

mod/ENCODE Transcription
Comparison Paper
Figure Slide Pack

11/30/2012

Transcriptome Comparison Group

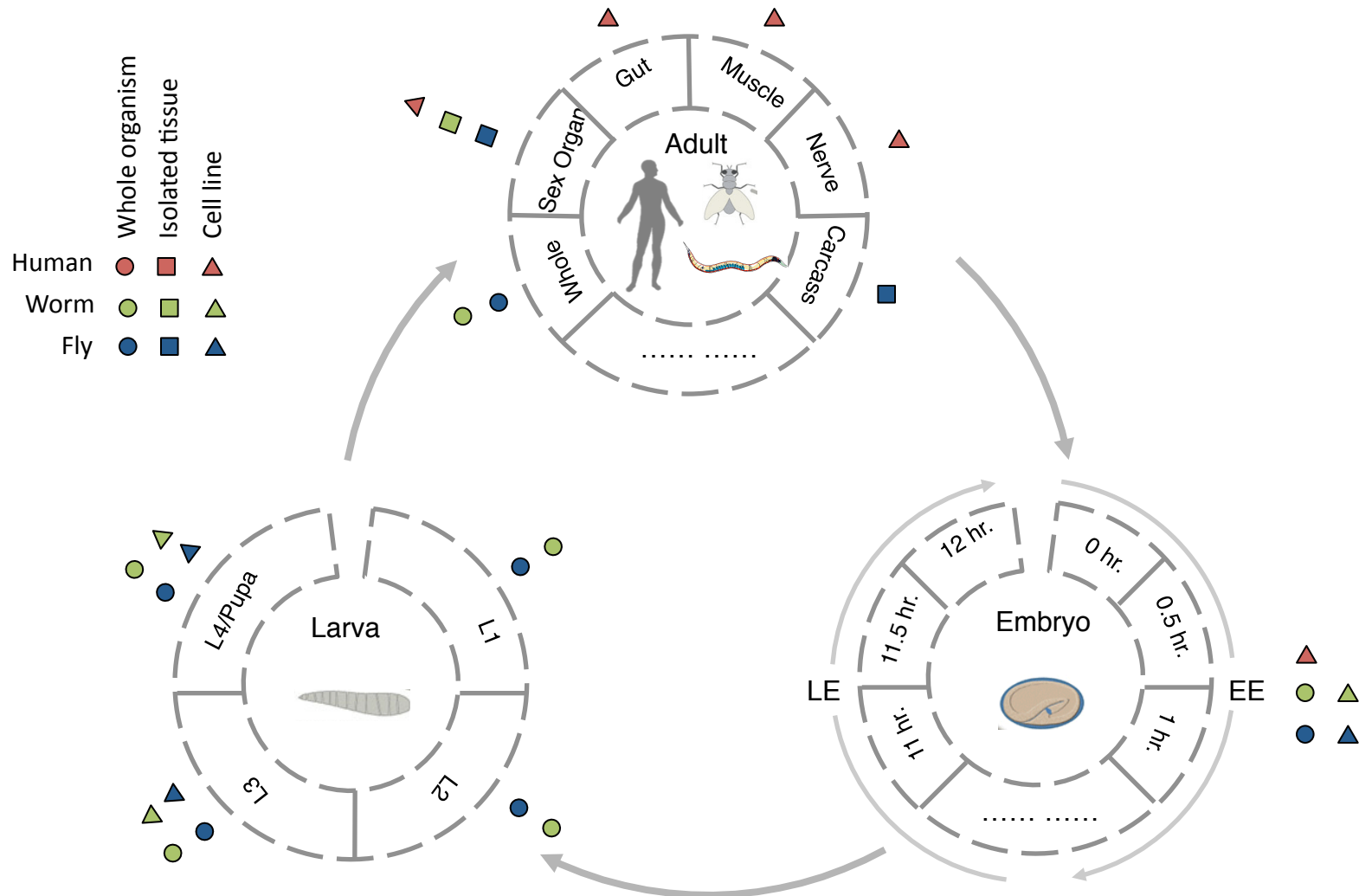


Exhibit 1: Data Sets

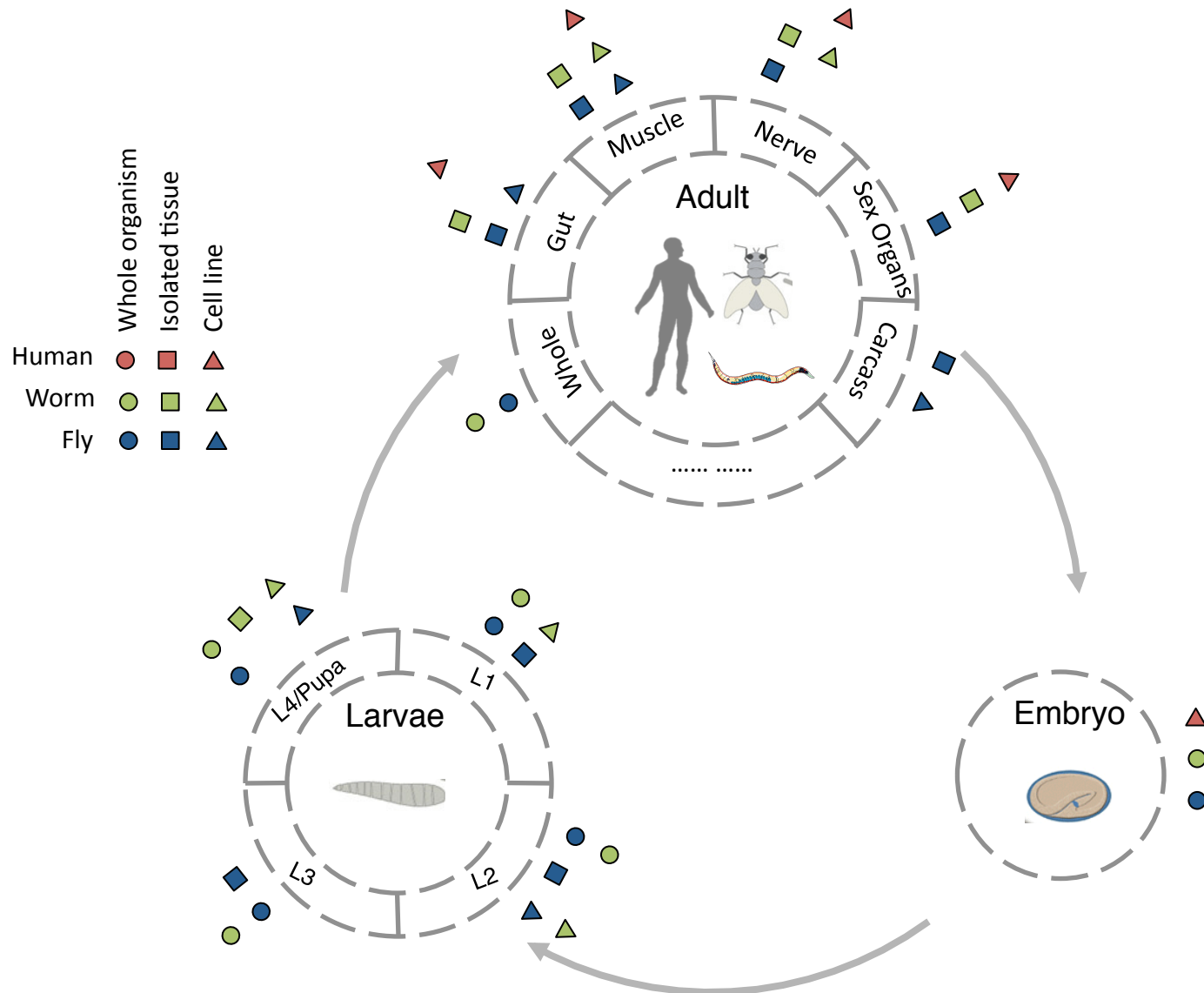


Exhibit 2: Summary Statistics for Protein Coding Genes

Organism	CDS			5' UTR			3' UTR			Intron			Intergenic*		
	nts	% of genome	% of CDS detected by PolyA RNAseq	nts	% of genome	% of 5' UTR detected by PolyA RNAseq	nts	% of genome	% of 3' UTR detected by PolyA RNAseq	nts	% of genome	% of Introns detected by PolyA RNAseq	nts	% of genome	% of Intergenic Sequence detected by PolyA RNAseq
Human	34,902,359	1.1	93	10,891,135	0.4	91	30,987,688	1.0	95	1,169,472,184	37.4	64	1,877,412,271	60.8	17
Fly	22,784,848	16.3	97.7	2,965,559	2.1	98.8	4,030,758	2.9	95.7	48,675,555	34.8	91.3	61,421,763	44.0	36.4
Worm	26,629,887	26.6		4,712,660	4.7		4,180,415	4.2		33,587,432	33.5		31,161,882	31.1	

