GWAS Explorer (GWAX) Formulae

Jeremy Yang

May 2019

1 GWAX Gene-Trait Association Scoring

1.1 RCRAS = Relative Citation Ratio (RCR) Aggregated Score

The purpose is to evaluate the evidence for a gene-trait association, by aggregating multiple studies and their corresponding publications. The iCite[1] RCR[1] is itself a statistic designed to evaluate the evolving empirical impact of a publication (in contrast to the non-empirical impact factor). Hence by aggregating RCRs we seek an corresponding measure of scientific community impact.

$$RCRAS_{gt} = \sum_{study} \left(\frac{1}{gc} \sum_{pub} \frac{log_2(RCR+1)}{sc} \right) \tag{1}$$

Where

$$\begin{aligned} study &= \text{GWAS (study\_accession)} \\ gc &= \text{gene count (in study)} \\ pub &= \text{publication (PubMed ID)} \\ sc &= \text{study count (in pub)} \end{aligned}$$

RCR $median = 2.0$ and $90\%ile = 8.5$. The $log_2()$ function is used with the belief that additional publications add diminishing evidence. Division by *spp* effects a partial count for papers associated with multiple studies. Since $RCR >= 0$, $log_2(RCR+1) >= 0$ and intuitively, when $RCR = 1$ and $sc = 1$, $log_2(RCR+1) = 1$, Similarly division by gc reflects a partial count since papers and studies may associate to few or many findings and reported genes.

From GWAS Catalog[2] and iCite PubMed statistics.

This approach informed by bibliometric methodology, including fractional counting, of Jensen et al. as employed in DISEASES and TIN-X.
As with loss functions, the absolute value is less important than the gradient, which solely

[1] https://icite.od.nih.gov/
[2] https://www.ebi.ac.uk/gwas/

determines ranking. For example, ten supporting publications may not be twice the evidence than five, but is certainly more.

References

[1] BI Hutchins, X Yuan, JM Anderson, GM Santangelo. *Relative Citation Ratio (RCR): A New Metric That Uses Citation Rates to Measure Influence at the Article Level*. PLoS Biol 14(9): e1002541, 2016, https://doi.org/10.1371/journal.pbio.1002541.