Supplementary figures and tables of "Evaluating preprocessing and differential expression combinations for Affymetrix GeneChip microarrays via spike-in, RT-PCR and cross-laboratory datasets"

Ya-Li Wang[1], Guan-Hua Huang[1]*

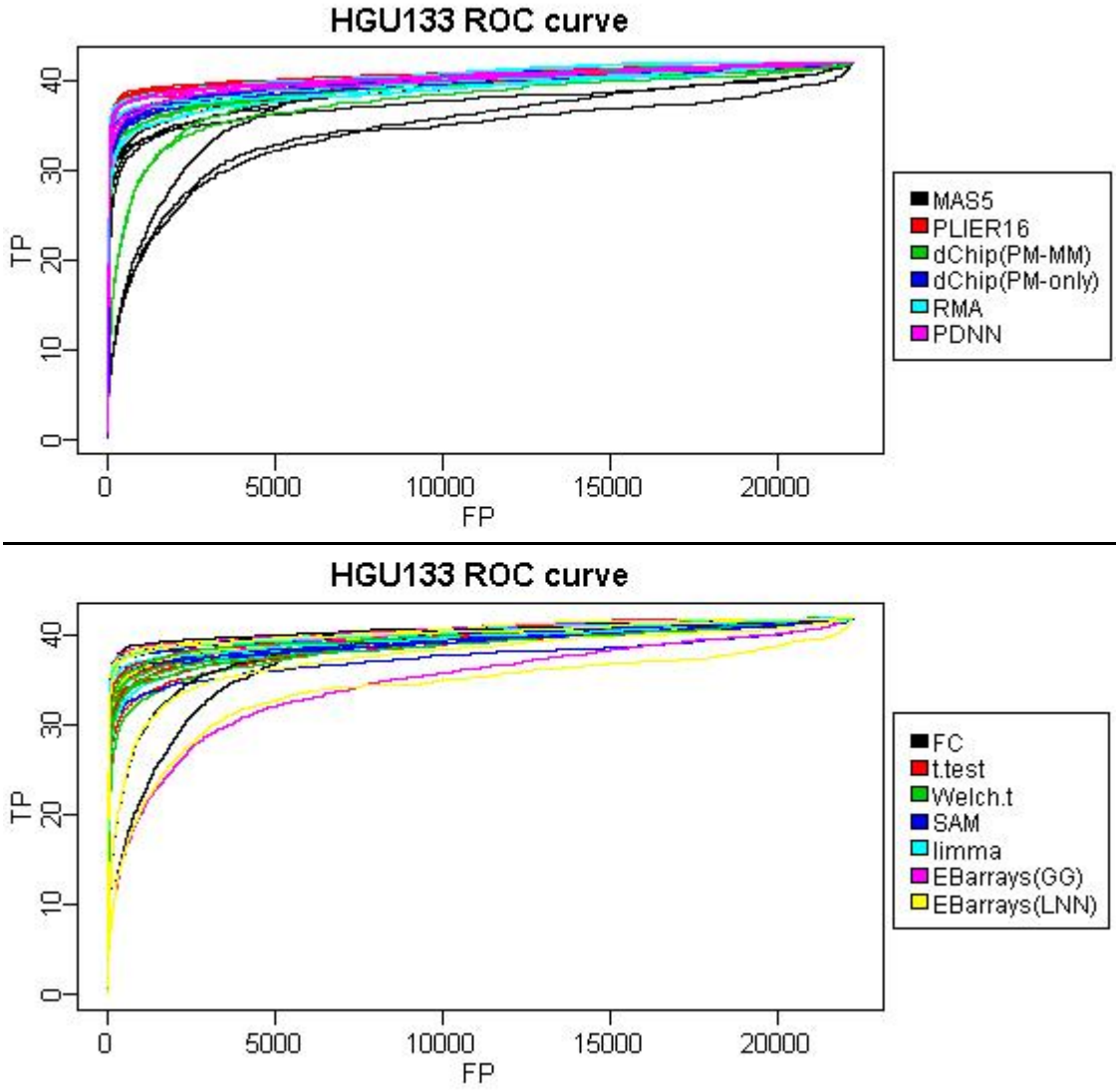
[1]Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan

