

Supplementary Table 4 Normalized connectivity of phenotypes from the training set and the prediction set. It is evident that phenotype vectors from the training set have markedly more high-scoring phenotype associations (in the interval 0.1-0.525, after which the difference evens out) than phenotype vectors representing phenotypes in the prediction set. This is because files describing diseases where a gene has been identified are often more meticulous in describing the phenotype, (less noisy), than files reporting a critical interval, and a phenotype, but no identified gene.

Phenotype association score (interval)	Normalized value training set	Normalized value prediction set
0.10-0.125	112.0577	62.95867
0.125-0.15	42.3953	27.84386
0.15-0.175	18.37393	14.02181
0.175-0.2	8.863248	7.644087
0.2-0.225	5.114672	4.523536
0.225-0.25	3.14886	2.963261
0.25-0.275	2.06339	1.949483
0.275-0.3	1.579772	1.352468
0.3-0.325	1.07906	0.995408
0.325-0.35	0.918803	0.774971
0.35-0.375	0.701567	0.647532
0.375-0.4	0.577635	0.495982
0.4-0.425	0.452279	0.407577
0.425-0.45	0.361111	0.342135
0.45-0.475	0.340456	0.280138
0.475-0.5	0.252849	0.19977
0.5-0.525	0	0
0.525-0.55	0.197293	0.19977
0.55-0.575	0.165954	0.156142
0.575-0.6	0.148148	0.107922
0.6-0.625	0.136752	0.153846
0.625-0.65	0.090456	0.106774
0.65-0.675	0.078348	0.082664
0.675-0.7	0.064103	0.065442
0.7-0.725	0.054843	0.055109
0.725-0.75	0.05057	0.052813
0.75-0.775	0.0349	0.06659
0.775-0.8	0.019231	0.041332
0.8-0.825	0.017094	0.025258
0.825-0.85	0.020655	0.025258
0.85-0.875	0.007123	0.025258
0.875-0.9	0.002849	0.027555
0.9-0.925	0.002849	0.021814
0.925-0.95	0	0.006889
0.95-0.975	0	0.011481
0.975-1	0.001425	0.006889