A Generic Parallel Framework for Inferring Large Scale Gene Regulatory Networks from Expression Profiles : Applications to Alzheimer's Disease Network

SUPPLEMENTARY -1

Algorithmic Representation of the Generic Parallel Framework for Inferring Large Scale Gene Regulatory Networks from Expression Profiles

We include the algorithmic representation of the generic parallel framework in this section.

Algorithm 1 Dataset partitioning

procedure PARTITIONING(p, c) ▷ Partitioning the dataset with c columns into p partitions
    count=1
    d=c/p ▷ d is the number of columns in each partition
    for every partition p **do**
        Extract the columns from count to d
        Store the extracted columns as a sub-matrix
        count=count+d
    end for
    Store all the sub-matrices in SMs
end procedure

Algorithm 2 Local sub-network inference

procedure INFERENCE(SMs, Algorithm-X) ▷ Inferring the sub-networks using an Algorithm X
    for every sub-matrix in SMs **do** parallely
        Infer the network using Algorithm-X
    end for
    Store all the inferred sub-networks in SNs
end procedure

Algorithm 3 Centrality Analysis of sub-networks

procedure FINDING-HUBGENES(SNs, t) ▷ Finding the hub genes in each sub-network
    for every sub-network in SNs **do** parallely
        Calculate the degree of every gene in the sub-network
        Sort the genes as per descending order of degree.
        Select all genes that are ranked top t and store it in HG
    end for
end procedure

Algorithm 4 Hub gene distribution

procedure APPENDING-HUBGENES(SNs,HGs) ▷ Distributing the hub genes to each sub-matrix
    for every sub-network in SNs **do** parallely
        Add the hub genes HGs of other sub-networks to each sub-matrix
    end for
    Store all the modified sub-matrices in MSMs
end procedure

Algorithm 5 Rewiring local sub-networks

procedure RE-INFERENCE($MSMs$, Algorithm-X) ▷ Re-inferring the modified sub-networks using an Algorithm X
    for every modified sub-matrix in MSMs **do** parallely
        Infer the network using Algorithm-X
    end for
    Store all the re-inferred sub-networks in MSNs
end procedure

Algorithm 6 Global Network Construction

procedure MERGE($MSNs$) ▷ Merge the modified sub-networks to get the global network
    Create an empty adjacency matrix, adj, of size c∗c
    for every modified sub-networks in MSNs **do**
        for every row r in the modified sub-network **do**
            for every column c in the modified sub-network **do**
                Copy the content of the cell $r * c$ and paste it in corresponding cell in adj
            end for
        end for
    end for
end procedure

Analysis of the effect of 'p'-the number of partitions into which the expression matrix is sliced on the results of the framework

Different methods respond differently to the variations in the number of partitions. Sensitivity analysis was performed to see the effect of the number of partitions on the execution time and accuracy of results. The number of partitions were varied from 3 to 100 except for the dataset with 100 genes for which the maximum number of partitions was 50. With the exception of five methods — CLR, MRNET, MRNETB, MutRank and GeneNet, the number of partitions did not affect the accuracy of the results. For the five methods mentioned above, the accuracy decreased when there was less than 10% of the total genes in each partition of the dataset. We also see that there is an increase in accuracy observed again when as less as 2% of the total genes is in one partition but having that less number of genes would render the purpose of the framework useless because of very expensive execution time and hence is ignored.

As far as the execution time is concerned, we have two observations. For datasets with less than 2000 genes, we observed that the execution time was more than the serial execution time when there was less than 2% of the total genes in each partitions. For the datsets with more than 2000 genes, the parallel execution time was always less than the serial execution time for all methods. The results of the sensitivity analysis of two methods in reported in **Table 1 and 2**

Table 1. Results of the sensitivity analysis of the number of partitions — p (method — MRNETB)

Method	Dataset	No. of genes	No. of divisions	Time (parallel)	Time (ori)	AUROC (serial)	AUROC (parallel)
MRNETB	E1	100	3	**0.2582793**	0.0304215	0.4700965	0.4751735
			4	0.2699604			0.4626734
			5	0.3001034			0.4460002
			10	0.4028945			0.4070058
			15	0.6074078			0.5061602
			20	0.87482			0.4750513
			25	1.46544			0.4708554
			50	2.201079			0.4837548
	E2	500	3	1.037702	1.437677	0.4700795	0.5047914
			4	0.8332999			0.4943463
			5	0.7474859			0.4876388
			10	**0.7468171**			0.4747045
			15	0.8918905			0.4199268
			20	1.232851			0.4445019
			25	2.241731			0.4324351
			30	2.999165			0.4405668
			40	9.017609			0.4449185
			50	13.87991			0.4428238
			100	75.83868			0.4644348
	E3	1000	4	2.87572	8.781887	0.5849953	0.6393649
			5	2.37909			0.6255273
			8	1.752395			0.6214211
			10	1.700114			0.6232417
			15	1.676256			0.5901537
			20	**1.625044**			0.5713419
			25	2.230592			0.5604388
			30	3.476223			0.5567747
			40	5.220152			0.5395399
			50	10.46326			0.5855421
			100	125.35788			0.5756264
	Y1	2000	5	11.48563	60.94098	0.4874883	0.4678863
			8	7.952174			0.4495279
			10	6.376798			0.4491476
			15	**5.366306**			0.4212134
			20	5.925039			0.4162524
			25	5.83725			0.4057272
			30	6.272249			0.4187293
			40	9.302474			0.4039421
			50	14.55863			0.4016726
			100	124.30506			0.4291133
			200	1822.1346			0.4515485
	Y2	3000	5	38.80129	232.97442	0.5226453	0.5476824
			10	16.12336			0.5268498
			15	12.0399			0.5102305
			20	10.47794			0.4817858
			25	10.76091			0.4741886
			30	**9.627196**			0.4764677
			40	13.33211			0.4644823
			50	21.44946			0.4555635
			100	144.6708			0.4654877
			200	1965.207			0.4841594
	Y3	4000	5	75.44892	596.8266	0.5143631	0.5389527
			10	34.00466			0.5204785
			15	22.85685			0.4998019
			20	20.78727			0.4940239
			25	19.4952			0.4763505
			30	**17.14795**			0.4741595
			40	21.71285			0.4662791
			50	24.88754			0.4609697
			100	116.17974			0.4425268

Table 2. Results of the sensitivity analysis of the number of partitions — p (method — GENIE3)