Domain structure complexity

***D. melanogaster* (10802)**

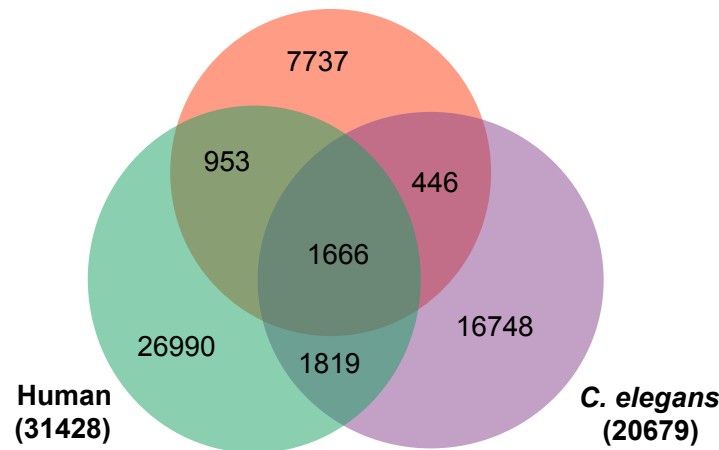


Exhibit 3: Splicing Figure

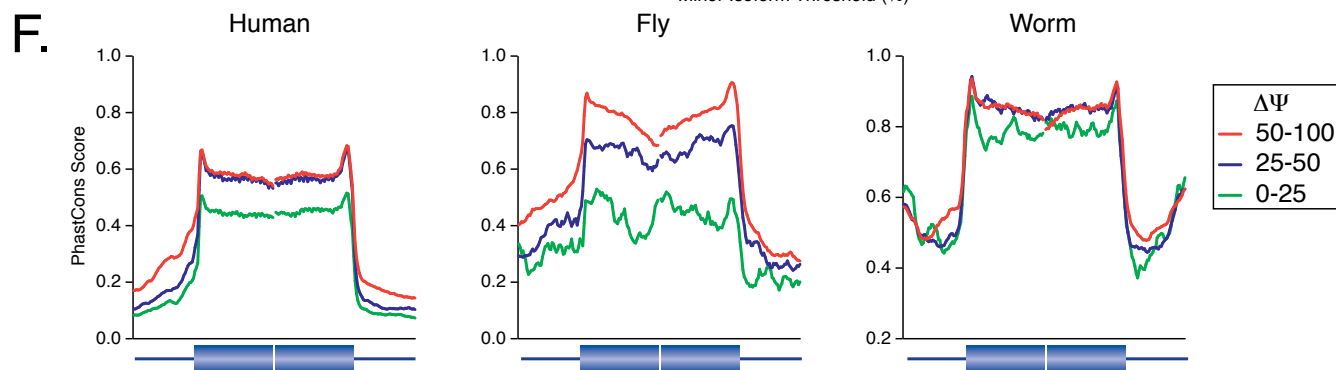
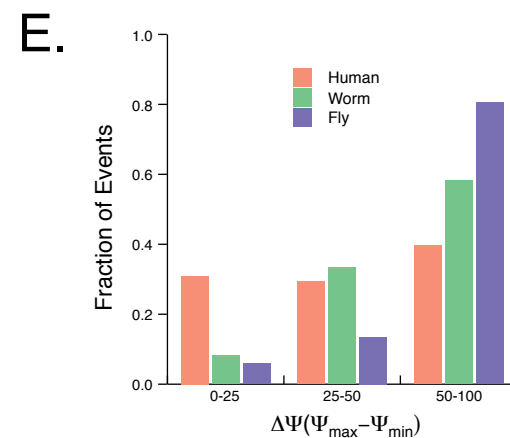
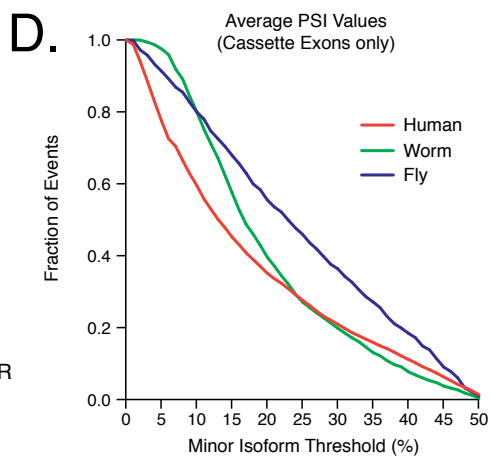
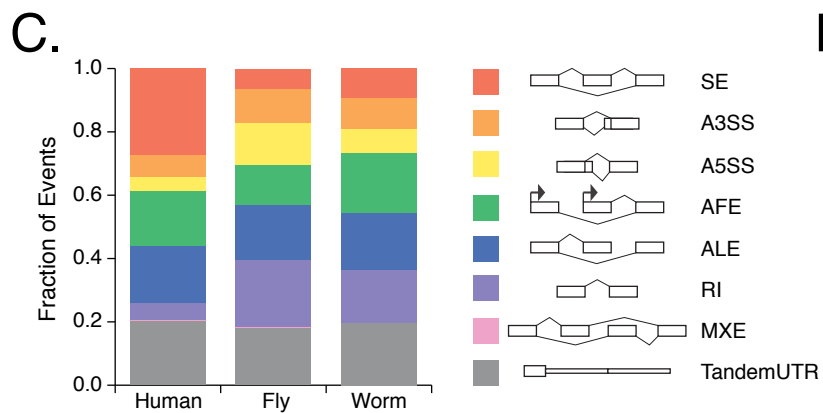
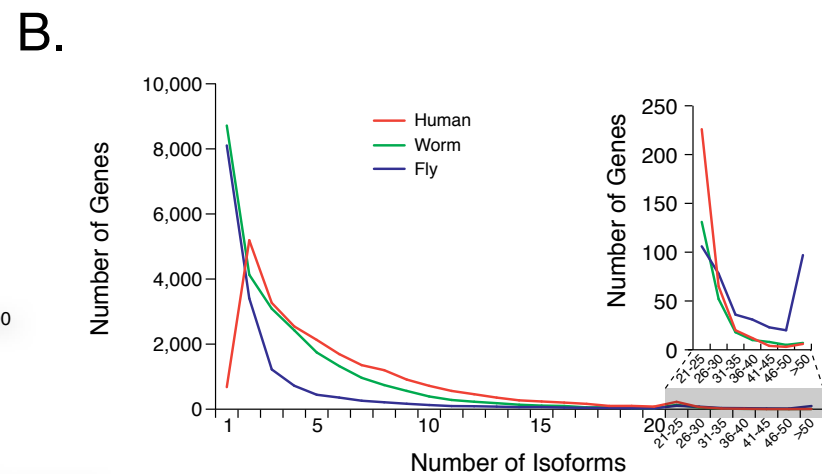
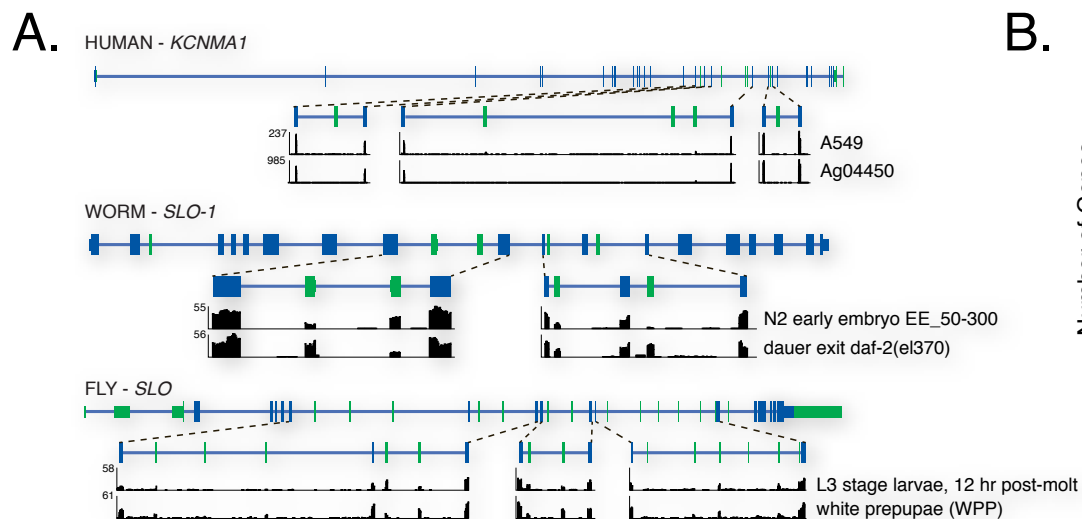
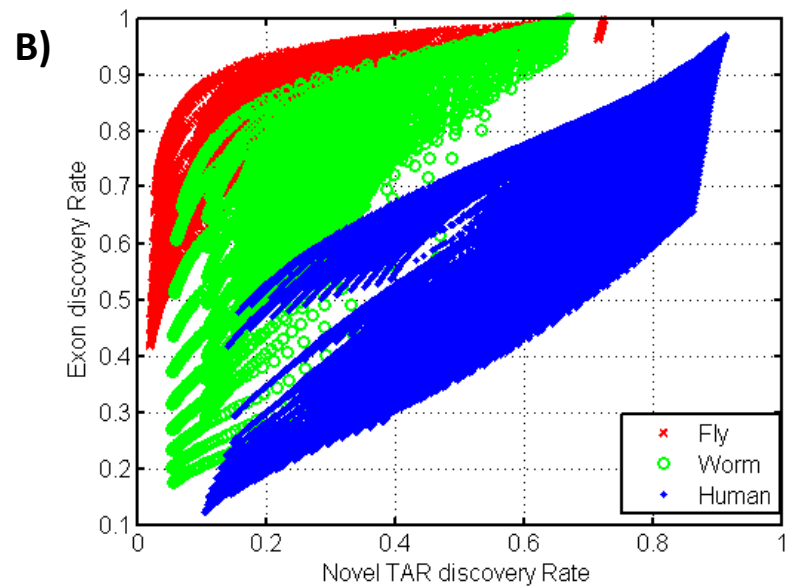
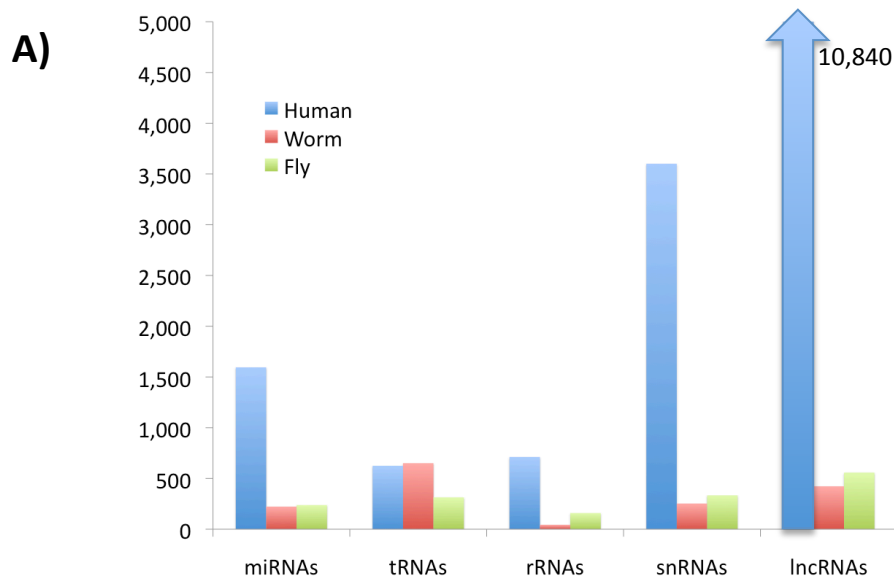


