Supplementary Material

Meta-Analysis Approaches to Combine Multiple Gene Set Enrichment Studies

Wentao Lu, Xinlei Wang*, Xiaowei Zhan and Adi Gazdar

In this supplementary material, we provide additional figures and tables mentioned in the main body of the paper.

1 About the size-adjusted KS statistic

For the size-adjusted KS statistic v_s proposed in this research, Figure 1 shows the empirical CDF (the black points) and the asymptotic CDF (the red curve) given by $F(z) = 1 - \exp(-2z^2)$, $z > 0$, where m denotes the number of genes in the gene set and n denotes the number out of the gene set. The empirical CDF is computed based on 1000 replicates in each (m, n) combination. Clearly, the curves are quite close, especially when $m \geq 30$. This explains why the distribution of v_s becomes (nearly) independent of the set size.

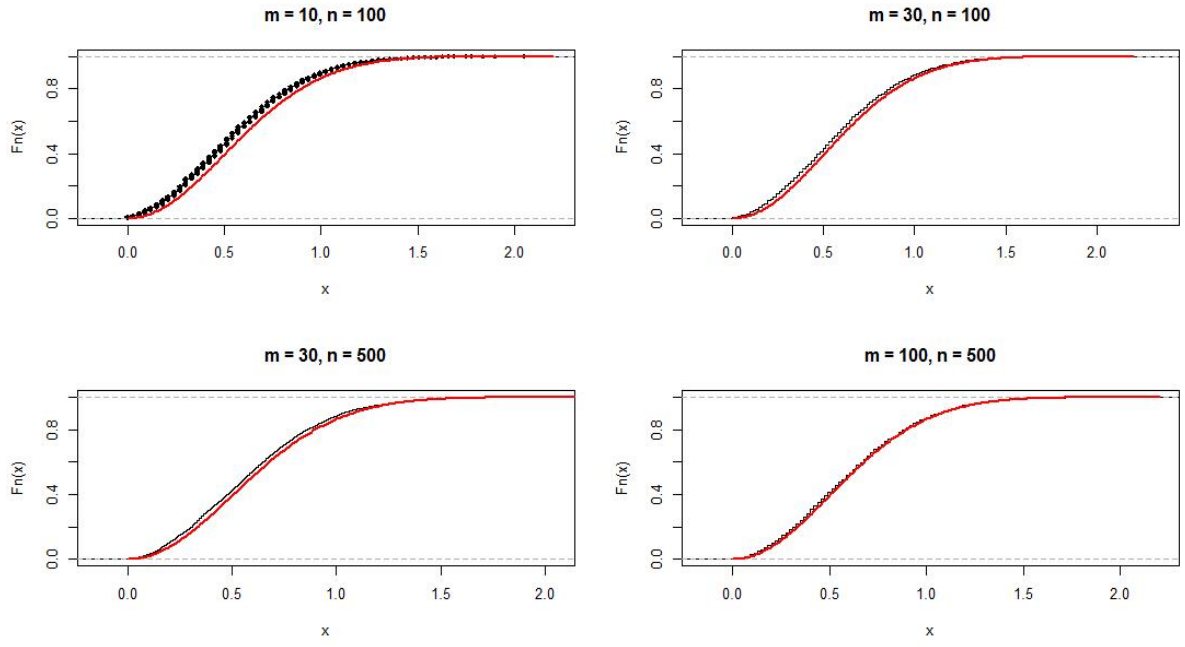

Figure 1: Comparison of the empirical CDF of v_s with the asymptotic CDF.

2 Additional simulation results for binary phenotypes

2.1 Comparing test sizes

As mentioned in the main body of the paper, we computed type I errors (i.e. test sizes) of the different methods by setting the enrichment signal $\omega = 0.2$ (i.e., H_0 : no enrichment in the gene set holds) and then compared them with the nominal significance level 0.05 under various settings.

The following three tables present the simulated test sizes from 1000 replicate datasets for a fixed sampling rate $\lambda = 0.5$ and a fixed value of the between-study heterogeneity $\tau = 1$. Apparently, the sizes of all the three iGSEA methods are consistently smaller than 0.05, meaning that they are conservative in rejecting the null. Among the three MAPE methods, the size of MAPE-G is close to 0.05, but sometimes below 0.05; the size of MAPE-P is usually larger than 0.05, especially for large γ, meaning that it is aggressive in rejecting the null; and the size of MAPE-I appears to be somewhere between MAPE-G and MAPE-P, and so sometimes above 0.05 and sometimes below 0.05. The patterns are quite similar in settings with other λ and τ values (results not reported for brevity).