Method	Dataset	No. of genes	No. of divisions	Time (parallel)	Time (ori)	AUROC (serial)	AUROC (parallel)
GENIE3	E1	100	3	**67.302**	101.9859	0.6122919	0.7450411
			4	55.92979			0.6501785
			5	68.61876			0.6016743
			10	195.47706			0.6616885
			15	506.79594			0.6242751
			20	815.481			0.6204374
			25	1079.7702			0.6573135
			50	2201.3382			0.6591835
	E2	500	3	620.4708	1101.2172	0.4547994	0.4563433
			4	488.78856			0.4587709
			5	556.53576			0.4622911
			10	**491.3475**			0.4624013
			15	504.69216			0.4561326
			20	595.6548			0.4613543
			25	872.2956			0.4602847
			30	1126.983			0.4580237
			40	3731.0688			0.4580121
			50	8115.336			0.4530124
			100	43827.012			0.4529339
	E3	1000	4	1365.5496	3028.029	0.4859839	0.4876607
			5	1441.9476			0.4892534
			8	1001.336			0.4900074
			10	959.1			0.4954473
			15	948.4194			0.5000924
			20	**1328.409**			0.4920679
			25	1530.747			0.4932211
			30	1593.7086			0.5000215
			40	2181.3426			0.4852749
			50	3902.292			0.4872499
			100	55956.456			0.4887764
	Y1	2000	5	4360.1256	8368.8876	0.3501581	0.3903841
			8	3670.3123			0.3905948
			10	3183.0498			0.3910653
			15	**2702.895**			0.3909497
			20	2715.6816			0.3931362
			25	2460.7878			0.3912922
			30	2873.8068			0.3921571
			40	3565.4142			0.3933395
			50	4395.4992			0.3914612
			100	14648.8068			0.3909949
			200	-			-
	Y2	3000	5	7811.2116	15444.7416	0.4196192	0.4296868
			10	4621.3884			0.4300033
			15	4091.2272			0.4303226
			20	3625.2972			0.4305626
			25	3742.7832			0.4312513
			30	**3845.9196**			0.4313225
			40	3995.028			0.4312531
			50	5419.296			0.4299473
			100	16049.9412			0.4292132
			200	-			-
	Y3	4000	5	12150.7488	23756.094	0.4146161	0.4156369
			10	7294.5072			0.4159663
			15	6240.6612			0.4161526
			20	5747.094			0.4163694
			25	6178.2228			0.4165883
			30	**5195.4696**			0.4168235
			40	5986.5624			0.4165706
			50	7237.3752			0.4164336
			100	18868.6584			0.4163032

Analysis of the effect of the number of hub genes considered in the framework

While performing experiments, the number of hub genes to be included in the phase 3 of the framework was decided on the basis of the results of the experiments performed by varying the number of hub genes on the basis of their degrees. We ranked the genes on the basis of the degree and then considered only those genes that had the highest degree. The execution time and the accuracy was observed. Then the top two highest degree possessing genes were considered. Similary, the experiment was done up till the top 10 highest degree possessing genes were included. The results are reported in Table 3. We can see that it is sufficient to include the top two highest degree possessing nodes as the best accuracy with less execution time is achieved in this situation.

Analysis of how the order of the datasets affects the execution time and accuracy

We shuffled the order of the genes in each dataset to see whether it affected the accuracy of the networks obtained. We inferred the networks using both the original methods serially as well as using our framework. We observed slight change in both the accuracy of the results and also the execution time. Since both the serial methods as well as the framework saw similar changes in accuracy and execution time we concluded that this was due to the methods. Otherwise there was no reason why the serial execution also showed changed in accuracy when the whole expression matrix was available in its case. The results of this analysis is reported in Table 4.

Analysis of the effect of the number of cores on the parallel execution time and actual speedup of the generic parallel framework

After reporting the theoretical speedup using Amdahl's law, it was necessary to see how the execution time and thereby the speedup varied with the increase in the number of cores. Since the system we used had six cores, we varied the number of cores by 1, 2, 4 and 6 for the datasets consisting of 1000 and 4000 genes. The results are reported in Table 5 and 6. We see that as expected, there is an increase in speedup as the number of cores is increased.

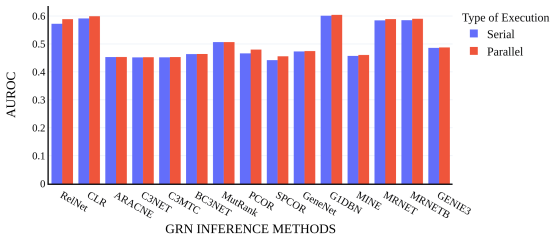

Fig. 1. Performance comparison of the GRN inference methods (benchmarked using gold standard network) for the dataset with 1000 genes

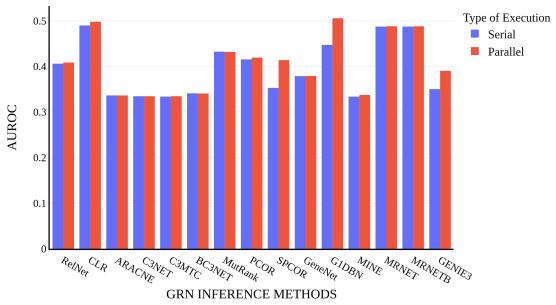

Fig. 2. Performance comparison of the GRN inference methods (benchmarked using gold standard network) for the dataset with 2000 genes

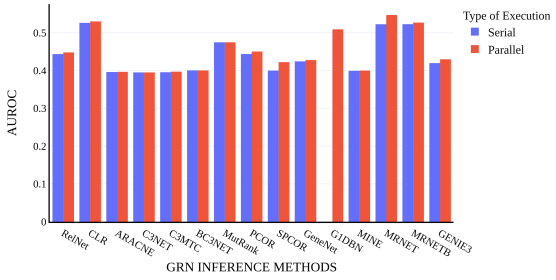

Fig. 3. Performance comparison of the GRN inference methods (benchmarked using gold standard network) for the dataset with 3000 genes

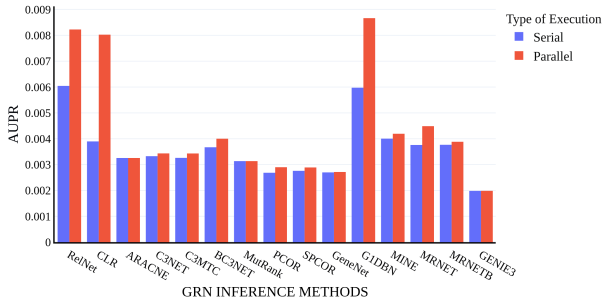

Fig. 4. Prediction accuracy comparison of the GRN inference methods (benchmarked using gold standard network) for the dataset with 1000 genes

Extended results of the enrichment analysis

We report here the complete results of the enrichment analysis performed on the 41 clusters obtained from the inferred AD network in Table 7. We see that for almost all clusters, a good number of genes are seen associated with brain related disorders and other pathways that were reported by the work that contributed the dataset. Despite using CLR, an algorithm that is just an average performance showcasing inference algorithm, the inferred network shows good performance. We have used

Table 3. Results from the experiments performed by varying the number of highest degree genes considered

