A
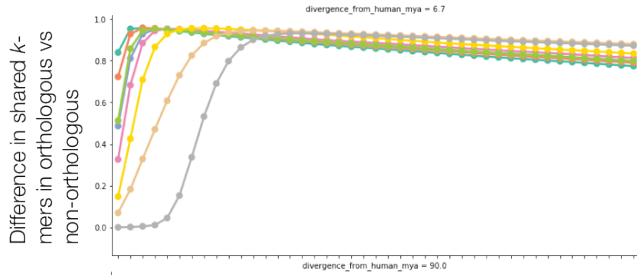
Divergence time
(millions of years ago, mya)

6.7 mya

Human vs Chimpanzee

90 mya

Human vs Mouse
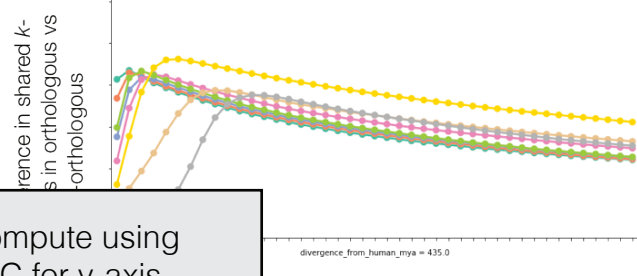
Recompute using
AUC for y-axis

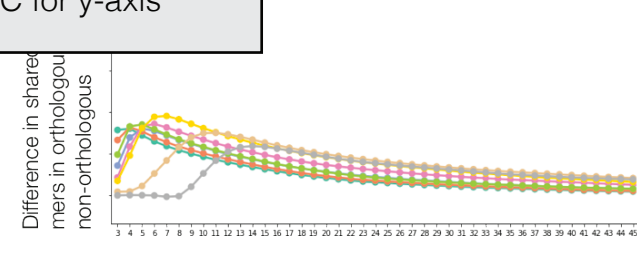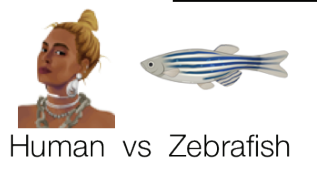435 mya

Human vs Zebrafish

Difference in shared *k*-mers in orthologous vs non-orthologous

divergence_from_human_mya = 6.7

divergence_from_human_mya = 90.0

divergence_from_human_mya = 435.0

encoding
- protein20
- hsdm17
- sdm12
- aa9
- botvinnik8
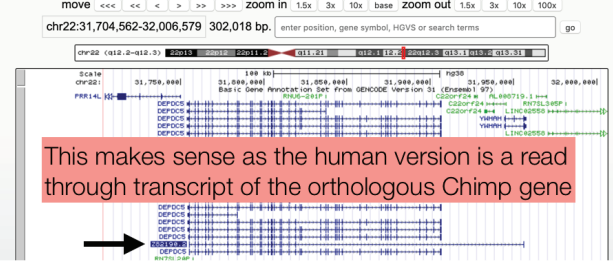- dayhoff6
- gbmr4
- hp2

B

Human gene: Z82190.2: novel protein,
DEPDC5-YWHAH readthrough

Chimp gene: DEPDC5

High jaccard (0.91) but not labeled as
homologous by ENSEMBL COMPARA

Dayhoff encoding, peptide ksize=16

Human gene: DLG3

Mouse gene: Dlg2

Medium jaccard similarity (0.41)

Not homologous by ENSEMBL COMPARA

Dayhoff encoding, peptide ksize=8

UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

chr22:31,704,562-32,006,579 302,018 bp. enter position, gene symbol, HGVS or search terms go

This makes sense as the human version is a read
through transcript of the orthologous Chimp gene

This makes sense as the human and mouse
genes are from the same gene family

C

Apply kmerslay compare-kmer-content to
finding putative orthologs