*Corresponding author

Email addresses:

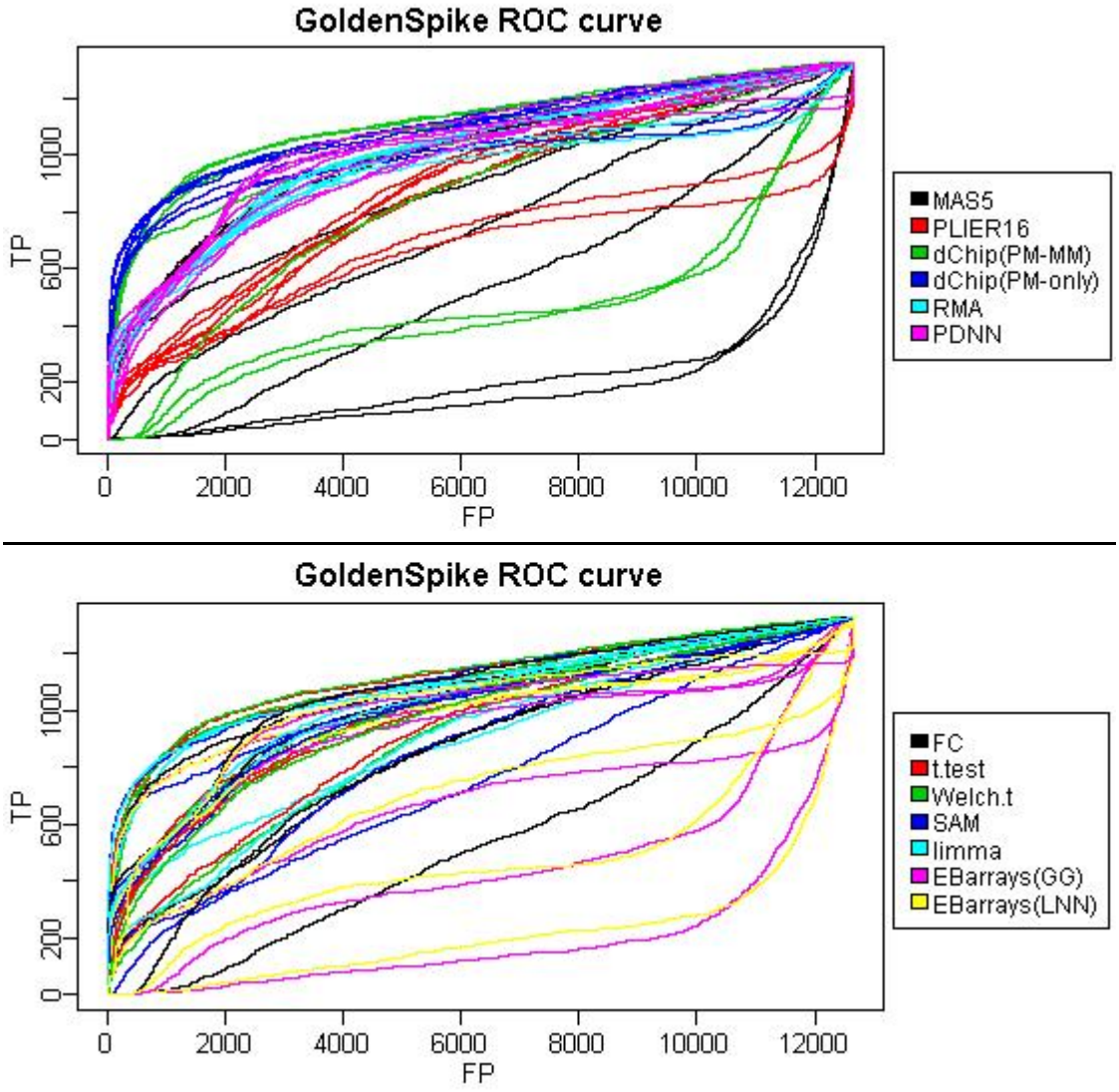
YLW: yali0629@hotmail.com

GHH: ghuang@stat.nctu.edu.tw