Parameters	iGSEA-FE	iGSEA-RE	iGSEA-AT	MAPE-G	MAPE-P	MAPE-I
$\gamma = 0$	0.05	0.04	0.04	0.05	0.06	0.06
$\gamma = 0.2$	0.04	0.05	0.04	0.04	0.04	0.04
$\gamma = 0.4$	0.05	0.04	0.04	0.05	0.07	0.06
$\gamma = 0.6$	0.04	0.04	0.04	0.04	0.06	0.05
$\gamma = 0.8$	0.04	0.03	0.03	0.04	0.06	0.05
$\gamma = 1$	0.03	0.03	0.02	0.05	0.07	0.05

Table 1: Type I errors of each method when $\lambda = 0.5$, $\mu = 0.3$ and $\tau = 1$.

Parameters	iGSEA-FE	iGSEA-RE	iGSEA-AT	MAPE-G	MAPE-P	MAPE-I
$\gamma = 0$	0.04	0.03	0.04	0.05	0.05	0.05
$\gamma = 0.2$	0.03	0.04	0.03	0.05	0.04	0.04
$\gamma = 0.4$	0.04	0.04	0.04	0.05	0.06	0.06
$\gamma = 0.6$	0.03	0.03	0.03	0.04	0.06	0.05
$\gamma = 0.8$	0.03	0.03	0.03	0.05	0.07	0.04
$\gamma = 1$	0.02	0.03	0.02	0.04	0.07	0.06

Table 2: Type I errors of each method when $\lambda = 0.5$, $\mu = 0.45$ and $\tau = 1$.

Parameters	iGSEA-FE	iGSEA-RE	iGSEA-AT	MAPE-G	MAPE-P	MAPE-I
$\gamma = 0$	0.04	0.04	0.03	0.05	0.05	0.05
$\gamma = 0.2$	0.03	0.04	0.03	0.04	0.04	0.04
$\gamma = 0.4$	0.04	0.04	0.03	0.05	0.06	0.05
$\gamma = 0.6$	0.04	0.03	0.03	0.04	0.05	0.05
$\gamma = 0.8$	0.03	0.03	0.03	0.04	0.07	0.04
$\gamma = 1$	0.04	0.03	0.02	0.04	0.1	0.06

Table 3: Type I errors of each method when $\lambda = 0.5$, $\mu = 0.6$ and $\tau = 1$.

2.2 Comparing test power

We simulated the power of the methods for all of the parameter combinations (i.e., $\alpha \in \{0.3, 0.4, 0.5\}$, $\mu \in \{0.3, 0.45, 0.6\}$, $\lambda \in \{0.5, 0.6, 0.7, 0.8, 0.9, 1\}$, $\gamma \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$, $\tau \in \{0.5^2, 1\}$). When α and μ get larger, iGSEA-FE, iGSEA-RE and iGSEA-AT all have (nearly) 100% power, so we only report detailed results for situations where these methods differ in power. The figures show the power values for the three proposed methods and maxMAPE (the maximum power of the three MAPE methods).

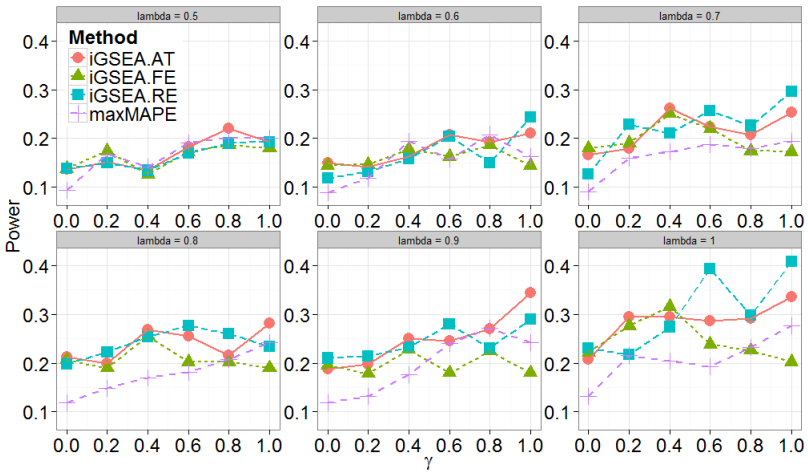

Figure 2: Power comparison for the settings with $\omega = 0.3$, $\mu = 0.3$ and $\tau = 0.5^2$.

