

mod/ENCODE Transcription
Comparison Paper
Figure Slide Pack

1 Mar 2013

Transcriptome Comparison Group

Exhibit 1: Data Sets

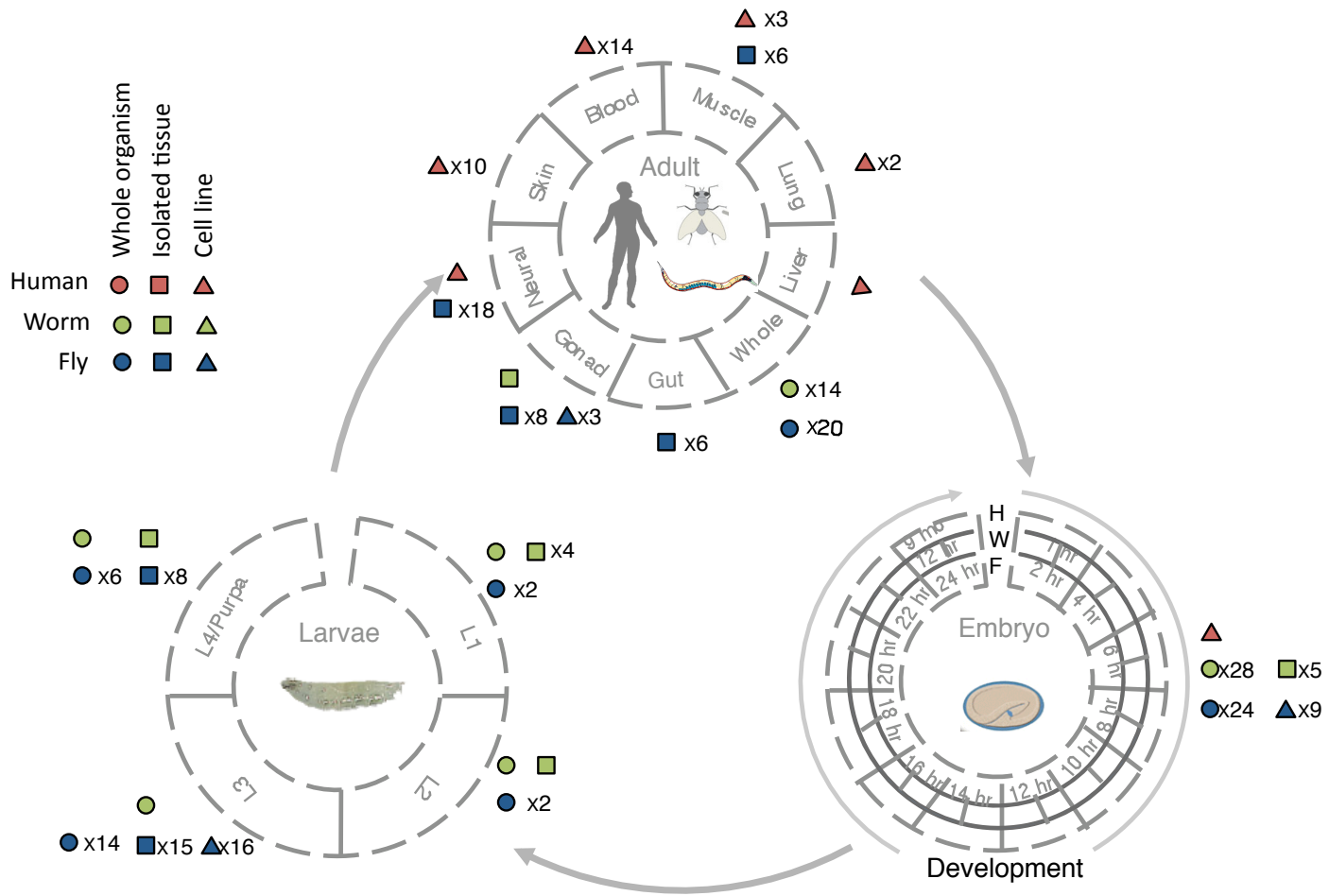


Exhibit 2: Summary Statistics for Protein Coding Genes

A

	Annotated Bases (Millions, % of genome)										
	mRNA								Intron		
	5' UTR		CDS		3' UTR		Total				
Human	17.6	0.6%	57.4	1.9%	7.3	0.2%	82.4	2.7%	1,409	46%	
Worm	4.2	4.2%	26.3	26%	4.1	4.1%	34.6	35%	33.5	33%	
Fly	6.9	5.0%	22.9	16%	6.3	4.5%	36.1	26%	51.4	37%	

	Mapped Reads (Millions, % of total)										
	mRNA								Intron		
	5' UTR		CDS		3' UTR		Total				
Human	1,384	13%	6,438	62%	292	2.8%	8,114	78%	1,031	9.9%	
Worm	107	4.7%	1,961	86%	156	6.8%	2,224	97%	21.0	0.9%	
Fly	378	5.5%	4,917	71%	1236	18%	6,532	94%	84.3	1.2%	