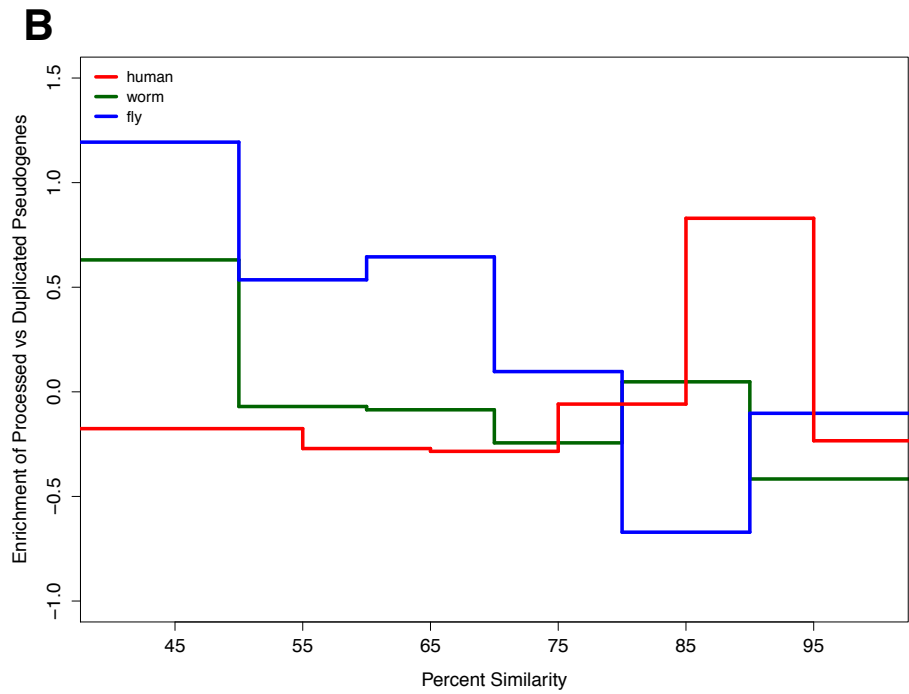
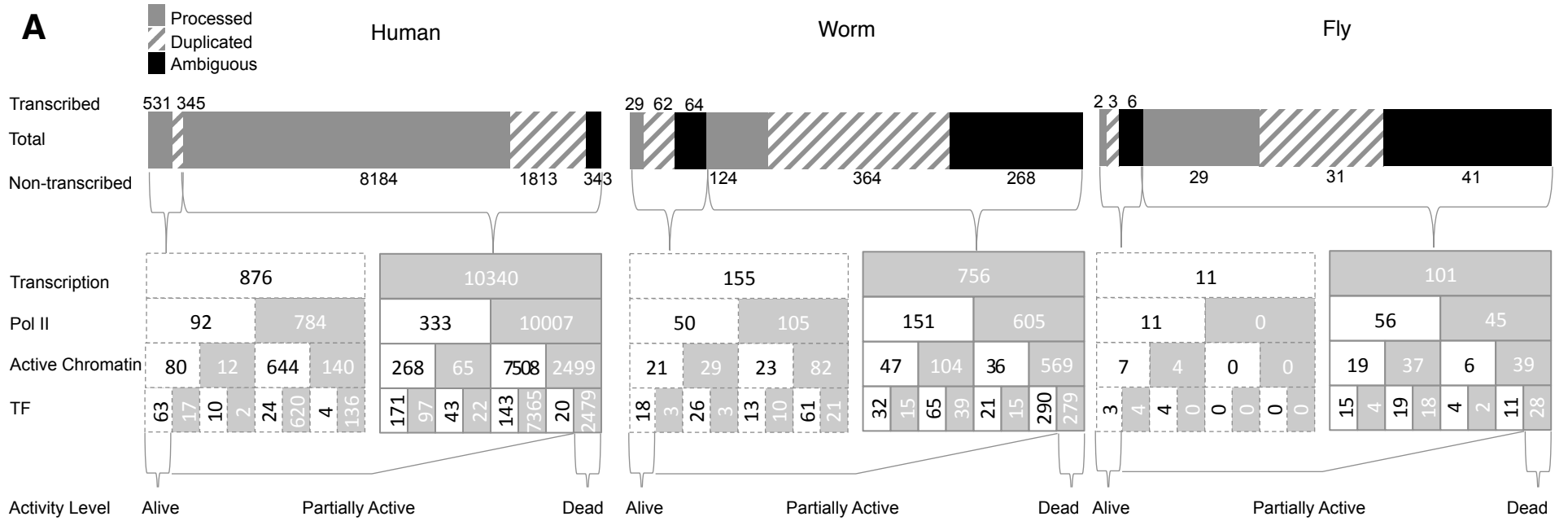
Exhibit 4: Comparison of ncRNAs



C)

			Human	Worm	Fly
Annotated ncRNAs			13,557 Kb (17,373)	301 Kb (1,594)	739 Kb (1,604)
Predicted ncRNAs	incRNA ncRNA Predictions	Type 1 (short)	36,586 Kb (273,215)	314 Kb (3,250)	19 Kb (96)
		Type 2 (long)	35,486 Kb (12,651)	446 Kb (2,554)	374 Kb (1,517)
	Unannotated TARs	90% (stringent)	386,114 Kb (626,623)	7,911 Kb (82,678)	14,151 Kb (196,929)
		95% (medium)	658,072 Kb (841,597)	20,200 Kb (132,243)	27,523 Kb (99,567)
		98% (inclusive)	949,644 Kb (1,578,963)	32,983 Kb (142,975)	43,494 Kb (92,868)
	Enhancer Associated RNAs		33,470	20,543	10,654
	HOT Region Associated RNAs		28,434	553	33,290
RNA-seq Reads	Exonic (CDS + UTR)		81.4%	96.9%	98.3%
	Intronic		14.4%	1.0%	1.1%
	Intergenic		4.2%	2.1%	0.6%

