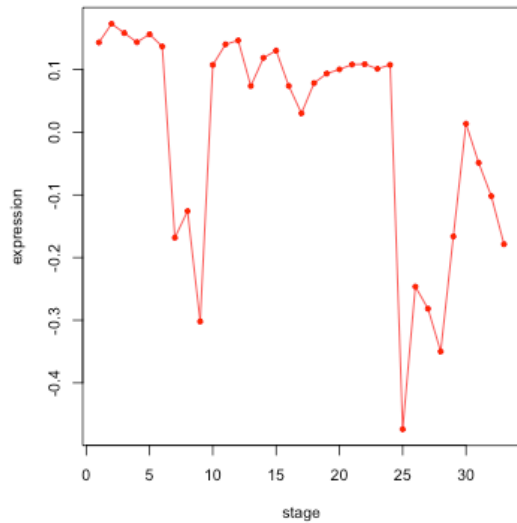


f-turquoise-miRNA-77

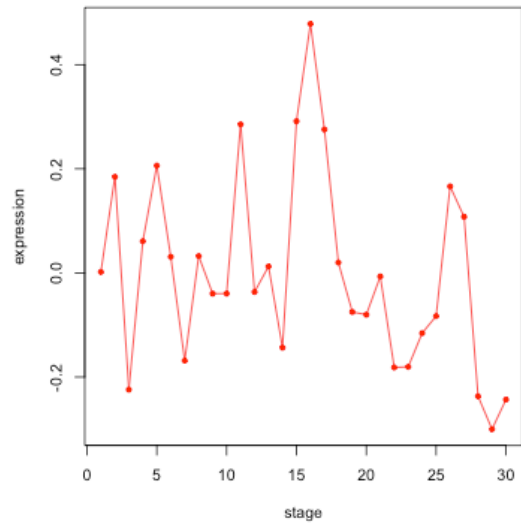


No.1 W-F-H module triplet

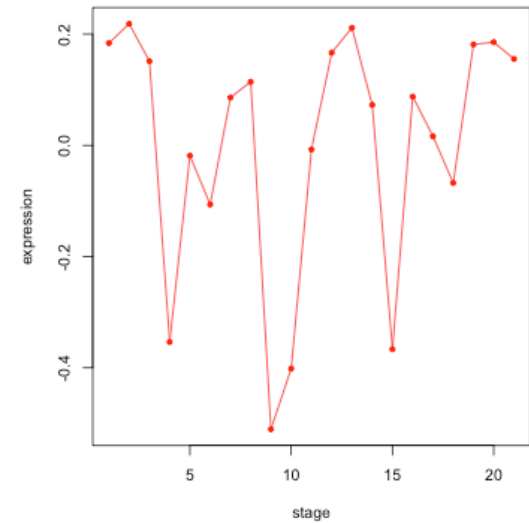
1st PC in worm magenta



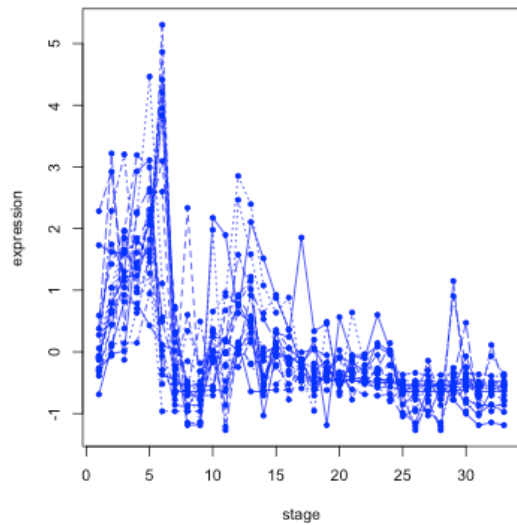
1st PC in fly salmon



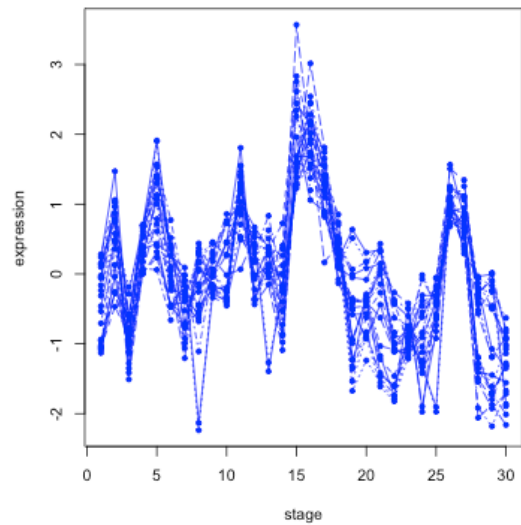
1st PC in human yellow



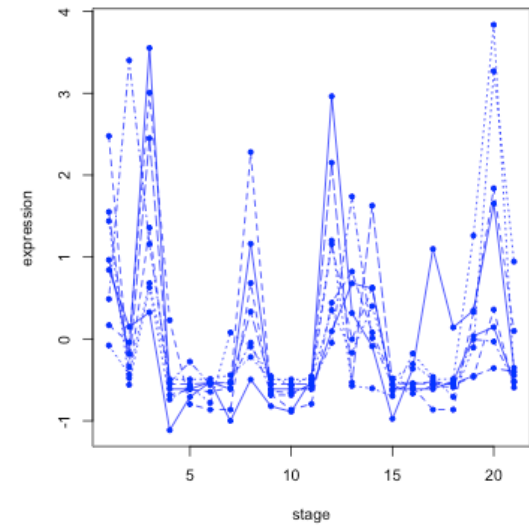
high cor worm ncRNAs(18)

