**Supplementary Materials**

**Methods**

***Comparison of Cross Validation, BIC and mBIC in model fitting***

**Notation.** We denote the true model ST = {1,…, p0}, containing p0 significant variables, while the full model SF= {1,…, p}, having p (p p0 ) variables. The definitions of overfitted and underfitted models follow that in Wang et al. (2007). Specifically, for any candidate model S, if ST S, S is referred to an underfitted model in the sense it misses at least one important variable, In contrast, any SST other than ST itself is referred to as an overfitted model in the sense it contains all significant variables, but also at least one insignificant variable.

**Simulation Studies.** The comparison of optimal penalty parameter selection among Cross validation (CV), BIC and mBIC are performed on the basis of 100 simulated datasets for each simulation scenario (sample size n = 50, 100, 200, 400 and 800; p = 40 and 300 genes). We don’t include the n = 10 scenario here simply because the sample size is too small to perform 5-fold CV, which we used for comparison in the study. For each simulation scenario, we use CV, BIC and mBIC for selecting the regularization penalty parameter λ1 and compare their performance in model fitting. The percentages of underfitted and overfitted models out of the 100 simulated datasets are summarized, as well as the precision, recall and the F-scores of the inferred models (Supplementary Table 4). Traditionally, CV includes several common forms, k-fold, leave-one-out, and the generalized cross-validation. Here we describe the 5-fold CV, which was used in our study. The steps are following: 1) Data consisting of n observations are divided at random into 5 mutually exclusive subsamples, known as 5-folds; 2) The partial correlation network structure is built using the data omitting one fold; 3) The fitted network is then used for the prediction of the omitted subsample and the prediction error is obtained against each choice of the tuning parameter λ1; 4) The value λ1 which minimizes the prediction error is considered the optimal choice of the tuning parameter. The details on BIC and mBIC for regularization parameter selection can be found in the manuscript section 2.4.