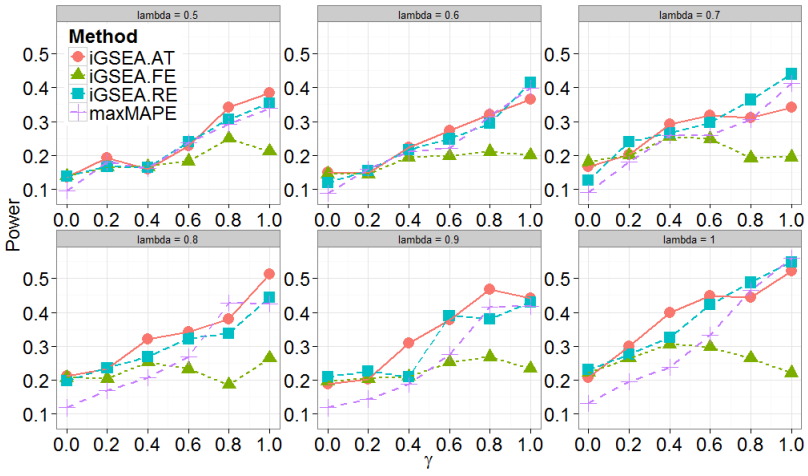

Figure 3: Power comparison for the settings with $\omega = 0.3$, $\mu = 0.3$ and $\tau = 1$.

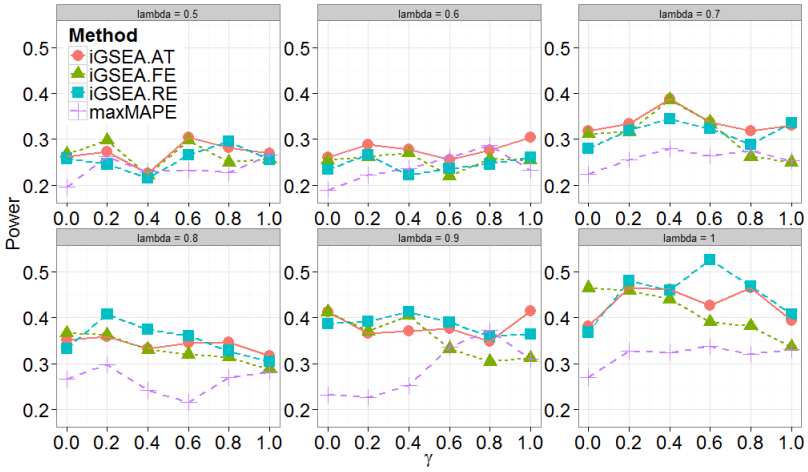

Figure 4: Power comparison for the settings with $\omega = 0.3$, $\mu = 0.45$ and $\tau = 0.5^2$.

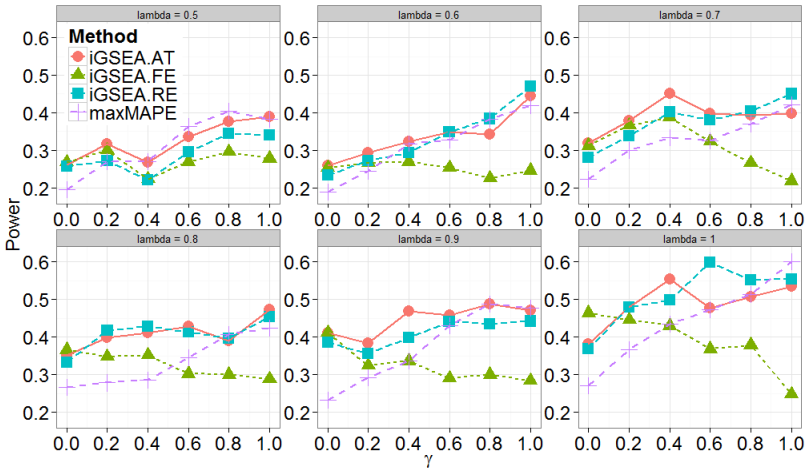

Figure 5: Power comparison for the settings with $\omega = 0.3$, $\mu = 0.45$ and $\tau = 1$.