Exhibit 5: Pseudogenes



C

Top Family	Human	Worm	Fly

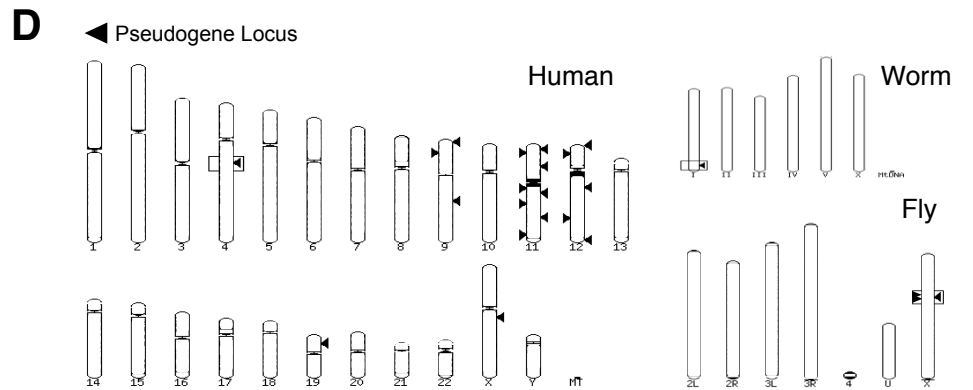


Exhibit 6: Histone and TF Models for Gene Expression

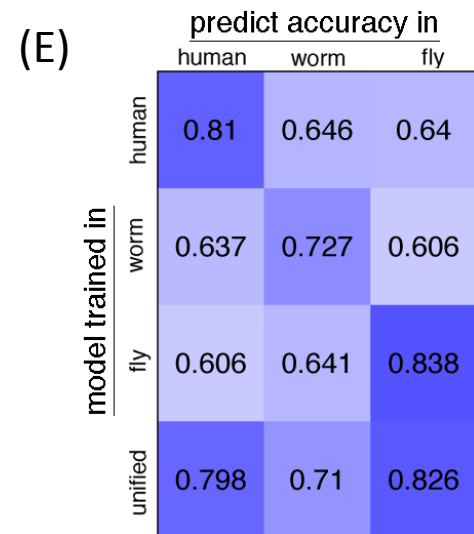
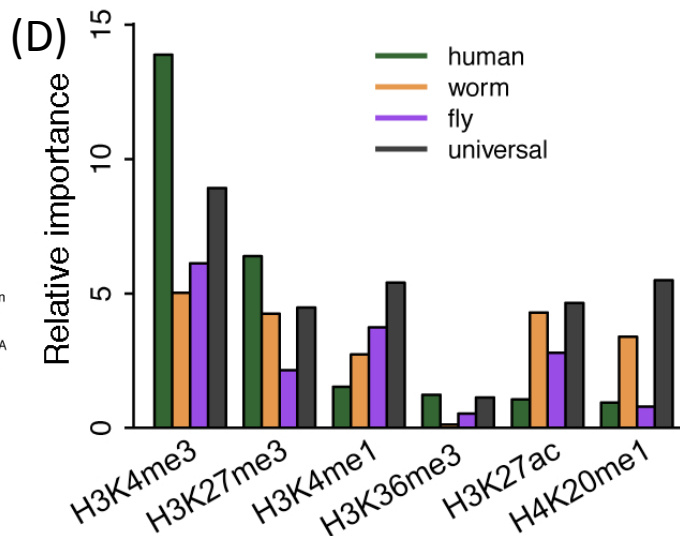
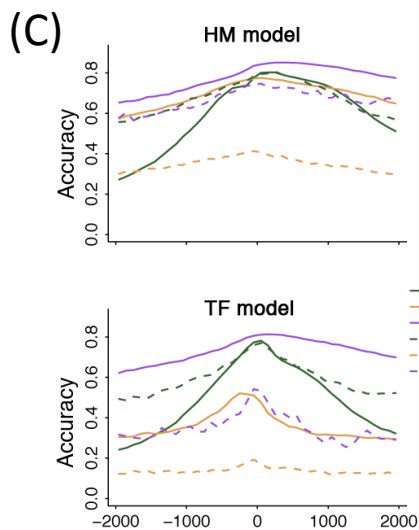
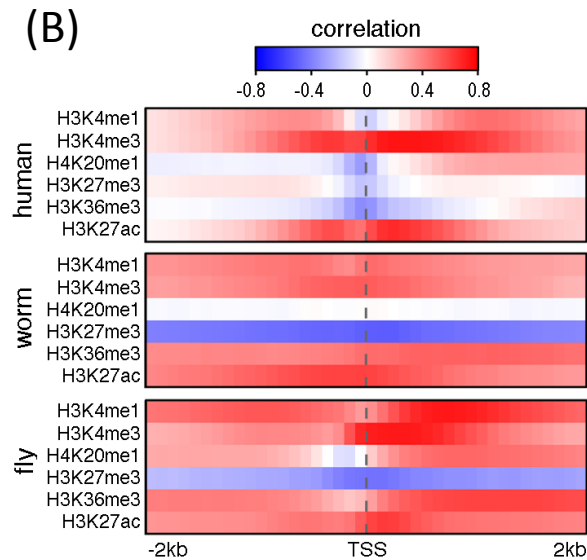
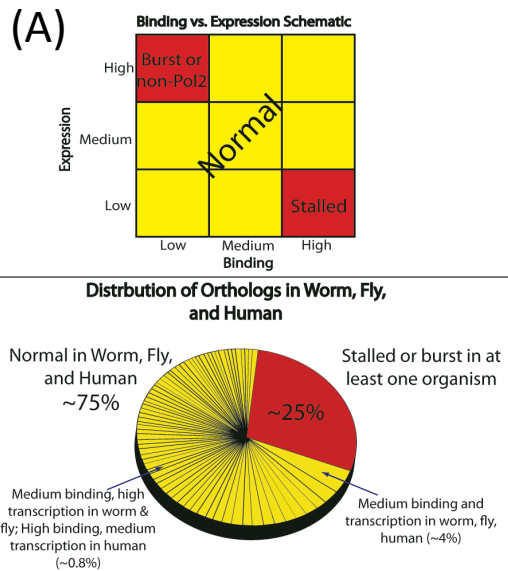
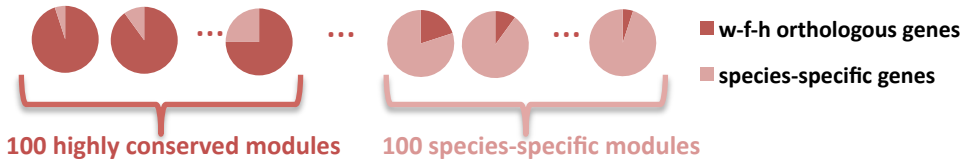
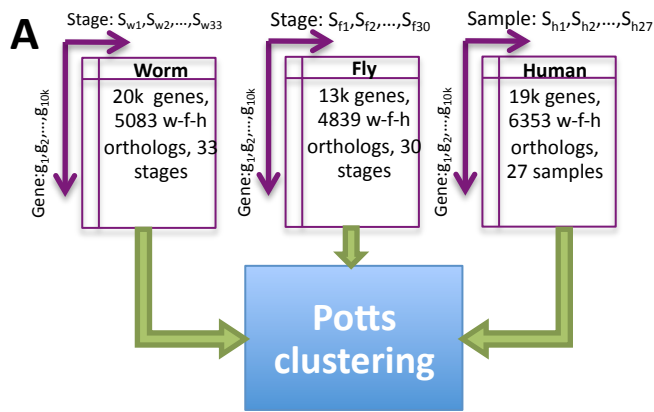
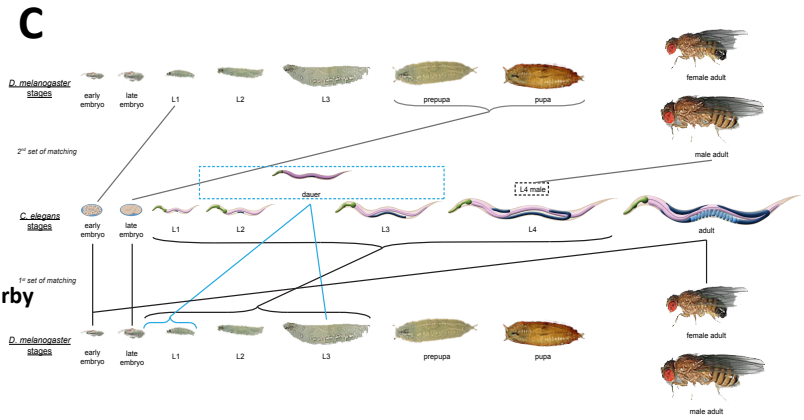
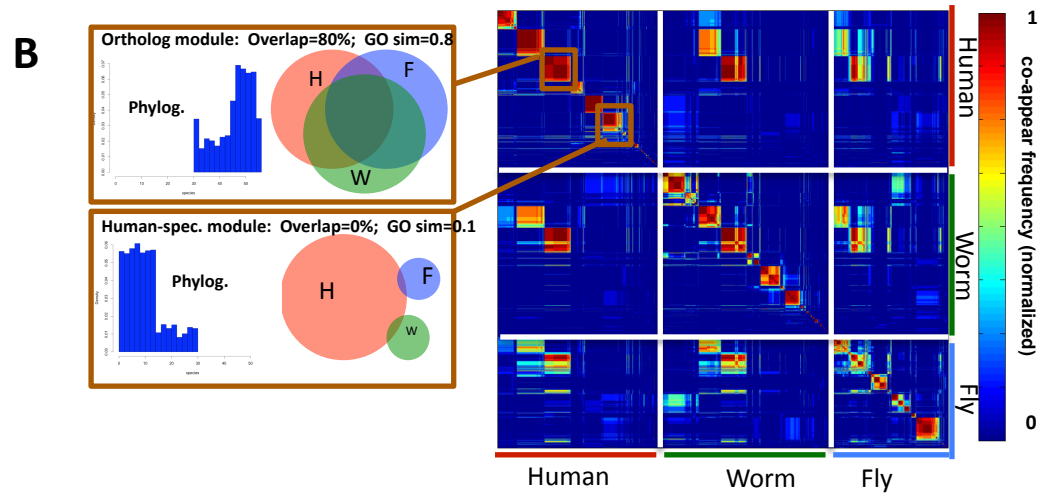
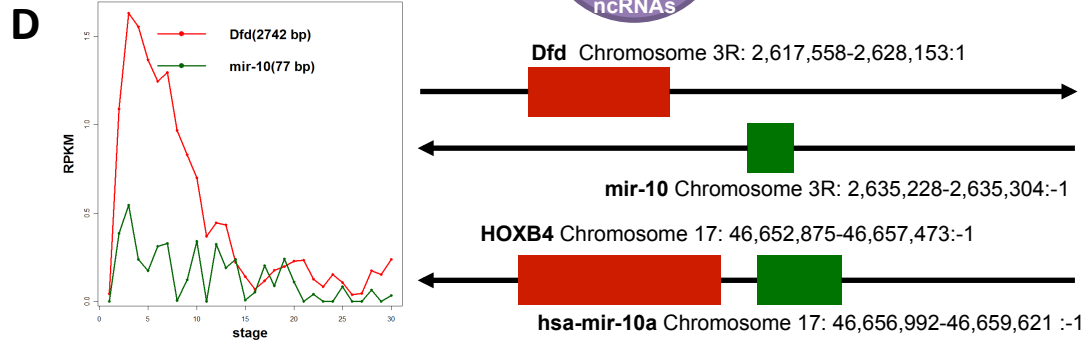
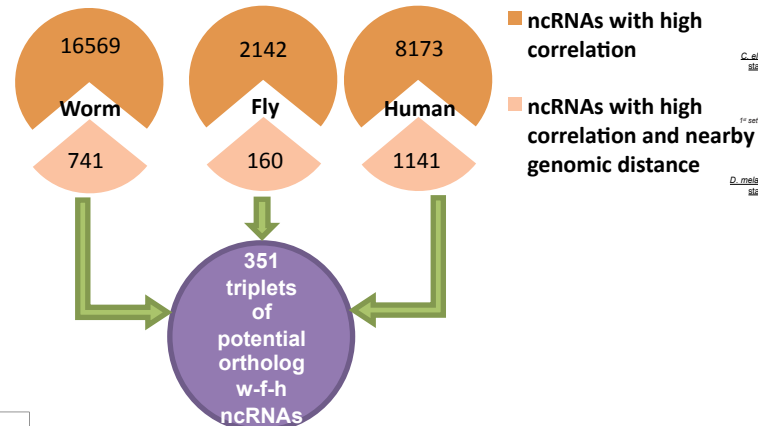


