

Pseudogenes

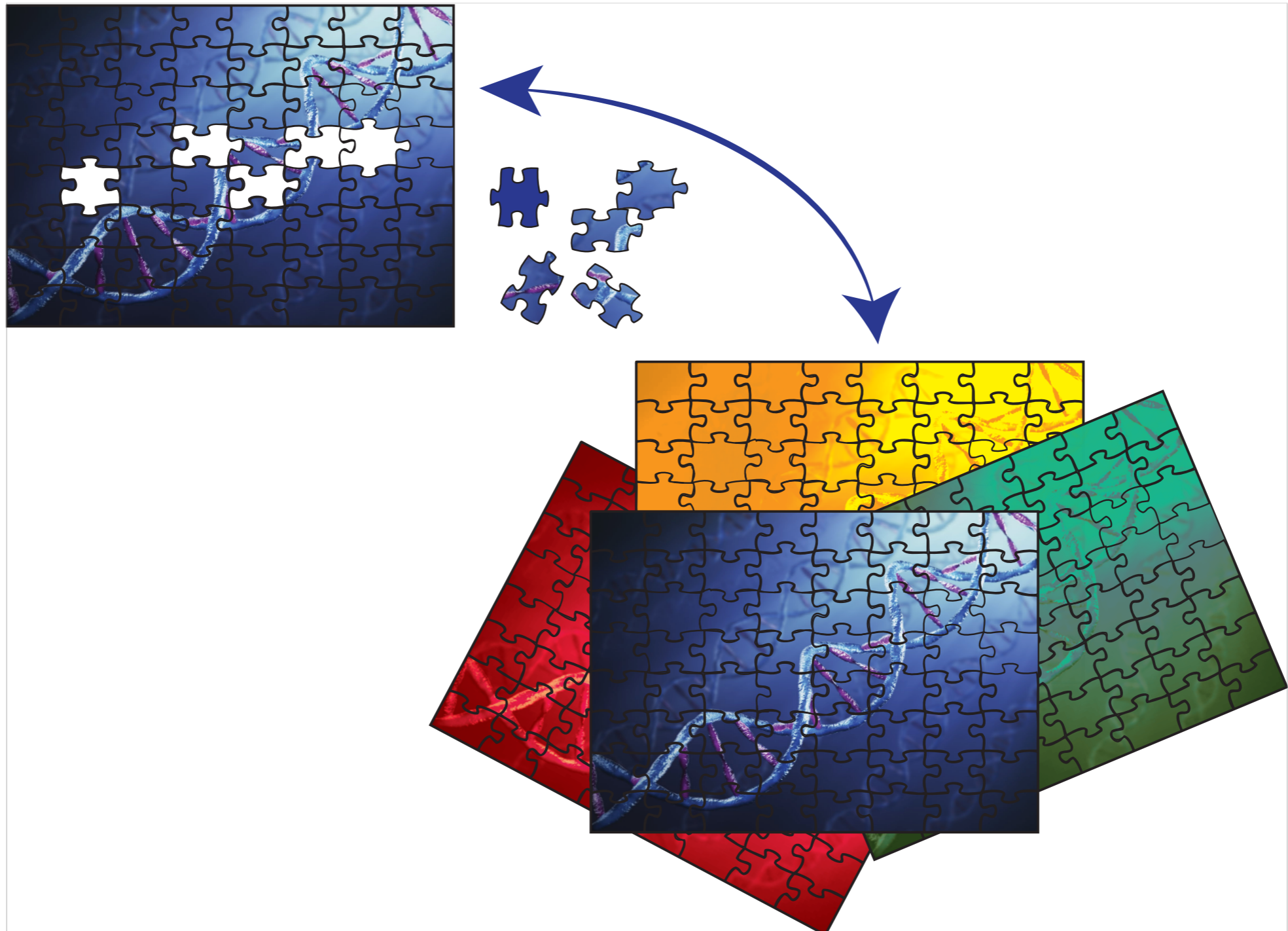
Gerstein Lab
Yale

GENCODE Meeting
21-22 June 2014

Comparison of pseudogenes across 3 phyla

Mark

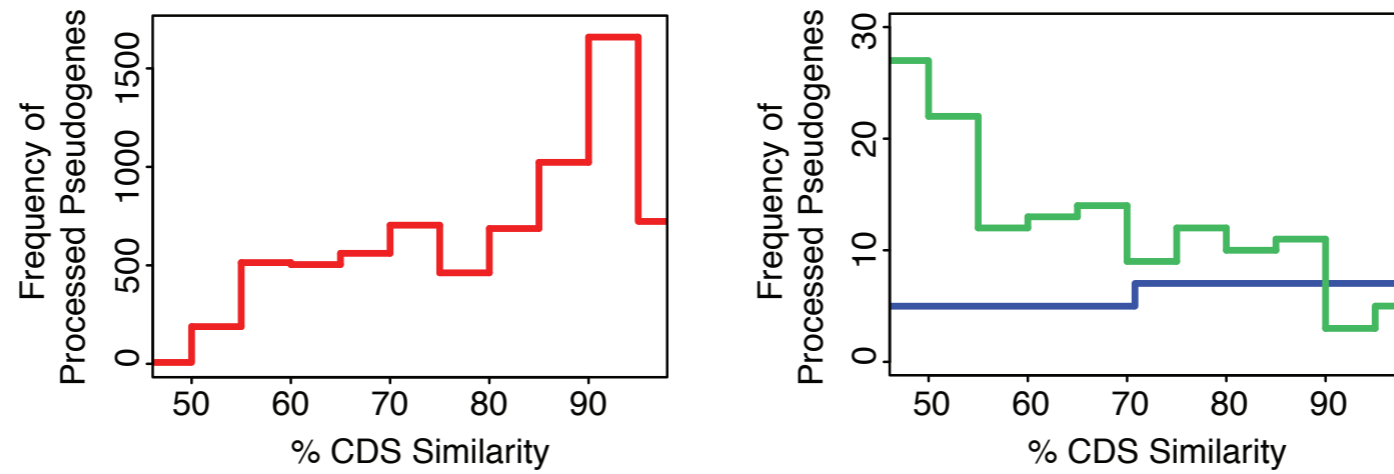
Comparing Pseudogenes