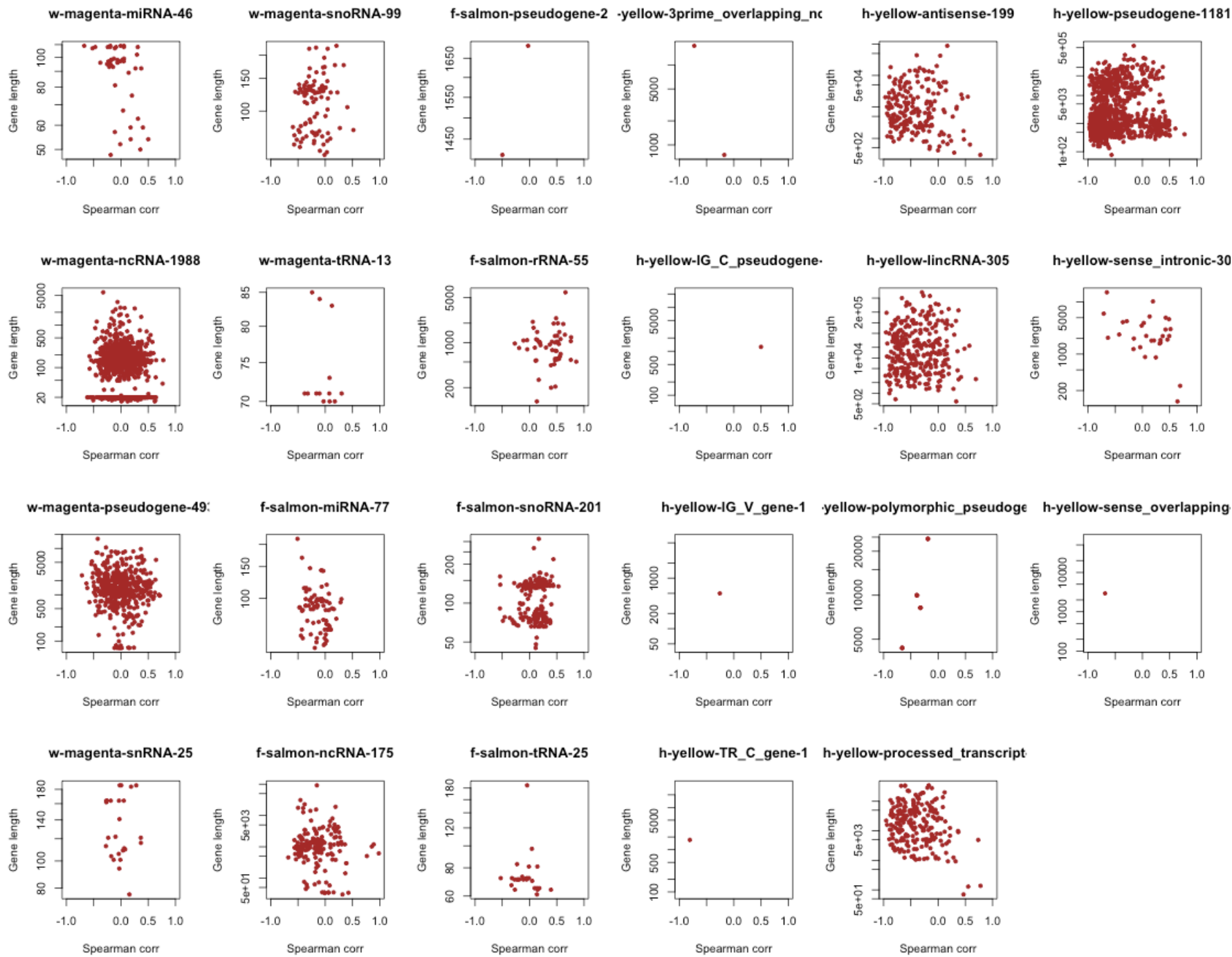


high cor fly ncRNAs(21)