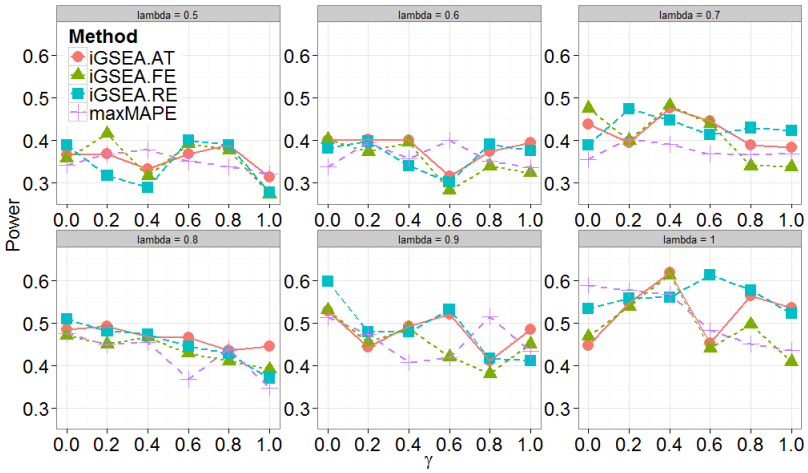

Figure 6: Power comparison for the settings with $\omega = 0.3$, $\mu = 0.6$ and $\tau = 0.5^2$.

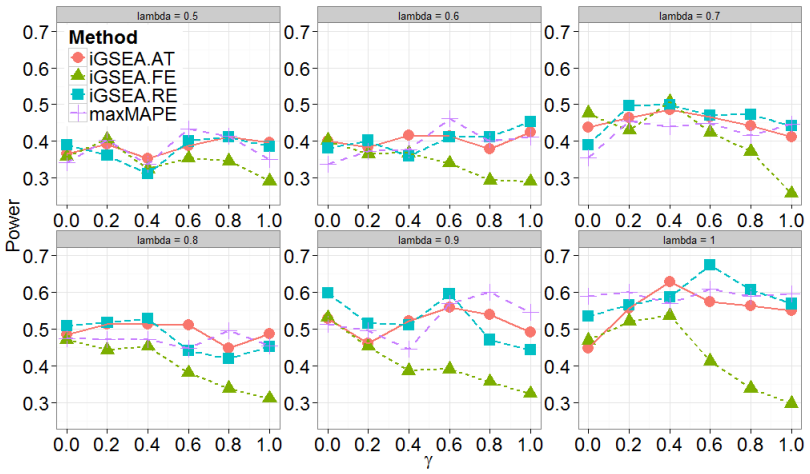

Figure 7: Power comparison for the settings with $\omega = 0.3$, $\mu = 0.6$ and $\tau = 1$.

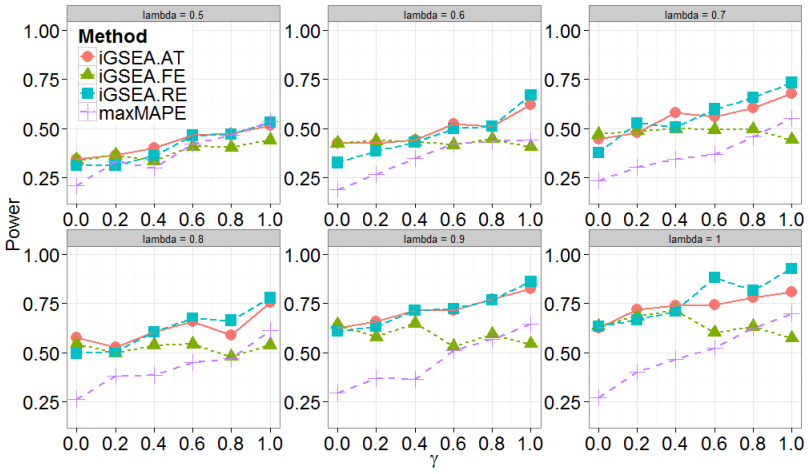

Figure 8: Power comparison for the settings with $\omega = 0.4$, $\mu = 0.3$ and $\tau = 0.5^2$.

