

Pseudogene Conservation

- Pseudogene conservation was calculated from divergence between human/chimp orthologs and human/mouse orthologs
- Pseudogenes with significantly low divergence are considered as under evolutionary constrain
- p-values of divergence were calculated from Poisson distribution, followed by adjust of multiple hypothesis testing with FDR approach (FDR = 0.05)
- Background substitution rate: 1.5% (human/chimp), 5% (human/mouse)

Example: Pseudogene ENST00000531871.1

Sequence identity between human and chimp: 7,218 / 7,268 (substitution: 50)

Background rate: 1.5% => Mean: 7,268 * 1.5% = 109.02

P-value: $\sum_{x \leq 50} \frac{(e^{-109.02})(109.02^x)}{x!} = 2.02 \times 10^{-10}$

Q-value: 1.13×10^{-6}

Pseudogene Conservation

Total constrained pseudogenes: 1,019 (chimp: 64, mouse: 972)

Conservation vs. Transcription

	Transcribed	Non-transcribed
Constrained	195	824
Total Pgenes	876	10,340
Percentage	22.3%	8.0%

Transcribed pgenes are enriched in the constrained pgene set, p-value = 1.19e-35

Conservation vs. Biotype

	Duplicated	Processed	Unitary	Other
Constrained	607	294	103	15
Total Pgenes	2,158	8,715	138	205
Percentage	28.1%	3.4%	74.6%	7.3%

Pseudogene Conservation

