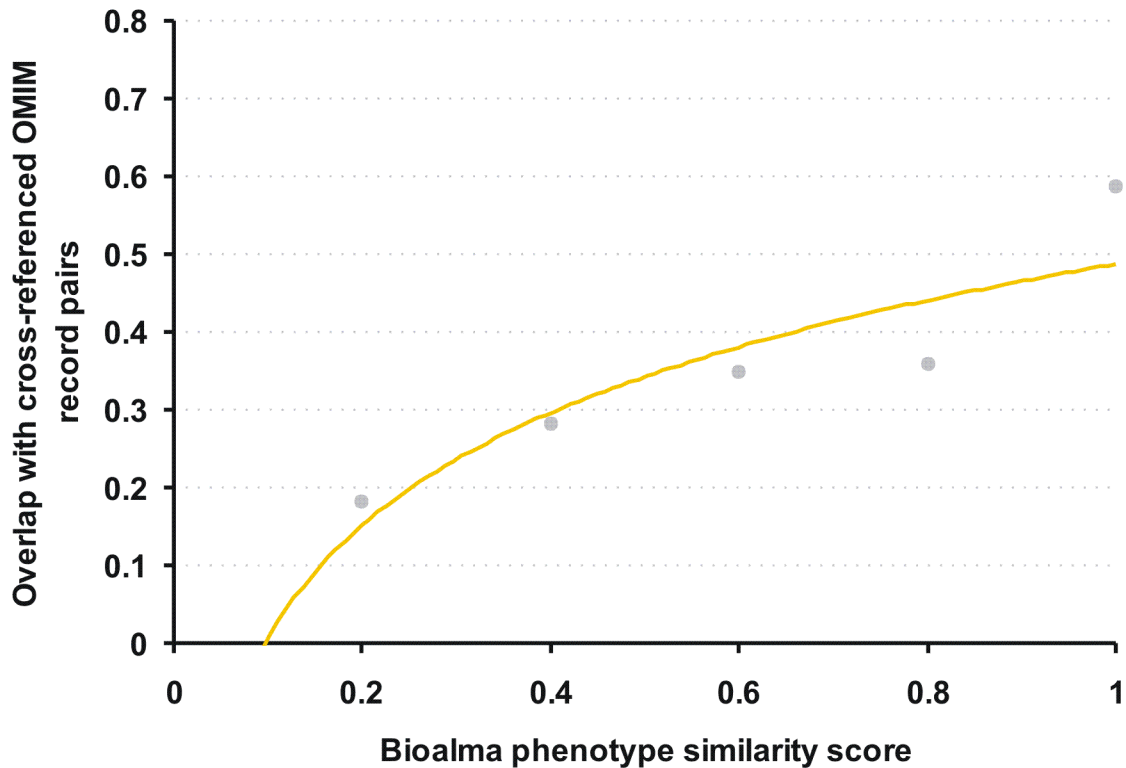


Correlation Bioalma phenotype scores and cross-referenced OMIM record pairs



Supplementary Figure 8 Robustness of the cosine phenotype similarity measure. We fitted a calibration curve of the phenotype association score against the overlap with cross-referenced OMIM record pairs with a high degree of phenotypic overlap. Like with the OMIM based phenotype association scores, the Bioalma scores directly correlate with the probability of overlap with the benchmarking file pairs. Hereby we show that the cosine similarity measure also works for phenotype vectors built from a different text source, extraction method, weighing method, and vocabulary. This is an indication that the phenotype vector cosine similarity method is a robust way of capturing phenotype similarity between records.