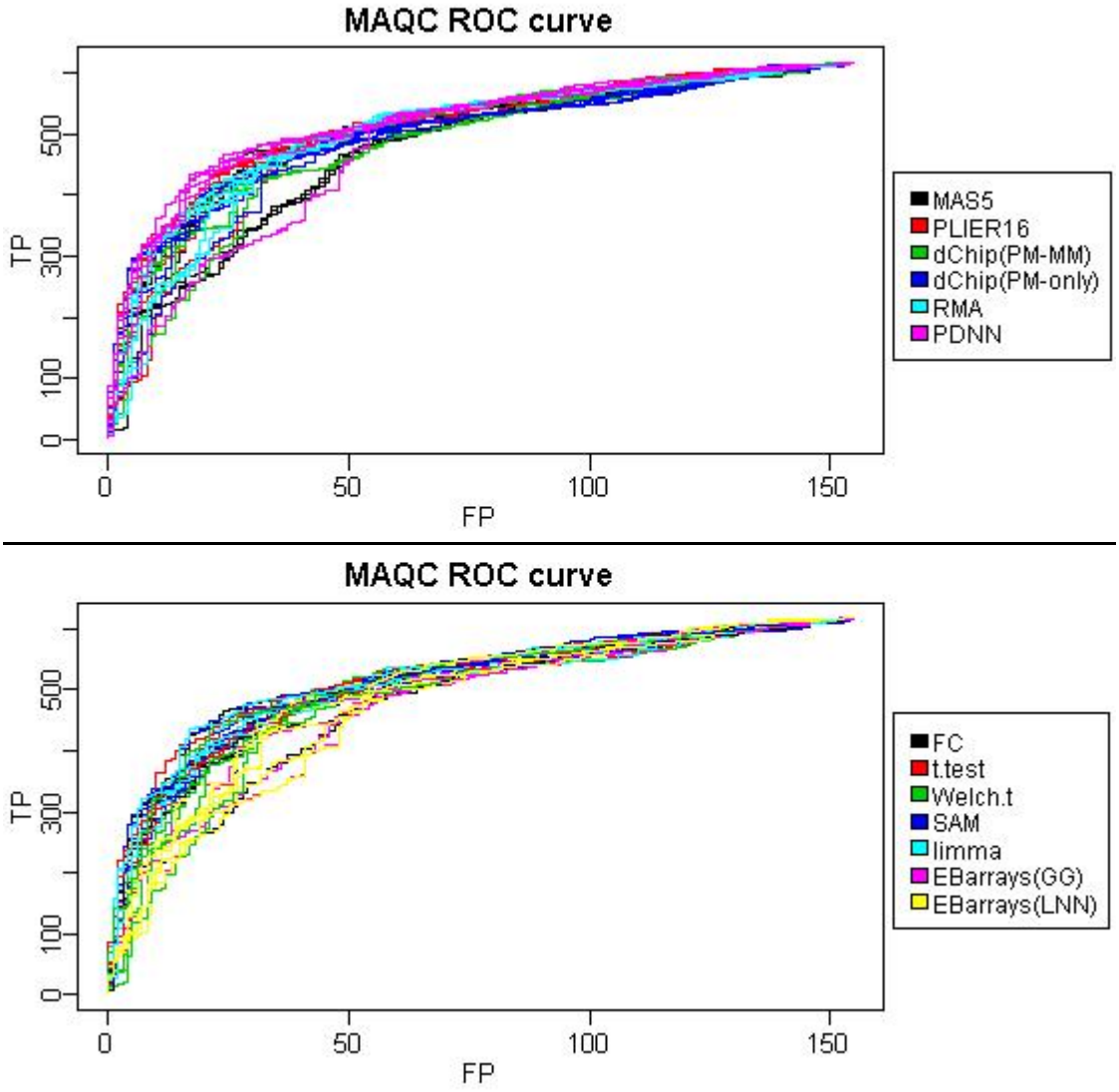
Supplementary figure 1 - ROC curves for the HGU133 Spike-in dataset, showing all FPs.