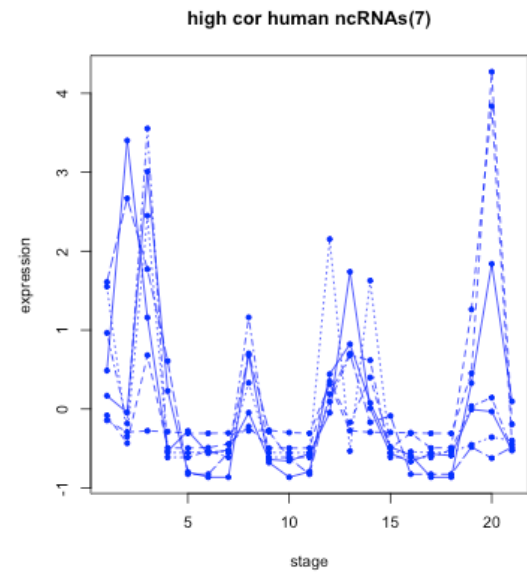
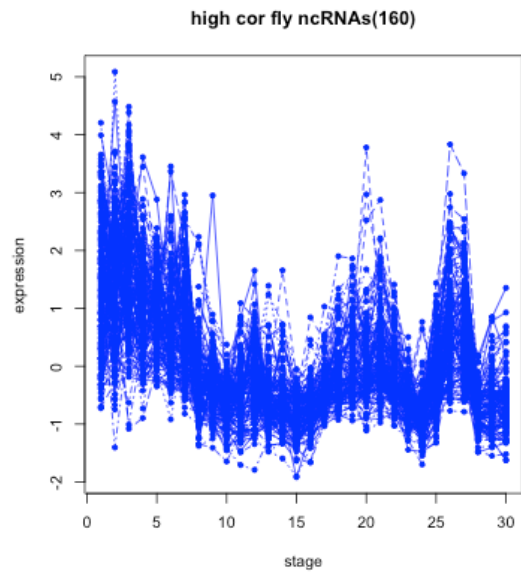
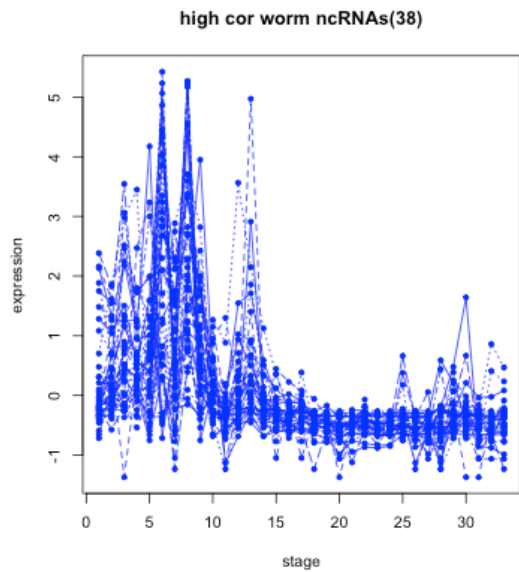