**Tables**

**Supplementary Table 1. Performance of pLasso with** **prior information provided at different precision levels.** µ, the multiplication term for getting the optimal λ2, where λ2 = µ\*λ1. pBIC, the optimal minimum BIC score corresponding to the selected λ2.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | ***Number of Nodes p= 40*** | | | | | | | | | |
|  | *n = 10* | | | | | *n = 100* | | | | |
| Precision of prior information | µ | pBIC | Precision | Recall | F-score | µ | pBIC | Precision | Recall | F-score |
| 0.1 | 0.67 | -0.78  ±0.72 | 0.21  ±0.04 | 0.16  ±0.03 | 0.18  ±0.03 | 0.98 | -12.42  ±4.55 | 0.69  ±0.05 | 0.69  ±0.06 | 0.69  ±0.05 |
| 0.2 | 0.53 | -1.16  ±0.95 | 0.24  ±0.05 | 0.21  ±0.05 | 0.22  ±0.05 | 0.75 | -14.20  ±3.64 | 0.68  ±0.05 | 0.71  ±0.05 | 0.69  ±0.03 |
| 0.3 | 0.50 | -1.61  ±0.99 | 0.29  ±0.05 | 0.26  ±0.05 | 0.27  ±0.05 | 0.59 | -14.33  ±3.88 | 0.66  ±0.05 | 0.75  ±0.05 | 0.70  ±0.04 |
| 0.4 | 0.42 | -2.43  ±1.06 | 0.34  ±0.04 | 0.33  ±0.05 | 0.33  ±0.04 | 0.46 | -16.77±3.53 | 0.66  ±0.04 | 0.78  ±0.04 | 0.71  ±0.03 |
| 0.5 | 0.35 | -2.90  ±0.93 | 0.40  ±0.04 | 0.40  ±0.05 | 0.40  ±0.04 | 0.42 | -20.90  ±4.63 | 0.67  ±0.04 | 0.80  ±0.05 | 0.73  ±0.04 |
| 0.6 | 0.34 | -3.98  ±0.98 | 0.46  ±0.06 | 0.47  ±0.07 | 0.46  ±0.06 | 0.33 | -24.35  ±2.76 | 0.68  ±0.03 | 0.83  ±0.04 | 0.75  ±0.03 |
| 0.7 | 0.31 | -4.83  ±0.94 | 0.55  ±0.06 | 0.56  ±0.07 | 0.55  ±0.06 | 0.26 | -28.27  ±3.67 | 0.70  ±0.04 | 0.88  ±0.03 | 0.78  ±0.03 |
| 0.8 | 0.26 | -6.08  ±0.91 | 0.65  ±0.04 | 0.67  ±0.07 | 0.66  ±0.05 | 0.22 | -32.10  ±4.56 | 0.76  ±0.04 | 0.91  ±0.02 | 0.83  ±0.02 |
| 0.9 | 0.25 | -6.99  ±1.14 | 0.73  ±0.04 | 0.76  ±0.05 | 0.74  ±0.03 | 0.20 | -36.13  ±3.38 | 0.81  ±0.04 | 0.96  ±0.02 | 0.88  ±0.02 |
| 1.0 | 0.25 | -8.02  ±0.89 | 0.78  ±0.05 | 0.83  ±0.08 | 0.80  ±0.05 | 0.16 | -39.92  ±4.00 | 0.88  ±0.04 | 0.98  ±0.02 | 0.93  ±0.02 |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | ***Number of Nodes p= 300*** | | | | | | | | | |
|  | *n = 10* | | | | | *n =100* | | | | |
| Precision of prior information | µ | pBIC | Precision | Recall | F-score | µ | pBIC | Precision | Recall | F-score |
| 0.1 | 0.85 | -3.81  ±0.34 | 0.04  ±0.01 | 0.05  ±0.01 | 0.04  ±0.01 | 0.93 | 6.87  ±0.77 | 0.36  ±0.02 | 0.27  ±0.02 | 0.31  ±0.02 |
| 0.2 | 0.81 | -3.86  ±0.31 | 0.05  ±0.01 | 0.06  ±0.02 | 0.06  ±0.02 | 0.79 | 6.45  ±0.55 | 0.38  ±0.02 | 0.30  ±0.02 | 0.33  ±0.02 |
| 0.3 | 0.75 | -4.08  ±0.32 | 0.07  ±0.02 | 0.08  ±0.03 | 0.08  ±0.02 | 0.58 | 5.05  ±0.57 | 0.41  ±0.02 | 0.37  ±0.03 | 0.39  ±0.02 |
| 0.4 | 0.74 | -4.12  ±0.26 | 0.08  ±0.02 | 0.08  ±0.03 | 0.08  ±0.03 | 0.42 | 2.81  ±0.61 | 0.43  ±0.01 | 0.36  ±0.01 | 0.44  ±0.02 |
| 0.5 | 0.67 | -4.42  ±0.40 | 0.11  ±0.02 | 0.12  ±0.03 | 0.11  ±0.03 | 0.33 | -0.34  ±0.78 | 0.47  ±0.01 | 0.45  ±0.02 | 0.50  ±0.01 |
| 0.6 | 0.65 | -4.77  ±0.37 | 0.16  ±0.02 | 0.17  ±0.03 | 0.17  ±0.03 | 0.25 | -4.20  ±1.09 | 0.51  ±0.01 | 0.54  ±0.02 | 0.56  ±0.01 |
| 0.7 | 0.55 | -5.13  ±0.32 | 0.20  ±0.04 | 0.23  ±0.05 | 0.22  ±0.05 | 0.20 | -8.60  ±0.84 | 0.56  ±0.01 | 0.63  ±0.03 | 0.63  ±0.01 |
| 0.8 | 0.50 | -5.74  ±0.48 | 0.26  ±0.03 | 0.30  ±0.04 | 0.28  ±0.04 | 0.15 | -13.51  ±0.97 | 0.62  ±0.01 | 0.71  ±0.01 | 0.70  ±0.01 |
| 0.9 | 0.47 | -6.32  ±0.31 | 0.32  ±0.05 | 0.37  ±0.07 | 0.34  ±0.06 | 0.12 | -18.80  ±1.20 | 0.70  ±0.01 | 0.80  ±0.02 | 0.79  ±0.02 |
| 1.0 | 0.42 | -7.06  ±0.39 | 0.39  ±0.05 | 0.47  ±0.07 | 0.42  ±0.06 | 0.10 | -23.29  ±0.90 | 0.77  ±0.01 | 0.98  ±0.01 | 0.86  ±0.01 |