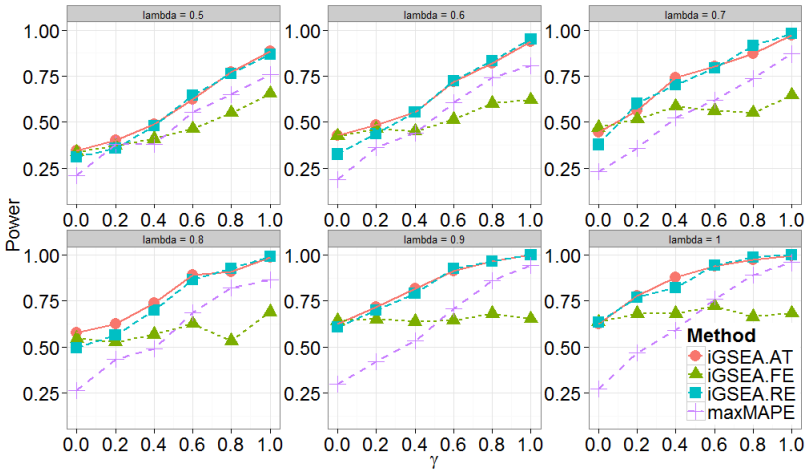

Figure 9: Power comparison for the settings with $\omega = 0.4$, $\mu = 0.3$ and $\tau = 1$.

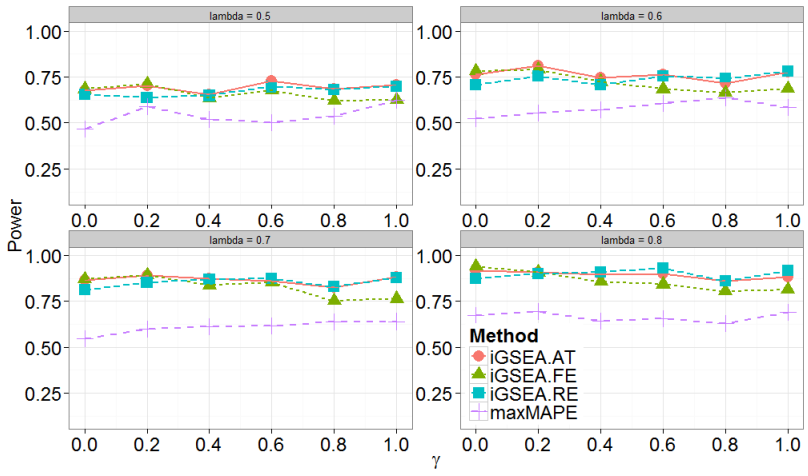

Figure 10: Power comparison for the settings with $\omega = 0.4$, $\mu = 0.45$ and $\tau = 0.5^2$.

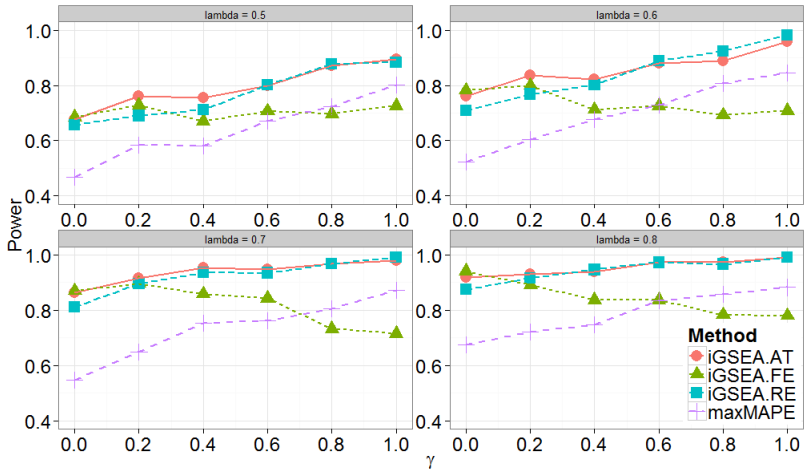

Figure 11: Power comparison for the settings with $\omega = 0.4$, $\mu = 0.45$ and $\tau = 1$.