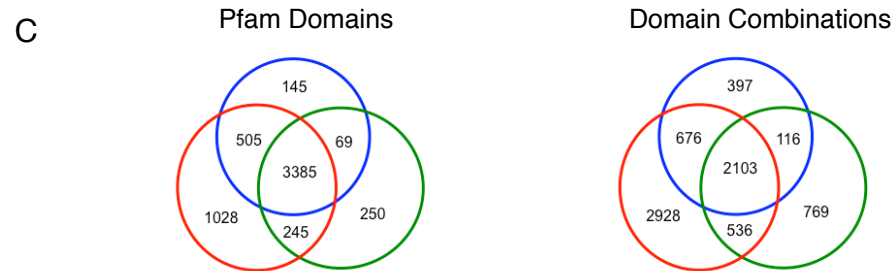
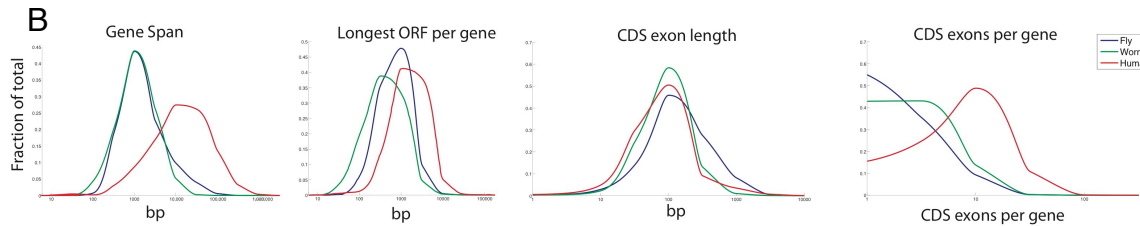


