Testing OutPredict using Arabidopsis and Maize NxT data

Ji Huang

2020-08-14

Dataset

- Arabidopsis root Time-course experiment.

- Expression matrix for network inference:
  - Root time-course – Matt's NC paper

- Network validation:
  - 29 Arabidopsis TARGET DE genes – Matt's NC paper

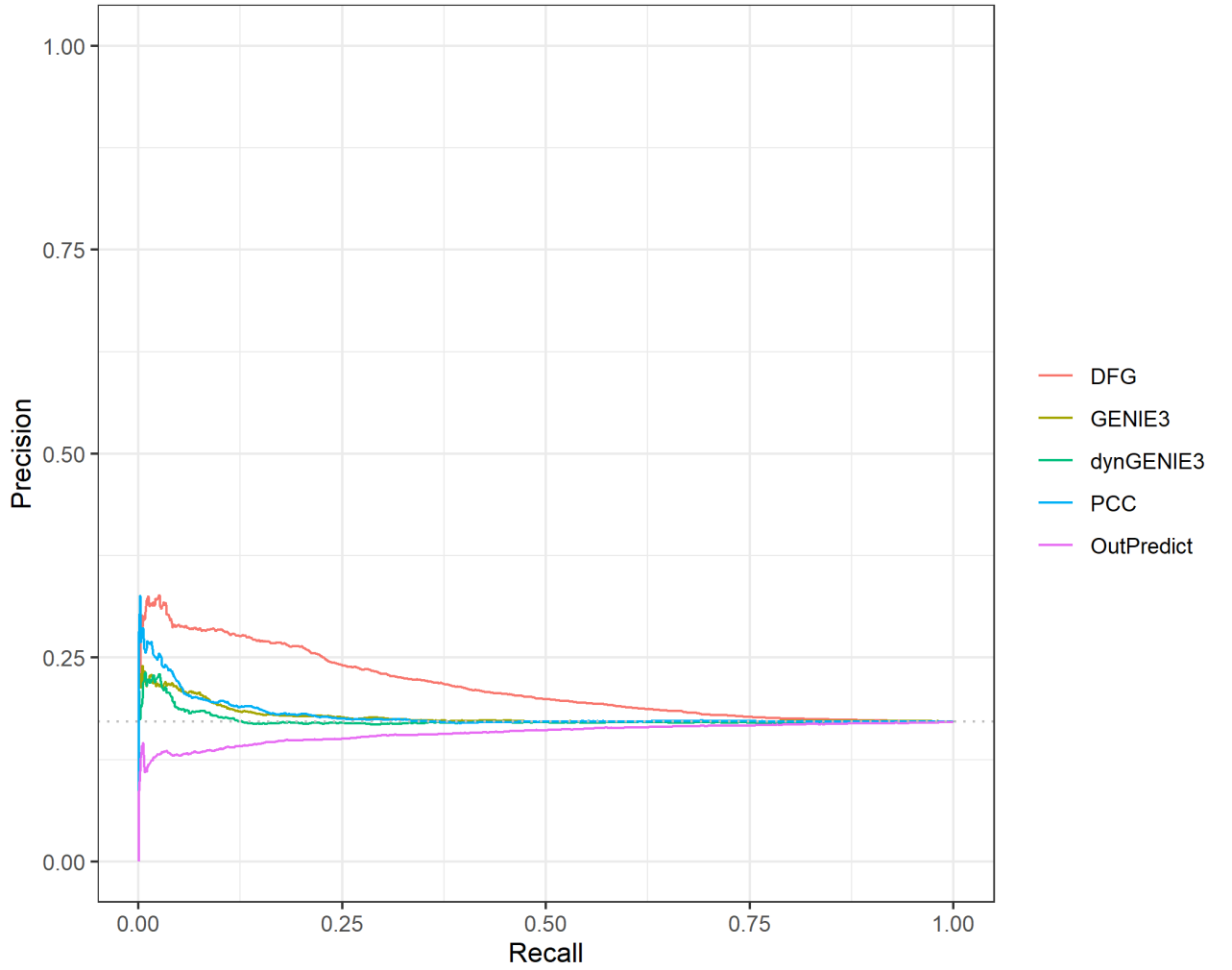
- Random network:
  - Randomly assign edge values and calculate AUPR. Repeat 10,000 times for p-values.

