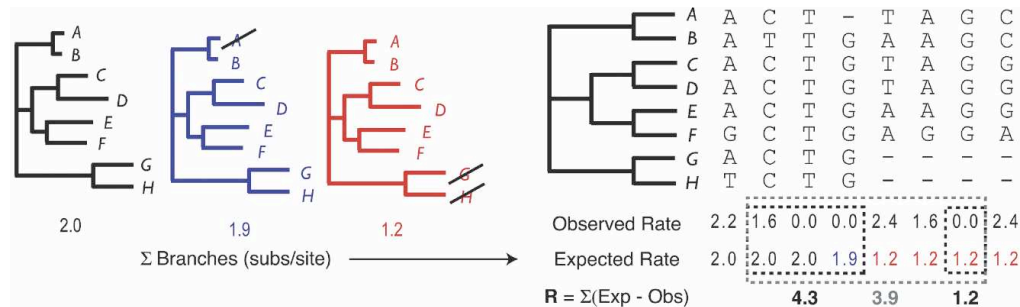


Constraint Regions in Pseudogene Annotations

Annotation Meeting
Baikang Pei
06/02/2011

Genomic Evolutionary Rate Profiling (GERP)



Objective: to find constraint regions in genome subject to purifying selection

Procedure overview:

- use evolutionary tree and multiple alignments to estimate rejected substitution scores (RS) on a column-by-column basis.
- Constrained elements are stretches of the multiple alignment where the sequences are highly conserved according to the previous score.

Data

Gene Annotation: from GENCODE v7 annotation file

- Pseudogenes: exons
- Protein coding genes: CDS, 5' UTR, 3' UTR and introns

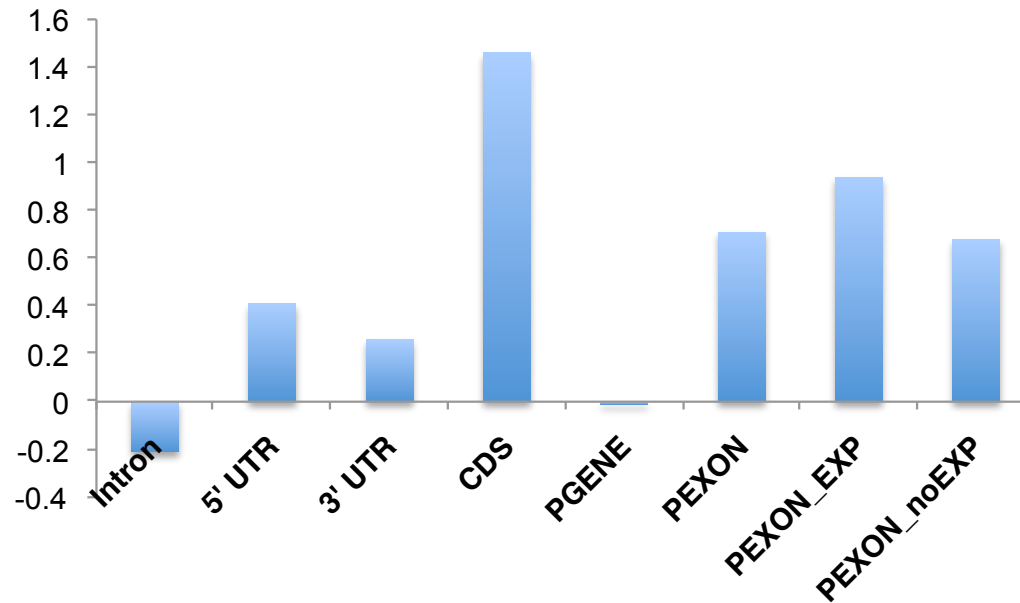
Constrain Annotation: download from ENCODE wiki

- EPO alignment of 33 mammals (of which 22 are 2x mammal)
- Evolutionary tree
- GERP scores and constrained elements