In the top panel, combinations using the same preprocessing method are assigned to the same color. In the bottom panel, combinations using the same differential expression method are assigned to the same color. There are 40 combinations in total.



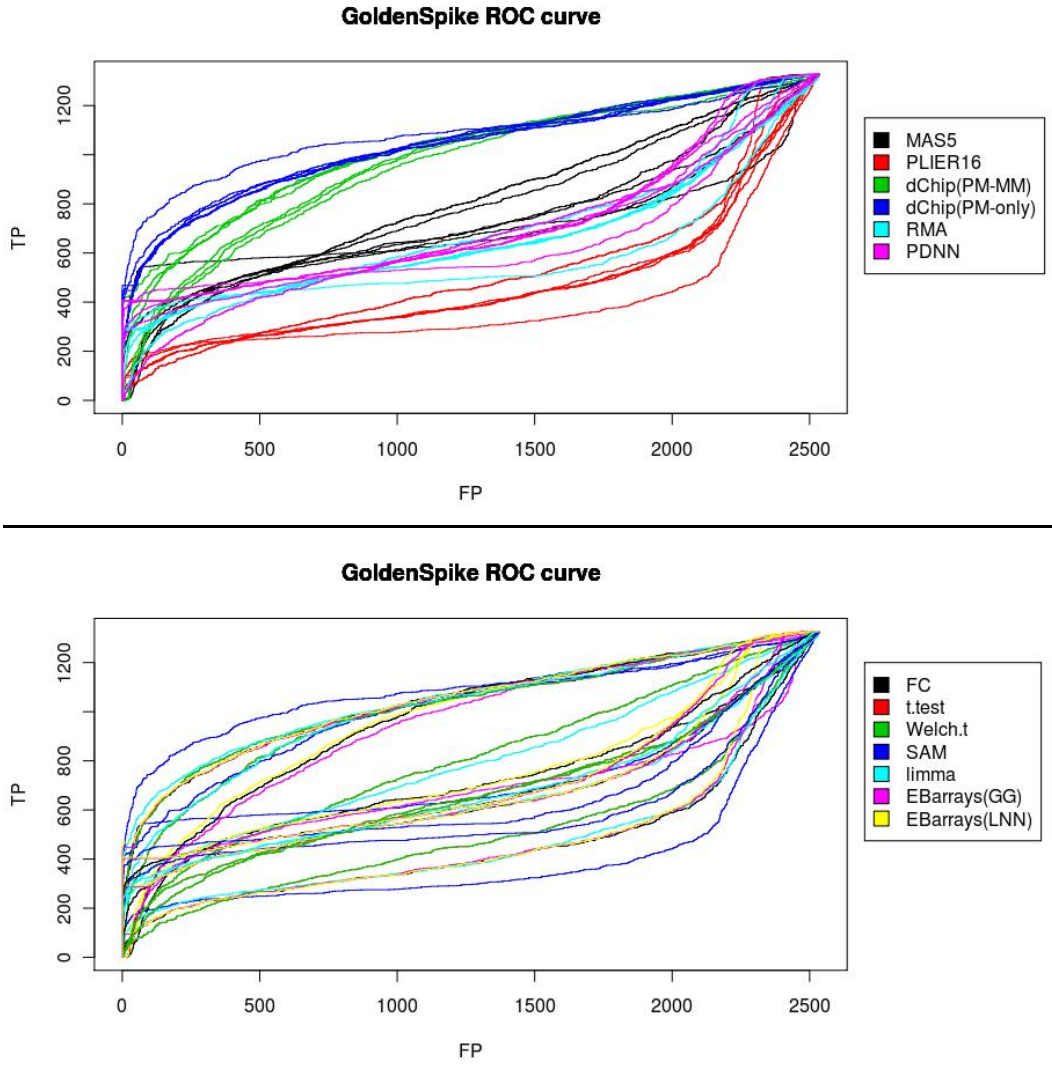
Supplementary figure 2 - ROC curves for the Golden Spike dataset, showing all FPs.