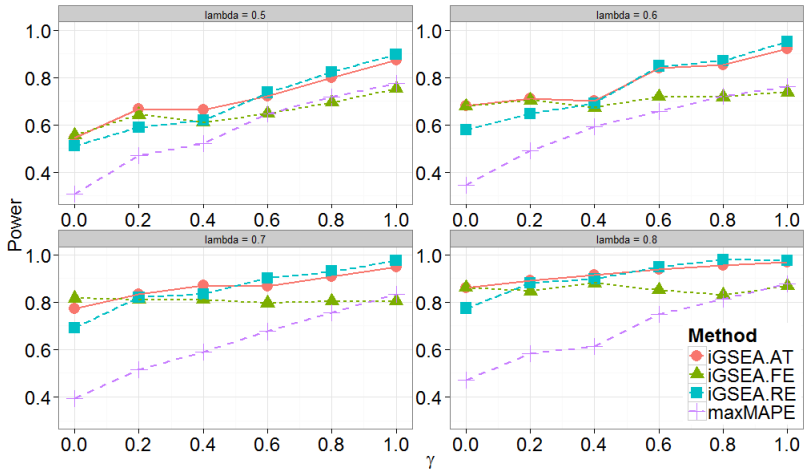

Figure 12: Power comparison for the settings with $\omega = 0.5$, $\mu = 0.3$ and $\tau = 0.5^2$.

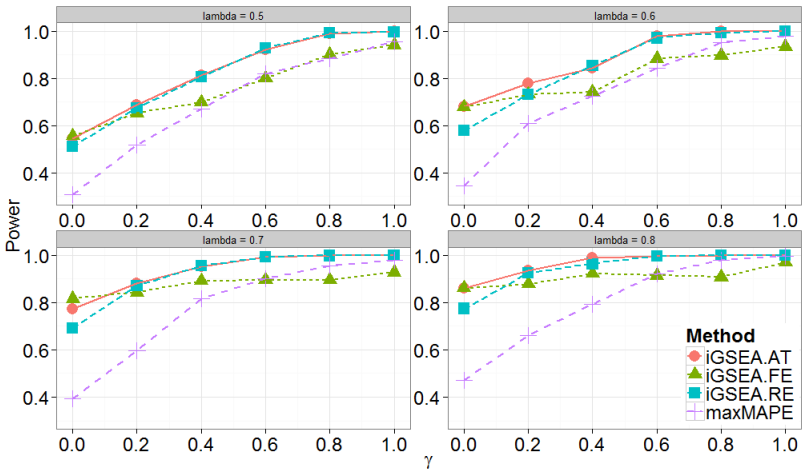

Figure 13: Power comparison for the settings with $\omega = 0.5$, $\mu = 0.3$ and $\tau = 1$.

2.3 Detail in generating multiple gene sets

To compare the sensitivity and specificity of the methods in Section 6.1, we generated 200 gene sets. Table 4 illustrates how different types of gene sets were constructed. Out of the 1000 genes in the genome, the first 100 genes are UR genes, the last 100 genes are DR genes, and the rest are EE genes. UR, DR and EE genes in each gene set are randomly chosen from the corresponding populations. Table 5 provides the detailed design about how to generate gene expression levels.

Set Type	UR genes	DR genes	EE genes
Enriched by UR genes	20%	10%	70%
Enriched by DR genes	10%	20%	70%
Non-enriched	10%	10%	80%

Table 4: Design detail in constructing gene sets; 30% of the gene sets are enriched by UR genes, another 30% are enriched by DR genes, and the remaining 40% are non-enriched.

Gene ID	Gene expression in each study	Effect size β of a RE gene	Effect size β of an FE gene
1-100	$N(\beta, 1)$	$N(0.45, 0.5^2)$	0.45
101-900	$N(0, 1)$	N/A	N/A
901-1000	$N(\beta, 1)$	$N(-0.45, 0.5^2)$	-0.45

Table 5: Design detail in generating expression levels for different types of genes. Genes 1-100 are UR genes, Genes 101-900 are EE genes, and Genes 901- 1000 are DR genes.

2.4 ROC curves

Figure 14 presents an example of ROC curves in each γ setting using a randomly generated dataset.