A

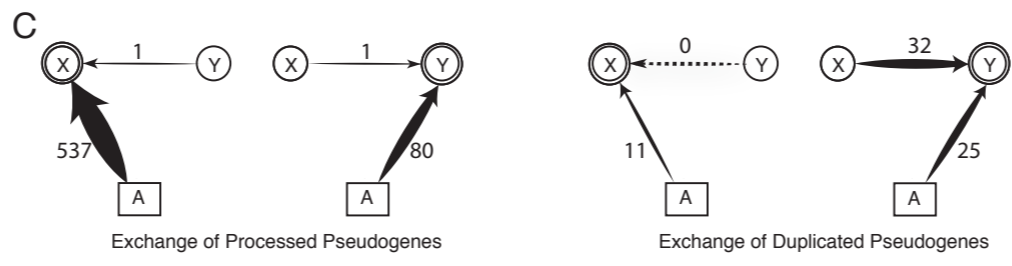
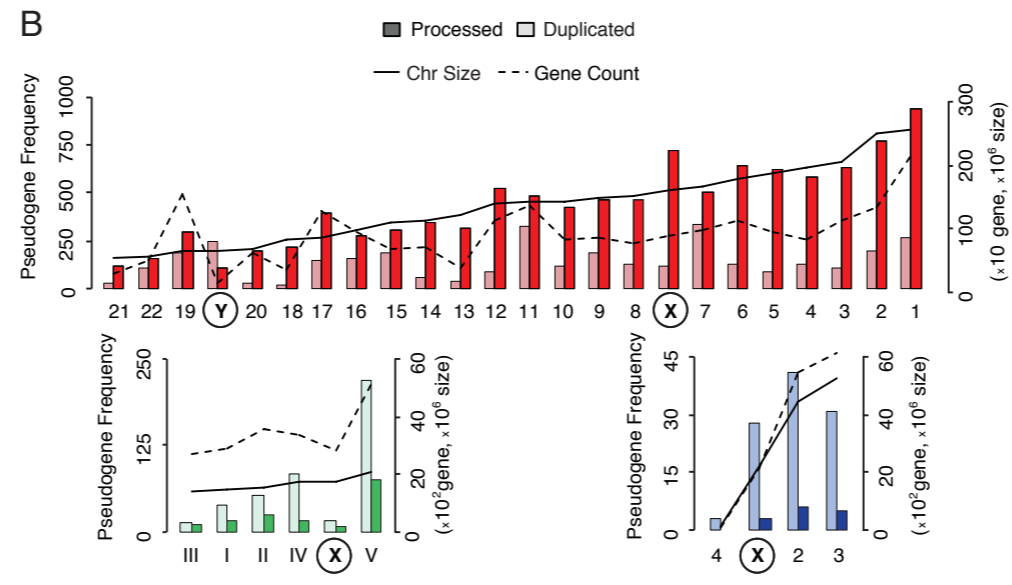
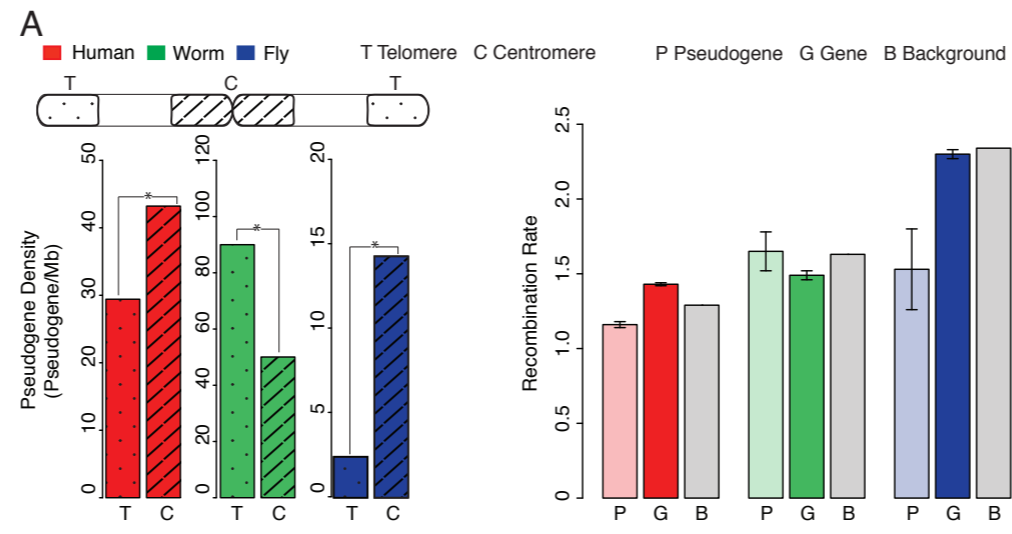
Organism	Total Pseudogenes	Biotype Distribution		ENCODE Functional Genomics Data	Completed Manual Annotation
		Processed	Duplicated		
Human	12,358	8908	2266	✓	✓
Worm	911	159	566	✓	✓
Fly	145	16	109	✓	✓
Zebrafish	229	21	177	✓	✓
Macaque	11,136	6570	1725	X	X
Mouse	13,169	7811	1827	✓	X