No. of genes	Method	Rank of hub gene degree[1]	Execution Time	%increase in time	AUROC
4000	CLR	1	56.19894	-	0.508316
		2	77.43394	27.42337533	**0.508492**
		3	87.18292	35.53904824	0.508031
		4	128.564	56.2871877	0.507786
		5	169.94232	66.93057974	0.507068
		6	175.53024	67.98332868	0.506776
		7	175.96728	68.06284668	0.506537
		8	283.44988	80.17323556	0.505534
		9	480.51426	88.30441785	0.505674
		10	789.45654	92.88131301	0.507823
3000	ARACNE	1	26.06695	-	0.506135
		2	27.50107	5.214778916	**0.506142**
		3	27.79961	6.232677365	0.503316
		4	45.44948	42.64631851	0.502632
		5	50.3376	48.21574727	0.502067
		6	53.81044	51.55782038	0.498136
		7	67.66704	61.47762633	0.497384
		8	79.31166	67.1335211	0.494322
		9	149.93004	82.61392447	0.491485
		10	150.92364	82.72838503	0.494049
2000	MRNET	1	24.70366	-	0.505361
		2	24.90474	0.807396504	**0.506317**
		3	25.28636	2.30440443	0.505751
		4	25.18159	1.897934165	0.505179
		5	25.50633	3.146944308	0.504698
		6	25.50846	3.155031703	0.503321
		7	35.37619	30.16868125	0.502513
		8	35.55297	30.51590345	0.502514
		9	35.82522	31.04394055	0.501631
		10	36.1349	31.63490144	0.500626
1000	MutRank	1	4.277107	-	0.508207
		2	4.68632	8.732075488	**0.509583**
		3	4.921183	13.08782868	0.506387
		4	4.941176	13.43949295	0.506346
		5	5.460856	21.67698617	0.505557
		6	5.653765	24.34940257	0.504989
		7	6.107507	29.96967502	0.503167
		8	6.894055	37.95948828	0.503161
		9	7.743583	44.76578865	0.504027
		10	8.967887	52.30641287	0.504382
100	Genie3	1	53.20957	-	0.639795
		2	56.54971	6.277329435	**0.647199**
		3	60.51642	13.7322102	0.638214
		4	63.09306	18.57464738	0.633848
		5	68.21982	28.20968108	0.640744
		6	68.43	28.60468521	0.642381
		7	84.26478	58.36395596	0.634011
		8	92.16348	73.20846607	0.629124
		9	103.44786	94.41589173	0.570448
		10	166.81758	213.5104832	0.581995

Table 4. Results from the experiments conducted to observe how the order of genes in a dataset affects the accuracy and execution time both in serial execution as well as using our generic parallel framework

Dataset	Method	Order	Serial execution (AUROC)	Parallel Execution (AUROC)	Serial execution (time)	Parallel Execution (Time)
1000	CLR	Not changed	0.5914037	0.6216419	2.013068	1.41811
		Shuffled	0.6144022	0.6353742	2.4156816	1.660474
	ARACNE	Not changed	0.4531609	0.4666568	3.6787	0.8885281
		Shuffled	0.4621106	0.4631829	4.04657	1.073504
2000	GeneNet	Not changed	0.3786474	0.3799406	801.3882	18.56008
		Shuffled	0.3568267	0.3557269	830.604762	20.62397
	Genie3	Not changed	0.3501581	0.3909497	8368.8876	2460.7878
		Shuffled	0.3552018	0.4027007	9353.84648	2873.8068
3000	MINE	Not changed	0.3991989	0.4074706	4225.73	50.25479
		Shuffled	0.4000889	0.4113608	4502.2064	58.26799
	MRNET	Not changed	0.5224756	0.5468827	172.38498	10.08324
		Shuffled	0.5135056	0.5213046	184.66974	10.5455
4000	PCOR	Not changed	0.5141001	0.5386138	8625.906	54.30579
		Shuffled	0.5150379	0.5278262	8675.799962	61.76574
	C3NET	Not changed	0.3954438	0.3971617	251.74806	9.96609
		Shuffled	0.3990732	0.3993559	258.68586	10.75217

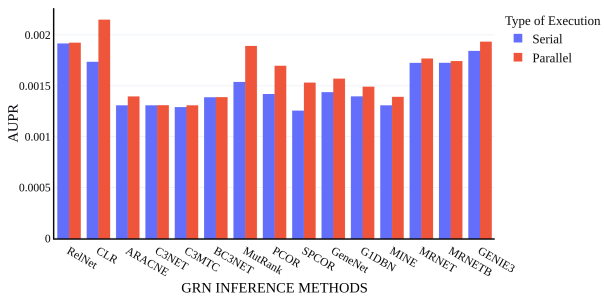

Fig. 5. Prediction accuracy comparison of the GRN inference methods (benchmarked using gold standard network) for the dataset with 2000 genes

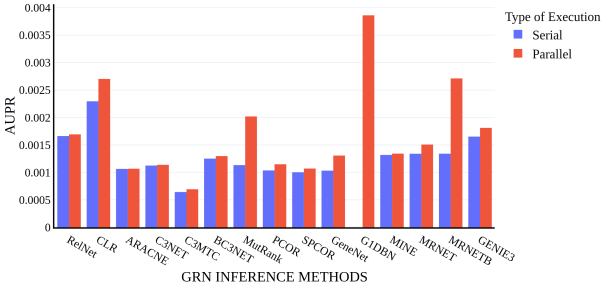

Fig. 6. Prediction accuracy comparison of the GRN inference methods (benchmarked using gold standard network) for the dataset with 3000 genes

the default parameters in all the algorithms so as to not meddle with its working. If the algorithm can be tuned with the best performing parameters then the framework is sure to perform better.

Additional results and discussion of the performance assessment of the generic parallel framework

This section includes the results of the performance assessment performed by comparing the resultant networks from the serial

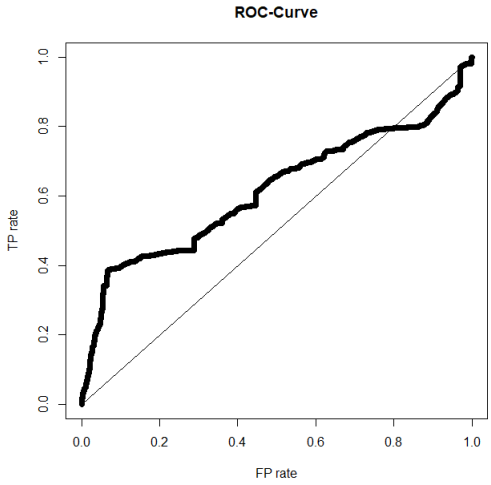

Fig. 7. ROC curve plot for G1DBN (Dataset: 1000)

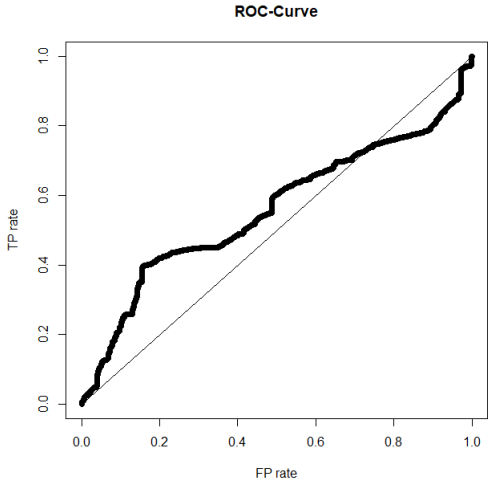

Fig. 8. ROC curve plot for SPCOR (Dataset: 1000)

and parallel execution through the framework for all the methods for the datasets having 1000, 2000, and 3000 genes as seen in Figures 1-6. From these figures too it is established