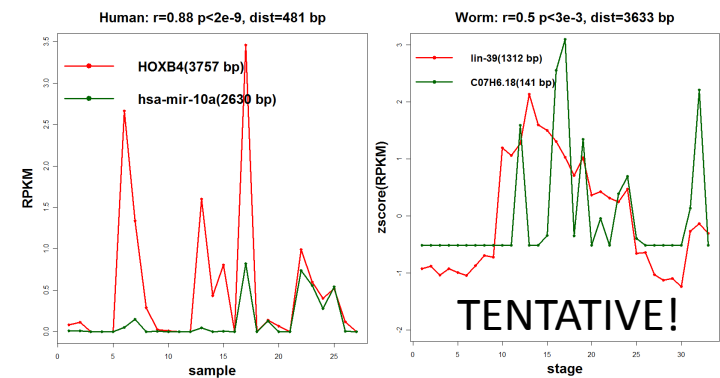
Exhibit 7: Expression Clustering of Coding Genes and ncRNAs

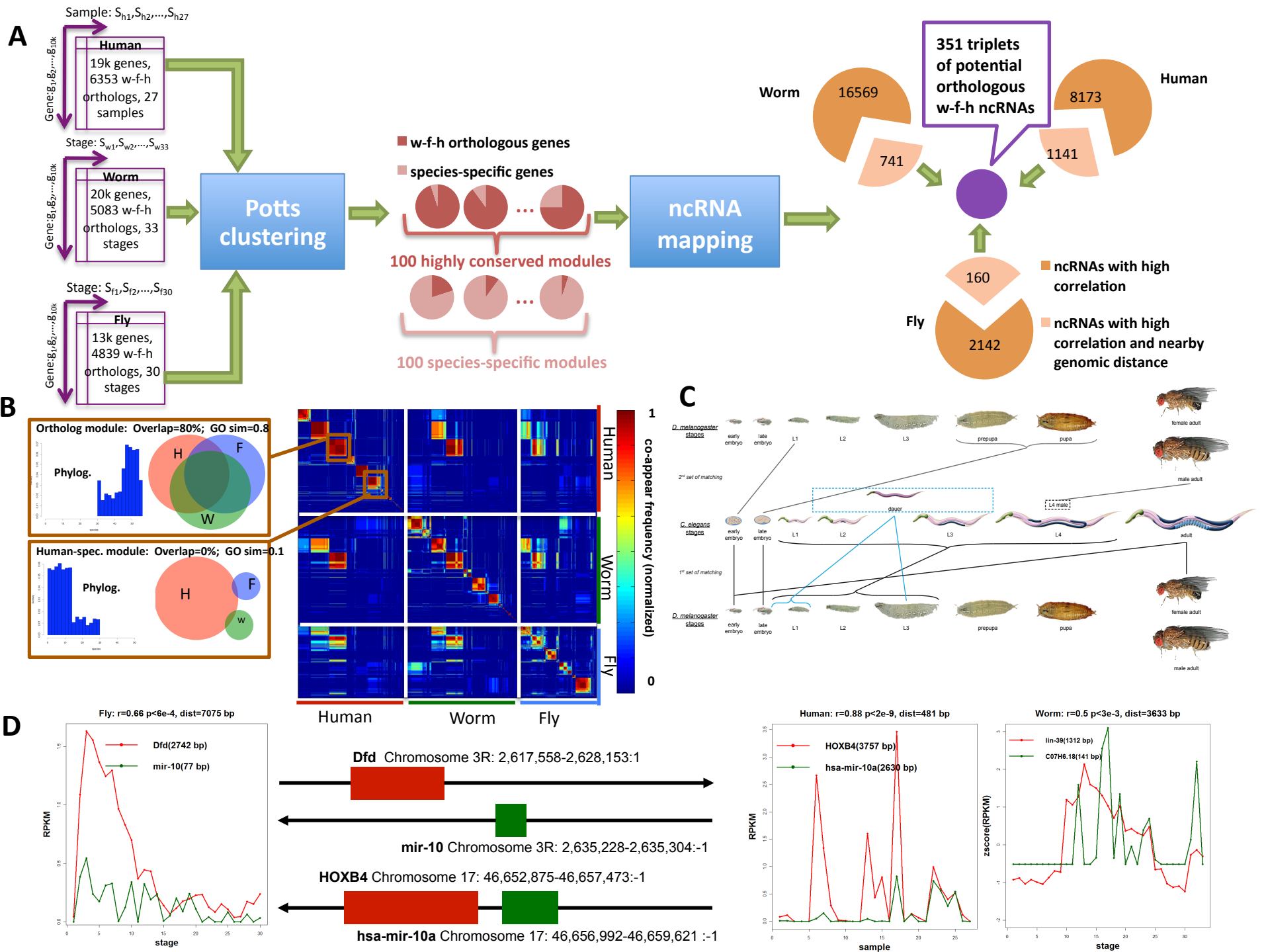


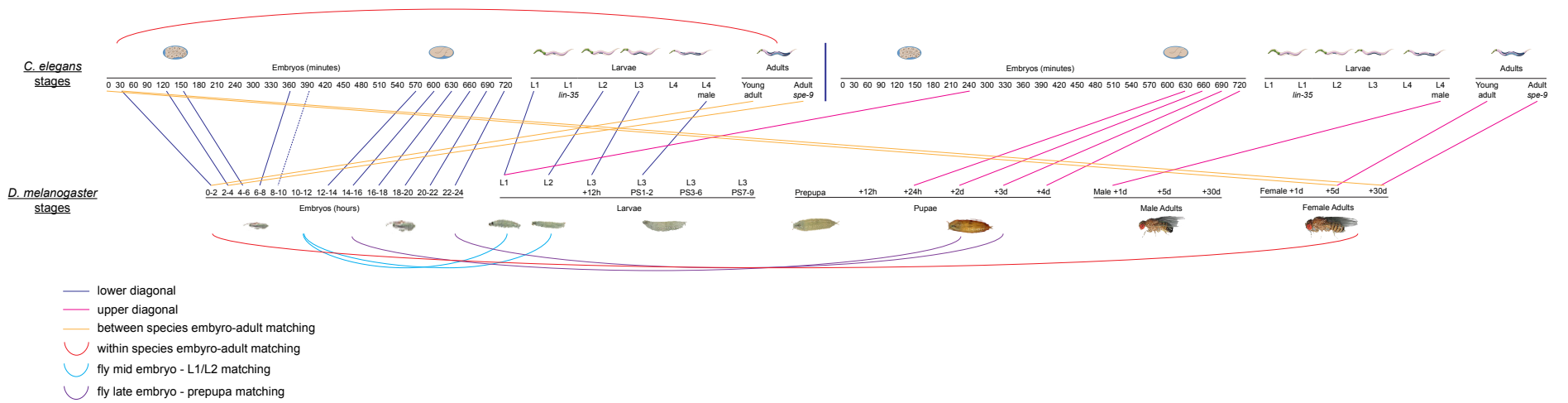
ncRNA mapping



