Inter-Species Pathway Perturbation Prediction via Data Driven Detection of Functional Homology

Christoph Hafemeister [1], Roberto Romero [2], Erhan Bilal [3], Pablo Meyer [3],
Raquel Norel [3], Kahn Rhrissorrakrai [3], Richard Bonneau [1,4] and
Adi L. Tarca [2,5*]

[1]Department of Biology, Center for Genomics & Systems Biology, New York University, New York, NY, USA
[2]Perinatology Research Branch, NICHD/NIH, USA
[3]IBM Thomas J. Watson Research Center, Yorktown Heights, NY, USA
[4]Computer Science Department, Courant institute of Mathematical Sciences, New York University, New York, NY, USA
[5]Department of Computer Science, Wayne State University, Detroit, MI, USA

Supplemental Material

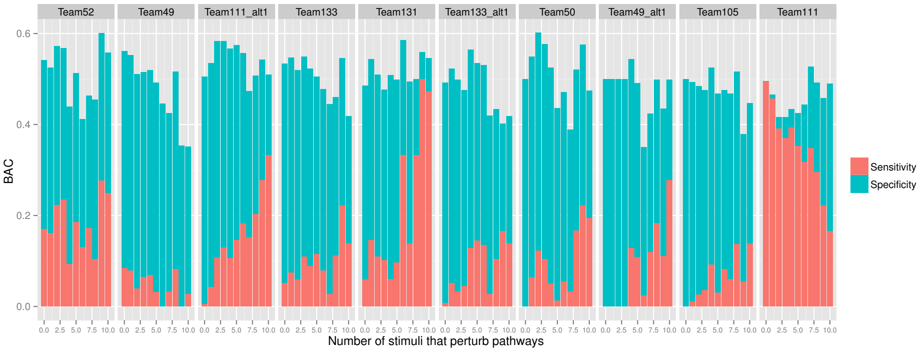

Figure S1: Team performance as a function of the number of stimuli that human pathways are perturbed by in the training data set.

---

*to whom correspondence should be addressed

Table S2: Specificity on the 70 pathways that are not perturbed in the human test stimuli.

	TN	FP	Specificity
Team49_alt1	1820	0	1.000
Team111_alt1	1811	9	0.995
Team50	1799	21	0.988
Team105	1797	23	0.987
Team49	1737	83	0.954
Team133	1735	85	0.953
Team133_alt1	1734	86	0.953
Team131	1566	254	0.860
Team52	1362	458	0.748
Team111	9	1811	0.005

Correlation Analysis of gene pairs

In this subsection we show how our response-ortholog pairs are correlated between rat and human in the training data, as well as the test data. For reference, we also show the correlations of sequence-based orthologs (HGNC).

The following plots summarize the spearman correlations of inter-species gene pairs. To calculate these correlations, we used the normalized ranks, which we derived from the moderated t-values as returned by LIMMA (see main method section Team49).