Table 5. Results from the experiments performed on the dataset with 1000 genes to analyze how the number of cores influences the T(p) and Speedup of the generic parallel framework

Method	Serial Execution Time	No. of Cores	Time taken	Speedup
RELNET	4.708791	1	3.305254	1.424638
		2	2.215338	2.125541
		4	2.066441	2.278696
		6	2.047779	2.299462
CLR	2.013068	1	1.891723	1.064145
		2	1.660474	1.212345
		4	1.589154	1.266755
		6	1.41811	1.419543
ARACNE	3.6787	1	3.380013	1.088369
		2	1.332889	2.759945
		4	1.177884	3.123143
		6	0.8885281	4.140218
MRNET	3.817024	1	5.747403	0.66413
		2	2.275635	1.677345
		4	1.711515	2.230202
		6	1.357832	2.811117
MRNETB	8.781887	1	8.046114	1.091445
		2	5.002226	1.755596
		4	3.857496	2.276577
		6	1.625004	5.404225
C3NET	4.78581	1	11.06605	0.432477
		2	3.49622	1.368853
		4	2.997953	1.596359
		6	2.73455	1.750127
C3MTC	1.159587	1	14.50754	0.07993
		2	2.420735	0.479023
		4	1.268201	0.914356
		6	1.08314	1.070579
BC3NET	62.65368	1	31.77093	1.972044
		2	8.373955	7.48197
		4	4.475174	14.00028
		6	3.876542	16.16226
MutRank	9.320698	1	20.19428	0.461551
		2	4.45259	2.093321
		4	4.124638	2.259761
		6	1.552822	6.002425
PCOR	88.12548	1	41.93982	2.101236
		2	8.155652	10.80545
		4	6.503647	13.55016
		6	5.607112	15.71673
SPCOR	171.48102	1	32.12222	5.338393
		2	12.58253	13.6285
		4	7.351351	23.32646
		6	5.55842	30.85068
Genie3	3028.029	1	3399.6708	0.890683
		2	1117.17	2.710446
		4	1005.927	3.010188
		6	948.419	3.192712
GeneNet	119.61564	1	49.1186	2.435241
		2	12.31467	9.713264
		4	11.05297	10.82204
		6	9.55122	12.5236
MINE	361.03002	1	25.4252	14.19969
		2	9.464289	38.14655
		4	9.233094	39.10174
		6	8.20522	44.00004
G1DBN	327495.83	1	6728.7852	48.67087
		2	2411.115	135.8275
		4	2315.5524	141.4331
		6	1332.399	245.7941

Table 6. Results from the experiments performed on the dataset with 4000 genes to analyze how the number of cores influences the T(p) and Speedup of the generic parallel framework

Method	Serial Execution Time	No. of Cores	Time taken	Speedup
RELNET	264.4341	1	73.93242	3.576700181
		2	43.69066	6.052417153
		4	38.41355	6.883875612
		6	35.50105	7.448627576
CLR	251.58714	1	67.60982	3.721162695
		2	31.08285	8.094082106
		4	19.05674	13.20200307
		6	16.5067	15.24151648
ARACNE	331.1806	1	73.04046	4.534207479
		2	23.38503	14.16207719
		4	7.499687	44.15925625
		6	6.470149	51.18593096
MRNET	418.08018	1	77.81466	5.372768833
		2	33.63054	12.4315631
		4	23.23	17.99742488
		6	18.25389	22.90362109
MRNETB	596.8266	1	75.44892	7.91033987
		2	34.00466	17.55131797
		4	20.78727	28.71115832
		6	17.14795	34.80454515
C3NET	251.748	1	85.65786	2.93899474
		2	25.46409	9.886392956
		4	11.10081	22.67834509
		6	9.96609	25.26045821
C3MTC	12.84364	1	10.33884	1.242270893
		2	7.191122	1.786041177
		4	6.720699	1.911057168
		6	5.02833	2.554255588
BC3NET	1010.1282	1	177.17256	5.70138062
		2	59.25709	17.04653738
		4	23.50392	42.97700979
		6	18.56242	54.41791534
MutRank	245.129	1	73.84638	3.319445043
		2	51.46657	4.762878117
		4	28.06297	8.734962835
		6	19.01626	12.89049477
PCOR	8625.906	1	746.8266	11.55007869
		2	114.56202	75.29463953
		4	76.38342	112.9290362
		6	61.68558	139.8366685
SPCOR	8146.2212	1	430.89678	18.90527286
		2	118.8702	68.53039029
		4	86.79114	93.8600553
		6	67.68558	120.3538656
Genie3	23756.094	1	11877.5376	2.000085775
		2	8030.9304	2.958074945
		4	5780.6532	4.109586439
		6	5195.47	4.572462934
GeneNet	13932.0072	1	778.0776	17.9056783
		2	71.91906	193.7178712
		4	46.02863	302.6813355
		6	44.4271	313.5925415
MINE	8833.67	1	169.8726	52.00173542
		2	105.7275	83.55129933
		4	93.3194	94.66059576
		6	88.2916	100.0510807

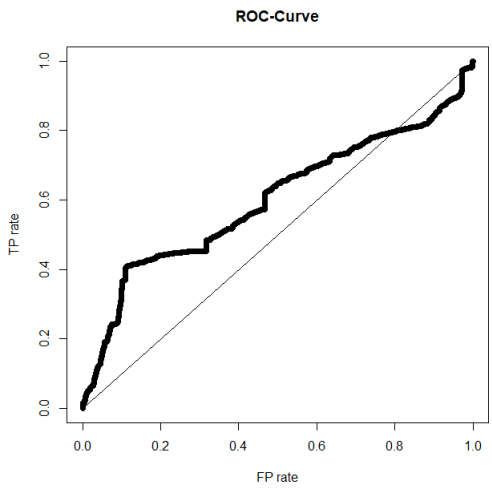

Fig. 9. ROC curve plot for PCOR (Dataset: 1000)

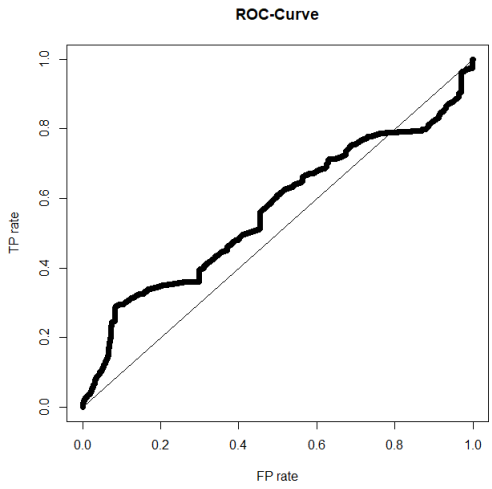

Fig. 10. ROC curve plot for MUTRANK (Dataset: 1000)