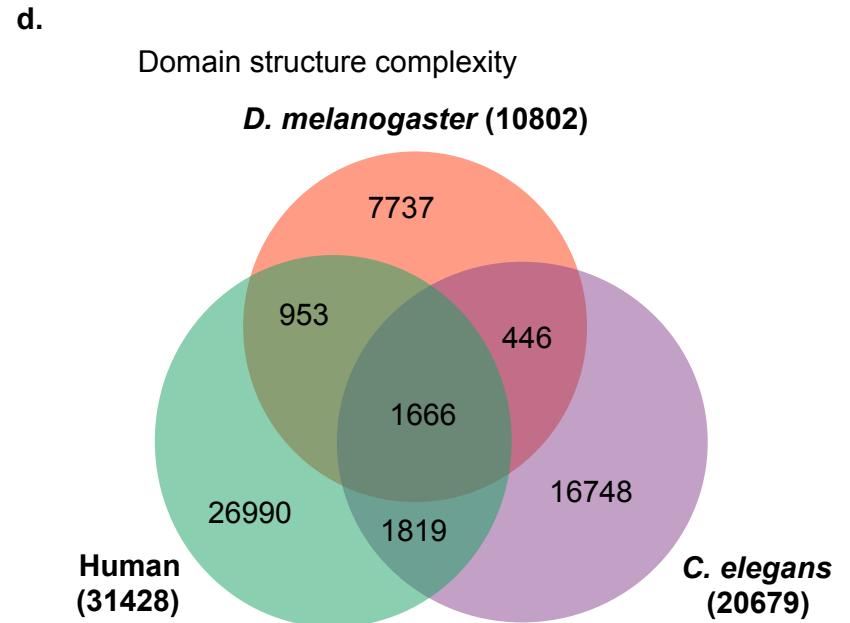
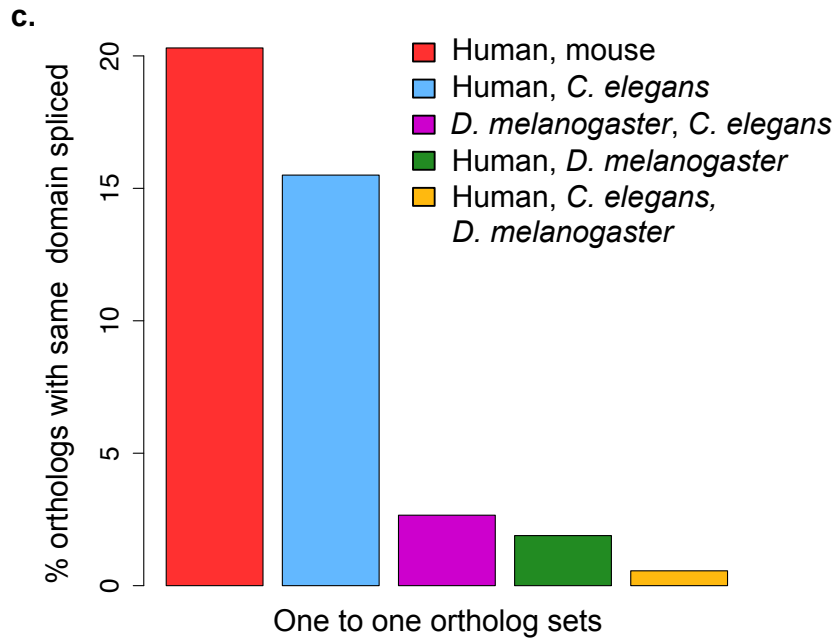
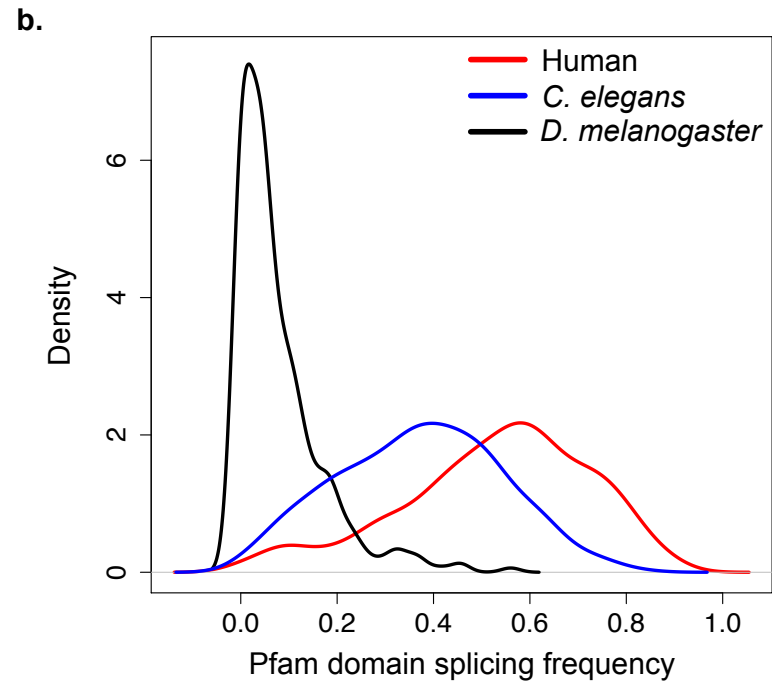
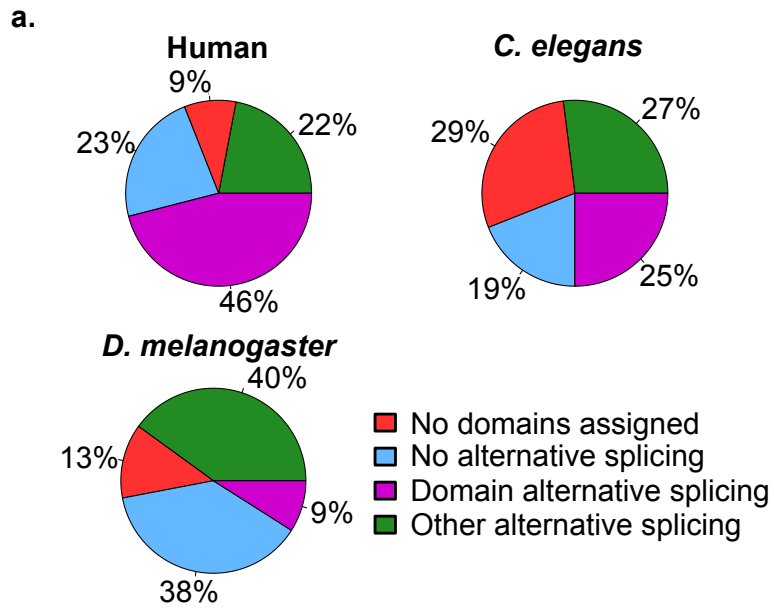
Brenner, Bickel, Huang, Li in UC Berkeley