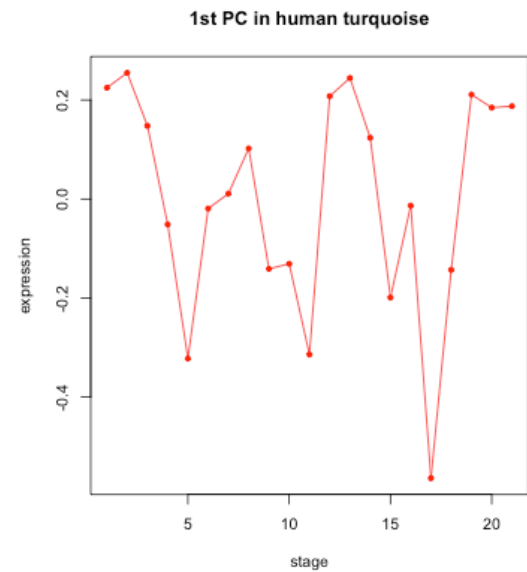
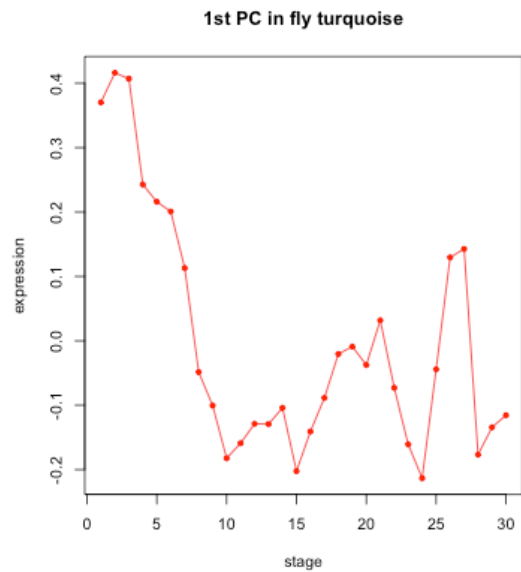
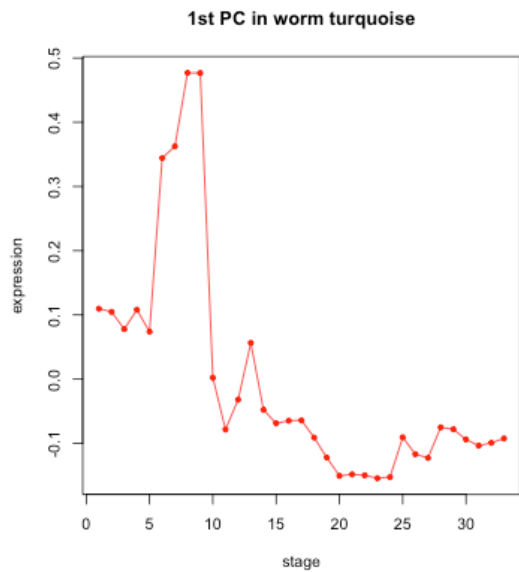