**Supplementary Table 2. The effects of estimated precision level on the method performance measured by the F-score.** The entries in bold represent the F-scores based on the true precision value.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Small Network  (p = 40) | n = 10 | F-score | | Estimated precision (α) | | | | | | | | | |
| 0.1 | 0.3 | 0.5 | | 0.6 | | 0.9 | | 1.0 | |
| True  Precision | 0.1 | **0.16 (0.03)** | 0.15 (0.02) | 0.12 (0.03) | | 0.11 (0.02) | | 0.10 (0.02) | | 0.10 (0.03) | |
| 0.3 | 0.24 (0.06) | **0.26 (0.05)** | 0.28 (0.04) | | 0.29 (0.04) | | 0.30 (0.03) | | 0.30 (0.03) | |
| 0.6 | 0.33 (0.07) | 0.36 (0.08) | 0.39 (0.05) | | **0.41 (0.05)** | | 0.43 (0.03) | | 0.43 (0.03) | |
| 0.9 | 0.55 (0.10) | 0.60 (0.10) | 0.67 (0.07) | | 0.68 (0.07) | | **0.73 (0.05)** | | 0.73 (0.05) | |
| n = 100 | F-score | | Estimated precision (α) | | | | | | | | | |
|  | | 0.1 | 0.3 | 0.5 | | 0.6 | | 0.9 | | 1.0 | |
| True  Precision | 0.1 | **0.70 (0.04)** | 0.70 (0.04) | 0.70 (0.05) | | 0.68 (0.05) | | 0.60 (0.05) | | 0.59 (0.04) | |
| 0.3 | 0.71 (0.03) | **0.71 (0.03)** | 0.68 (0.03) | | 0.67 (0.04) | | 0.63 (0.03) | | 0.62 (0.03) | |
| 0.6 | 0.77 (0.03) | 0.77 (0.03) | 0.76 (0.05) | | **0.76 (0.05)** | | 0.75 (0.03) | | 0.74 (0.03) | |
| 0.9 | 0.88 (0.02) | 0.88 (0.02) | 0.88 (0.02) | | 0.88 (0.02) | | **0.88 (0.02)** | | 0.88 (0.02) | |
| Large Network  (p = 300) | n = 10 | F-score | | Estimated precision (α) | | | | | | | | | |
| 0.1 | 0.3 | | 0.5 | | 0.6 | | 0.9 | | 1.0 |
| True  Precision | 0.1 | **0.037 (0.004)** | 0.035 (0.004) | | 0.032 (0.004) | | 0.031 (0.003) | | 0.028 (0.003) | | 0.026 (0.002) |
| 0.3 | 0.06 (0.02) | **0.08 (0.02)** | | 0.10 (0.02) | | 0.12 (0.02) | | 0.14 (0.02) | | 0.14 (0.02) |
| 0.6 | 0.11 (0.02) | 0.13 (0.03) | | 0.17 (0.03) | | **0.20 (0.03)** | | 0.23 (0.03) | | 0.23 (0.04) |
| 0.9 | 0.14 (0.03) | 0.18 (0.04) | | 0.26 (0.03) | | 0.31 (0.05) | | **0.37 (0.04)** | | 0.37 (0.06) |
| n = 100 | F-score | | Estimated precision (α) | | | | | | | | | |
| 0.1 | 0.3 | 0.5 | | 0.6 | | 0.9 | | 1.0 | |
| True  Precision | 0.1 | **0.30 (0.01)** | 0.30 (0.01) | 0.30 (0.01) | | 0.28 (0.02) | | 0.26 (0.01) | | 0.26 (0.01) | |
| 0.3 | 0.36 (0.02) | **0.39 (0.01)** | 0.40 (0.01) | | 0.40 (0.01) | | 0.38 (0.01) | | 0.38 (0.01) | |
| 0.6 | 0.52 (0.03) | 0.55 (0.03) | 0.56 (0.05) | | **0.56 (0.01)** | | 0.56 (0.01) | | 0.56 (0.01) | |
| 0.9 | 0.73 (0.02) | 0.75 (0.02) | 0.76 (0.01) | | 0.77 (0.01) | | **0.78 (0.01)** | | 0.78 (0.01) | |