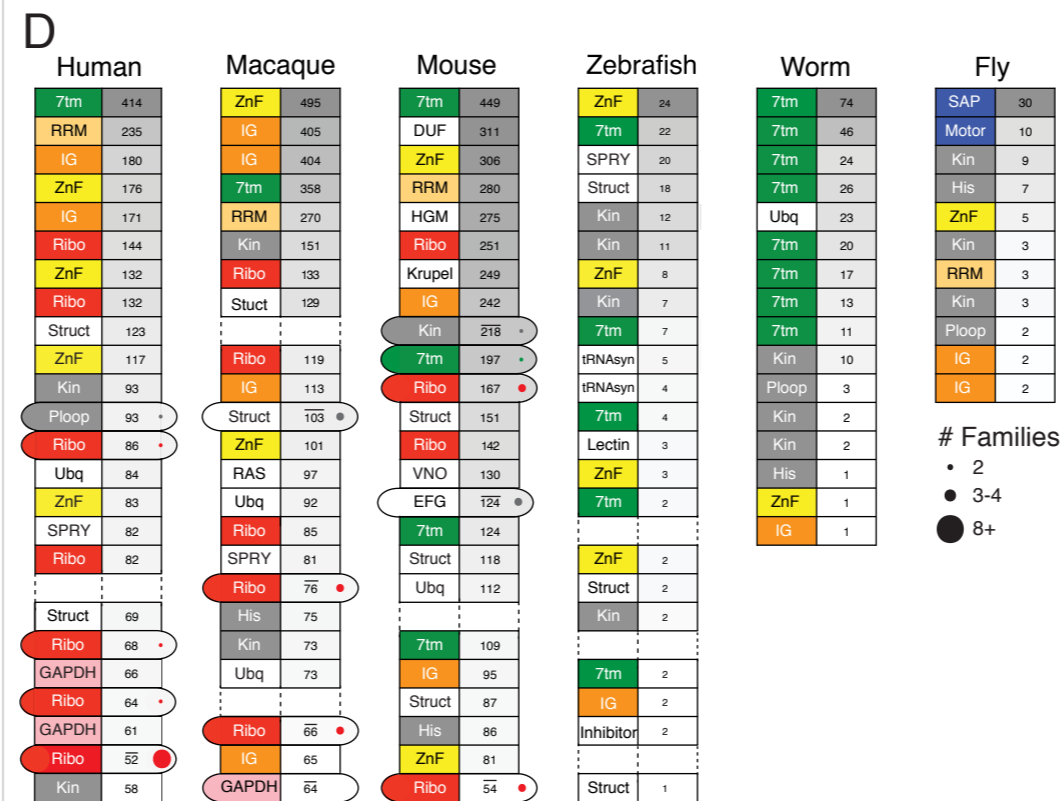
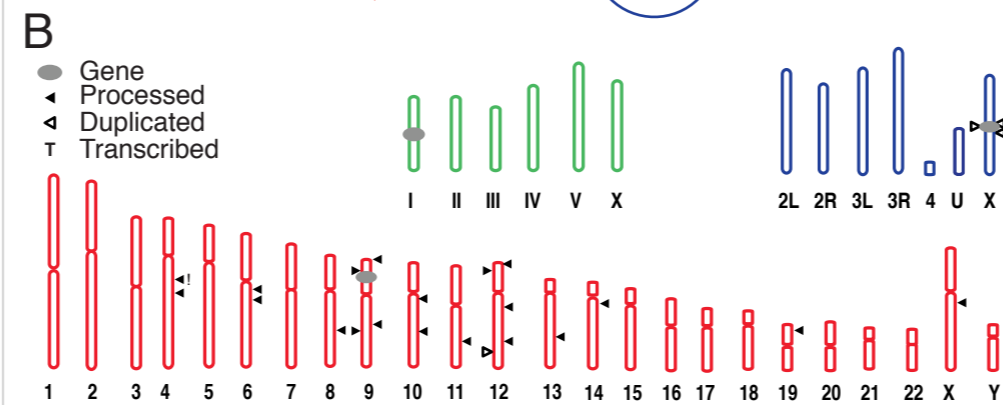
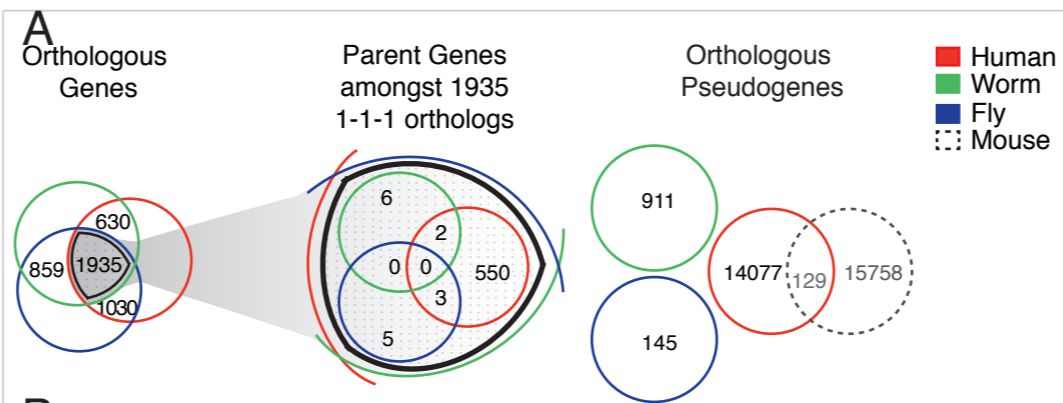
B