Exhibit 3: Splicing Figure

Figure 2 - 2/5/13

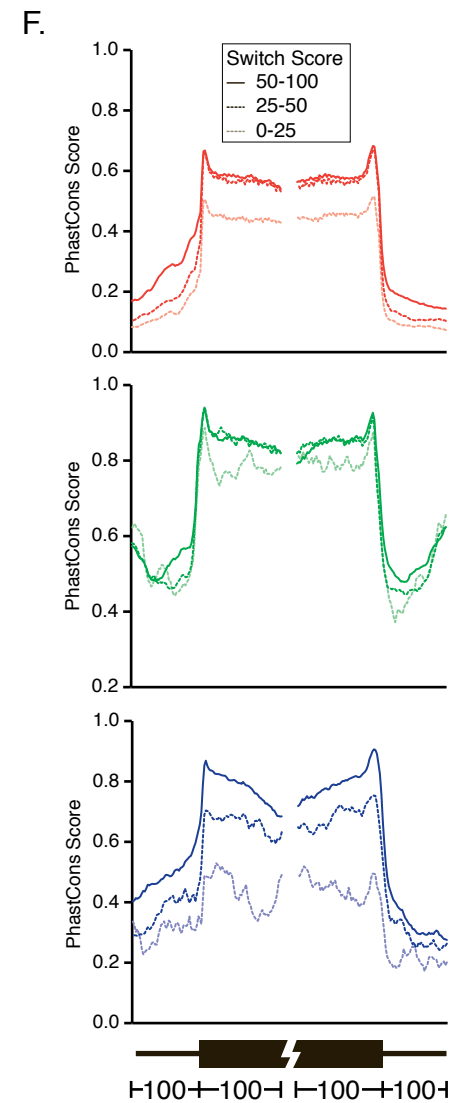
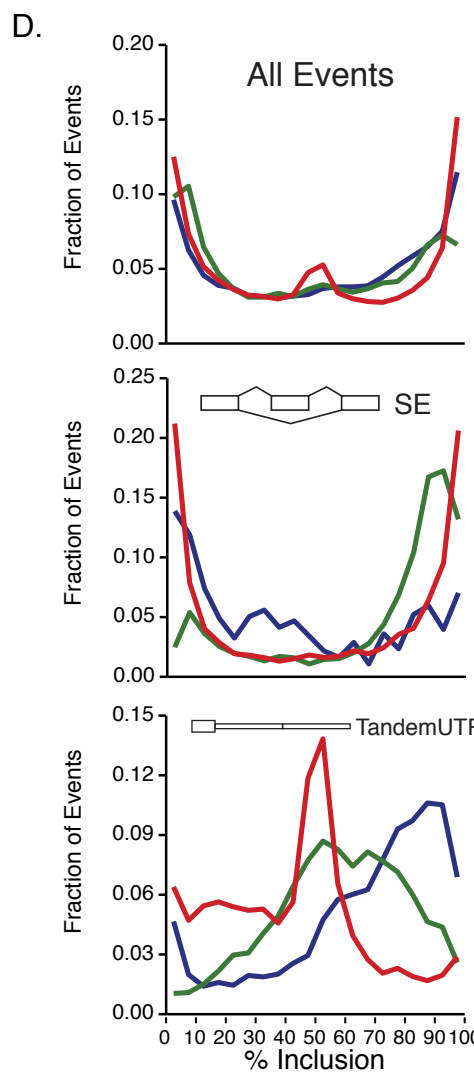
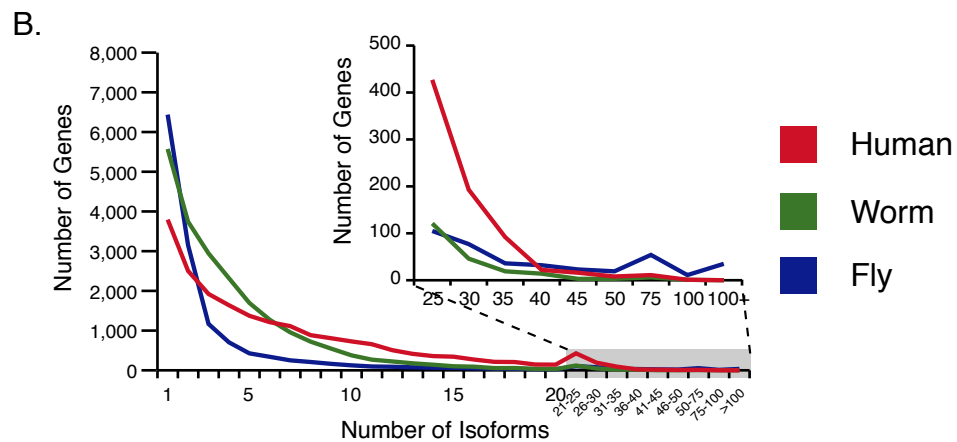
