CAGI 2010 Assessments on submissions to the CBS dataset, Assessor Pauline Ng

To assess prediction performance on the CBS substitutions, we used 2 different tests:

1) Spearman's rank correlation, a nonparametric test that measures if the predicted score rankings correlate with the ranked observed activities

2) Area under the curve, where if protein activity was < 50% in high pyridoxine, the substitution was considered to affect protein function.Otherwise the substitution was considered to be neutral.
The 50% threshold was selected based on the distribution of the experimental data where there was a clear distinction between substitutions that gave 0% activity versus > 50% activity.

Submission ID	Algorithm*	High Pyridoxine, Spearman's rank correlation	Low Pyridoxine, Spearman's rank correlation	AUC High Pyroxidine	Notes
1		0.5687012	0.598783	0.8866782	
2		0.381131	0.3643787	0.6638655	
3		-0.1795474	-0.2011665	0.410084	
5		0.5341949	0.5203259	0.894958	
7		0.1591937	0.1667553	0.5907563	
8		0.6227359	0.6549464	0.8966387	
9	SIFT	0.5039223	0.5684457	0.7931034	
10		0.3749921	0.4327666	0.7731092	
13		0.531259	0.6023679	0.8369748	
14		0.4437076	0.3934007	0.7071429	
15		0.4267311	0.4911984	0.7870968	
18		0.5773559	0.5485727	0.8479167	
19		0.1483933	0.2601728	0.6063218	
20		0.08356909	0.07079727	0.6521008	
22		0.6091524	0.2900731	0.8344538	
23		0.2687158	0.2323854	0.611828	
4		0.4081997	0.4174941	0.7378151	late submission
6					late submission, submitted only raw output from algorithm
11	Align-GVGD + SIFT, with curated alignments	0.5394849	0.5850496	0.8680672	late submission
12	Align-GVGD, with curated alignments	0.4387335	0.4527056	0.8016807	late submission
16		0.315134	0.3441739	0.6473118	late submission
17					late submission, submitted only raw output from algorithm
21					late submission, submitted only raw output from algorithm

Algorithm* = The predictors have de-anonymized themselves, and allowed CAGI organizers to disclose the name of the algorithm.
The late submissions with no submitted relative growth rate values were not assessed with Spearman's rank correlation and AUC measures.

R commands

```
---------------------------------------------------------------
#high pyrodoxine, Spearman's rank
cor.test (x[,2], x[,7], method="spearman", na.rm=TRUE)
---------------------------------------------------------------
#low pyrox, Spearman's ranknonparametric
cor.test (x[,4], x[,9], method="spearman", na.rm=T)
---------------------------------------------------------------
# Area Under the Curve
obs <- ifelse(x[,2] < 50, 0, 1)
 #  if high proxidine in experimnetal is < 50, consider it damaging

highpyroxpred <- x[,7]
# Area under the Curve (asumes binary affect fxn/neutral
m <- somers2(highpyroxpred, obs, na.rm=T)
```