In the top panel, combinations using the same preprocessing method are assigned to the same color. In the bottom panel, combinations using the same differential expression method are assigned to the same color. There are 42 combinations in total.



Supplementary figure 3 - ROC curves for the MAQC RT-PCR dataset, showing all FPs.

In the top panel, combinations using the same preprocessing method are assigned to the same color. In the bottom panel, combinations using the same differential expression method are assigned to the same color. Genes that are significantly differentially expressed in TaqMan arrays are treated as the true positives. There are 42 combinations in total.



Supplementary figure 4 - ROC curves for the Golden Spike dataset using genes spiked-in at equal levels as true negatives, showing all FPs.
In the top panel, combinations using the same preprocessing method are assigned to the same color. In the bottom panel, combinations using the same differential expression method are assigned to the same color. Only genes that are spiked-in at equal levels are used as the true negatives. There are 42 combinations in total.

Supplementary table 1 - Area under ROC curve (FP<100) for the HGU133 Spike-in dataset

HGU133	Preprocessing	Differential expression	AUC (FP<100)
1	RMA	EBarrays(LNN)	0.775545
2	RMA	EBarrays(GG)	0.775209
3	RMA	FC	0.745778
4	PLIER16	EBarrays(GG)	0.737474
5	RMA	SAM	0.735601
6	PLIER16	SAM	0.730937
7	PLIER16	EBarrays(LNN)	0.728083
8	RMA	limma	0.724682
9	PLIER16	FC	0.720157
10	PLIER16	limma	0.717157
11	PDNN	SAM	0.715714
12	PDNN	limma	0.710198
13	dChip(PM-only)	SAM	0.678571
14	PDNN	FC	0.670846
15	PDNN	EBarrays(LNN)	0.670143
16	dChip(PM-only)	limma	0.666364
17	PDNN	t.test	0.643946
18	dChip(PM-only)	t.test	0.639840
19	dChip(PM-MM)	SAM	0.628897
20	dChip(PM-MM)	limma	0.620192
21	RMA	t.test	0.607460
22	dChip(PM-MM)	t.test	0.600749
23	PLIER16	t.test	0.593802
24	PDNN	Welch.t	0.588472
25	dChip(PM-only)	Welch.t	0.573446
26	RMA	Welch.t	0.556811
27	MAS5	SAM	0.554397
28	dChip(PM-only)	EBarrays(GG)	0.549018
29	dChip(PM-MM)	Welch.t	0.542362
30	dChip(PM-only)	FC	0.537517
31	dChip(PM-only)	EBarrays(LNN)	0.535917
32	PLIER16	Welch.t	0.533251
33	MAS5	limma	0.533016
34	MAS5	t.test	0.516221
35	MAS5	Welch.t	0.445383
36	dChip(PM-MM)	EBarrays(LNN)	0.217948
37	dChip(PM-MM)	FC	0.217485
38	MAS5	EBarrays(GG)	0.155624
39	MAS5	FC	0.154372
40	MAS5	EBarrays(LNN)	0.153766