**Supplementary Table 3. The effects of estimated precision level on the selection of optimal µ.** The entries in bold represent the optimal µ based on the true precision value.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Small Network  (p = 40) | n = 10 | Optimal µ | | Estimated precision (α) | | | | | |
| 0.1 | 0.3 | 0.5 | 0.6 | 0.9 | 1.0 |
| True  Precision | 0.1 | **0.7** | 0.6 | 0.4 | 0.4 | 0.3 | 0.2 |
| 0.3 | 0.7 | **0.5** | 0.4 | 0.4 | 0.3 | 0.2 |
| 0.6 | 0.6 | 0.5 | 0.4 | **0.3** | 0.3 | 0.3 |
| 0.9 | 0.5 | 0.4 | 0.4 | 0.3 | **0.3** | 0.2 |
| n = 100 | Optimal µ | | Estimated precision (α) | | | | | |
|  | | 0.1 | 0.3 | 0.5 | 0.6 | 0.9 | 1.0 |
| True  Precision | 0.1 | **0.9** | 0.9 | 0.7 | 0.6 | 0.3 | 0.2 |
| 0.3 | 0.7 | **0.6** | 0.5 | 0.5 | 0.2 | 0.1 |
| 0.6 | 0.4 | 0.4 | 0.3 | **0.3** | 0.2 | 0.2 |
| 0.9 | 0.3 | 0.3 | 0.2 | 0.2 | **0.2** | 0.2 |
| Large Network  (p = 300) | n = 10 | Optimal µ | | Estimated precision (α) | | | | | |
| 0.1 | 0.3 | 0.5 | 0.6 | 0.9 | 1.0 |
| True  Precision | 0.1 | **0.8** | 0.7 | 0.6 | 0.6 | 0.4 | 0.4 |
| 0.3 | 0.8 | **0.8** | 0.6 | 0.6 | 0.4 | 0.4 |
| 0.6 | 0.8 | 0.8 | 0.6 | **0.6** | 0.4 | 0.4 |
| 0.9 | 0.8 | 0.7 | 0.6 | 0.6 | **0.5** | 0.4 |
| n = 100 | Optimal µ | | Estimated precision (α) | | | | | |
| 0.1 | 0.3 | 0.5 | 0.6 | 0.9 | 1.0 |
| True  Precision | 0.1 | **0.9** | 0.8 | 0.6 | 0.4 | 0.1 | 0.1 |
| 0.3 | 0.8 | **0.6** | 0.4 | 0.3 | 0.2 | 0.1 |
| 0.6 | 0.5 | 0.4 | 0.3 | **0.3** | 0.2 | 0.1 |
| 0.9 | 0.3 | 0.3 | 0.2 | 0.2 | **0.1** | 0.1 |