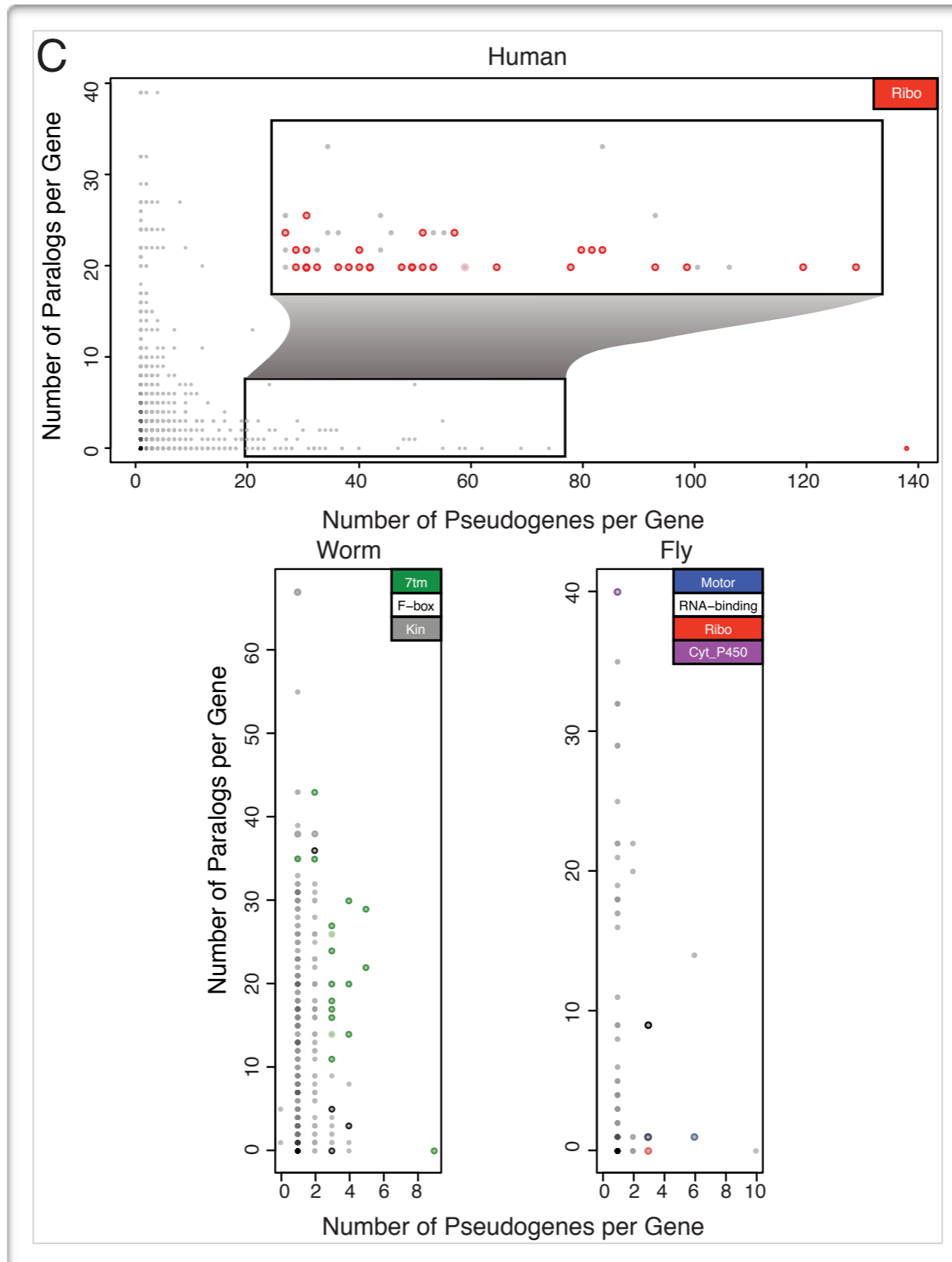


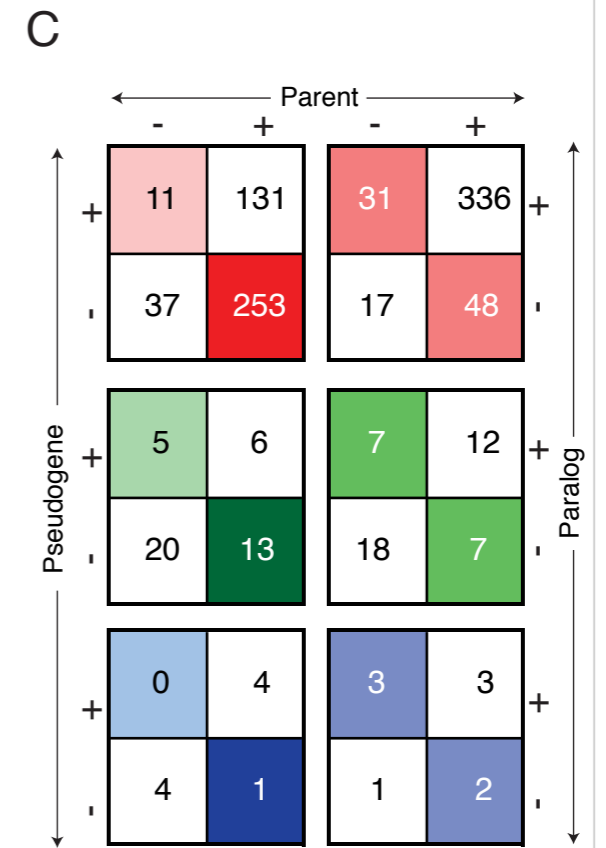
