Additional file 2 – Supplementary Figures

Supplementary Figure 1 – DT degree plot.
The plot highlights the line corresponding to the 95th percentile, showing that most DTs have lower degree.

Supplementary Figure 2 – Comparative histogram of DTs and non-DTs over a degree rank.

Supplementary Figure 3 - Over-representation of drug targets over a BC ranking of nodes in the full spoke PIN, Binary (B), N-ary (N), Spoke-represented (S), BioGRID and Rual+Stelzl data sets.
Proteins were grouped into bins according to their BC. The width of each bin represents the number of proteins in that bin while the height (-log of the p-value of the hypergeometric test) represents how over-represented drug targets are in that bin. Each bin contains at least 200 proteins. Over-represented bins (p-value < 0.05) are highlighted in red. Drug targets are over-represented in high-degree bins and some middle-degree bins for the full PIN (a) and the B data set (b), while this trend is largely lost in the N (c), S (d), BioGRID (e) and the Rual-Stelzl (f) data sets.

Supplementary Figure 4 – ROC curve for BC as DT predictor (full PIN and reliable subsets).
Plot of False Positive Rate versus True Positive Rate for a BC rank of the full PIN and five subsets considered as containing higher-confidence interactions: non-predicted interactions include all interactions except those coming from orthologous transfer; LTP includes interactions with an lpr score < 22; MI-IntAct includes interactions with MI-IntAct scores > 0.6; MI-psicquic includes interactions with MI-psicquic scores > 0.7; and B includes the true binary interactions (i.e., potential spoke-represented n-ary data is removed).

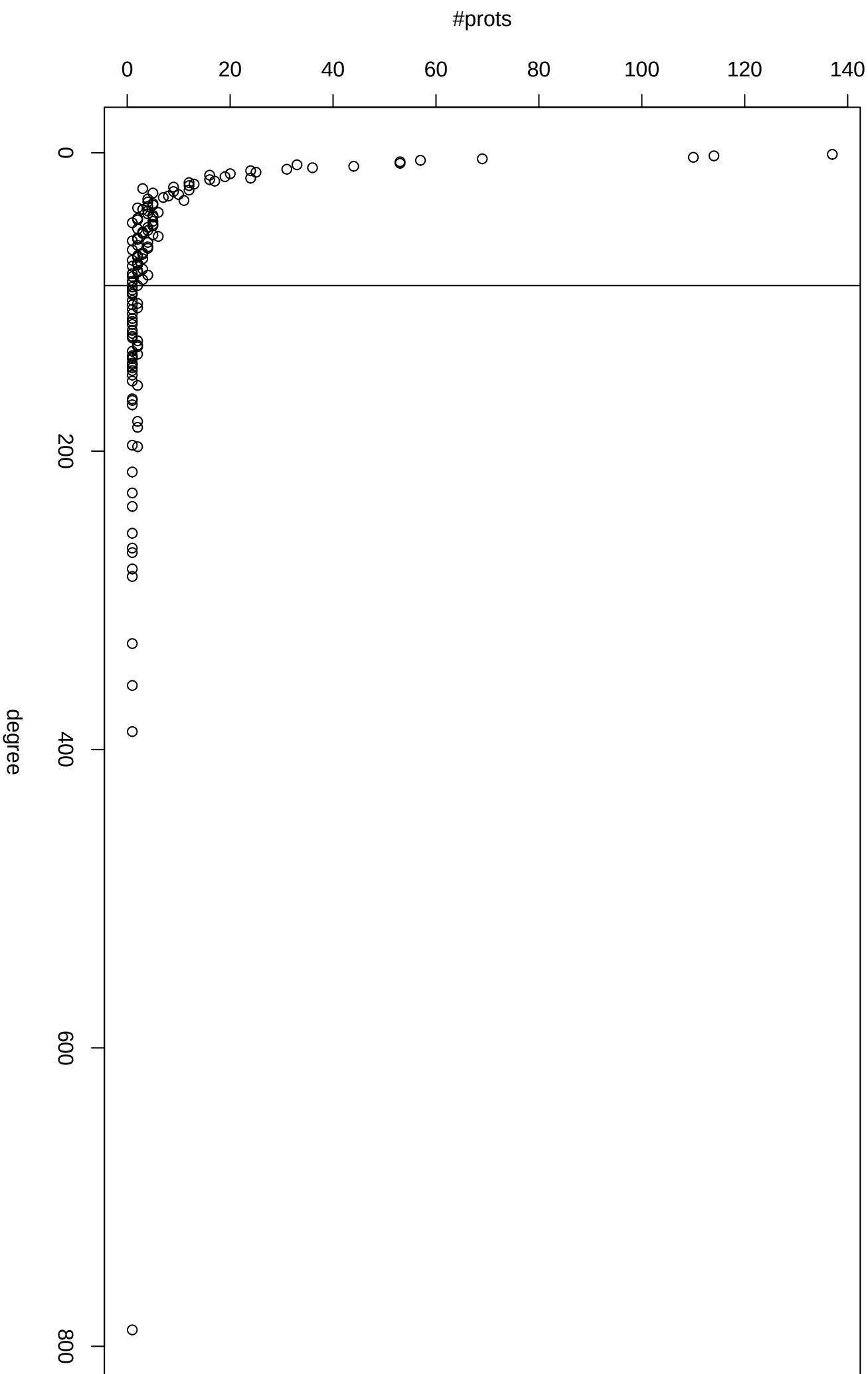
Supplementary Figure 5 - Over-representation of drug targets along a degree rank and a BC rank for the full PIN with a matrix representation of protein complexes.
Proteins were grouped into bins according to their degree (a) and their BC (b), for a matrix model of the full PIN. The width of each bin represents the number of proteins in that bin while the height (-log of the p-value of the hypergeometric test) represents how over-represented drug targets are in that bin. Each bin contains at least 200 proteins. Over-represented bins (p-value < 0.05) are highlighted in red. Drug targets are over-represented in high-degree bins and some middle-degree bins, just as in the spoke representation.