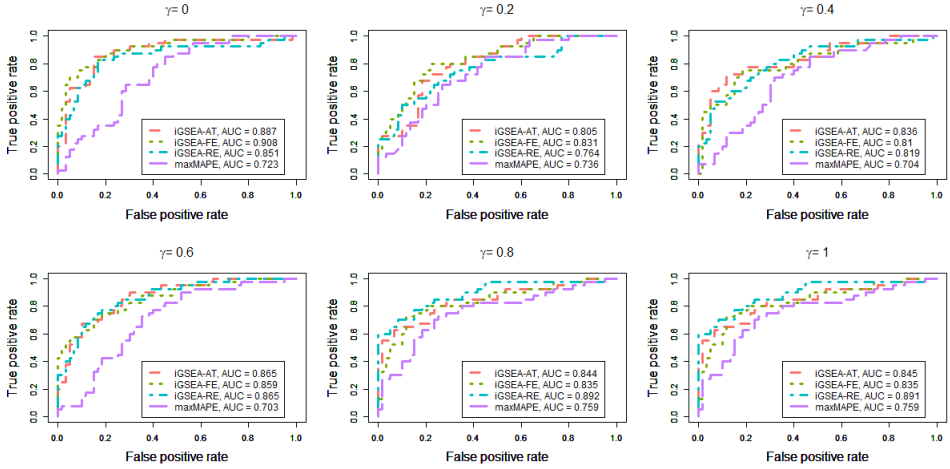

Figure 14: ROC curves for detecting multiple enriched gene sets using iGSEA-FE, iGSEA-RE, iGSEA-AT, and maxMAPE. The purple curve on the plots actually represents the performance of MAPE-P because it is the best of its kind.

2.5 Comparing FDRs

In Table 6, we report empirical FDRs for the six methods under the settings described in the

Parameters	iGSEA-FE	iGSEA-RE	iGSEA-AT	MAPEG	MAPEP	MAPEI
$\gamma = 0$	0.05	0.07	0.06	0.25	0.08	0.10
$\gamma = 0.2$	0.06	0.06	0.06	0.21	0.09	0.10
$\gamma = 0.4$	0.06	0.06	0.05	0.20	0.09	0.09
$\gamma = 0.6$	0.07	0.06	0.06	0.18	0.08	0.09
$\gamma = 0.8$	0.07	0.07	0.06	0.19	0.09	0.09
$\gamma = 1$	0.08	0.07	0.07	0.21	0.08	0.09

Table 6: Empirical FDRs of each method in the multiple-gene-set simulation for binary phenotypes.

multiple-gene-set simulation of Section 5.1, and compare their values with the targeted level $\delta = 0.05$. Note that when calculating the Q-values, our iGSEA methods obtain $\hat{\pi}_0$ using the R package "*qvalue*" (Bass et al., 2015) while the MAPE methods always set $\hat{\pi}_0 = 1$ (Shen & Tseng 2010), as mentioned in Section 4. We find that the three iGSEA methods have FDRs pretty close to 0.05, although they are slightly inflated. By contrast, the inflation from the three MAPE methods is

more severe, especially for MAPE-G. Also, we observe that MAPE-I falls between MAPE-G and MAPE-P, but it is closer to the better one MAPE-P.

3 About data application

Data Set Name	Type	Number of Controls	Number of Cases
GSE14814 (Zhu et al. 2010)	Microarray	7	21
CL (Shedden et al. 2008)	Microarray	17	65
Moff (Shedden et al. 2008)	Microarray	27	52
NCI_U133A (Shedden et al. 2008)	Microarray	18	86
GSE37764 (Kim et al. 2013)	RNA-seq	6	6

Table 7: Lung adenocarcinoma datasets involved in data analysis

We used five real datasets from microarray and NGS experiments to evaluate the methods. The four microarray datasets were pre-processed following the steps in Chen et al. (2013). The RNA-seq dataset was downloaded from `http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE37765`. Table 7 provides the detail of each dataset, including the name, source, type of experiment and sample size.

