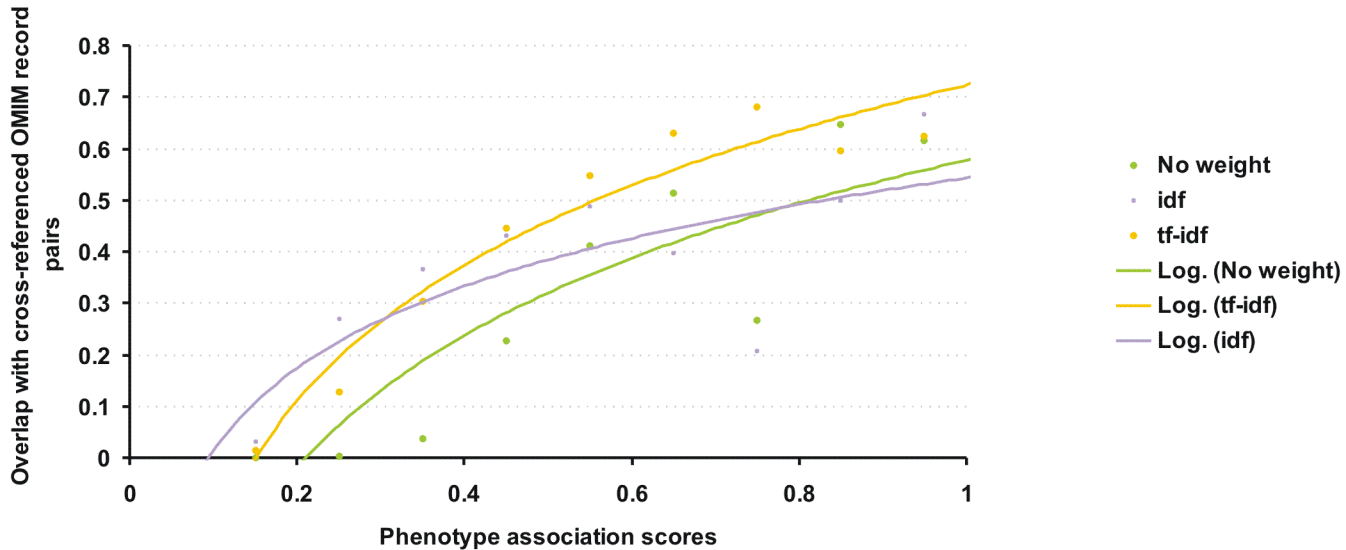


Benchmark of phenotype association scores, weighted and unweighted



**Supplementary Figure 7** Influence of tf-idf weight on phenotype similarity scoring scheme. The weighing procedure markedly improves the predictive power of the phenotype association score. We benchmarked phenotype association vectors constructed without the weight (green), only using the idf weight (purple), and using the full tf-idf weight (orange). When evaluated over the whole interval of scores the vectors constructed with the tf-idf weight markedly outperforms the other vectors. We also see that the vectors constructed using the idf weight perform worse than the vectors constructed with the tf-idf weight, but better than the unweighted vectors. Hereby, we show the impact of the weight on the predictive power of the procedure, and that we have chosen the optimal way of weighing the vectors.