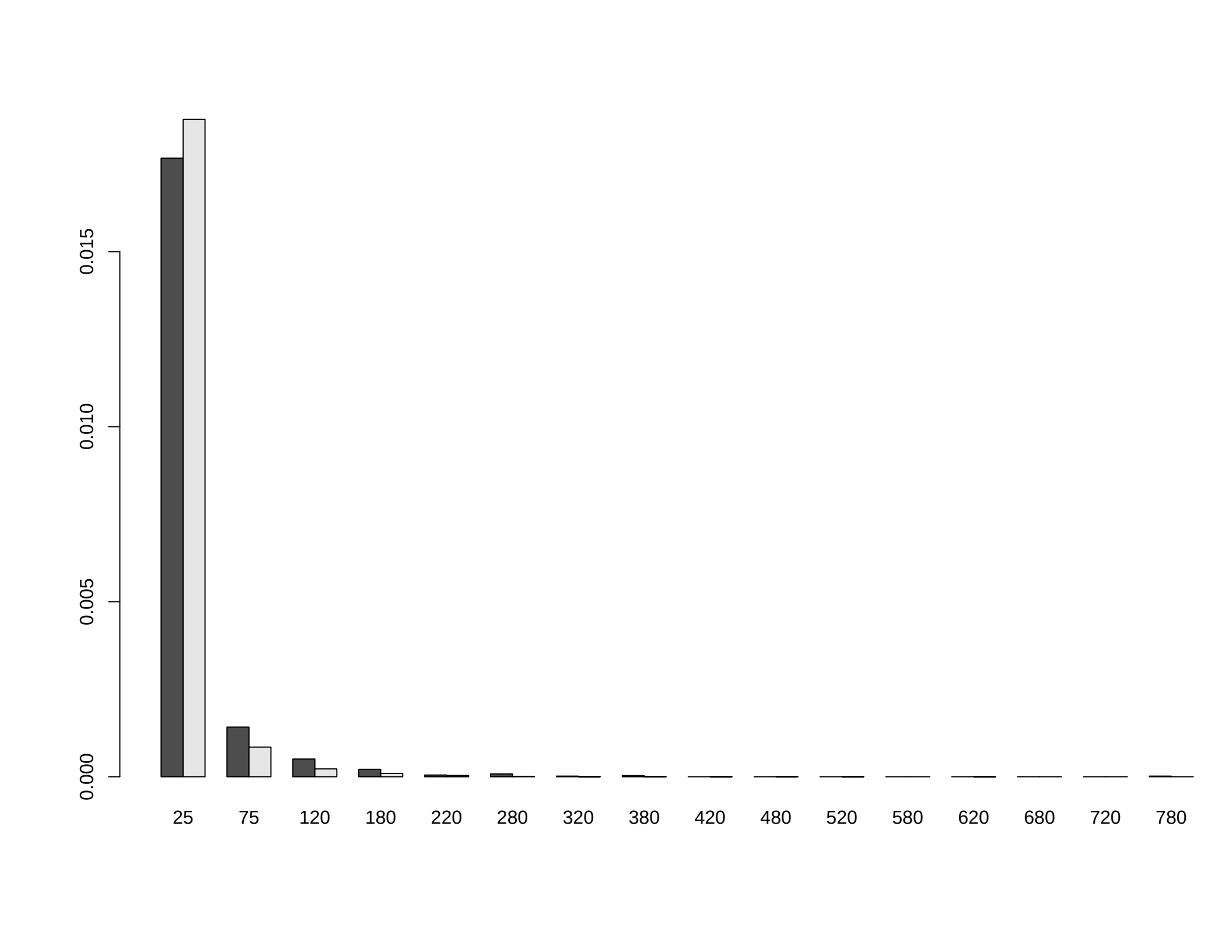
Supplementary Figure 6 – ROC curve for degree (spoke and matrix), and BC (matrix).
Plot of False Positive Rate versus True Positive Rate for a degree rank of the full PIN following a spoke and a matrix representation of protein complexes, and a BC rank of the full PIN following a matrix representation of protein complexes.