Supplementary table 2 - Area under ROC curve (FPR<0.1) for the Golden Spike dataset

Golden S	Preprocessing	Differential expression	AUC (FPR<0.1)
1	dChip(PM-only)	limma	0.566140
2	dChip(PM-only)	SAM	0.565670
3	dChip(PM-only)	t.test	0.549932
4	dChip(PM-only)	Welch.t	0.537779
5	dChip(PM-MM)	t.test	0.523752
6	dChip(PM-MM)	Welch.t	0.515406
7	dChip(PM-only)	FC	0.510864
8	dChip(PM-MM)	limma	0.504039
9	dChip(PM-only)	EBarrays(GG)	0.499712
10	dChip(PM-only)	EBarrays(LNN)	0.496237
11	dChip(PM-MM)	SAM	0.484404
12	PDNN	FC	0.366185
13	PDNN	limma	0.346782
14	RMA	limma	0.338811
15	RMA	SAM	0.337653
16	MAS5	t.test	0.335464
17	RMA	FC	0.334335
18	RMA	EBarrays(GG)	0.329024
19	RMA	EBarrays(LNN)	0.327159
20	PDNN	SAM	0.323364
21	MAS5	Welch.t	0.315065
22	PDNN	EBarrays(LNN)	0.312910
23	PDNN	EBarrays(GG)	0.312688
24	RMA	t.test	0.307614
25	RMA	Welch.t	0.295380
26	PDNN	t.test	0.292734
27	MAS5	limma	0.282221
28	PDNN	Welch.t	0.260818
29	PLIER16	limma	0.195952
30	PLIER16	t.test	0.192417
31	PLIER16	FC	0.179153
32	PLIER16	SAM	0.174857
33	PLIER16	EBarrays(GG)	0.166193
34	PLIER16	Welch.t	0.166022
35	PLIER16	EBarrays(LNN)	0.164703
36	MAS5	SAM	0.108998
37	dChip(PM-MM)	FC	0.059322
38	dChip(PM-MM)	EBarrays(LNN)	0.034430
39	dChip(PM-MM)	EBarrays(GG)	0.016426
40	MAS5	FC	0.006431
41	MAS5	EBarrays(LNN)	0.004667
42	MAS5	EBarrays(GG)	0.004140

Supplementary table 3 - Area under ROC curve (FP<50) for the MAQC RT-PCR dataset

MAQC	Preprocessing	Differential expression	AUC (FP<50)
1	PDNN	limma	0.650065
2	PDNN	SAM	0.649417
3	PDNN	t.test	0.648641
4	PDNN	FC	0.648058
5	PLIER16	SAM	0.632718
6	PLIER16	limma	0.631974
7	PLIER16	t.test	0.627605
8	PDNN	Welch.t	0.619159
9	PLIER16	FC	0.614401
10	MAS5	SAM	0.610971
11	dChip(PM-only)	limma	0.60835
12	MAS5	limma	0.60754
13	dChip(PM-only)	SAM	0.605922
14	dChip(PM-only)	t.test	0.605275
15	RMA	SAM	0.60301
16	RMA	limma	0.601327
17	RMA	FC	0.600712
18	MAS5	t.test	0.596408
19	dChip(PM-only)	FC	0.594563
20	dChip(PM-MM)	SAM	0.592783
21	dChip(PM-MM)	limma	0.588964
22	RMA	t.test	0.587411
23	dChip(PM-MM)	t.test	0.584045
24	PLIER16	Welch.t	0.579061
25	dChip(PM-only)	Welch.t	0.565437
26	dChip(PM-MM)	FC	0.555469
27	MAS5	Welch.t	0.555275
28	RMA	EBarrays(LNN)	0.542977
29	RMA	EBarrays(GG)	0.54165
30	RMA	Welch.t	0.536958
31	dChip(PM-MM)	EBarrays(GG)	0.534207
32	dChip(PM-MM)	EBarrays(LNN)	0.531197
33	dChip(PM-only)	EBarrays(LNN)	0.509482
34	dChip(PM-only)	EBarrays(GG)	0.509385
35	PLIER16	EBarrays(LNN)	0.506505
36	PLIER16	EBarrays(GG)	0.50644
37	dChip(PM-MM)	Welch.t	0.489741
38	MAS5	EBarrays(GG)	0.479061
39	MAS5	EBarrays(LNN)	0.472006
40	MAS5	FC	0.470259
41	PDNN	EBarrays(GG)	0.435113
42	PDNN	EBarrays(LNN)	0.433883