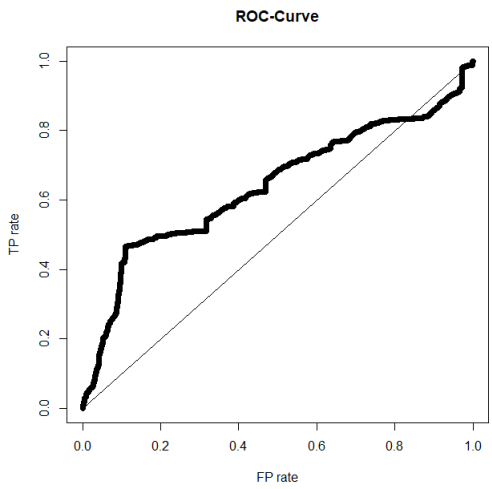

Fig. 11. ROC curve plot for MRNET (Dataset: 1000)

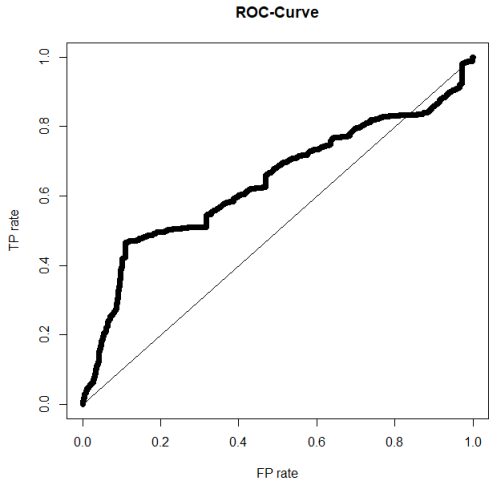

Fig. 12. ROC curve plot for MRNETB (Dataset: 1000)

that the generic parallel framework consistently performs well for all methods with datasets consisting of varying number of genes. The framework outputs a network that has the same accuracy as that of the original network and sometimes even better.

The AUROC scores for some of the methods are slightly below 0.5 for some of the datasets. This however has nothing to do with our framework. **We have used the methods as a blackbox, without any modifications. We were not intending to improve the qualitative aspect of the original algorithms. However, our generalized framework did not perform inferior either in comparison to original algorithms, which was one of the actual targets. However, the performance of the serial methods may be improved by appropriate parameter tuning, which is beyond the scope of this work at present.** The accuracy obtained is at par with that obtained when the methods are executed as is in the serial environment. The benefit of our framework is that it aids those methods incapable of inferring networks with more than 10000 nodes, to now infer networks with more than 25,000 genes on the same machine. With this achievement, we now can focus only on improving the methods in terms of accuracy without worrying on its scalability since our framework takes care of it.

We have also included the ROC and PR curve plots for 8 algorithms for the performance assessment against gold networks in Figures 7-22.

DoParallel R package

The doParallel package seeks to provide a "parallel backend" for the foreach package [2], i.e., it helps us execute the foreach loops in parallel. We first register the number of cores in the machine we are using and then pass it to the doParallel package. The machine should have multiple processors or cores or even both. We code the task that we require to be executed parallelly in the foreach loop. A detailed explanation of the working of this package is available at https://cran.r-project.org/web/packages/doParallel/doParallel.pdf.

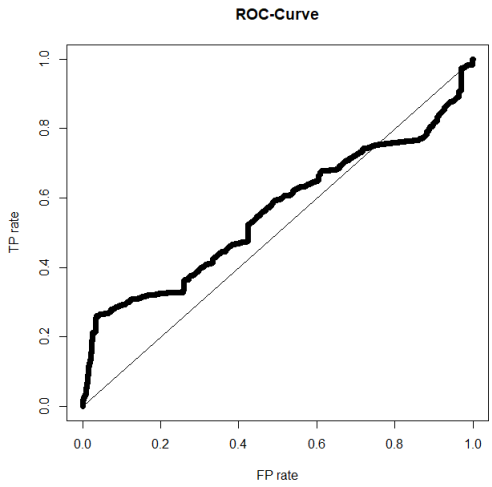

Fig. 13. ROC curve plot for RELNET (Dataset: 1000)

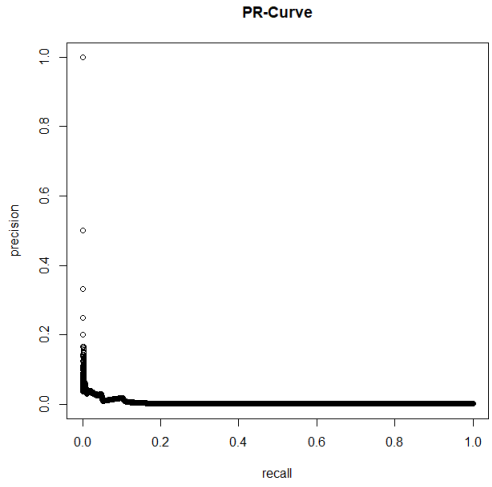

Fig. 14. ROC curve plot for ARACNE (Dataset: 1000)

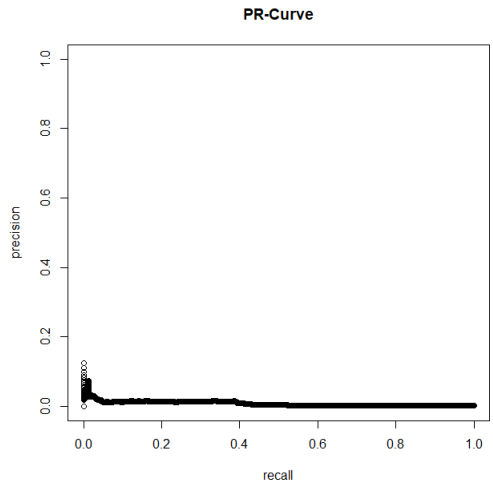

Fig. 15. PR curve plot for G1DBN (Dataset: 1000)

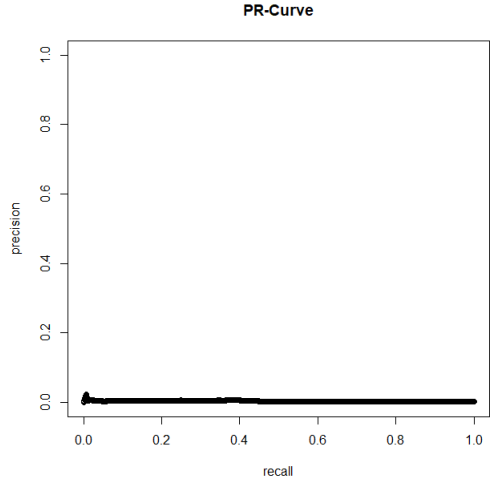

Fig. 16. PR curve plot for SPCOR (Dataset: 1000)

References

1. Da Wei Huang, Brad T Sherman, and Richard A Lempicki. Systematic and integrative analysis of large gene lists using david bioinformatics resources. *Nature protocols*, 4(1):44–57, 2009.
2. Steve Weston and Rich Calaway. Getting started with doparallel and foreach. *Available on https://cran. r-project. org/web/packages/doParallel/vignettes/gettingstartedParallel. pdf*, 2015.