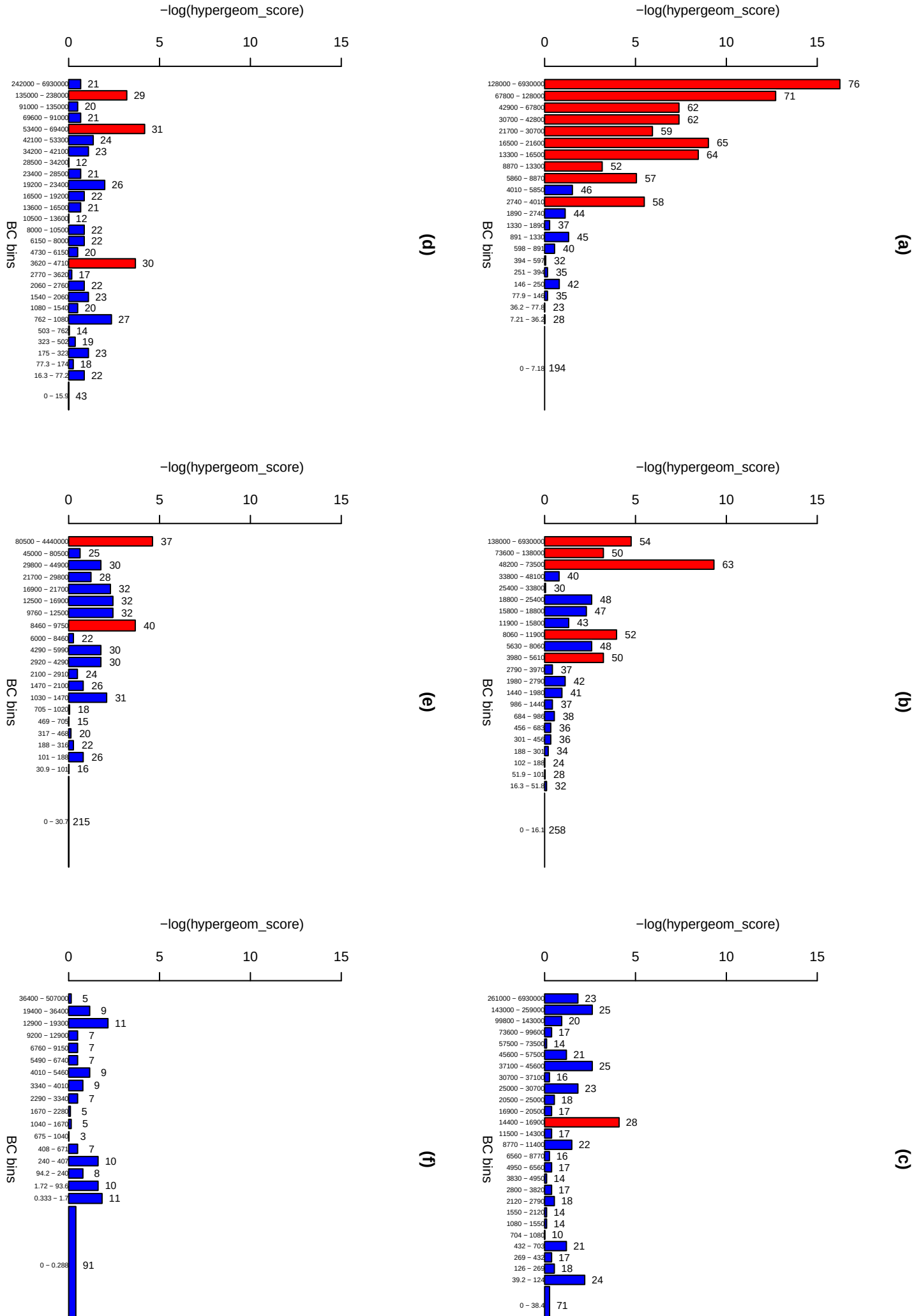
Supplementary Figure 7 - Over-representation of drug targets according to pathway centrality.
Proteins were grouped into bins according to their pathway centrality. The width of each bin represents

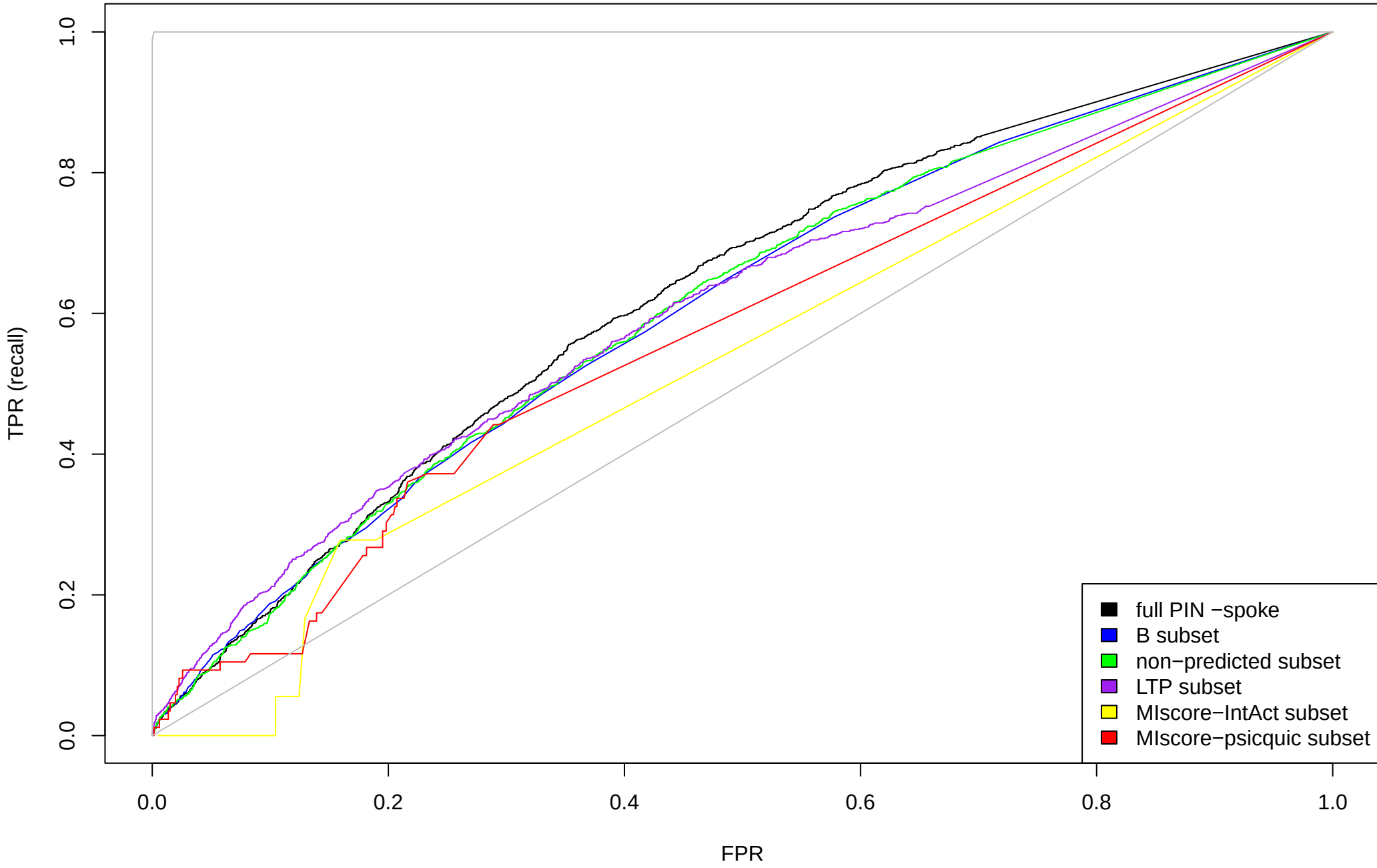
the number of proteins in that bin while the height (-log of the p-value of the hypergeometric test) represents how over-represented drug targets are in that bin. Each bin contains at least 200 proteins. Over-represented bins (p-value < 0.05) are highlighted in red. Drug targets are highly over-represented in all pathway centrality bins greater than zero and are not over-represented in proteins outside the pathway databases (pathway centrality = 0 – not shown). This observation is valid for all databases: PID (a), Reactome (b) and KEGG (c). A positive correlation occurs between increasing pathway centrality and over-representation only for the PID database.