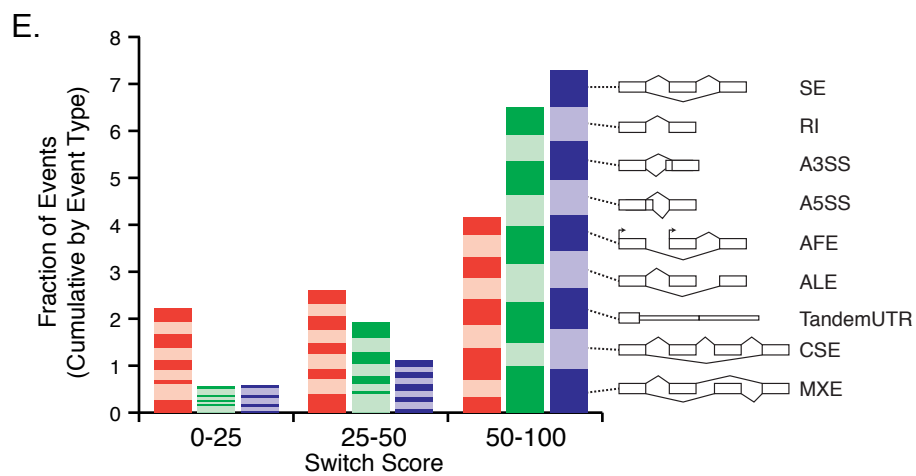
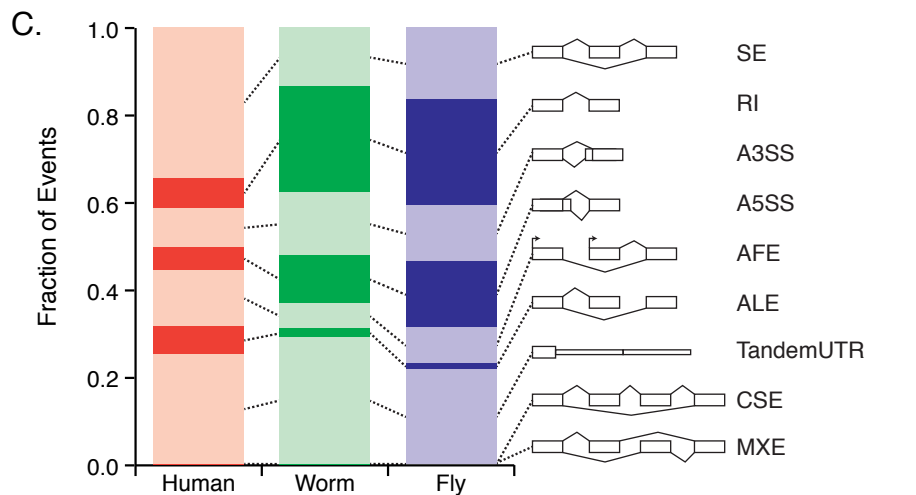
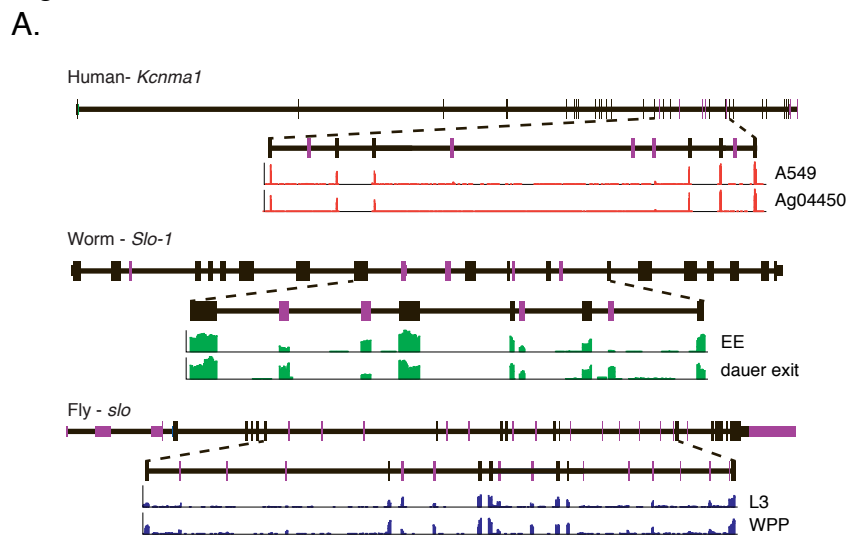
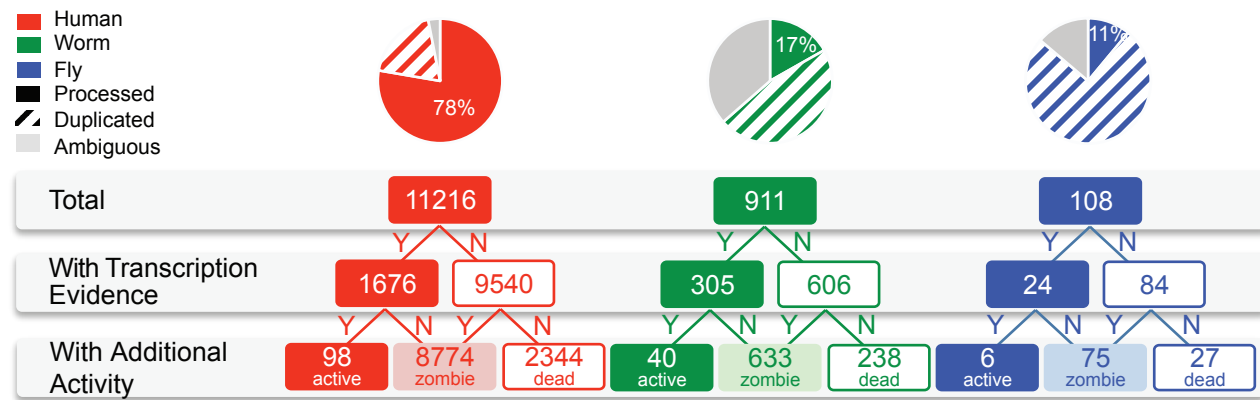
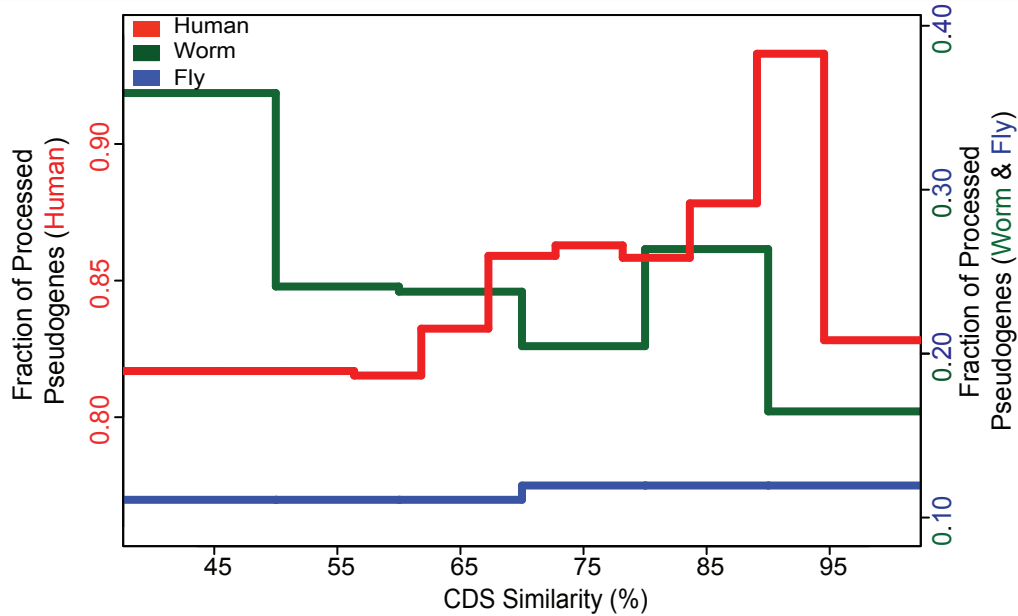