Three tests

- Test 1: Use Nitrogen-treated expression only.
- Matrix contains:
  - 10 time points (0min – 120min)
  - 1 condition (N-treated)
  - 3 replicates per condition per time points
  - In total 30 columns.

Method	AUPR	P-value
DFG	0.21402	< 1E-5
GENIE3	0.177556	5E-4
dynGENIE3	0.173174	0.1642
PCC	0.179174	< 1E-5
OutPredict	0.157609	1

```
ath_root_Nonly_output_RF_Ksqrt_ntrees1000_datatypeTS_LOdataTS
MSE Pen. Value Naive Time-Series Test set: 0.21071197750546147
MSE Time-Series Test Set of the best model is: 0.4185381328207922
```

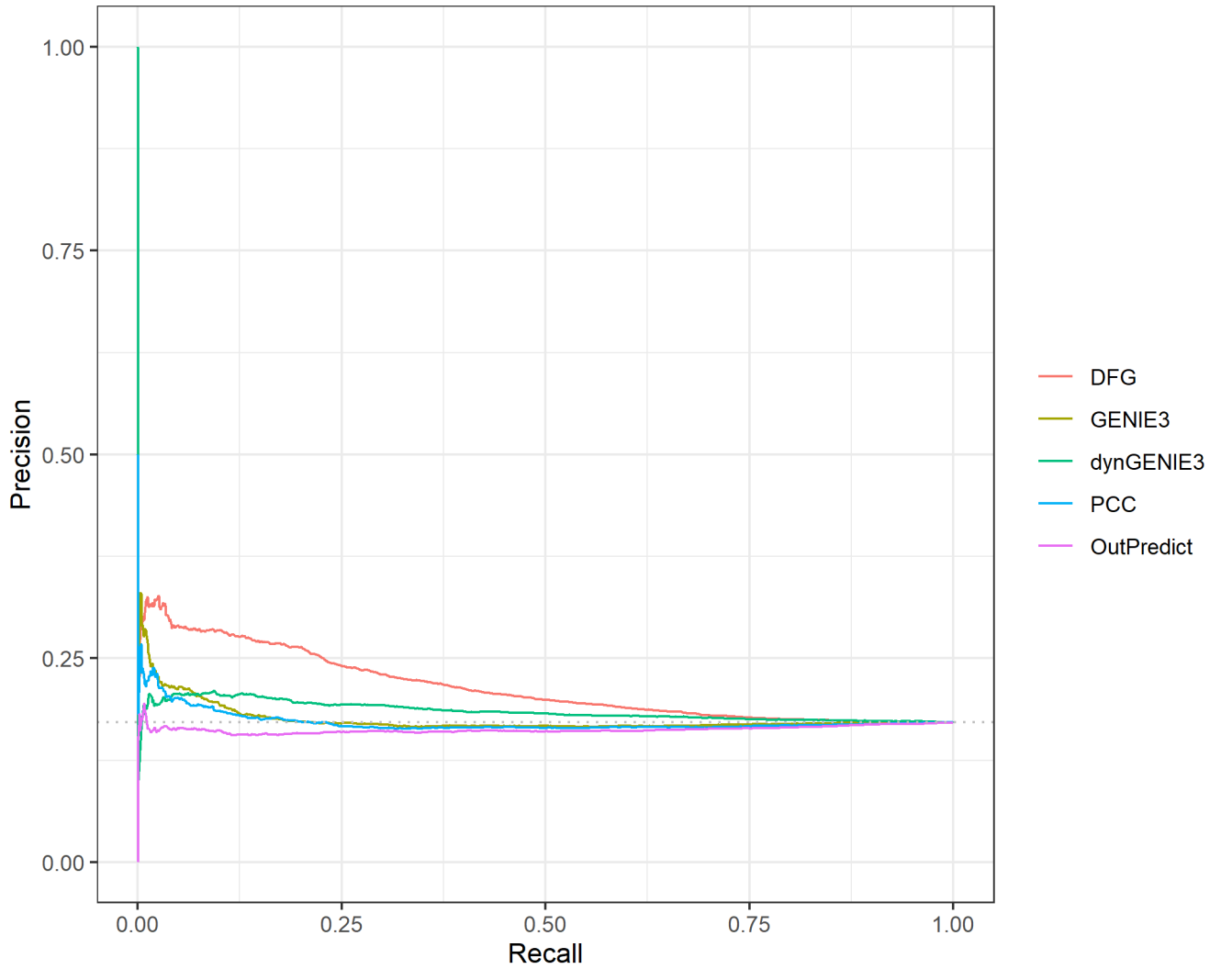


Three tests

- Test 2: Use Nitrogen-treated AND control expression.
- Matrix contains:
  - 10 time points (0min – 120min)
  - 2 condition (N-treated and KCl control)
  - 3 replicates per condition per time points
  - In total 60 columns.

Method	AUPR	p-value
DFG	0.2140199	< 1E-5
GENIE3	0.17553211	8.6E-3
dynGENIE3	0.18506776	< 1E-5
PCC	0.1718386	0.41
OutPredict	0.16269649	1

```
ath_root_all_output_RF_Ksqrt_ntrees1000_datatypeTS_LOdataTS
MSE Pen. Value Naive Time-Series Test set: 0.21022595619850626
MSE Time-Series Test Set of the best model is: 0.3735043465031758
```



Three tests

- Test 3: Use Nitrogen-treated AND control expression, plus Matt 33 TF expression as Steady-state.

- The Steady-state is the same data where the validation sets are from. I chose it on purpose to make sure AUPR is better than random. Also, I'm expecting (1) the best model MSE from OutPredict will be lower than Pen.Value. (2) the OutPredict AUPR will be higher than GENIE3 and dynGENIE3.