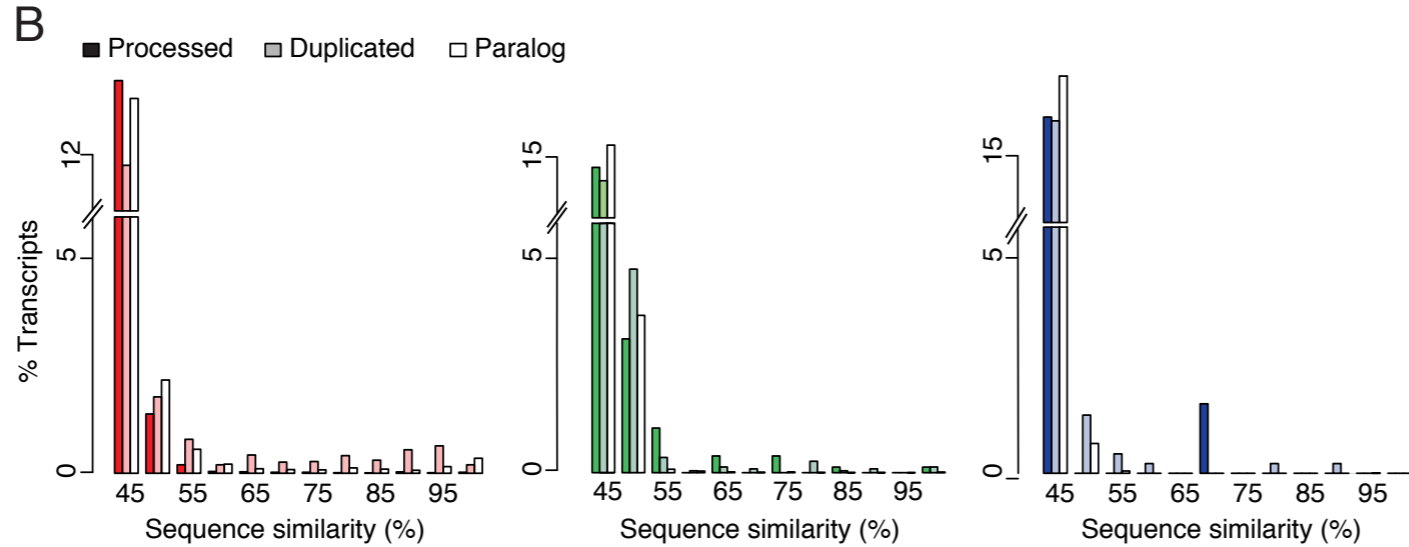
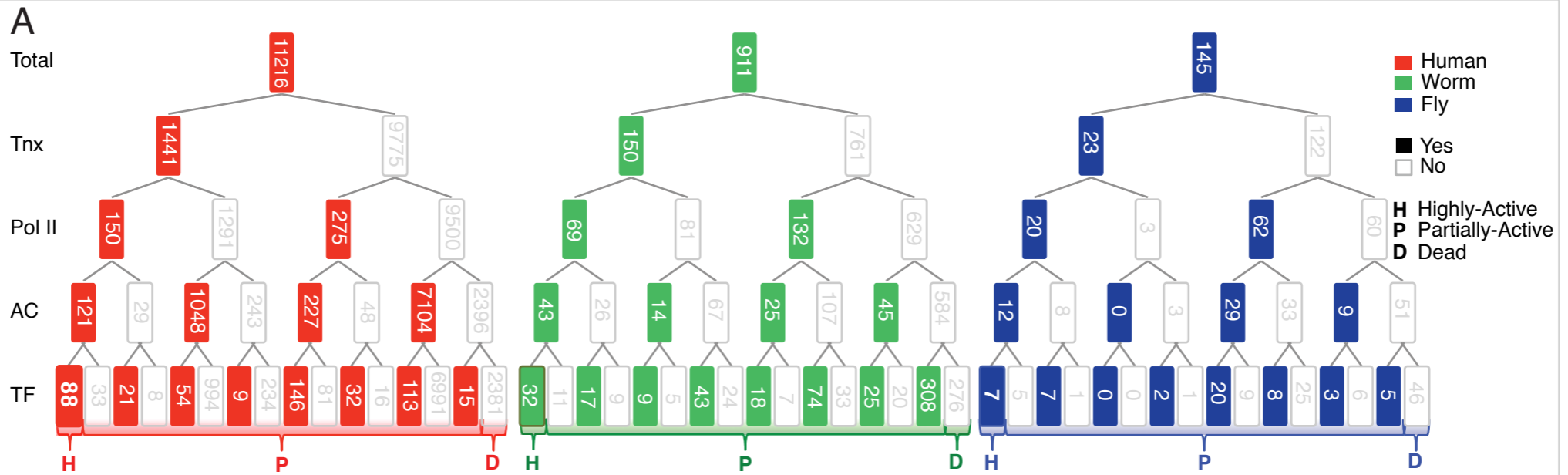
C