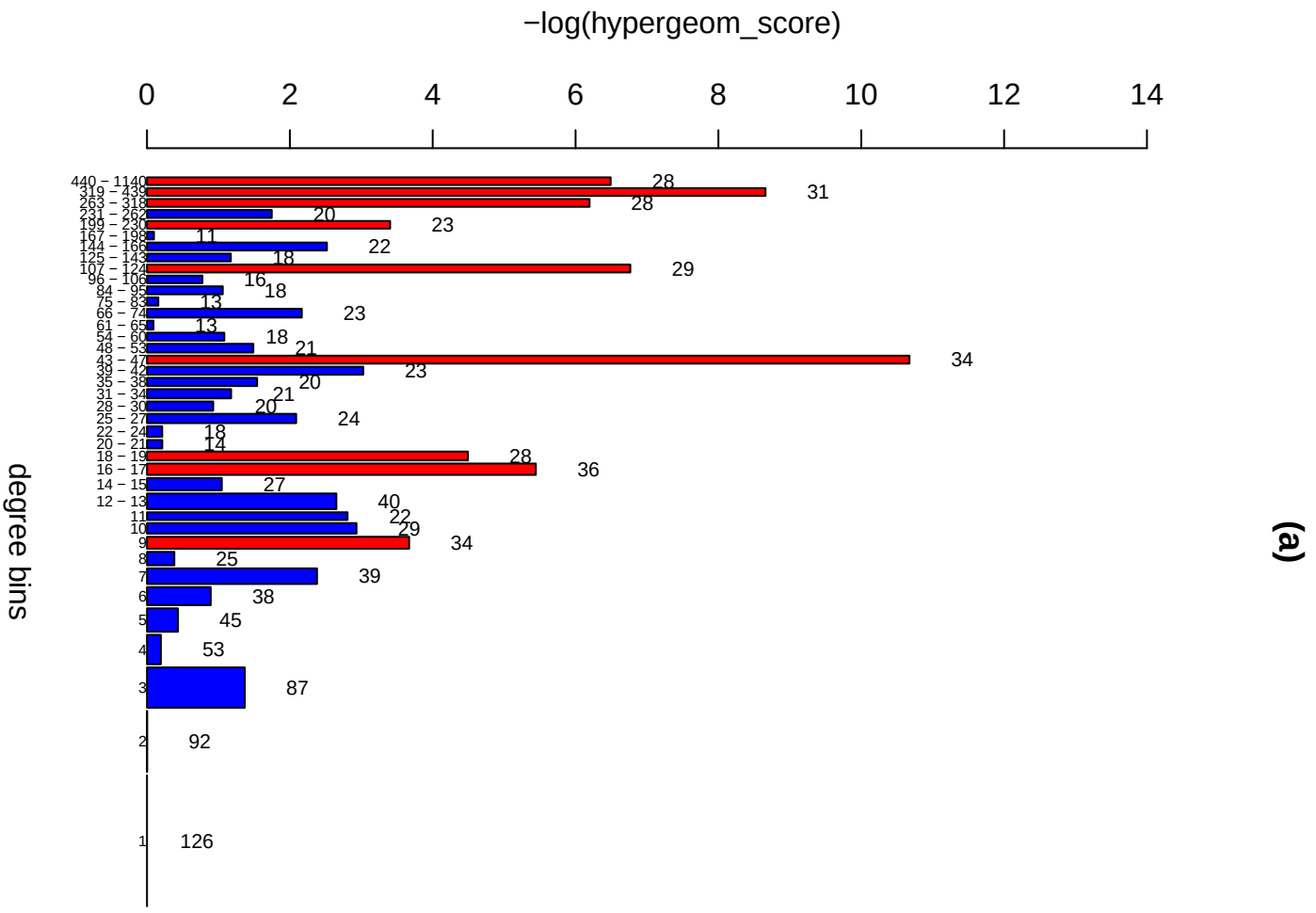
Supplementary Figure 8 – ROC curve for pathway centrality over different pathway databases. Plot of False Positive Rate versus True Positive Rate for a pathway centrality rank of the Reactome, PID and KEGG pathway databases.