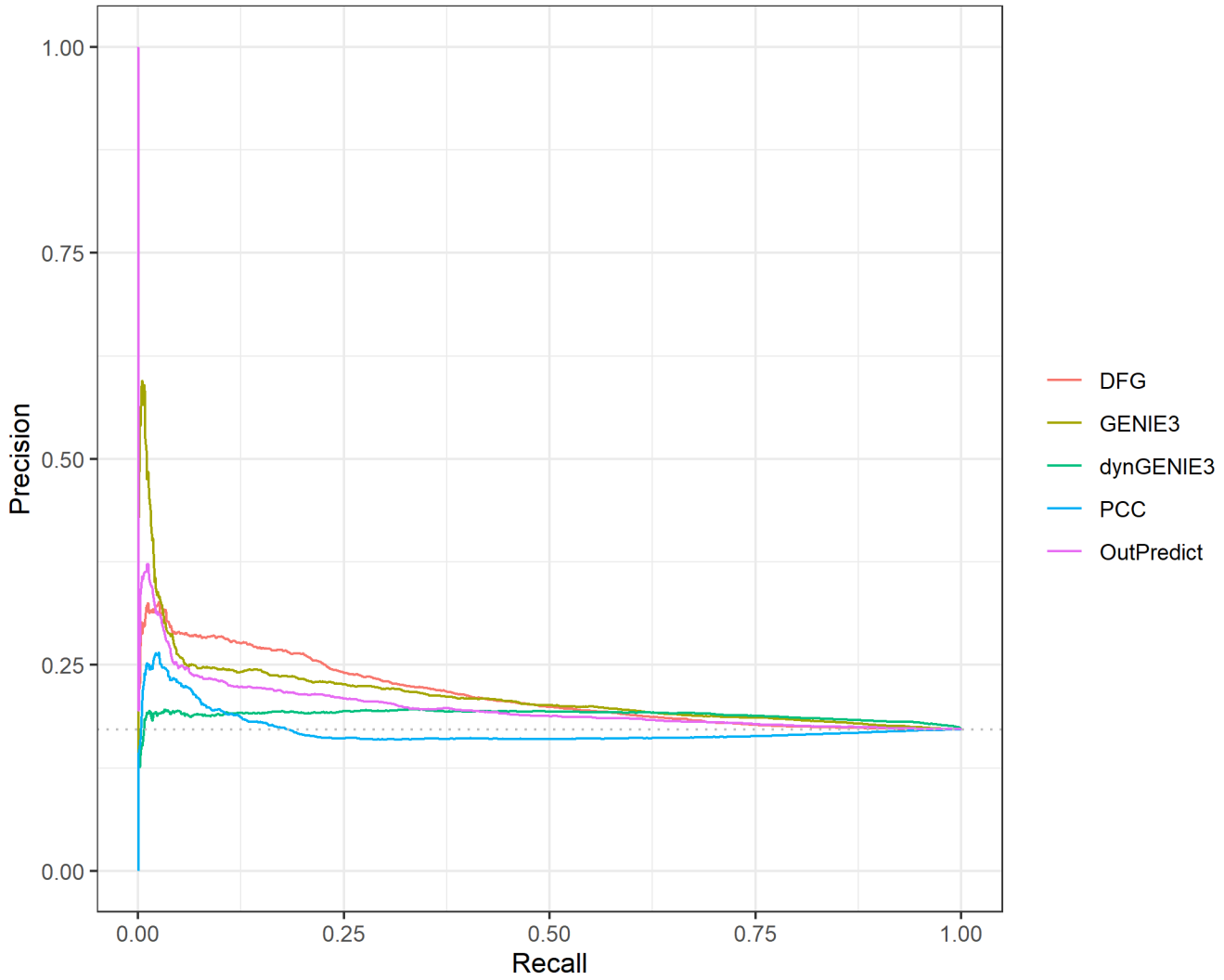
- Matrix contains:
  - 10 time points (0min – 120min)
  - 2 condition (N-treated and KCl control)
  - 3 replicates per condition per time points
  - 115 libraries from Matt's 33 TF expression
  - In total 172 columns.

Three tests

- Test 3: Use Nitrogen-treated AND control expression, plus Matt 33 TF expression as Steady-state.

Method	AUPR	p-value
DFG	0.214019896	< 1E-5
GENIE3	0.213003612	< 1E-5
dynGENIE3	0.18981872	< 1E-5
PCC	0.170787807	0.64
OutPredict	0.198647435	< 1E-5

```
Write output to file _summary_results.txt inside the output folder
ath_root_withSS_output_RF_Ksqrt_ntrees1000_datatypeTS-SS_LOdataTS
MSE Pen. Value Naive Time-Series Test set: 0.20195833162062307
MSE Time-Series Test Set of the best model is: 0.39386162426903776
```



Maize Dataset

- Maize shoot Time-course experiment.