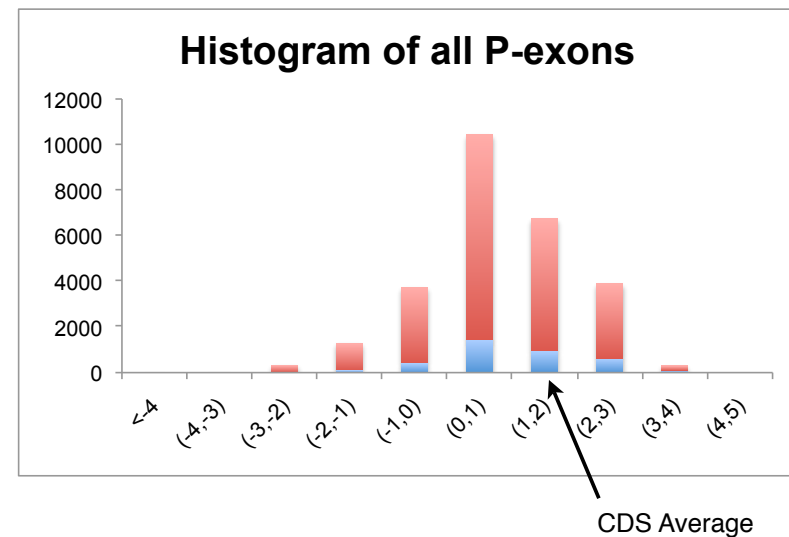
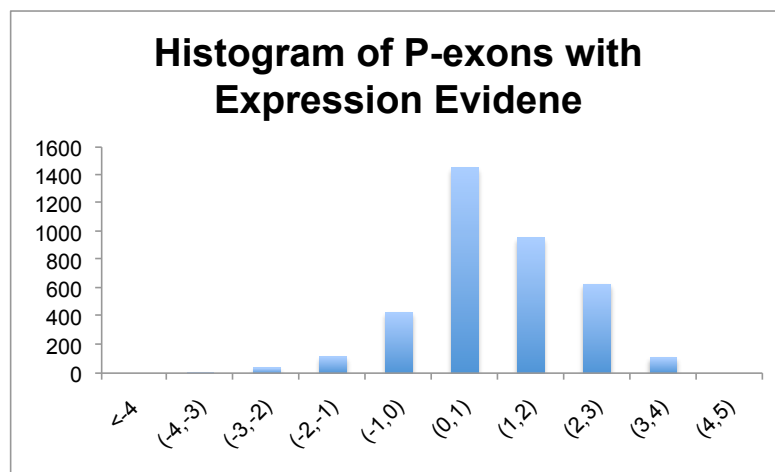
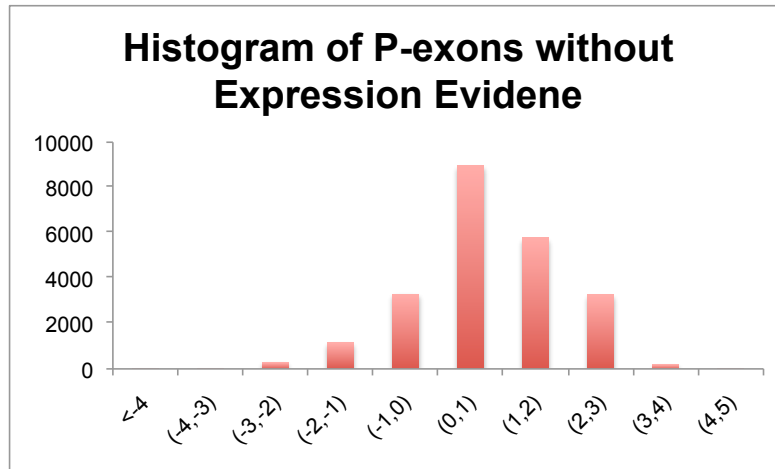
GERP: v2.1 from Sidow lab

- Get same results when using multiple alignment and evolutionary tree from ENCODE wiki