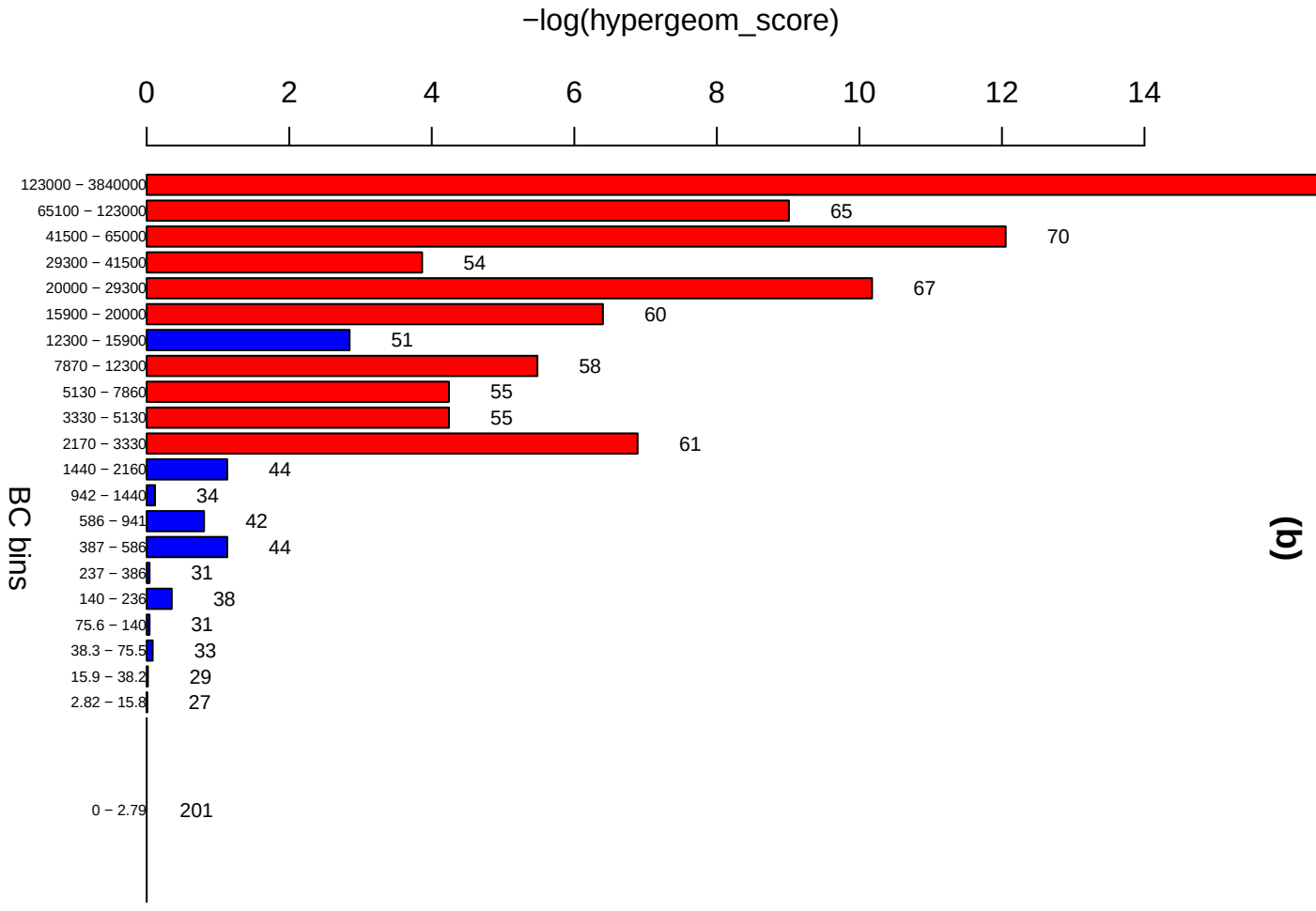

#prots
0
20
40
60
80
100
120
140
degree
0
200
400
600
800

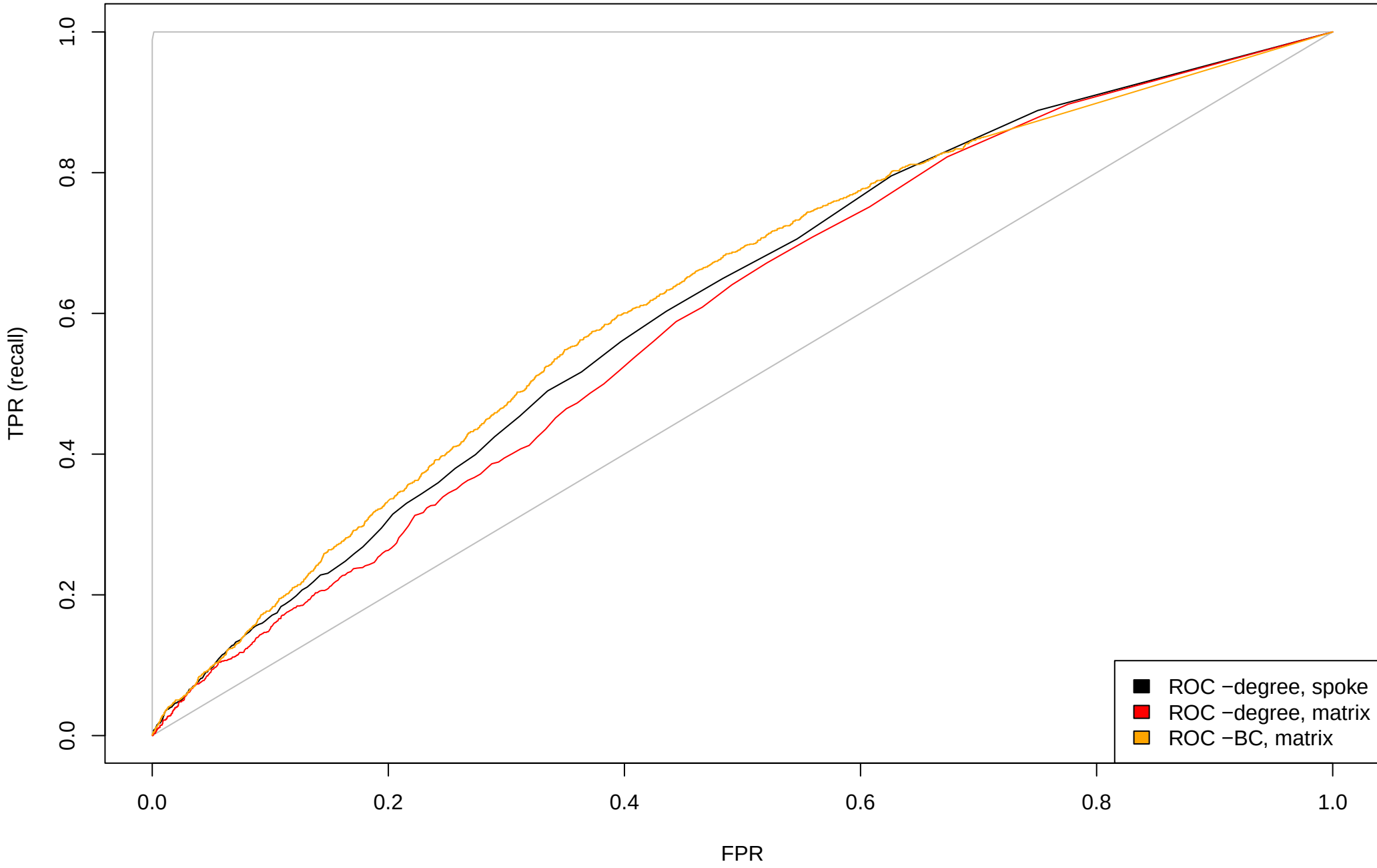

0.000
0.005
0.010
0.015
25
75
120
180
220
280
320
380
420
480
520
580
620
680
720
780

