PCC	$r = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)\left(\frac{y_i - \bar{y}}{\sigma_y}\right)$
Uncentered Correlation Coefficient	(same as PCC, except that sample means are set to 0)
Frequency Dot Product (FDP) (note: we've been calling this idf_unnorm) 1.	$FDP = \sum_{i=1}^{n}\left[(x_i)(y_i)\left(\log_2\frac{D}{count_i}\right)^2\right]$ • x_i and y_i are the i[th] phenotypes in the binary phenotype vectors $\vec{x}$ and $\vec{y}$. • n = total number of phenotypes • D = total number of binary phenotype vectors • $count_i$ = count of number of times the phenotype i appears in the data
Inverse Document Frequency (IDF)	$IDF = \frac{FDP}{\lvert\vec{x}\rvert * \lvert\vec{y}\rvert}$ • IDF is the FDP normalized by the lengths of the two vectors.
Euclidean Distance	$d = \sum_{i=1}^{n}(x_i - y_i)^2$
Jaccard Similarity Coefficient	$J = \frac{M_{11}}{M_{01} + M_{10} + M_{11}}$ • M_{11} = # of times 1 is observed in both vectors • $M_{01} + M_{10}$ = # of times 1 is observed in exactly one of the vectors in the pair.
Mutual Information	$I(X;Y) = \sum_{y\in Y}\sum_{x\in X} p(x,y)\log\left(\frac{p(x,y)}{p(x)\,p(y)}\right)$
Residual IDF (RIDF)	$RIDF = IDF - \log_2\frac{1}{1 - Poisson(0;\lambda_i)}$ • RIDF is the difference between the actual IDF and the inverse document frequency predicted by a Poisson distribution. • λ_i is the Poisson parameter, the average number of occurrences of the phenotype in each phenotype vector.

Table of measures tested and their mathematical formulas.