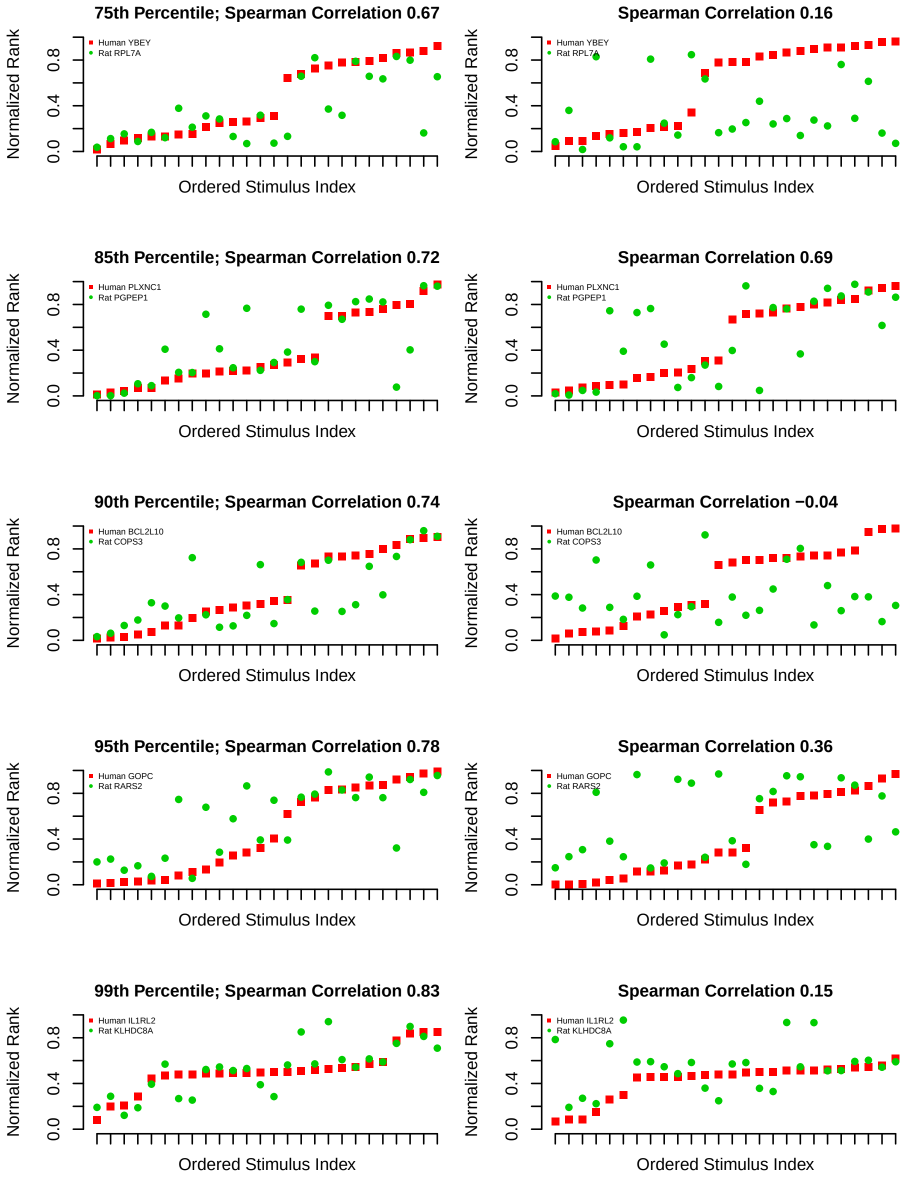

Figure S2: Example pairs for the Team49 method. The left column shows gene ranking in training stimuli, while the right column shows gene ranking in test stimuli. The gene pairs shown have the 75th, 80th, 90th, 95th and 99th percentile correlation in the training data.