KEGG Pathways	iGSEA			MAPE-G	MAPE-P	MAPE-I
	FE	RE	AT			
ABC_Transporters	0.065	**0.019**	0.060	0.415	0.464	0.515
Alpha_Linolenic_Acid_Metabolism	0.066	0.053	**0.041**	0.419	0.481	0.514
Apoptosis	**0.014**	**0.009**	**0.011**	0.349	**0**	**0.003**
Ascorbate_And_Aldarate_Metabolism	**0.014**	**0.045**	**0.011**	0.402	0.294	0.369
Basal_Cell_Carcinoma	0.316	0.109	0.301	0.361	**0.039**	0.079
Ether_Lipid_Metabolism	0.063	**0.036**	**0.041**	0.420	0.169	0.258
Glycosaminoglycan_Biosynthesis_Keratan_Sulfate	0.311	0.177	**0.043**	0.420	0.486	0.520
Glycosaminoglycan_Degradation	0.065	0.074	**0.033**	0.413	0.135	0.217
Glyoxylate_And_Dicarboxylate_Metabolism	**0**	**0**	**0**	0.155	0.106	0.078
Huntingtons_Disease	0.156	**0.024**	**0.045**	0.411	0.516	0.520
Hypertrophic_Cardiomyopathy_HCM	0.063	**0.023**	**0.025**	0.276	0.055	0.088
Long_Term_Depression	**0**	**0**	**0**	0.447	0.533	0.331
Lysine_Degradation	**0.041**	**0.034**	**0.023**	0.333	0.102	0.163
Oxidative_Phosphorylation	0.166	**0.045**	0.115	0.413	0.529	0.526
Primary_Immunodeficiency	0.197	0.206	**0.038**	0.470	0.536	0.531
Type_I_Diabetes_Mellitus	**0.021**	0.083	0.065	0.420	0.129	0.209

Table 8: The estimated Q-values of identified KEGG pathways

We tested KEGG pathways. Gene sets with Q-values <5% were identified as enriched in the analysis. Table 8 provides the estimated Q-values of all the KEGG pathways identified by at least one of the six methods, with those <5% bolded. The pathways that were only identified by iGSEA-AT are marked in red.

References

Bass, J., Dabney, A., & Robinson, D. (2015). qvalue: Q-value estimation for false discovery rate control. r package version 2.8.0.

Chen, M., Zang, M., Wang, X., & Xiao, G. (2013). A powerful bayesian meta-analysis method to integrate multiple gene set enrichment studies. *Bioinformatics*, 29, 862–869.

Kim, S., Jung, Y., Park, J., Cho, S., Seo, C., Kim, J., Kim, P., Park, J., Seo, J., Kim, J., Park, S., Jang, I., Kim, N., Yang, J. O., Lee, B., Rho, K., Jung, Y., Keum, J., Lee, J., Han, J., Kang, S., Bae, S., Choi, S.-J., Kim, S., Lee, J.-E., Kim, W., Kim, J., & Lee, S. (2013). A high-dimensional, deep-sequencing study of lung adenocarcinoma in female never-smokers. *PLoS ONE*, 8, e55596.

Shedden, K., Taylor, J. M. G., Enkemann, S. A., Tsao, M.-S., Yeatman, T. J., Gerald, W. L., Eschrich, S., Jurisica, I., Giordano, T. J., Misek, D. E., Chang, A. C., Zhu, C. Q., Strumpf, D., Hanash, S., Shepherd, F. A., Ding, K., Seymour, L., Naoki, K., Pennell, N., Weir, B., Verhaak, R., Ladd-Acosta, C., Golub, T., Gruidl, M., Sharma, A., Szoke, J., Zakowski, M., Rusch, V., Kris, M., Viale, A., Motoi, N., Travis, W., Conley, B., Seshan, V. E., Meyerson, M., Kuick, R., Dobbin, K. K., Lively, T., Jacobson, J. W., & Beer, D. G. (2008). Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nature Medicine*, 14, 822–827.

Shen, K. & Tseng, G. C. (2010). Meta-analysis for pathway enrichment analysis when combining multiple genomic studies. *Bioinformatics*, 26, 1316–1323.

Zhu, C. Q., Ding, K., Strumpf, D., Weir, B. A., Meyerson, M., Pennell, N., Thomas, R. K., Naoki, K., Ladd-Acosta, C., Liu, N., Pintilie, M., Der, S., Seymour, L., Jurisica, I., Shepherd, F. A., ,

& Tsao, M. S. (2010). Prognostic and predictive gene signature for adjuvant chemotherapy in resected non-small-cell lung cancer. *Journal of Clinical Oncology*, 28, 4417–4424.