Supplementary table 4 - Area under ROC curve (all FPs) for the Golden Spike dataset using genes spiked-in at equal levels as true negatives

Golden S	Preprocessing	Differential expression	AUC
1	dChip(PM-only)	SAM	0.810309
2	dChip(PM-only)	limma	0.785189
3	dChip(PM-only)	EBarrays(LNN)	0.784673
4	dChip(PM-only)	EBarrays(GG)	0.781759
5	dChip(PM-only)	FC	0.781732
6	dChip(PM-only)	t.test	0.776872
7	dChip(PM-only)	welch.t	0.776872
8	dChip(PM-MM)	limma	0.766347
9	dChip(PM-MM)	t.test	0.763207
10	dChip(PM-MM)	welch.t	0.763207
11	dChip(PM-MM)	SAM	0.758200
12	dChip(PM-MM)	EBarrays(LNN)	0.731621
13	dChip(PM-MM)	FC	0.730505
14	dChip(PM-MM)	EBarrays(GG)	0.716040
15	MAS5	t.test	0.603740
16	MAS5	welch.t	0.603740
17	MAS5	limma	0.600205
18	MAS5	EBarrays(LNN)	0.560834
19	MAS5	SAM	0.548898
20	PDNN	EBarrays(LNN)	0.541970
21	MAS5	FC	0.539195
22	PDNN	EBarrays(GG)	0.538945
23	PDNN	FC	0.528292
24	PDNN	limma	0.515343
25	RMA	EBarrays(LNN)	0.509057
26	MAS5	EBarrays(GG)	0.507160
27	RMA	welch.t	0.503501
28	RMA	t.test	0.503501
29	RMA	EBarrays(GG)	0.500148
30	PDNN	welch.t	0.499310
31	PDNN	t.test	0.499310
32	RMA	limma	0.494678
33	RMA	FC	0.492670
34	PDNN	SAM	0.490104
35	RMA	SAM	0.439609
36	PLIER16	welch.t	0.384444
37	PLIER16	t.test	0.384444
38	PLIER16	EBarrays(LNN)	0.365771
39	PLIER16	EBarrays(GG)	0.360097
40	PLIER16	limma	0.357202
41	PLIER16	FC	0.348342
42	PLIER16	SAM	0.300208

Supplementary table 5 - Average log2-transformed signal intensity for various benchmark datasets preprocessed with different methods

		Benchmark dataset	HGU133 Spike-in	Golden Spike	MAQC RT-PCR
Pre-processing methods	MAS5	Positives[a]	9.58	11.01	8.95
		Negatives[a]	7.28	5.34	7.30
	PLIER16	Positives	7.74	10.99	7.66
		Negatives	6.35	7.08	6.61
	dChip(PM-MM)	Positives	7.72	10.95	7.57
		Negatives	5.59	5.73	6.07
	dChip(PM only)	Positives	8.60	11.52	8.36
		Negatives	7.31	8.61	7.44
	RMA	Positives	7.94	11.25	7.71
		Negatives	6.05	7.55	6.45
	PDNN	Positives	9.32	15.28	11.46
		Negatives	7.94	12.63	10.62

[a] Positives=genes that were treated as truly differentially expressed; Negatives=genes that were treated as truly non-differentially expressed.