**Supplementary Table 4. Comparison of Lasso performance with optimal λ1 value determined by 5-fold Cross validation (CV), BIC and mBIC.** The percentage of underfitted models and overfitted models were calculated based on 100 simulated datasets.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Nodes | Sample size | Criteria | Underfitted (%) | Overfitted (%) | Precision |  | Recall | F-score |
| 40 | 50 | CV | 100 | 0 | 0.38(0.04) |  | 0.67(0.05) | 0.48(0.05) |
|  |  | BIC | 100 | 0 | 0.70(0.05) |  | 0.38(0.03) | 0.49(0.05) |
|  |  | mBIC | 100 | 0 | 0.59(0.04) |  | 0.49(0.04) | 0.53(0.04) |
|  | 100 | CV | 100 | 0 | 0.37(0.04) |  | 0.90(0.02) | 0.52(0.04) |
|  |  | BIC | 100 | 0 | 0.69(0.04) |  | 0.71(0.04) | 0.69(0.04) |
|  |  | mBIC | 100 | 0 | 0.68(0.04) |  | 0.72(0.04) | 0.69(0.04) |
|  | 200 | CV | 68 | 32 | 0.37(0.05) |  | 0.98(0.03) | 0.53(0.04) |
|  |  | BIC | 99 | 1 | 0.72(0.03) |  | 0.91(0.03) | 0.80(0.03) |
|  |  | mBIC | 99 | 1 | 0.71(0.03) |  | 0.91(0.04) | 0.80(0.03) |
|  | 400 | CV | 8 | 92 | 0.38(0.04) |  | 1.00(0.01) | 0.55(0.03) |
|  |  | BIC | 63 | 37 | 0.73(0.03) |  | 0.99(0.02) | 0.84(0.03) |
|  |  | mBIC | 63 | 37 | 0.73(0.03) |  | 0.99(0.02) | 0.84(0.03) |
|  | 800 | CV | 1 | 99 | 0.43(0.04) |  | 1.00(0.01) | 0.60(0.02) |
|  |  | BIC | 13 | 87 | 0.77(0.03) |  | 1.00(0.01) | 0.87(0.02) |
|  |  | mBIC | 14 | 86 | 0.77(0.03) |  | 1.00(0.01) | 0.87(0.02) |
| 300 | 50 | CV | 100 | 0 | 0.27(0.04) |  | 0.08(0.02) | 0.11(0.03) |
|  |  | BIC | 100 | 0 | 0.42(0.04) |  | 0.04(0.01) | 0.07(0.02) |
|  |  | mBIC | 100 | 0 | 0.19(0.03) |  | 0.12(0.01) | 0.15(0.02) |
|  | 100 | CV | 100 | 0 | 0.27(0.02) |  | 0.34(0.03) | 0.29(0.03) |
|  |  | BIC | 100 | 0 | 0.73(0.04) |  | 0.10(0.02) | 0.16(0.03) |
|  |  | mBIC | 100 | 0 | 0.39(0.02) |  | 0.24(0.02) | 0.29(0.03) |
|  | 200 | CV | 100 | 0 | 0.27(0.02) |  | 0.67(0.03) | 0.38(0.02) |
|  |  | BIC | 100 | 0 | 0.88(0.03) |  | 0.26(0.02) | 0.39(0.02) |
|  |  | mBIC | 100 | 0 | 0.62(0.03) |  | 0.44(0.03) | 0.52(0.03) |
|  | 400 | CV | 100 | 0 | 0.25(0.02) |  | 1.00(0.01) | 0.55(0.03) |
|  |  | BIC | 100 | 0 | 0.80(0.03) |  | 0.71(0.02) | 0.75(0.03) |
|  |  | mBIC | 100 | 0 | 0.76(0.03) |  | 0.73(0.03) | 0.75(0.03) |
|  | 800 | CV | 99 | 1 | 0.25(0.02) |  | 0.99(0.01) | 0.60(0.02) |
|  |  | BIC | 100 | 0 | 0.76(0.03) |  | 0.95(0.02) | 0.85(0.02) |
|  |  | mBIC | 100 | 0 | 0.77(0.03) |  | 0.95(0.02) | 0.85(0.02) |

**Supplementary Table 5. Comparison of the performance of pLasso with different prior information: KEGG alone, Pathway Commons (PC) alone and a prior that combines KEGG and PC.** Network coverage summarizes the nodes and edges included in the inferred networks corresponding to patients with metastases (breast cancer) and patients with progressed disease (ovarian cancer). Sensitivity and specificity measure the prediction accuracy of the network hubs in the independent validation datasets.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  | KEGG  (3,045 genes,  18,294 edges) | PC  (10,518 genes, 83,297 edges) | Combined (11,211 genes, 97,218 edges) |
| Breast Cancer | Network Coverage | Nodes | 2,089 | 5,095 | 5,187 |
| Edges | 12,265 | 28,864 | 29,821 |
| Clinical significance of hubs | Significantly associated with prognosis (%) | 16 | 18 | 18 |
| Sensitivity | 0.76 | 0.78 | 0.81 |
| Specificity | 0.47 | 0.47 | 0.48 |
| Ovarian Cancer | Network Coverage | Nodes | 2,553 | 4,094 | 4,636 |
| Edges | 14,137 | 24,474 | 25,836 |
| Clinical significance of hubs | Significantly associated with prognosis (%) | 19 | 20 | 22 |
| Sensitivity | 0.63 | 0.65 | 0.69 |
| Specificity | 0.53 | 0.52 | 0.54 |

**REFERENCE**

Wang, H., Li, R. and Tsai, C.L. (2007) Tuning parameter selectors for the smoothly clipped absolute deviation method, *Biometrika*, **94**, 553-568.