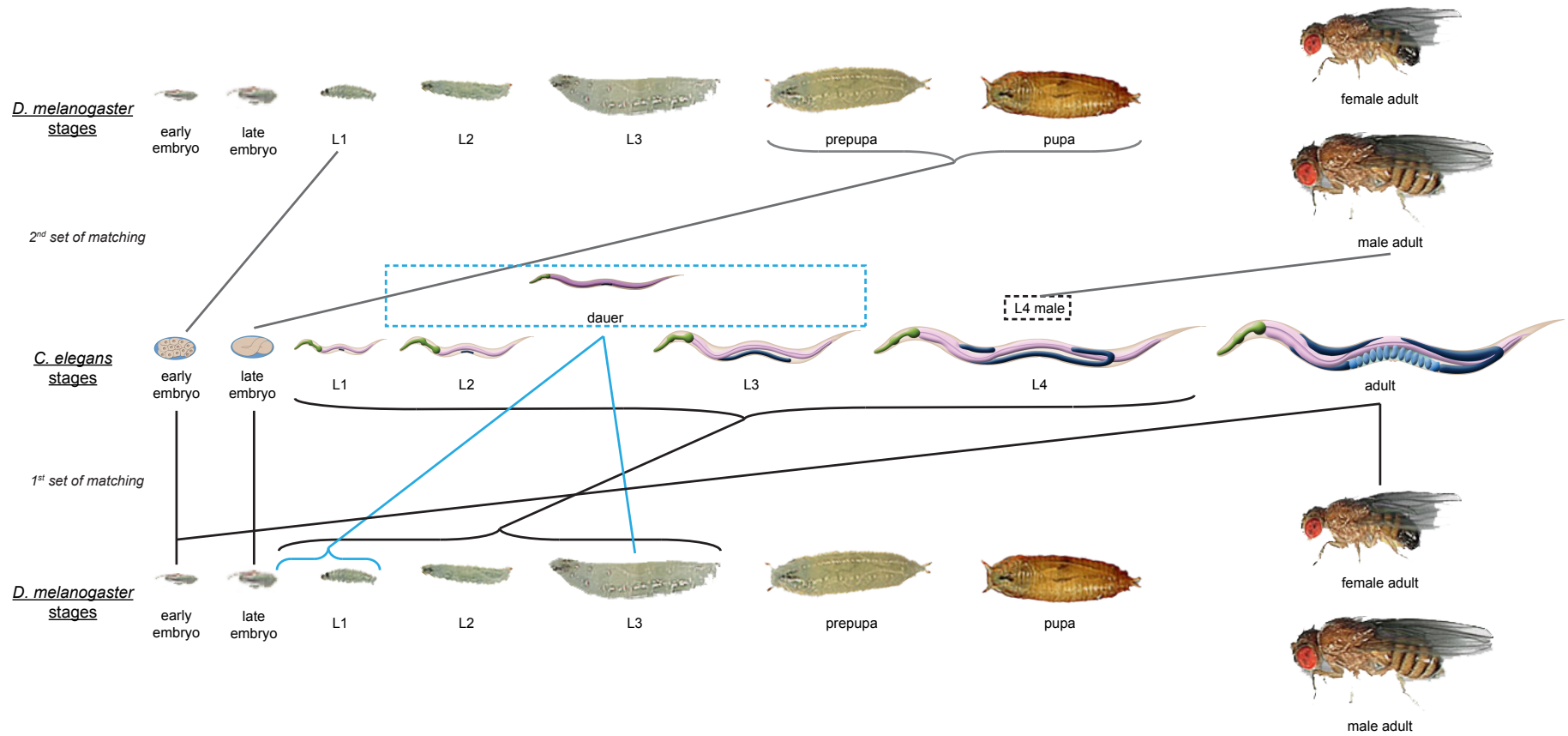


Extra/Older Slides

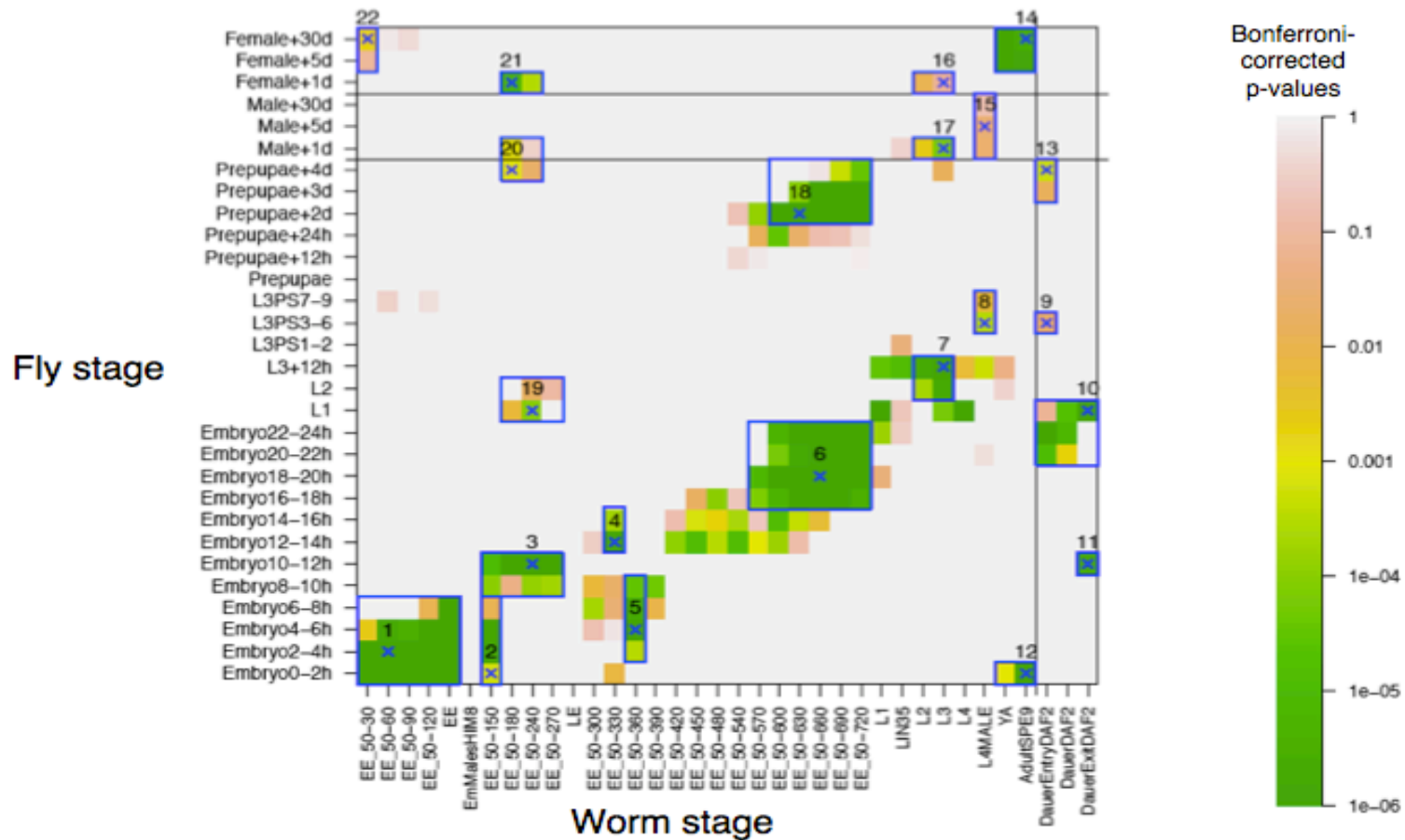


			Human (GENCODE v10)	Worm (WS220)	Fly (FlyBase R5.45)
Annotations	miRNAs		129,854 bps (1,595)	19,484 bps (223)	22,085 bps (238)
	tRNAs		45,468 bps (625)	45,553 bps (651)	22,998 bps (314)
	rRNAs		96,155 bps (712)	22,740 bps (44)	79,813 bps (160)
	snRNAs		385,163 bps (3,601)	29,127 bps (253)	33,522 bps (334)
	lncRNAs		12,900,062 bps (10,840)	184,123 bps (423)	580,956 bps (558)
In silico Predictions	incRNA ncRNA Predictions	Type 1 (short)	36,586,300 bps (273,215)	313,700 bps (3,250)	19,200 bps (96)
		Type 2 (long)	35,485,600 bps (12,651)	446,300 bps (2,554)	374,300 bps (1,517)
	Unannotated TARs	90%	386,114,409 bps (626,623)	7,910,961 bps (82,678)	14,150,725 bps (196,929)
		95%	658,071,990 bps (841,597)	20,200,403 bps (132,243)	27,523,315 bps (99,567)
		98%	949,644,074 bps (1,578,963)	32,983,410 bps (142,975)	43,493,772 bps (92,868)
	Enhancer RNAs (TARs overlapping enhancers)		33,470	20,543	10,654
	HOT Region Associated RNAs		28,434	553	33,290
	RNA-seq Reads	Exonic (CDS + UTR)		81.4%	96.9%
Intronic		14.4%	1.0%	1.1%	
Intergenic		4.2%	2.1%	0.6%	