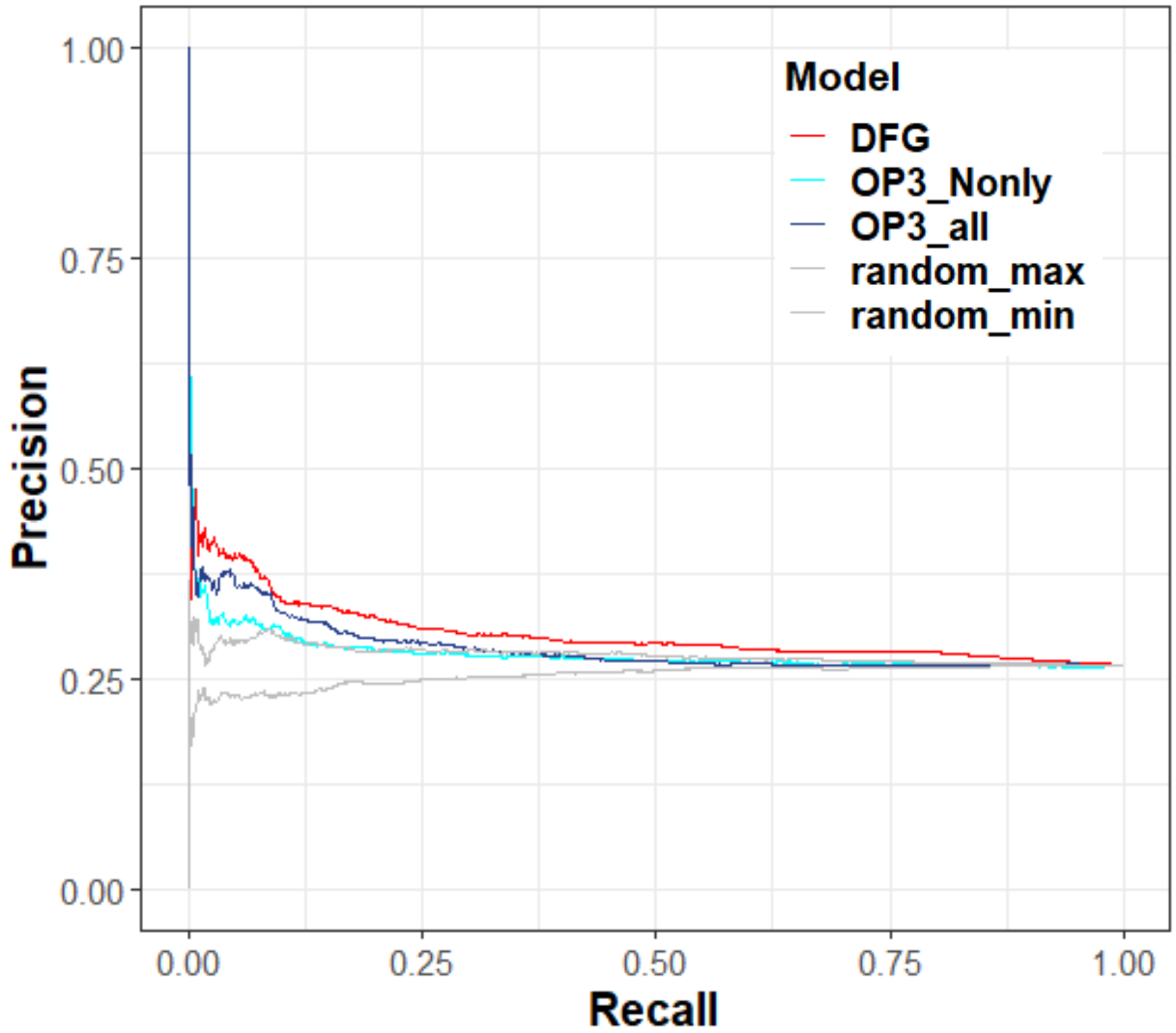
- Expression matrix for network inference:
  - Shoot time-course

- Network validation:
  - 7 Maize TARGET DE genes.
  - Tried different FDR cutoff for DE gene calling.

- Random network:
  - Repeat 1,000 times for p-values.