Exhibit 4: Pseudogenes

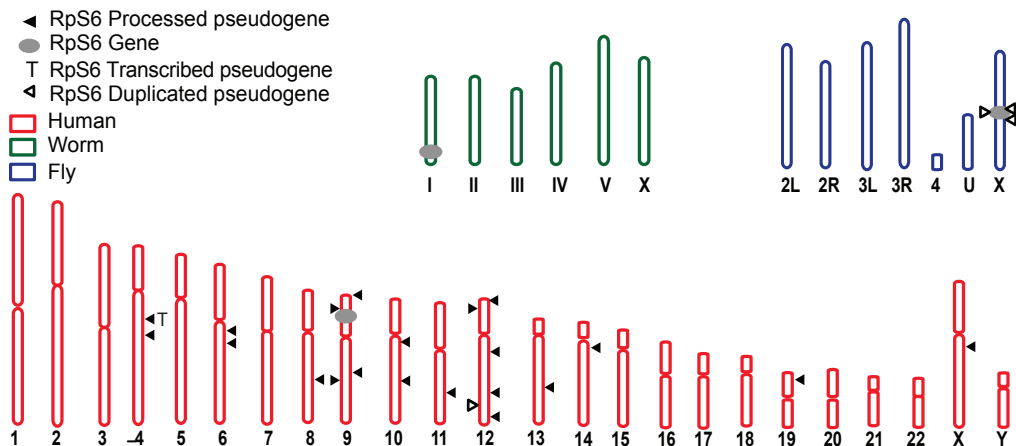
A



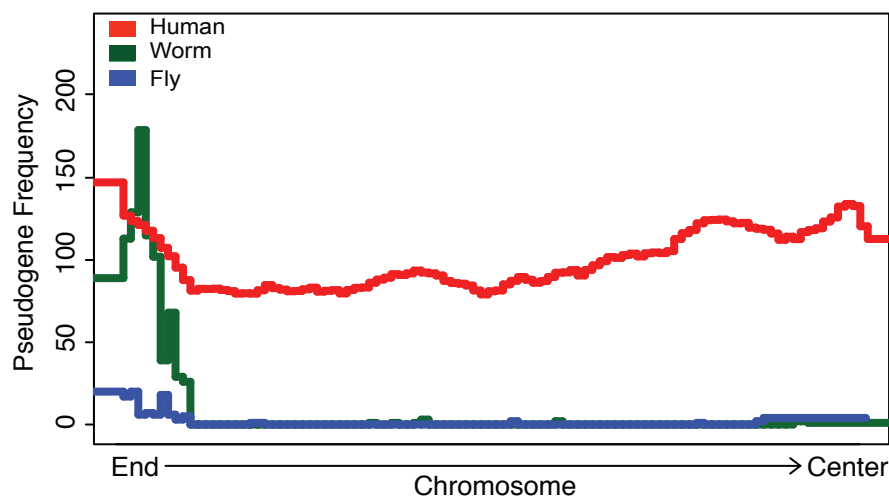
B



C



D



E

Fly	Worm	Human	Type
		414	7tm ◀
3		235	RRM
2	1	180	IG ◀
5	1	176	TF
2		171	IG
		144	Ribo
		132	TF
		132	Ribo
		123	Struct
		117	TF
3	2	93	Kin
		93	Ploop
		93	Ploop
		86	Ribo
		86	Ribo
		84	Ubiq
		83	TF
		82	•
		82	Ribo
		76	•
		74	•
		71	•
3		69	H
7	1	68	Ribo
		68	Ribo
		66	GDH
1		64	Ribo
1		64	Ribo
		62	Struct
		61	GDH
		60	Ribo
		59	Ribo
		58	Kin
2	1	58	•
		58	Ploop
		57	Ribo
	1	55	Ubiq ◀
		55	Ribo
		54	Ubiq
		54	Ribo
		53	Struct
		50	TF ◀
		49	Struct
		49	Struct
		48	Ribo
		48	Ribo
		47	TF
		46	•
2		45	Ploop
2	3	44	Ploop
		44	•
	1	43	Ploop
		43	Ribo
1		42	Ploop
		42	•
		41	•
		41	Struct
		22	•
3	9	12	Kin ◀
	10	12	Kin ◀
	6	9	Ubiq
	17	8	•
	24	2	7tm ◀
	26	1	7tm ◀
	74		7tm ◀
	46		7tm ◀
	23		Ubiq ◀
	20		7tm ◀
	17		7tm ◀
	16		Struct
	13		7tm ◀
	11		7tm ◀
	7		•
	6		•
	6		•
30		7	SAP ◀
10		6	Motor ◀
10			•
9			Kin
8		1	•

Exhibit 5: Comparison of ncRNAs

A)

		Human			Worm			Fly			
		Elements	Genome Coverage (Kb [%])	Mapped Reads (fraction)	Elements	Genome Coverage (Kb [%])	Mapped Reads (fraction)	Elements	Genome Coverage (Kb [%])	Mapped Reads (fraction)	
Annotated ncRNAs	Comparable ncRNAs	piRNA loci	88	1,272 [0.04]	3.2 e-4	35,322	449 [0.45]	6.7 e-3	21	1,604 [1.2]	1.6 e-3
		pri-miRNA	58	1,158 [0.04]	3.6 e-4	44	16 [0.02]	6.6 e-6	43	316 [0.25]	1.7 e-4
		pre-miRNAs	1,756	162 [0.006]	2.7 e-3	221	20 [0.02]	2.1 e-4	235	22 [0.02]	7.2 e-5
		tRNAs	624	47 [0.002]	3.0 e-4	638	45 [0.04]	1.3 e-5	308	22 [0.02]	1.4 e-6
		snoRNAs	1,521	168 [0.006]	3.3 e-4	141	16 [0.02]	2.9 e-4	287	34 [0.03]	2.9 e-4
		snRNAs	1,944	210 [0.007]	4.6 e-5	114	14 [0.01]	4.9 e-5	47	7 [0.006]	8.5 e-5
		lncRNAs	10,840	10,581 [0.37]	2.2 e-2	417	184 [0.18]	8.3 e-4	558	868 [0.68]	1.3 e-2
	Other ncRNAs	9,281	2,624 [0.091]	3.9 e-3	2,935	530 [0.53]	2.9 e-3	398	381 [0.30]	2.1 e-3	
Total ncRNAs		26,112	16,103 [0.56]	2.9 e-2	39,832	1,268 [1.26]	1.1 e-2	1,897	3,022 [2.35]	1.6 e-2	
Transcriptionally Unannotated Regions		294,956	2,749,660 [96.1]	2.0 e-1	146,913	64,232 [64.1]	5.1 e-2	60,267	89,630 [69.8]	5.5 e-2	
	TARs	718,537	923,774 [32.3]	2.0 e-1	235,915	37,482 [37.4]	4.6 e-2	76,757	47,542 [37.0]	5.5 e-2	
	Supervised Predictions	104,001	13,844 [0.51]	1.2 e-2	2,509	390 [0.77]	2.5 e-4	599	164 [0.004]	2.4 e-5	
	Overlapping TEs	1,633,407	455,803 [15.9]	8.0 e-2	65,600	5,621 [5.6]	2.3 e-3	38,595	7,414 [5.8]	2.9 e-2	
	Overlapping Introns	1,300,035	360,778 [12.6]	1.2 e-1	173,253	18,347 [18.3]	2.5 e-2	79,012	28,857 [22.5]	1.9 e-2	