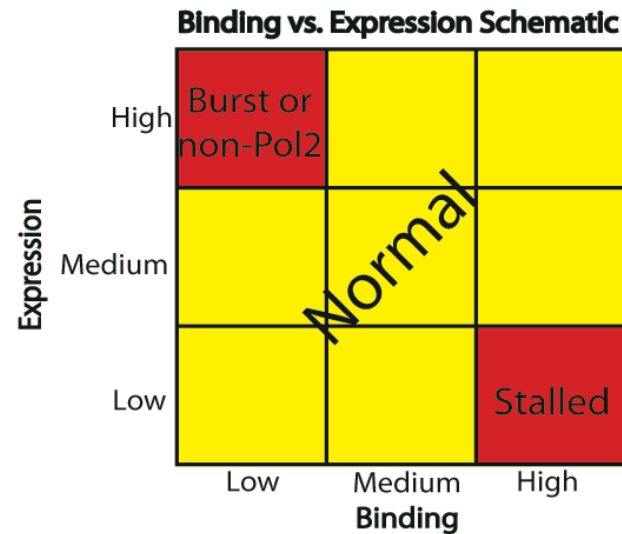
* Transcriptionally active regions as identified by the uniform processing for Worm, Fly, and Human RNA-seq data



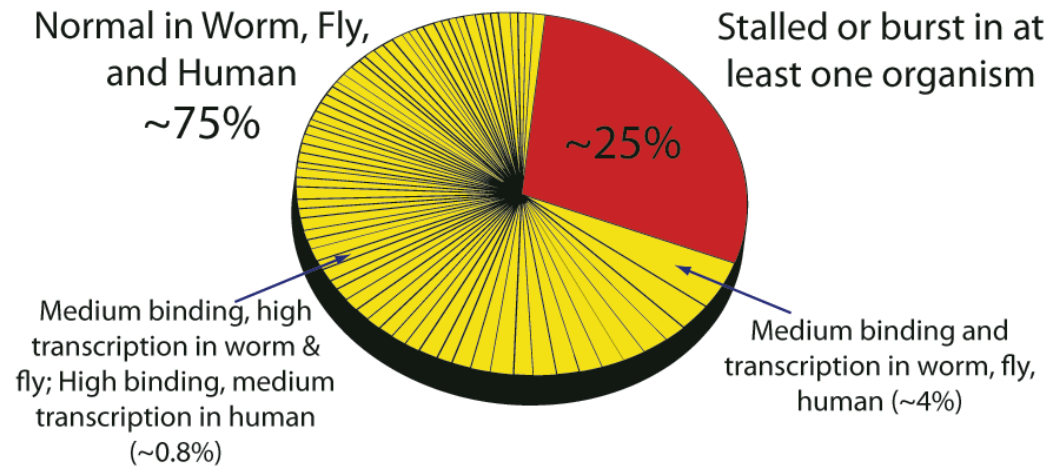
Developmental stage mapping between worm and fly based on co-expression clustering of orthologs

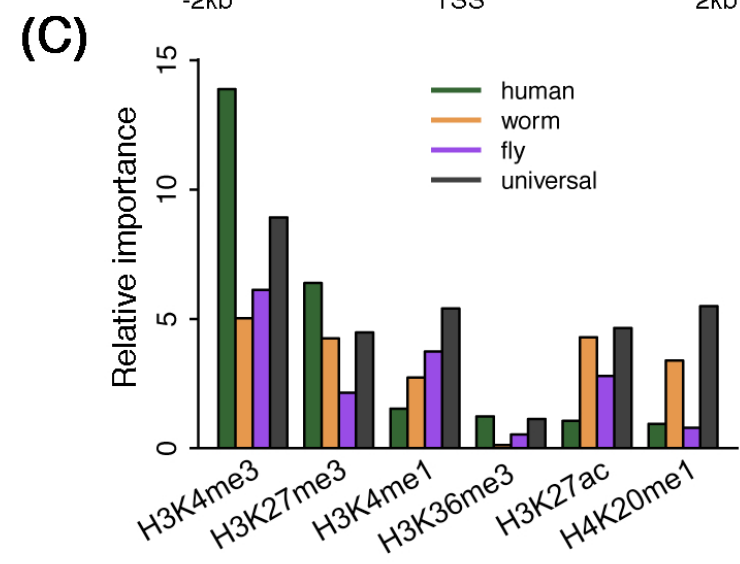
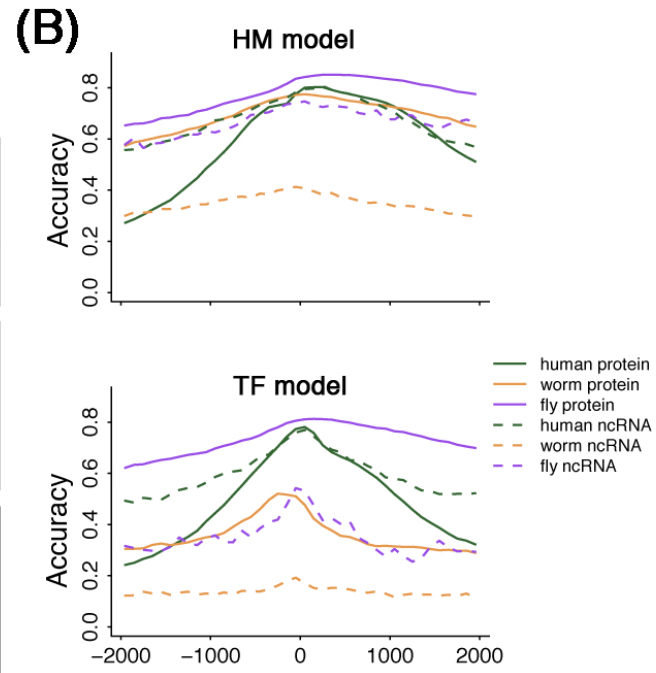
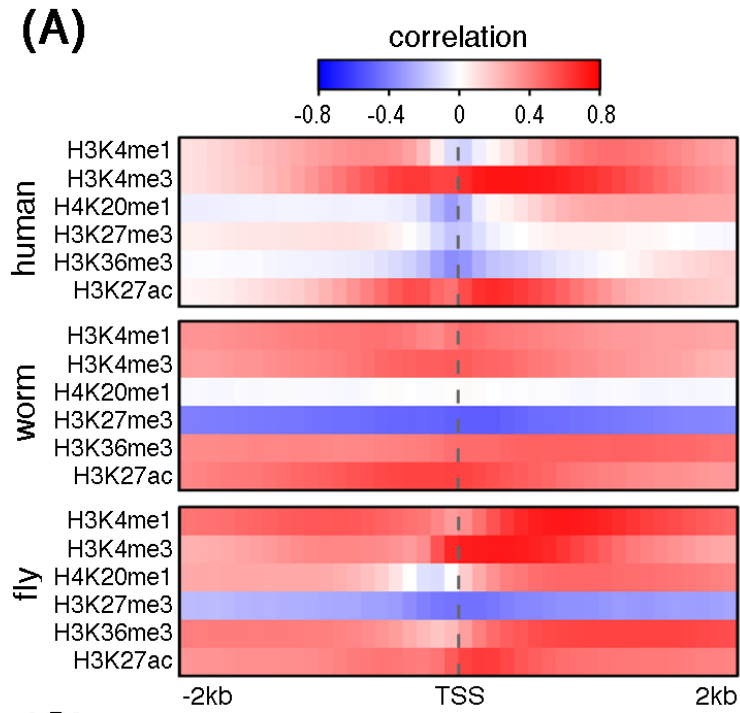


Brenner, Bickel, Huang, Li in UC Berkeley



Distribution of Orthologs in Worm, Fly, and Human





(D)

predict accuracy in

	human	worm	fly
human	0.81	0.646	0.64
worm	0.637	0.727	0.606
fly	0.606	0.641	0.838
unified	0.798	0.71	0.826

model trained in