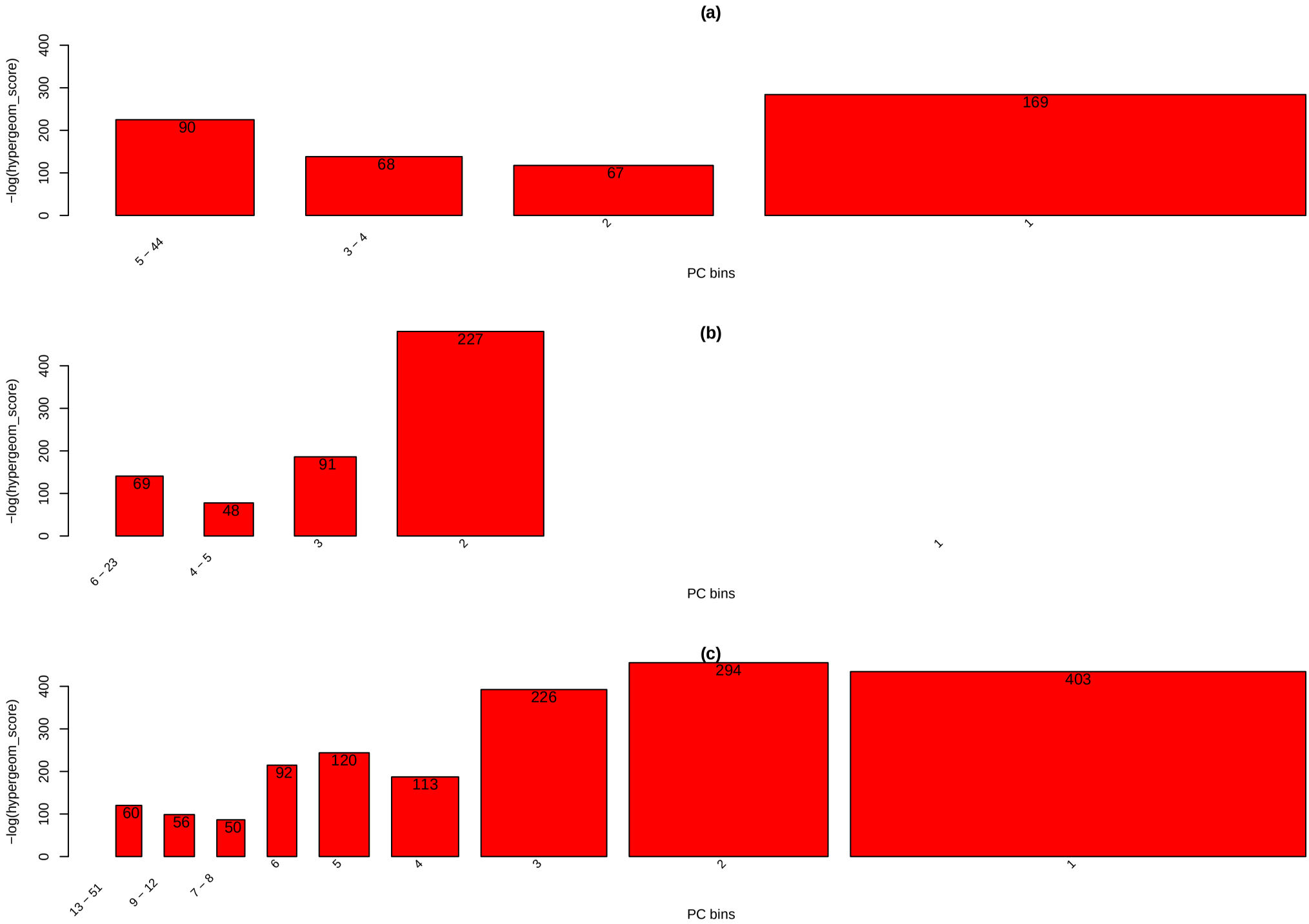

−log(hypergeom_score)
0
5
10
15
BC bins
242000 – 6930000 21
135000 – 238000 29
91000 – 135000 20
69600 – 91000 21
53400 – 69400 31
42100 – 53300 24
34200 – 42100 23
28500 – 34200 12
23400 – 28500 21
19200 – 23400 26
16500 – 19200 22
13600 – 16500 21
10500 – 13600 12
8000 – 10500 22
6150 – 8000 22
4730 – 6150 20
3620 – 4710 30
2770 – 3620 17
2060 – 2760 22
1540 – 2060 23
1080 – 1540 20
762 – 1080 27
503 – 762 14
323 – 502 19
175 – 323 23
77.3 – 174 18
16.3 – 77.2 22
0 – 15.9 43
(d)
−log(hypergeom_score)
0
5
10
15
BC bins
128000 – 6930000 76
67800 – 128000 71
42900 – 67800 62
30700 – 42800 62
21700 – 30700 59
16500 – 21600 65
13300 – 16500 64
8870 – 13300 52
5860 – 8870 57
4010 – 5850 46
2740 – 4010 58
1890 – 2740 44
1330 – 1890 37
891 – 1330 45
598 – 891 40
394 – 597 32
251 – 394 35
146 – 250 42
77.9 – 146 35
36.2 – 77.8 23
7.21 – 36.2 28
0 – 7.18 194
(a)
−log(hypergeom_score)
0
5
10
15
BC bins
80500 – 4440000 37
45000 – 80500 25
29800 – 44900 30
21700 – 29800 28
16900 – 21700 32
12500 – 16900 32
9760 – 12500 32
8460 – 9750 40
6000 – 8460 22
4290 – 5990 30
2920 – 4290 30
2100 – 2910 24
1470 – 2100 26
1030 – 1470 31
705 – 1020 18
469 – 705 15
317 – 468 20
188 – 316 22
101 – 188 26
30.9 – 101 16
0 – 30.7 215
(e)
−log(hypergeom_score)
0
5
10
15
BC bins
138000 – 6930000 54
73600 – 138000 50
48200 – 73500 63
33800 – 48100 40
25400 – 33800 30
18800 – 25400 48
15800 – 18800 47
11900 – 15800 43
8060 – 11900 52
5630 – 8060 48
3980 – 5610 50
2790 – 3970 37
1980 – 2790 42
1440 – 1980 41
986 – 1440 37
684 – 986 38
456 – 683 36
301 – 456 36
188 – 301 34
102 – 188 24
51.9 – 101 28
16.3 – 51.8 32
0 – 16.1 258
(b)
−log(hypergeom_score)
0
5
10
15
BC bins
36400 – 507000 5
19400 – 36400 9
12900 – 19300 11
9200 – 12900 7
6760 – 9150 7
5490 – 6740 7
4010 – 5460 9
3340 – 4010 9
2290 – 3340 7
1670 – 2280 5
1040 – 1670 5
675 – 1040 3
408 – 671 7
240 – 407 10
94.2 – 240 8
1.72 – 93.6 10
0.333 – 1.7 11
0 – 0.288 91
(f)
−log(hypergeom_score)
0
5
10
15
BC bins
261000 – 6930000 23
143000 – 259000 25
99800 – 143000 20
73600 – 99600 17
57500 – 73500 14
45600 – 57500 21
37100 – 45600 25
30700 – 37100 16
25000 – 30700 23
20500 – 25000 18
16900 – 20500 17
14400 – 16900 28
11500 – 14300 17
8770 – 11400 22
6560 – 8770 16
4950 – 6560 17
3830 – 4950 14
2800 – 3820 17
2120 – 2790 18
1550 – 2120 14
1080 – 1550 14
704 – 1080 10
432 – 703 21
269 – 432 17
126 – 269 18
39.2 – 124 24
0 – 38.4 71
(c)