Organism	Defect / Pseudogene x MB		
	Insertion	Deletion	Stop
Human	4.4	4.9	2.4
Worm	25.8	7.45	2.5
Fly	7.9	12.7	1.1









D

Translation Candidates	Biotype	Coexp Coef	pVal	% Similarity CDS / UTR	Defect	Txn	Pol II	AC	TF
SLIT-ROBO Rho GTPase activating protein 2B pseudogene	Duplicated	0.80	5.9E-7	58 / 50	ins	✓	-	✓	-
PRKY-004, Y-linked protein kinase pseudogene	Duplicated	-0.14	0.42	96 / 51	ins / del	✓	✓	-	✓
Fer-1-like 4 (C. elegans), pseudogene	Unitary	-0.38	0.03	62 / 32	ins / del	✓	-	✓	-

The pseudo-verminator

Cristina

Mouse - where are we now

- previous release vs current release
- pipeline consistency
- data available

Functional data integration

- transcription
- translation
- activity

Vertebrate context

- mouse strains
- rat, primates, other mammals
- orthologs, syntenic regions
- pseudogene phylogenetic tree vs gene tree

ToDo's

Loss of function

Suganthi

Acknowledgements

Gerstein Lab Pseudogene Subgroup

Rachel & Mark D

Adam, Tim & Jen