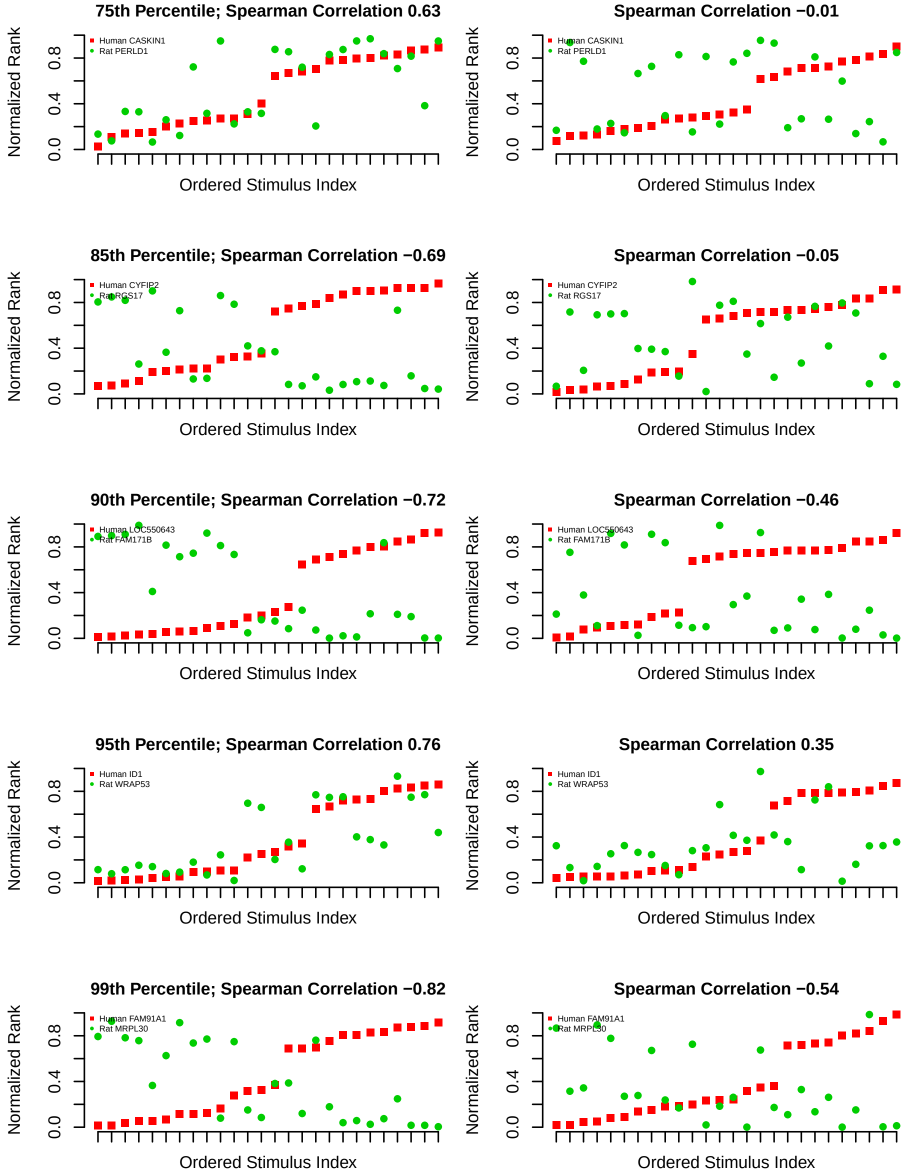

Figure S3: Example pairs for the Team133_alt1 method. The left column shows gene ranking in training stimuli, while the right column shows gene ranking in test stimuli. The gene pairs shown have the 75th, 80th, 90th, 95th and 99th percentile correlation in the training data.