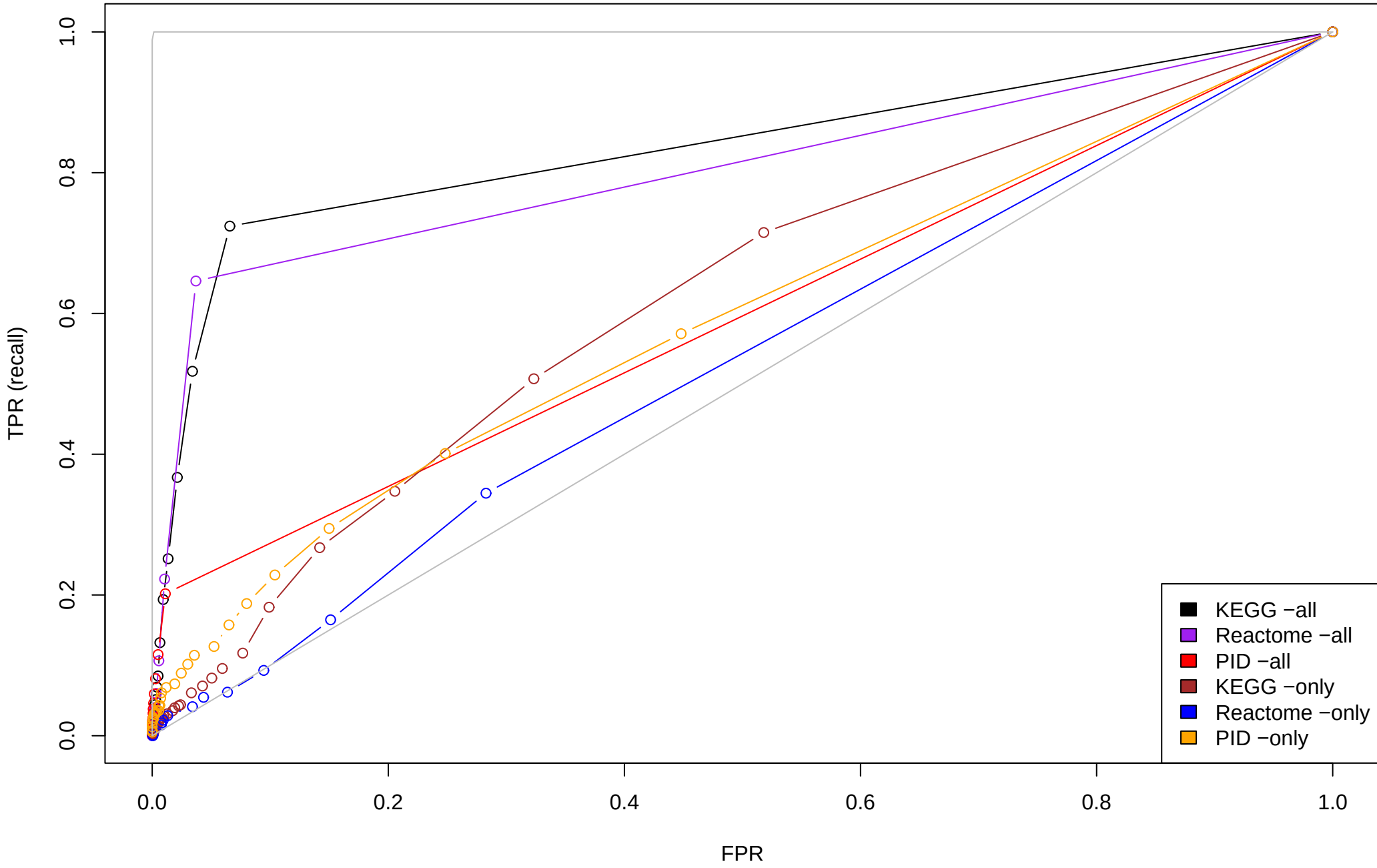

TPR (recall)
FPR
0.0
0.2
0.4
0.6
0.8
1.0
full PIN −spoke
B subset
non−predicted subset
LTP subset
MIscore−IntAct subset
MIscore−psicquic subset


−log(hypergeom_score)
0
2
4
6
8
10
12
14
degree bins
440 − 1140
319 − 439
263 − 318
231 − 262
199 − 230
167 − 198
144 − 166
125 − 143
107 − 124
96 − 106
84 − 95
75 − 83
66 − 74
61 − 65
54 − 60
48 − 53
43 − 47
39 − 42
35 − 38
31 − 34
28 − 30
25 − 27
22 − 24
20 − 21
18 − 19
16 − 17
14 − 15
12 − 13
11
10
9
8
7
6
5
4
3
2
1
28
31
28
20
23
11
22
18
29
16
18
13
23
13
18
21
34
23
20
21
20
24
18
14
28
36
27
40
22
29
34
25
39
38
45
53
87
92
126
(a)


−log(hypergeom_score)
0
2
4
6
8
10
12
14
BC bins
123000 − 3840000
65100 − 123000
41500 − 65000
29300 − 41500
20000 − 29300
15900 − 20000
12300 − 15900
7870 − 12300
5130 − 7860
3330 − 5130
2170 − 3330
1440 − 2160
942 − 1440
586 − 941
387 − 586
237 − 386
140 − 236
75.6 − 140
38.3 − 75.5
15.9 − 38.2
2.82 − 15.8
0 − 2.79
65
70
54
67
60
51
58
55
55
61
44
34
42
44
31
38
31
33
29
27
201
(b)


TPR (recall)
0.0
0.2
0.4
0.6
0.8
1.0
FPR
0.0
0.2
0.4
0.6
0.8
1.0
ROC −degree, spoke
ROC −degree, matrix
ROC −BC, matrix


(a)
−log(hypergeom_score)
0
100
200
300
400
90
68
67
169
5 – 44
3 – 4
2
1
PC bins
(b)
−log(hypergeom_score)
0
100
200
300
400
69
48
91
227
6 – 23
4 – 5
3
2
1
PC bins
(c)
−log(hypergeom_score)
0
100
200
300
400
60
56
50
92
120
113
226
294
403
13 – 51
9 – 12
7 – 8
6
5
4
3
2
1
PC bins


TPR (recall)
FPR
0.0
0.2
0.4
0.6
0.8
1.0
KEGG −all
Reactome −all
PID −all
KEGG −only
Reactome −only
PID −only