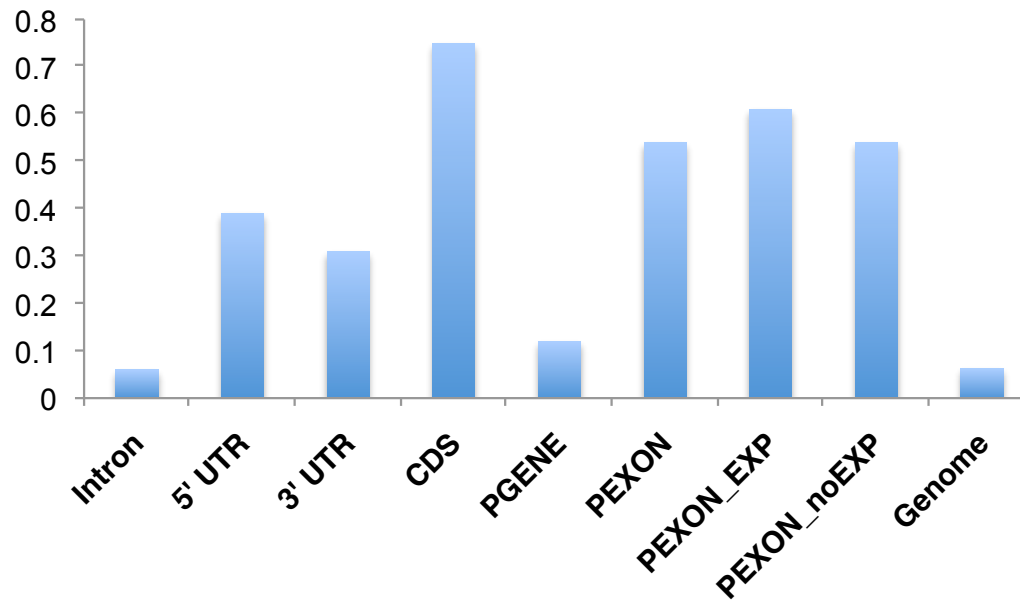
Average RS Scores in Different Regions



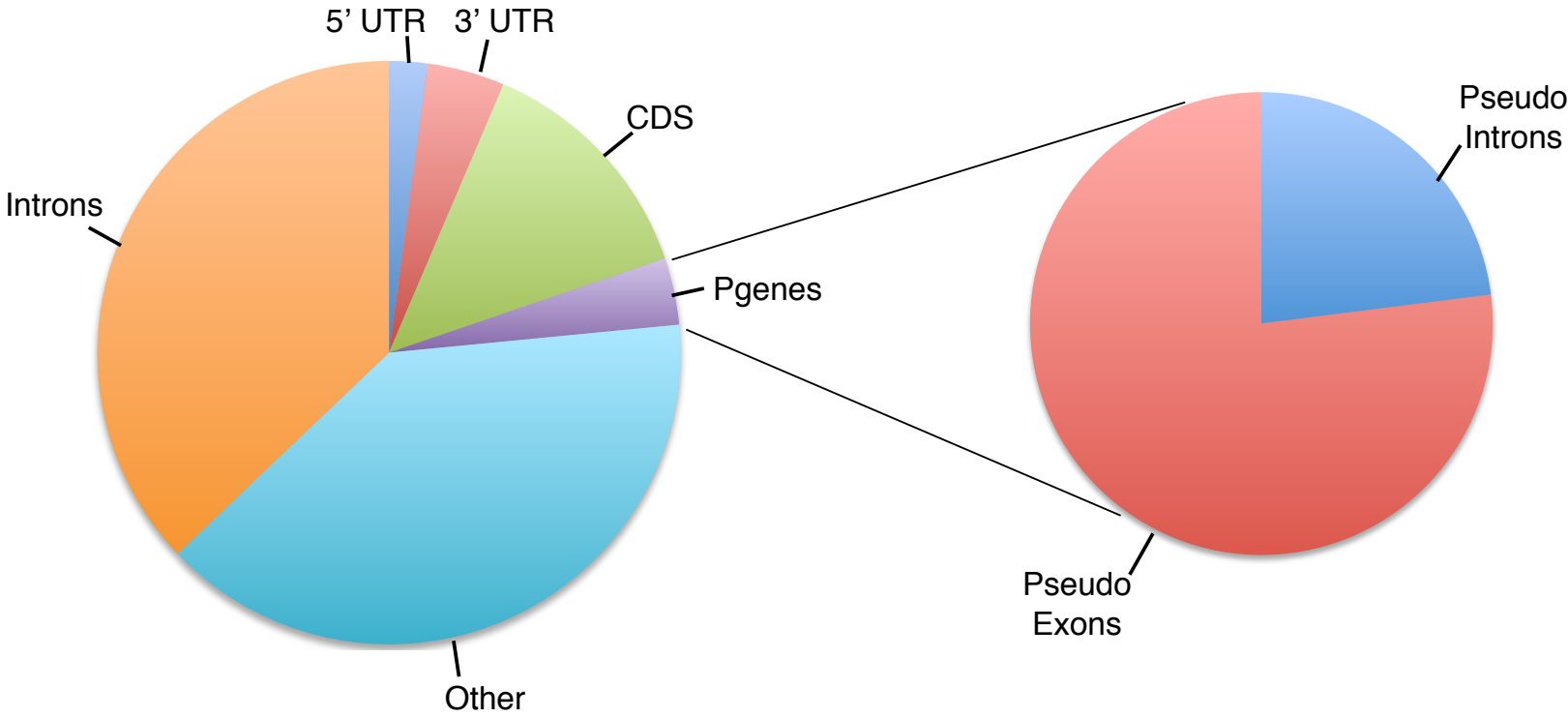
Distribution of Average RS Scores for Pseudo-exons