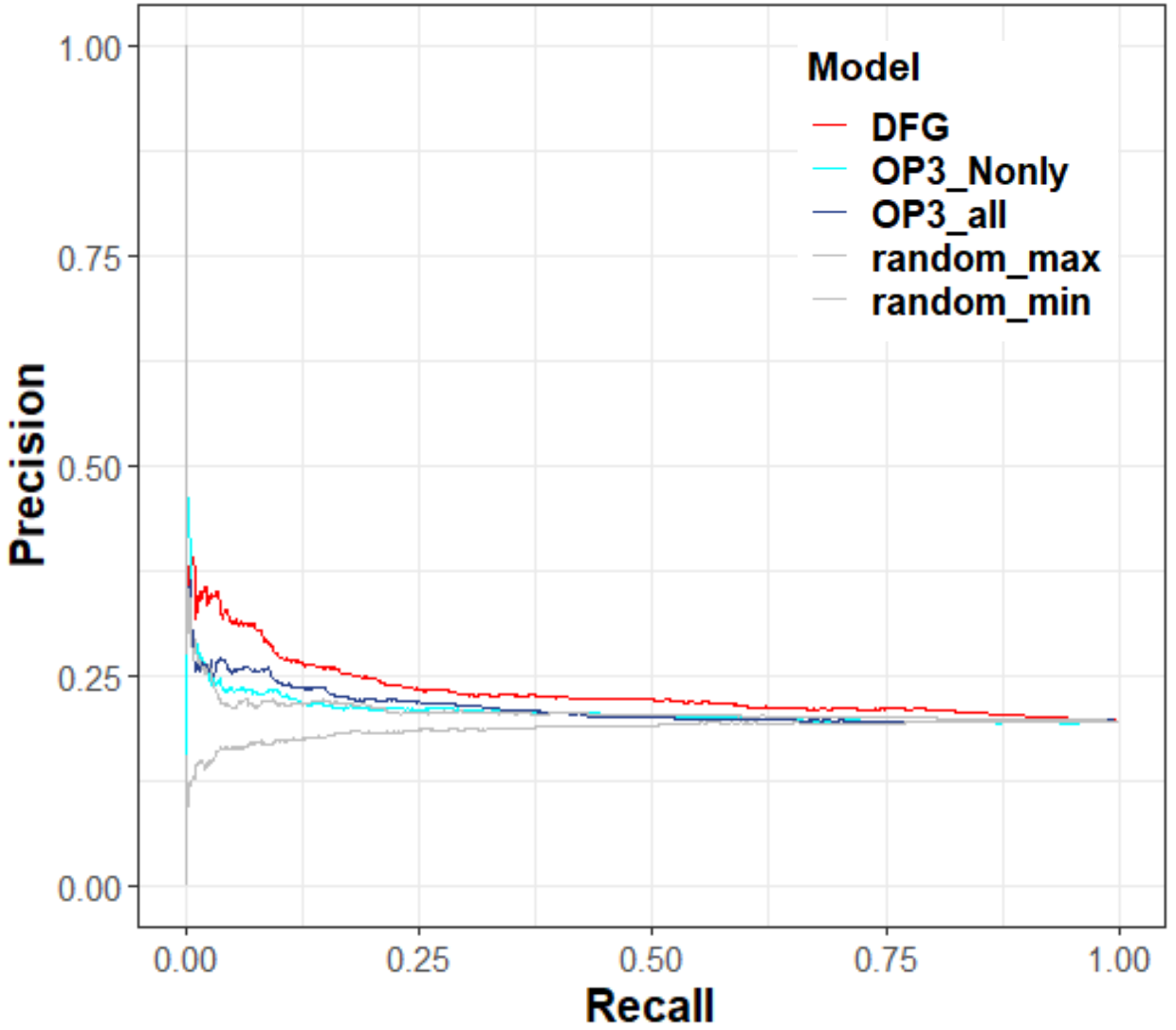
DE gene FDR cutoff 0.05

Method	AUPR	P-value
DFG	0.303	< 1E-3
OutPredict_Nonly	0.280	< 1E-3
OutPredict_all	0.286	< 1E-3
Random_max	0.279	
Random_min	0.254	



DE gene FDR cutoff 0.01

Method	AUPR	P-value
DFG	0.231	< 1E-3
OutPredict_Nonly	0.207	< 1E-3
OutPredict_all	0.211	< 1E-3
Random_max	0.207	
Random_min	0.187	



DEgene FDR cutoff 0.001

Method	AUPR	P-value
DFG	0.169	< 1E-3
OutPredict_Nonly	0.146	0.015
OutPredict_all	0.152	< 1E-3
Random_max	0.149	
Random_min	0.132