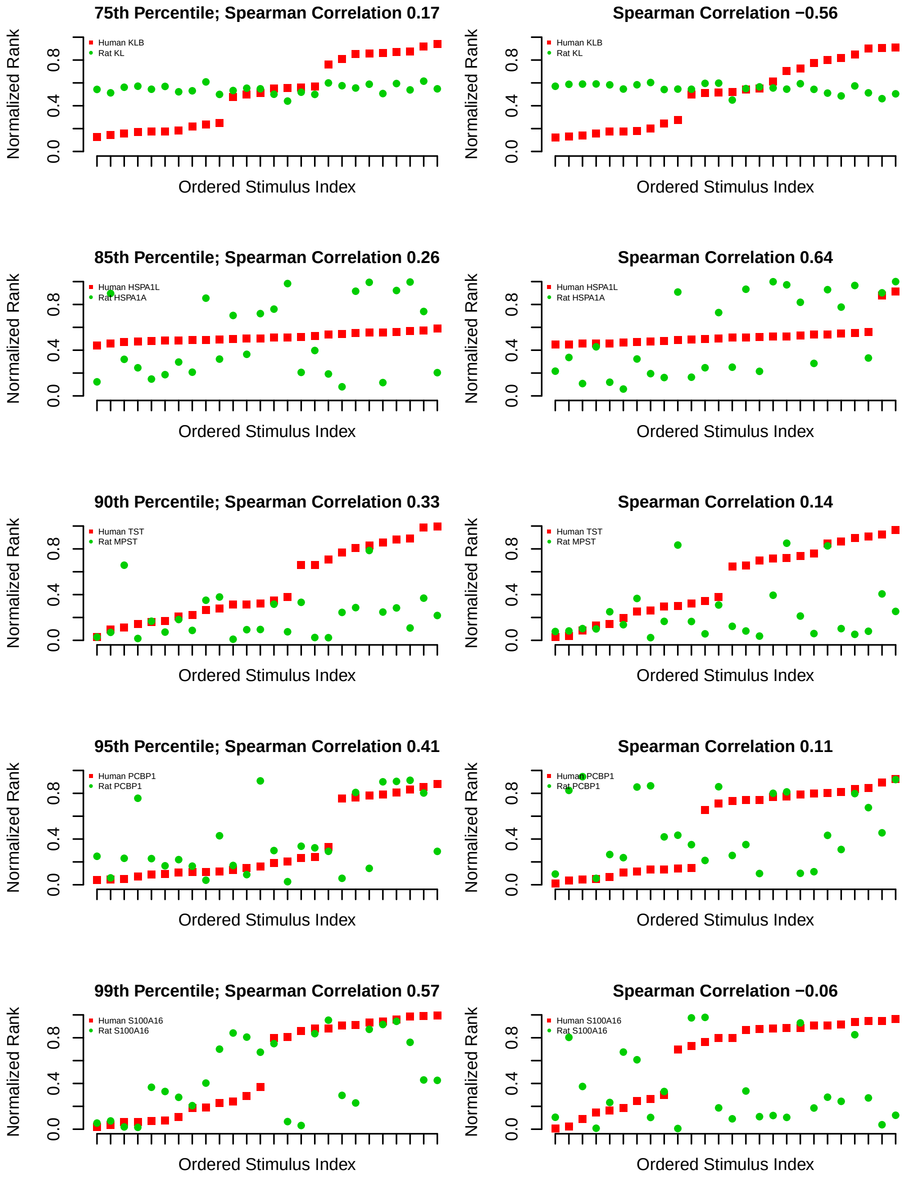

Figure S4: Example pairs for the sequence-based ortholog pairs (HGNC). The left column shows gene ranking in training stimuli, while the right column shows gene ranking in test stimuli. The gene pairs shown have the 75th, 80th, 90th, 95th and 99th percentile correlation in the training data.

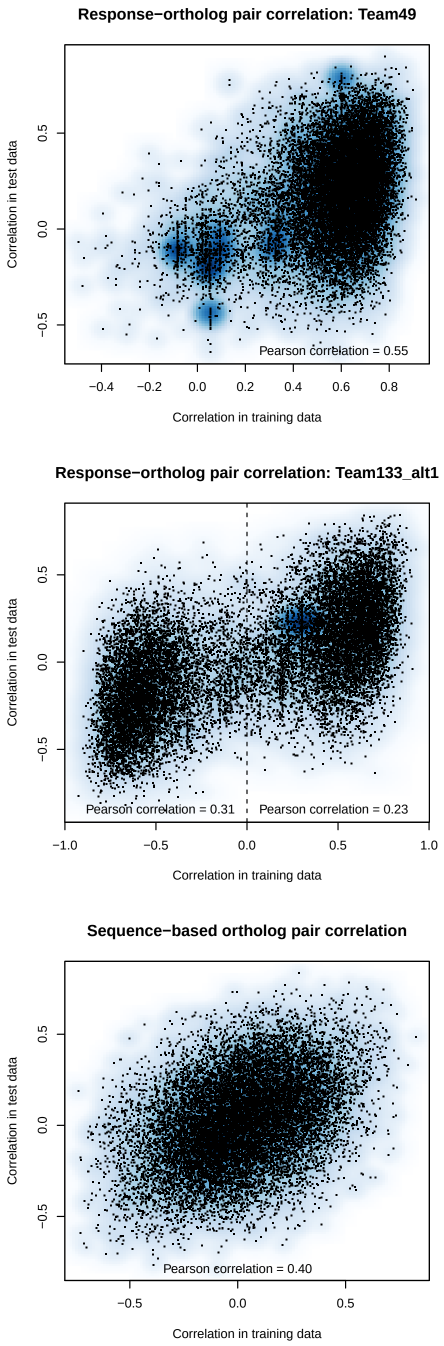

Figure S5: Gene pair correlation in training data vs test data. In all cases the correlation was significant ($p<0.0001$).