Percentage of Genomic Region Under Constraint

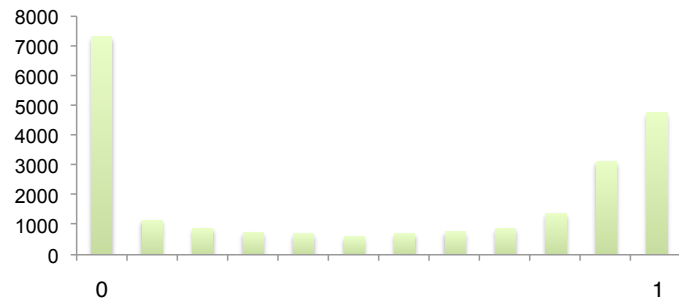


Composition of Constraint Elements

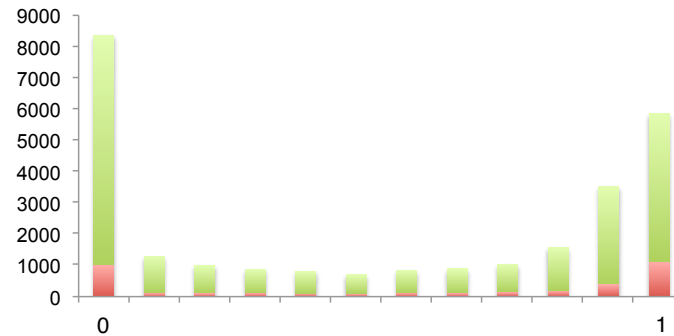


Distribution of Constraint Region Percentage for Pseudo-exons

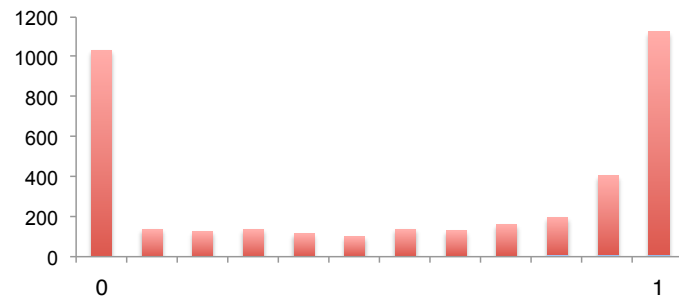
P-exons without Expression Evidence



All p-exons



P-exons with Expression Evidence



Multiple Alignment Example

chr4:3925799-3926001, ENSG00000253917.2, ENST00000513501.1

```
>Homo_sapiens      TGGAAGGTGGTGGGTGCCACACAAGTTAAGCTGTCTTCCCTCTGGCAGGC
>Gorilla_gorilla  TGGAAGGTGGTGGGTGCCACACAAGTTAAGCTGTTTTCCCTCTGGGAGGC
>Callithrix_jacchus TGAAAAGTGGTGGGTGCCACACAGGTTAAGCTGTCTTCCCTCTGGGAGGC
>Tarsius_syrichta TGGAAGGTGGTGGGTGCCACACAGGTTAAGCTGTCTTCCCTCTGGGTGGC
>Microcebus_murinus TGAAAGGTGGTGGGTGCCACACAGGTTAAGCTGTCTTCCCTCTGGGAGGC
>Otolemur_garnettii TGGAAGGTGGTGGGTGCCACACAGGTTAAGCTGTCTTCCCTCTGGGAGGC
>Tupaia_belangeri  TGCAAGGTGGTGGGTGCCACACAGGTTAAGCCGTCTTCCCTCTGGGAGGC
>Cavia_porcellus   TGCAAGGTGGTGGGTGCCACACAGTTTAAGCTGTCTTCCCTCTGGGAGTC
>Ochotona_princeps TGCAAGGTGGTGGGTGCCACACAGGCTAAGTGGTCTTCCCTCTG-----
>Oryctolagus_cuniculus TGCAAGGTGGTGGGTGCCACACAGGCTAAGCTGTCTTCCCTCTCGGGGGC
>Erinaceus_europaeus TGCAATTGTGGTGGGTGCCACACAGCGTAGGCTGCCTTCCCTCCCAGGAGC
>Felis_catus       TGCAAGGTGGTGGGTGCCACACAGGTTAAGCTGTCTTCCCTCTGGGAGGC
>Pteropus_vampyrus TGCAACGTGGTGGGCGCCACACAGGTTAAGCCATCTTCCCTCTGGGAGGC
>Tursiops_truncatus TGCAAGGTGGTGGGCGCCACACAGGTTAAGCTGTCTTCTCTCTGGGAGGC
>Choloepus_hoffmanni TGCAACGAGGTGGGTGCCACGCATATCAAGCTGCCCTTCTTCTGGGGAGC
>Dasypus_novemcinctus TGCAACGTGGTGGGCGCCGACAGGTCAGGCCGGCTGTCTTCTGGGGTGC
>Loxodonta_africana TGCAACGTGGTGGATGCTACACAGGTTAGGCTGTCTTCCCTCTGGGATGC
>Procavia_capensis TGCAAGGTGGTGGATGCCACACAGTGTGCGGCTGTCTTCCCTCTGGGATAC
```