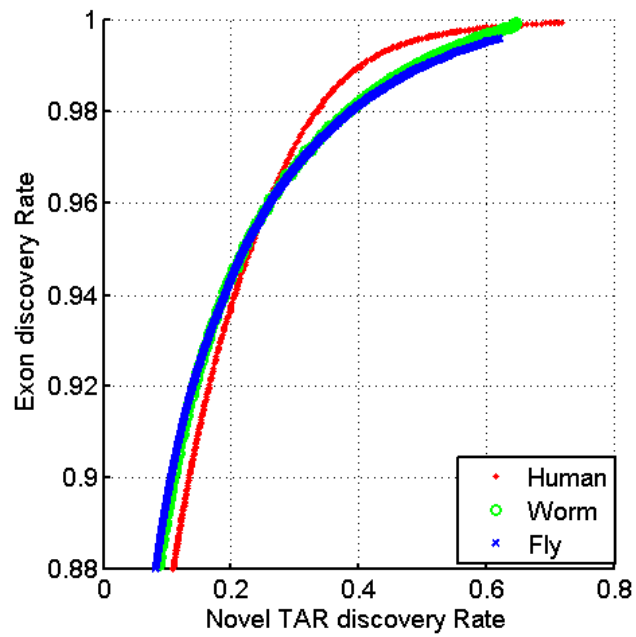
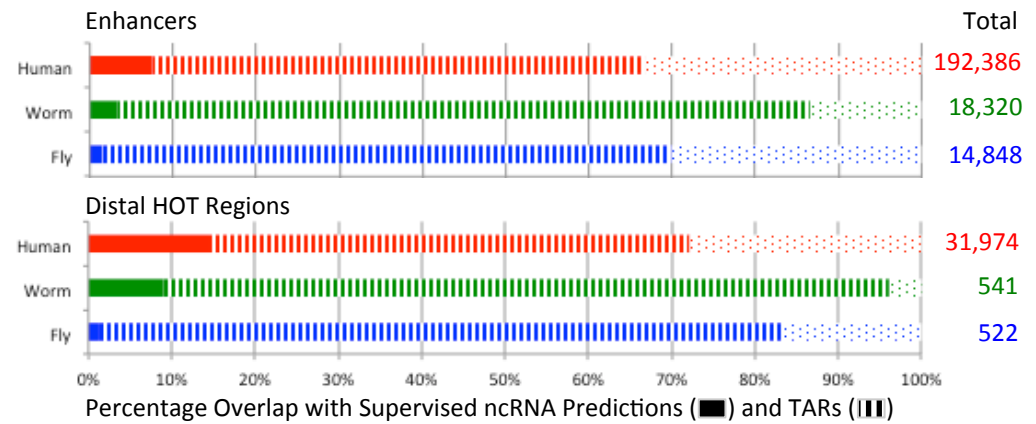
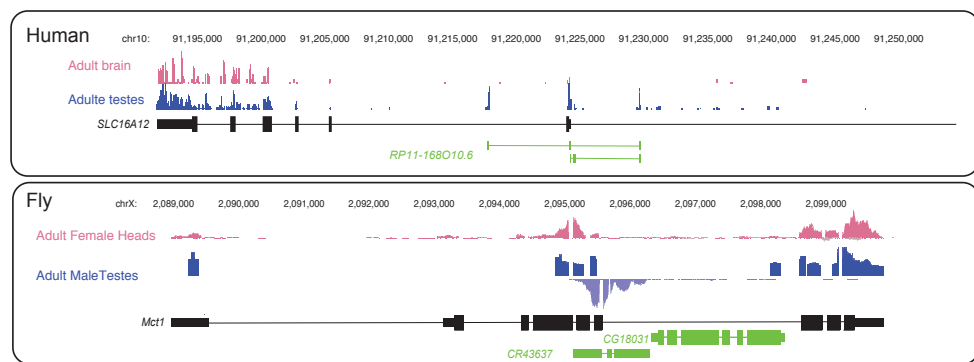
B)**C)****D)**