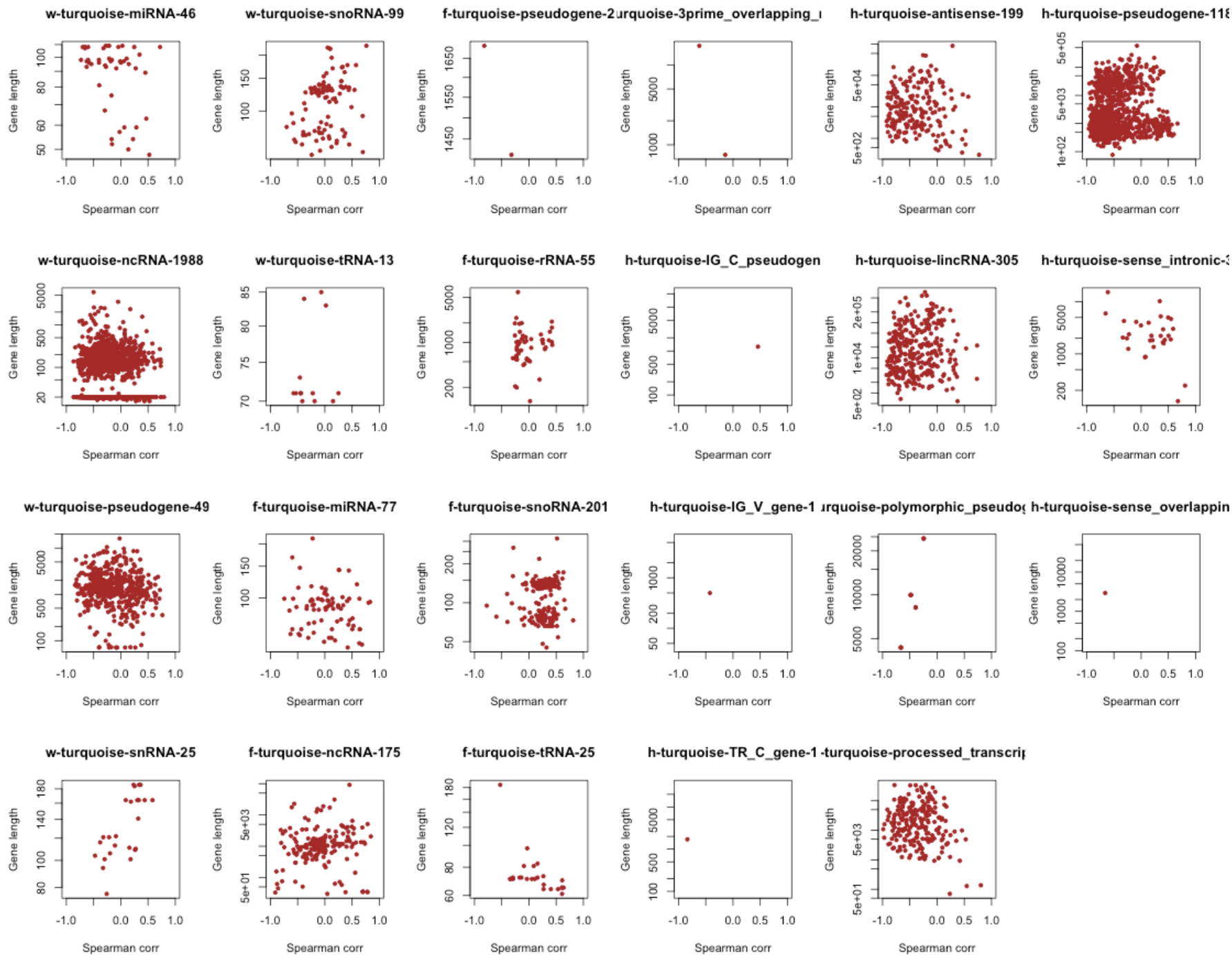
high cor human ncRNAs(8)





No.2 W-F-H module triplet





Future work

- Annotations of modules that “orthologous” ncRNAs highly correlate
- KKY’s model
 - Expect to increase overlap among worm, fly and human modules
- Other characteristics of ncRNAs besides length
 - vs. correlations with orthologs
- Any other suggestions...