Multiple Alignment Example

chr3:12045877-12046428 ENSG00000157152.11, ENST00000426379.2

```
>Homo_sapiens      TTGGGATGACTCCTTTCGTTCTCACCTCCGCATAGGGAGGCTGACACCTC
>Pan_troglodytes  TTGGAATGACTCCTTTCGTTCTCACCTCCGCATAGGGAGGCTGACACCTC
>Gorilla_gorilla  TTGGGATGACTGCTTTCGTTCTCACCTCCGCATAGGGAGGCTGACACCTC
>Pongo_pygmaeus   TTGGGATGACTCCTTTCGTTCTCACCTCCGCATAGGGAGGCTGACACCTC
>Macaca_mulatta   TTGAGATGACTCCTTTAGTTCTCACCTCCGCGTAGGGAGGCTGACACCTC
>Callithrix_jacchus TTGGGATGACTCCTTTCGTTCTCACCTCTGGGTAGGGAGGCTGACACCTC
>Tarsius_syrichta GAGGGGGGGACCCCTT--TCTTCACC-CCGGGTAGGAAGGCTGACACCTC
>Tupaia_belangeri ACGGATTGCTCCTTTTCTTTCCAGCCTCAGGGTCAGAAGGCTGACACTTC
>Cavia_porcellus  gtgtgCACACGTCCCTTTTCTCCCTCCAGGGC--GGAGGCTGATCCCTC
>Mus_musculus     CAGGTGTAGCTCTT--TTCTCTCACCTCAGGGTAGGGAAGCTGACACCTC
>Rattus_norvegicus CGGGTGTAGCTCTT-TTTCTCTCACCTCAGGGTAGGGAAGCTGACACCTC
>Oryctolagus_cuniculus GGGGGCTGGCTCCC----TTCTACCCCAGCATAGGT-GGCTGACACTTC
>Bos_taurus       GAGGAATGGCTCCT----CTCTCACCTCAGGGTGGGGAAGCTGACATTTT
>Tursiops_truncatus GGGGAATGGC----TTCGTTCACACCTCAGGGTGGGGAGGTTGACACCTT
>Sus_scrofa       AGGTTATGGCTTCT----CCCTCACCTTAGGGTAGGAAGGCTGACACCTC
>Felis_catus      CGGGACTGGCTTTTTTC---CTCACCTCATGGTGGGGAGGCTGACACCTC
>Equus_caballus   CGGGAACGGCTCCT----TTCTCACCTCAGGGTAGGGGGGCTGACACCTC
>Myotis_lucifugus CCGGGATGGCTCCT----TTCTCACCT-----TTAACACCTC
>Pteropus_vampyrus TAGGAATGG-----TCCTTTCCCAACTCAGAGCAGGGAGGTTAACACCTT
>Loxodonta_africana GGGAGATGGCTCTTTTCTTCTCACCTCAGGGCAGAGAAGCTGACGCCTC
>Procavia_carpensis GGAGGATGGGTTTCTTTTCTCTCACCATAGGGTAGGGAGGCTGATGCCTC
```