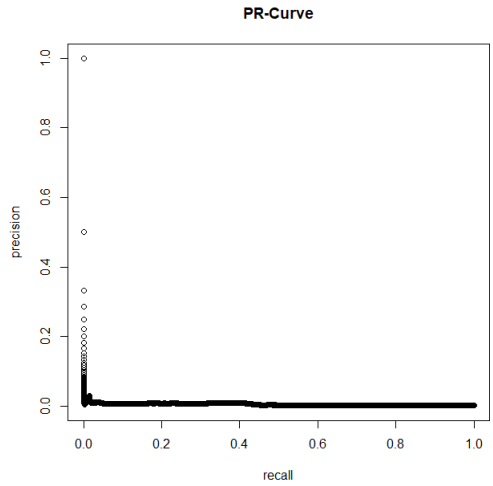

Fig. 17. PR curve plot for PCOR (Dataset: 1000)

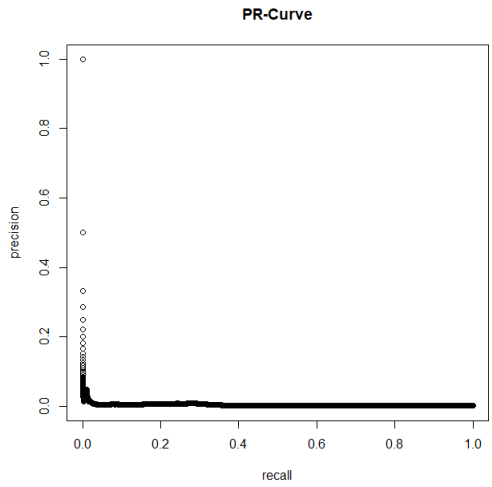

Fig. 18. PR curve plot for MUTRANK (Dataset: 1000)

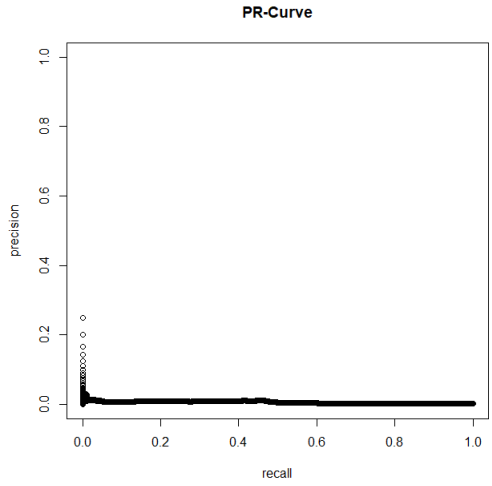

Fig. 19. PR curve plot for MRNET (Dataset: 1000)

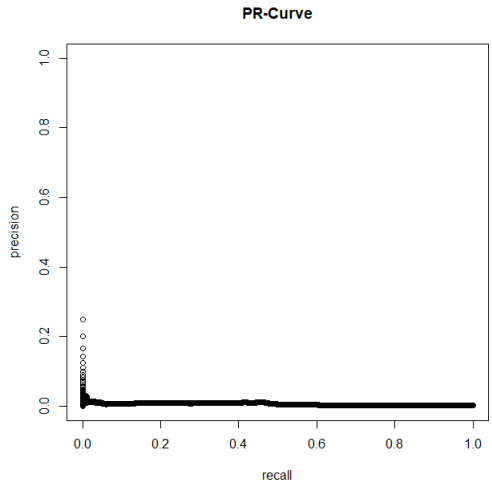

Fig. 20. PR curve plot for MRNETB (Dataset: 1000)

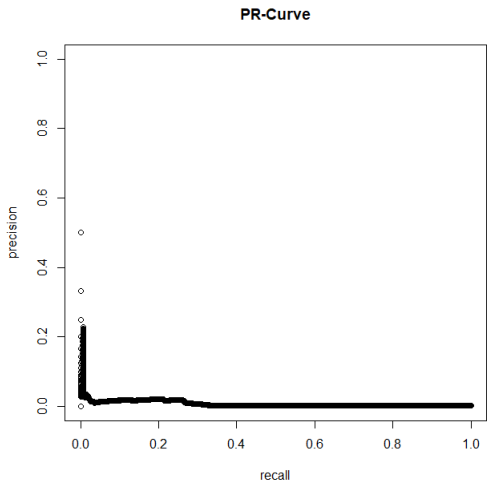

Fig. 21. PR curve plot for RELNET (Dataset: 1000)

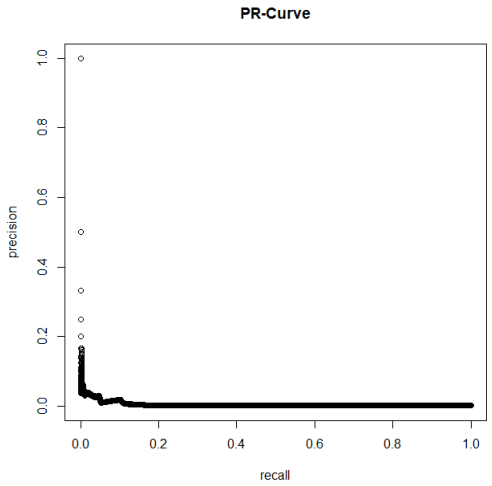

Fig. 22. PR curve plot for ARACNE (Dataset: 1000)

Table 7: Results of the enrichment analysis of the 41 clusters obtained from DAVID [1].

Cluster No.	Size of Cluster	Enrichment Score and Disease	Genes	P_value	Benjamini
7	1001	Enrichment Score: 7.37			
		Amyotrophic lateral sclerosis	51	3.90E-10	6.10E-08
		Parkinson disease	41	9.40E-10	9.60E-08
		Pathways of neurodegeneration - multiple diseases	57	4.80E-09	3.10E-07
		Huntington disease	43	5.00E-09	3.10E-07
		Prion disease	39	1.60E-08	8.40E-07
		Alzheimer disease	46	2.50E-07	8.50E-06
1	1000	Enrichment Score: 7.94	Count	P_Value	Benjamini
		Amyotrophic lateral sclerosis	54	3.40E-11	4.80E-09
		Parkinson disease	44	4.70E-11	4.80E-09
		Pathways of neurodegeneration - multiple diseases	60	6.60E-10	5.10E-08
		Huntington disease	45	1.10E-09	6.90E-08
		Prion disease	41	3.30E-09	1.70E-07
		Alzheimer disease	47	2.10E-07	8.20E-06
2	5064	Enrichment Score: 0.6	Count	P_Value	Benjamini
		Pathways of neurodegeneration - multiple diseases	48	1.30E-01	7.50E-01
		Amyotrophic lateral sclerosis	38	1.50E-01	7.70E-01
		Alzheimer disease	39	1.70E-01	7.70E-01
		Huntington disease	29	3.30E-01	9.40E-01
		Parkinson disease	24	4.70E-01	1.00E+00
		Prion disease	24	5.00E-01	1.00E+00
3	1000	Enrichment Score: 1.98	Count	P_Value	Benjamini
		Prion disease	28	1.50E-04	4.80E-02
		Alzheimer disease	33	1.10E-03	9.60E-02
		Pathways of neurodegeneration - multiple diseases	38	1.50E-03	9.60E-02
		Parkinson disease	25	1.50E-03	9.60E-02
		Huntington disease	23	2.70E-02	3.60E-01
		Amyotrophic lateral sclerosis	26	4.30E-02	4.40E-01
4	1000	Enrichment Score: 0.14	Count	P_Value	Benjamini
		Parkinson disease	12	6.20E-01	1.00E+00
		Prion disease	12	6.40E-01	1.00E+00
		Alzheimer disease	16	7.20E-01	1.00E+00
		Amyotrophic lateral sclerosis	15	7.50E-01	1.00E+00
		Huntington disease	12	7.80E-01	1.00E+00
		Pathways of neurodegeneration - multiple diseases	16	9.30E-01	1.00E+00
5	999	Enrichment Score: 0.1	Count	P_Value	Benjamini
		Prion disease	13	6.20E-01	1.00E+00
		Pathways of neurodegeneration - multiple diseases	22	6.50E-01	1.00E+00
		Huntington disease	12	8.50E-01	1.00E+00
		Alzheimer disease	15	8.80E-01	1.00E+00
		Amyotrophic lateral sclerosis	14	9.00E-01	1.00E+00
		Parkinson disease	9	9.40E-01	1.00E+00
8	998	Enrichment Score: 0.42	Count	P_Value	Benjamini
		Alzheimer disease	23	1.70E-01	7.50E-01
		Prion disease	17	1.80E-01	7.50E-01
		Pathways of neurodegeneration - multiple diseases	24	4.30E-01	9.60E-01
		Amyotrophic lateral sclerosis	19	4.50E-01	9.60E-01
		Parkinson disease	12	6.90E-01	1.00E+00
		Huntington disease	13	7.50E-01	1.00E+00
9	1000	Enrichment Score: 1.05	Count	P_Value	Benjamini
		Huntington disease	21	7.80E-03	3.90E-01
		Amyotrophic lateral sclerosis	24	9.10E-03	3.90E-01
		Alzheimer disease	23	2.50E-02	6.10E-01