Exhibit 6: Expression Clustering of Coding Genes and ncRNAs

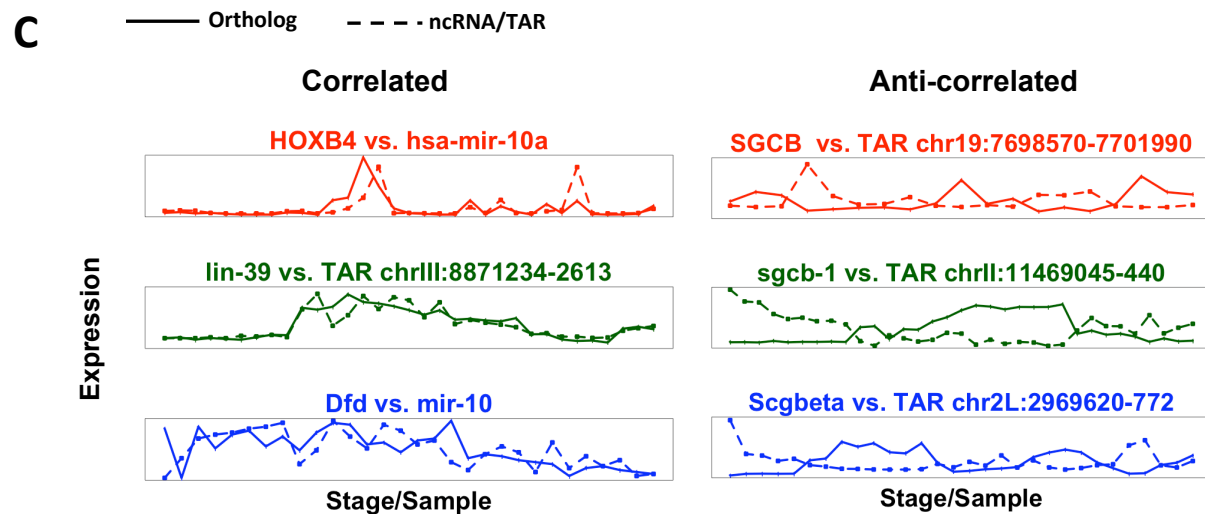
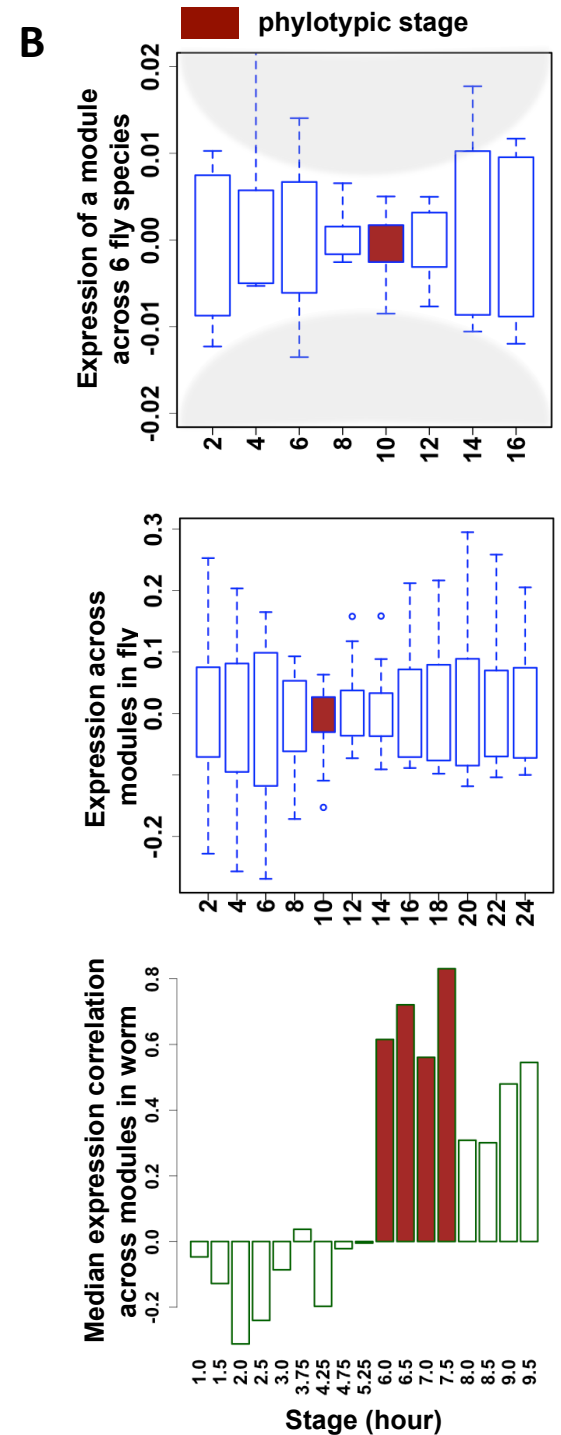
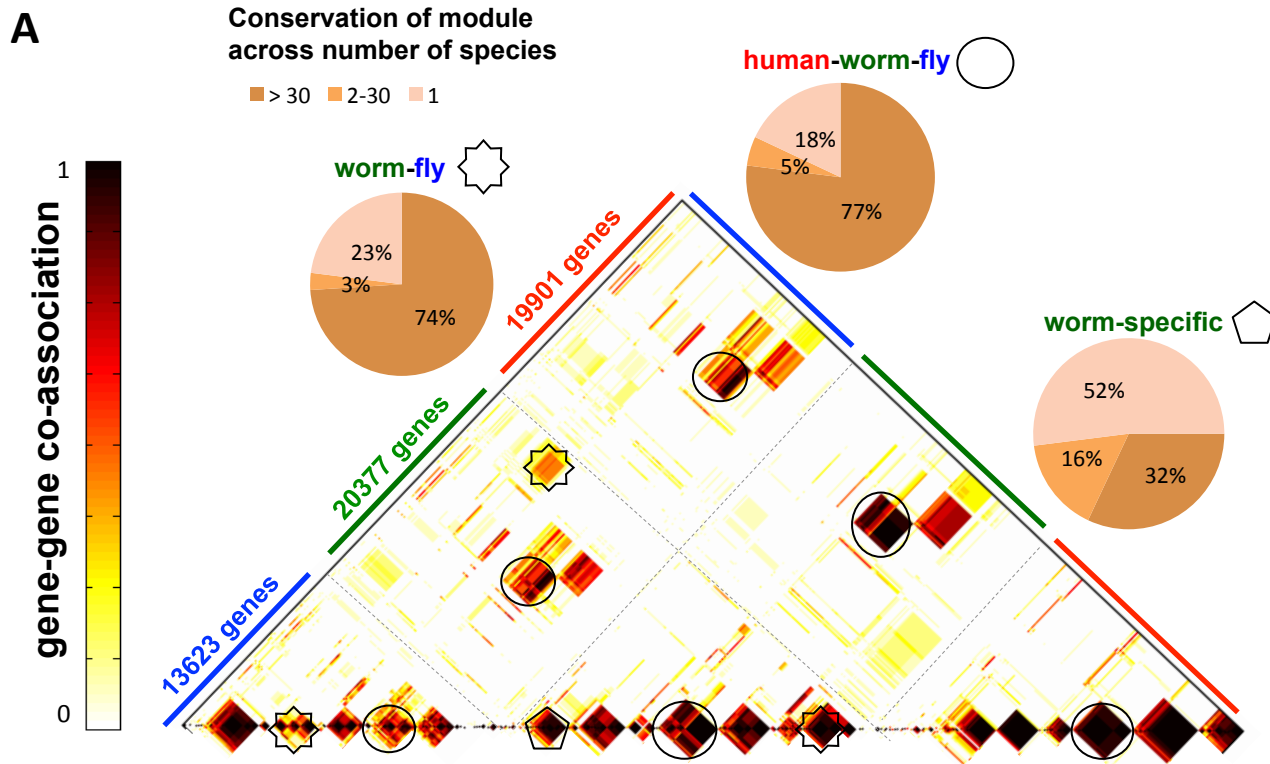


