WeightedPvalue: A software for integrative genomic data analysis via p-value weight adjustment

This is a software for integrative genomic data analysis via p-value weight adjustment (1). The method is designed to identify $X->Z->Y$, for example: miRNA -> mRNA -> Disease. For each triplet, the p-value (p) is first calculated based on the associations between the variables in X and Y. Then weights (w^*) are derived according to the strength of associations between X, Z, and Z and Y. The weighted p-value (p^*) is calculated as:

$$p^* = \frac{p}{w^*}.$$

The weighted p-value can then be used to control the family-wise error rate through the Bonferroni correction or the Holm's generalized sequential rejective procedures.

There are three files: Source.R, RunSim.R, and RunRealData.R.

- Source.R contains the source code for calculating weighted p-value.

- RunSim.R contains the r code for running simulation analysis in the paper. Need to source the Source.R file in the beginning. The algorithm produces several output values:

1. bon.raw: a scalar value of 0 or 1 indicating whether any significance (1: at least 1 X declares significance) is declared based on the original p-value (p) via Bonferroni correction procedure
2. bon.li: a scalar value of 0 or 1 indicating whether any significance is declared based on the weighted p-value proposed by Li et al. (2) via the Bonferroni correction procedure
3. bon.minp: a scalar value of 0 or 1 indicating whether any significance is declared based on our proposed weighted p-value (p^*) via the Bonferroni correction procedure
4. bon.minsc: a scalar value of 0 or 1 indicating whether any significance is declared based on our proposed weighted p-value (p^*) with a sample splitting approach via the Bonferroni correction procedure
5. hom.raw: a scalar value of 0 or 1 indicating whether any significance is declared based on the original p-value via the Holm's generalized sequential rejective procedure
6. hom.li: a scalar value of 0 or 1 indicating whether any significance is declared based on the weighted p-value proposed by Li et al. (2) via the Holm's generalized sequential rejective procedure
7. hom.min: a scalar value of 0 or 1 indicating whether any significance is declared based on our proposed weighted p-value (p^*) via the Holm's generalized sequential rejective procedure
8. hom.scp: a scalar value of 0 or 1 indicating whether any significance is declared based on our proposed weighted p-value (p^*) with a sample splitting approach via the Holm's generalized sequential rejective procedure

- RunRealData.R contains the r code for analyzing a dataset. Need to source the Source.R file in the beginning. The method is designed to identify $X->Z->Y$, for example: miRNA -> mRNA -> Disease. The column names of the dataset need to contain X, Y, Z to indicate the types of data. The main output is stored in the object ans which contains two columns:

1. minp: weighted p-values (p^*) derived using our proposed method
2. h.min: a vector of 0 or 1 indicating whether the corresponding Xs are significant, based on the Holm's generalized sequential rejective procedure

Reference

1. Ho Y.-Y., Ma C., Zhang W., Huang H.-H., Habiger J.D., Nho R. (2020) A powerful integrative genomics approach to incorporate both prior knowledge and multiple data types through p-value weight adjustment (Under Review).

2. Li, L., Kabesch, M., Bouzigon, E., Demenais, F., Farrall, M., Moffatt, M. F., Lin, X., and Liang, L. (2013). Using eQTL weights to improve power for genome-wide association studies: a genetic study of childhood asthma. Frontiers in Genetics 4, 103.