Table 7 continued from previous page

		Parkinson disease	16	6.20E-02	9.20E-01
		Prion disease	15	1.20E-01	9.70E-01
		Pathways of neurodegeneration - multiple diseases	22	2.10E-01	1.00E+00
		Enrichment Score: 0.43	Count	P_Value	Benjamini
		Pathways of neurodegeneration - multiple diseases	23	1.80E-01	1.00E+00
		Prion disease	14	2.20E-01	1.00E+00
12	1000	Parkinson disease	13	3.00E-01	1.00E+00
		Alzheimer disease	17	3.90E-01	1.00E+00
		Amyotrophic lateral sclerosis	14	6.50E-01	1.00E+00
		Huntington disease	10	8.20E-01	1.00E+00
		Enrichment Score: 1.31	Count	P_Value	Benjamini
		Amyotrophic lateral sclerosis	22	6.00E-03	6.50E-01
		Parkinson disease	15	3.80E-02	1.00E+00
13	1000	Prion disease	13	1.40E-01	1.00E+00
		Alzheimer disease	17	1.50E-01	1.00E+00
		Huntington disease	14	1.50E-01	1.00E+00
		Pathways of neurodegeneration - multiple diseases	20	1.60E-01	1.00E+00
		Enrichment Score: 0.18	Count	P_Value	Benjamini
		Spinocerebellar ataxia	5	5.00E-01	1.00E+00
		Alzheimer disease	11	5.40E-01	1.00E+00
14	1000	Prion disease	8	5.50E-01	1.00E+00
		Huntington disease	8	6.80E-01	1.00E+00
		Amyotrophic lateral sclerosis	9	7.50E-01	1.00E+00
		Parkinson disease	6	8.20E-01	1.00E+00
		Pathways of neurodegeneration - multiple diseases	10	8.70E-01	1.00E+00
		Enrichment Score: 0.44	Count	P_Value	Benjamini
		Amyotrophic lateral sclerosis	11	1.70E-01	1.00E+00
		Pathways of neurodegeneration - multiple diseases	12	2.90E-01	1.00E+00
15	1000	Alzheimer disease	10	3.10E-01	1.00E+00
		Parkinson disease	6	5.70E-01	1.00E+00
		Prion disease	6	5.90E-01	1.00E+00
		Huntington disease	5	8.40E-01	1.00E+00
		Enrichment Score: 0.2	Count	P_Value	Benjamini
		Spinocerebellar ataxia	6	1.70E-01	1.00E+00
		Huntington disease	7	6.00E-01	1.00E+00
16	999	Alzheimer disease	8	6.90E-01	1.00E+00
		Pathways of neurodegeneration - multiple diseases	9	7.70E-01	1.00E+00
		Amyotrophic lateral sclerosis	6	8.90E-01	1.00E+00
		Parkinson disease	4	9.20E-01	1.00E+00
		Prion disease	4	9.20E-01	1.00E+00
		Enrichment Score: 0.35	Count	P_Value	Benjamini
		Amyotrophic lateral sclerosis	17	2.30E-01	1.00E+00
17	999	Pathways of neurodegeneration - multiple diseases	18	5.00E-01	1.00E+00
		Parkinson disease	10	5.90E-01	1.00E+00
		Prion disease	10	6.10E-01	1.00E+00
		Alzheimer disease	12	8.00E-01	1.00E+00
		Enrichment Score: 0.64	Count	P_Value	Benjamini
18	1000	Alzheimer disease	23	2.40E-02	5.50E-01
		Parkinson disease	10	6.30E-01	1.00E+00
		Prion disease	8	8.70E-01	1.00E+00
		Enrichment Score: 0.59	Count	P_Value	Benjamini
		Huntington disease	15	7.90E-02	1.00E+00
		Amyotrophic lateral sclerosis	16	1.60E-01	1.00E+00
19	1000	Pathways of neurodegeneration - multiple diseases	19	1.90E-01	1.00E+00

