scalnet: Scalable network estimation with *L0* penalty

Junghi Kim[1], Hongtu Zhu[2], Xiao Wang[3], Kim-Anh Do[4]

[1]DBVIII, Office of Biostatistics, CDER, FDA, [2]Department of Biostatistics, Univ of North Carolina, [3]Department of Statistics, Purdue Univ, [4]Department of Biostatistics, MD Anderson Cancer Center

1.Background

- Estimating a large and sparse precision matrix for genomic data has garnered substantial research attention for its direct application to classification analysis and graphical models.
- Most existing methods either use a lasso-type penalty that may lead to biased estimators or are computationally intensive, which prevents their application to very large graphs.

2. Method

- We propose using an *L*0 penalty to estimate an ultra-large precision matrix (*scalnetL0*), which includes two main procedures:
  1) convert the original problem into many small-scale linear regression problems.
  2) adopt an *L*0 penalty method to solve these linear regression problems.
- Software Installation from GitHub

```
devtools::install_github('BIG-S2/SCALNET')
```

3. Application

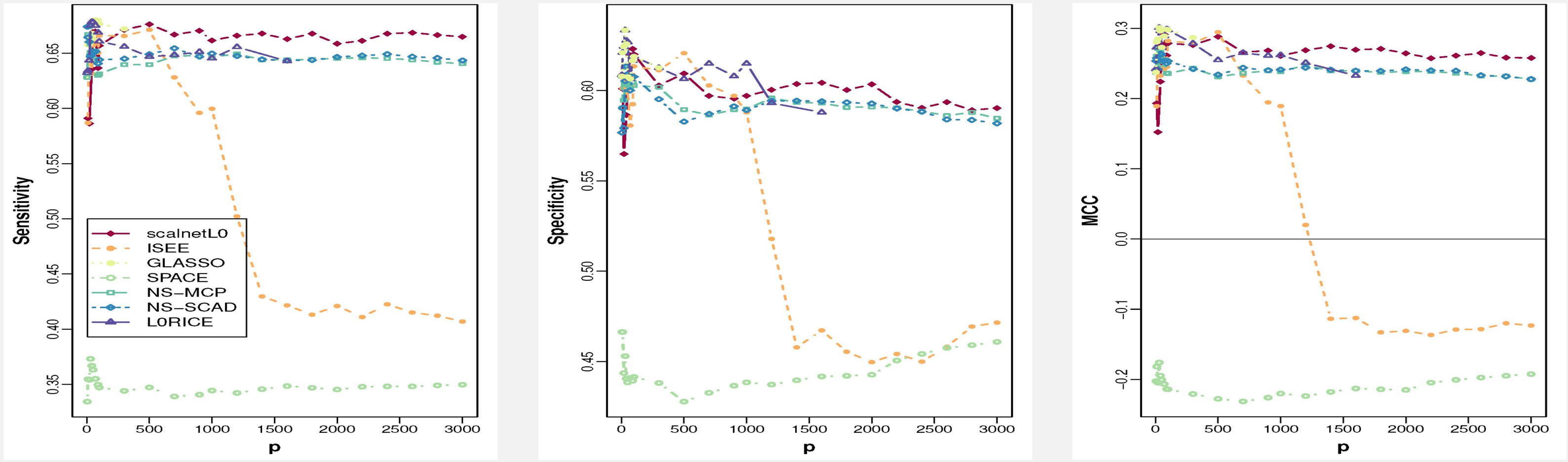
3.1. Linear Discriminant Analysis (LDA)

Using gene expression data, we perform a linear discriminant analysis for patients' survival times to predict whether individual patients with breast cancer will experience long or short survival times using their gene expression profiles.

3.2. Results

The *scalnetL0* can handle a larger number of genes than many existing methods (Figure 1).

Figure 1. Classification results for survival times in the TCGA breast cancer data, with the x-axis representing the number of expression profiles.



4. Acknowledgement The present study was funded in part by an appointment to the postdoctoral fellowship program at the University of Texas MD Anderson Cancer Center. The simulation studies, data analyses, and manuscript drafts were completed during JK's postdoctoral training program at the MD Anderson Cancer Center.

5. Disclaimer The present study reflects the views of the authors and should not be constructed to represent the views or policies of the US Food and Drug Administration.