Exhibit 7: Stage Mapping of Worm and Fly Developmental Timecourses

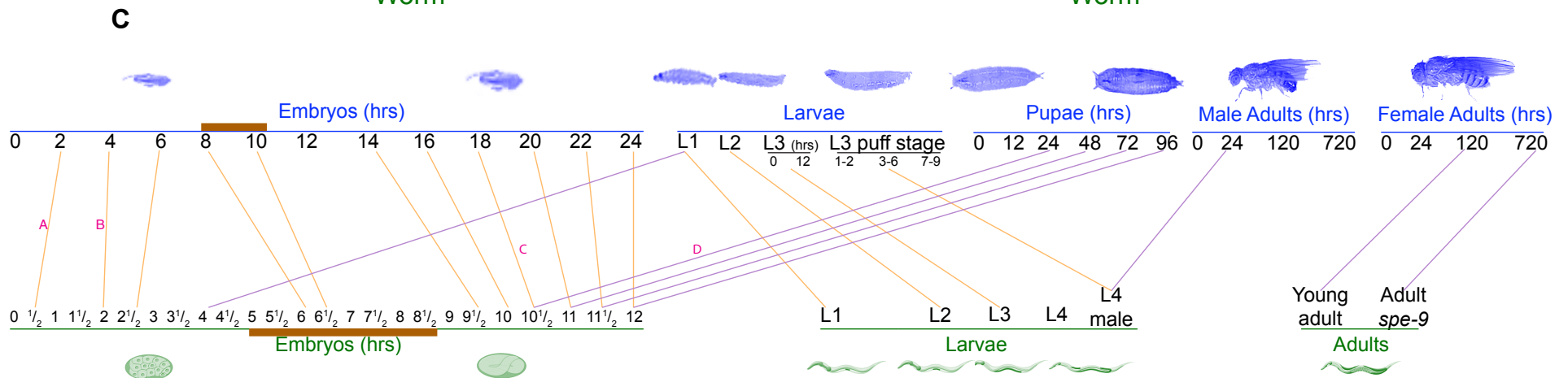
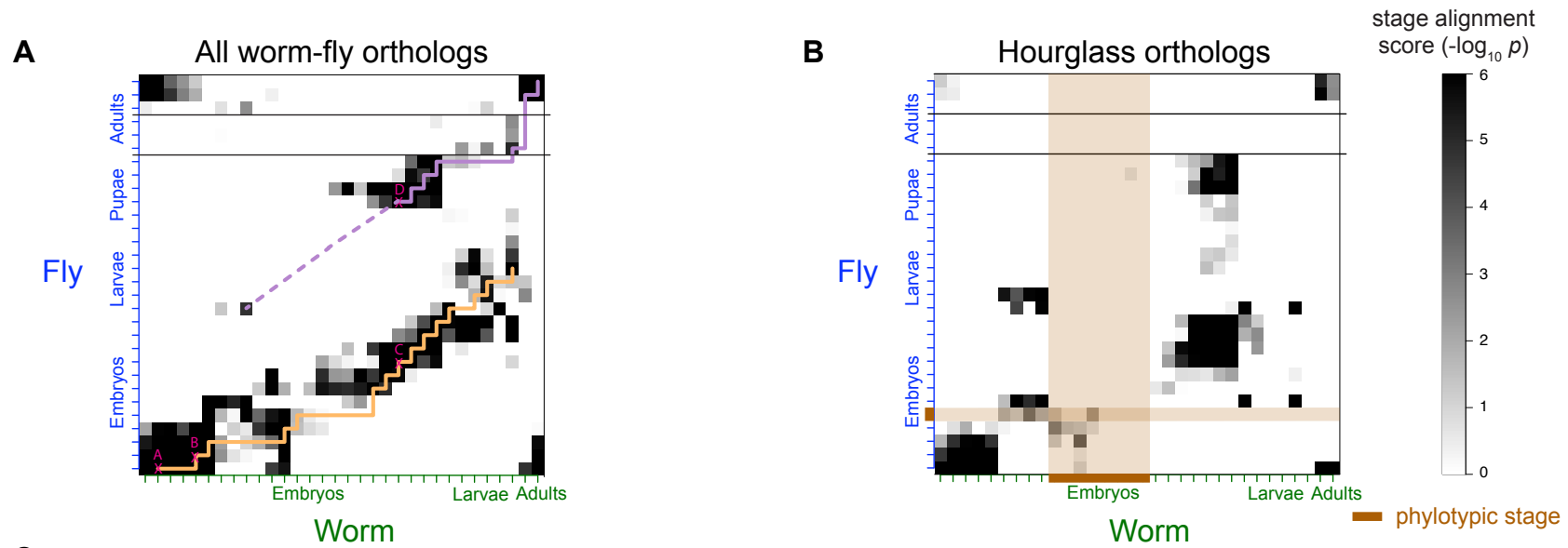


Exhibit 8: Histone and TF Models for Gene Expression