Table 7 continued from previous page

		Alzheimer disease	13	5.00E-01	1.00E+00
		Parkinson disease	9	5.60E-01	1.00E+00
		Prion disease	9	5.70E-01	1.00E+00
20	1000	Enrichment Score: 2.36	Count	P_Value	Benjamini
		Pathways of neurodegeneration - multiple diseases	35	2.50E-04	3.30E-02
		Parkinson disease	21	2.70E-03	9.30E-02
		Alzheimer disease	27	3.20E-03	9.70E-02
		Amyotrophic lateral sclerosis	26	3.90E-03	1.00E-01
		Huntington disease	21	1.20E-02	1.80E-01
		Prion disease	19	1.50E-02	2.00E-01
		Spinocerebellar ataxia	12	1.90E-02	2.30E-01
21	1000	Enrichment Score: 0.63	Count	P_Value	Benjamini
		Spinocerebellar ataxia	14	5.40E-04	1.50E-01
		Amyotrophic lateral sclerosis	19	4.70E-02	1.00E+00
		Alzheimer disease	18	1.00E-01	1.00E+00
		Pathways of neurodegeneration - multiple diseases	20	1.70E-01	1.00E+00
		Huntington disease	13	2.50E-01	1.00E+00
		Prion disease	10	4.80E-01	1.00E+00
		Parkinson disease	8	7.40E-01	1.00E+00
22	1000	Enrichment Score: 0.52	Count	P_Value	Benjamini
		Huntington disease	14	7.60E-02	1.00E+00
		Spinocerebellar ataxia	7	2.00E-01	1.00E+00
		Parkinson disease	10	3.10E-01	1.00E+00
		Alzheimer disease	12	4.90E-01	1.00E+00
		Pathways of neurodegeneration - multiple diseases	14	5.50E-01	1.00E+00
		Prion disease	8	6.20E-01	1.00E+00
23	999	Enrichment Score: 0.32	Count	P_Value	Benjamini
		Neurotrophin signaling pathway	6	1.30E-01	1.00E+00
		Alzheimer disease	6	9.30E-01	1.00E+00
24	999	Enrichment Score: 0.79	Count	P_Value	Benjamini
		Alzheimer disease	9	3.70E-01	1.00E+00
25	999	Enrichment Score: 0.49	Count	P_Value	Benjamini
		Amyotrophic lateral sclerosis	11	1.40E-01	1.00E+00
		Pathways of neurodegeneration - multiple diseases	11	3.70E-01	1.00E+00
		Prion disease	7	3.70E-01	1.00E+00
		Alzheimer disease	8	5.70E-01	1.00E+00
26	999	Enrichment Score: 0.31	Count	P_Value	Benjamini
		Spinocerebellar ataxia	4	3.60E-01	1.00E+00
		Alzheimer disease	6	7.00E-01	1.00E+00
28	997	Enrichment Score: 0.28	Count	P_Value	Benjamini
		Alzheimer disease	4	7.80E-01	1.00E+00
29	1000	Enrichment Score: 0.27	Count	P_Value	Benjamini
		Parkinson disease	6	3.80E-01	1.00E+00
		Amyotrophic lateral sclerosis	7	5.00E-01	1.00E+00
		Alzheimer disease	7	5.40E-01	1.00E+00
		Pathways of neurodegeneration - multiple diseases	8	5.90E-01	1.00E+00
		Prion disease	5	5.90E-01	1.00E+00
		Huntington disease	5	6.90E-01	1.00E+00
30	1000	Enrichment Score: 3.35	Count	P_Value	Benjamini
		Parkinson disease	32	9.60E-06	3.00E-03
		Pathways of neurodegeneration - multiple diseases	46	2.50E-05	3.10E-03
		Prion disease	30	8.50E-05	6.70E-03

Table 7 continued from previous page

		Huntington disease	32	1.30E-04	8.20E-03
		Alzheimer disease	36	4.70E-04	1.90E-02
		Amyotrophic lateral sclerosis	35	4.90E-04	1.90E-02
31	1000	Enrichment Score: 0.27	Count	P_Value	Benjamini
		Pathways of neurodegeneration - multiple diseases	26	3.30E-01	1.00E+00
		Amyotrophic lateral sclerosis	20	4.10E-01	1.00E+00
		Prion disease	14	5.30E-01	1.00E+00
		Alzheimer disease	17	7.50E-01	1.00E+00
		Huntington disease	13	7.90E-01	1.00E+00
33	1000	Enrichment Score: 0.43	Count	P_Value	Benjamini
		Pathways of neurodegeneration - multiple diseases	27	2.00E-01	1.00E+00
		Alzheimer disease	22	2.40E-01	1.00E+00
		Amyotrophic lateral sclerosis	20	3.40E-01	1.00E+00
		Prion disease	14	4.70E-01	1.00E+00
		Parkinson disease	13	5.60E-01	1.00E+00
		Huntington disease	14	6.40E-01	1.00E+00
34	1000	Enrichment Score: 0.36	Count	P_Value	Benjamini
		Prion disease	14	2.60E-01	1.00E+00
		Pathways of neurodegeneration - multiple diseases	21	4.00E-01	1.00E+00
		Parkinson disease	12	4.70E-01	1.00E+00
		Alzheimer disease	16	5.50E-01	1.00E+00
		Amyotrophic lateral sclerosis	15	6.00E-01	1.00E+00
35	999	Enrichment Score: 0.56	Count	P_Value	Benjamini
		Prion disease	14	2.30E-02	4.80E-01
		Alzheimer disease	14	2.00E-01	8.60E-01
		Parkinson disease	10	2.50E-01	8.80E-01
		Pathways of neurodegeneration - multiple diseases	14	4.50E-01	9.90E-01
		Huntington disease	9	5.30E-01	9.90E-01
		Amyotrophic lateral sclerosis	10	6.30E-01	1.00E+00
36	1000	Enrichment Score: 0.4	Count	P_Value	Benjamini
		Alzheimer disease	10	3.00E-01	1.00E+00
		Amyotrophic lateral sclerosis	8	5.50E-01	1.00E+00
		Parkinson disease	6	5.60E-01	1.00E+00
		Prion disease	6	5.80E-01	1.00E+00
		Pathways of neurodegeneration - multiple diseases	9	6.70E-01	1.00E+00
		Huntington disease	6	6.80E-01	1.00E+00
37	999	Enrichment Score: 0.56	Count	P_Value	Benjamini
		Pathways of neurodegeneration - multiple diseases	24	1.30E-01	8.20E-01
		Parkinson disease	15	1.30E-01	8.20E-01
		Huntington disease	16	1.70E-01	8.80E-01
		Alzheimer disease	18	2.90E-01	9.90E-01
		Amyotrophic lateral sclerosis	14	6.50E-01	1.00E+00
		Prion disease	9	8.00E-01	1.00E+00
38	1000	Enrichment Score: 1.03	Count	P_Value	Benjamini
		Amyotrophic lateral sclerosis	22	9.80E-03	9.50E-01
		Huntington disease	18	2.10E-02	1.00E+00
		Alzheimer disease	20	4.70E-02	1.00E+00
		Parkinson disease	15	5.30E-02	1.00E+00
		Prion disease	14	1.00E-01	1.00E+00
		Pathways of neurodegeneration - multiple diseases	20	2.00E-01	1.00E+00
39	999	Enrichment Score: 0.66	Count	P_Value	Benjamini
		Amyotrophic lateral sclerosis	13	8.10E-02	9.50E-01
		Huntington disease	10	1.70E-01	1.00E+00
		Pathways of neurodegeneration - multiple diseases	14	1.80E-01	1.00E+00

Table 7 continued from previous page

		Alzheimer disease	5	9.60E-01	1.00E+00
40	1000	Enrichment Score: 0.7	Count	P_Value	Benjamini
		Spinocerebellar ataxia	8	3.80E-03	8.70E-01
		Alzheimer disease	12	1.90E-02	1.00E+00
		Neurotrophin signaling pathway	6	2.80E-02	1.00E+00
		Pathways of neurodegeneration - multiple diseases	13	3.40E-02	1.00E+00
		Huntington disease	9	6.40E-02	1.00E+00
		Parkinson disease	7	1.70E-01	1.00E+00
41	999	Enrichment Score: 0.98	Count	P_Value	Benjamini
		Pathways of neurodegeneration - multiple diseases	23	5.40E-02	4.80E-01
		Huntington disease	16	6.70E-02	5.10E-01
		Amyotrophic lateral sclerosis	15	3.10E-01	8.80E-01