

## **Supplementary material to “An efficient method to identify differentially expressed genes in microarray experiments”**

Huaizhen Qin<sup>1</sup>, Tao Feng<sup>1,3</sup>, Scott A. Harding<sup>2</sup>, Chung-Jui Tsai<sup>2</sup> and Shuanglin Zhang<sup>1,3\*</sup>

<sup>1</sup>Department of Mathematical Sciences and <sup>2</sup>Biotechnology Research Center, School of Forest Resources and Environmental Science, Michigan Technological University, Houghton, MI, 49931, USA and <sup>3</sup>Department of Mathematics, Heilongjiang University, Harbin 150080, China

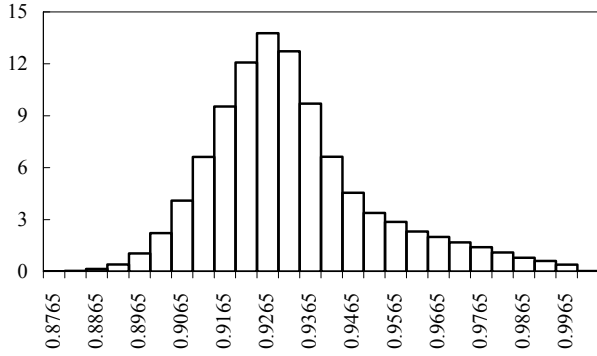
---

**Table S1.** FDR's of BH95 and FCPC under random dependency

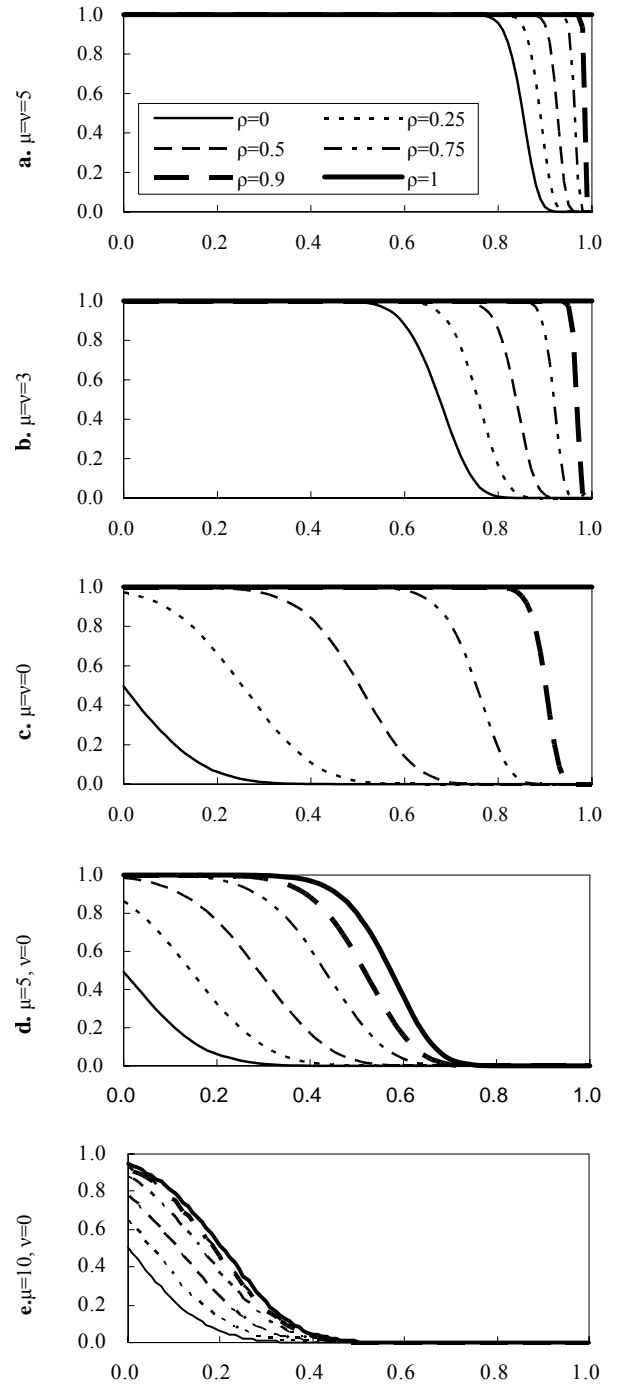
Method	Scenario		Number of genes (in thousand)									
	$m$	$\rho$	1	2	3	4	5	6	7	8	9	10
BH95	1	0.25	0.030	0.032	0.034	0.036	0.034	0.036	0.037	0.045	0.033	0.037
FCPC			0.016	0.021	0.028	0.035	0.038	0.042	0.048	0.058	0.046	0.055
BH95		0.5	0.028	0.028	0.029	0.030	0.034	0.033	0.038	0.045	0.032	0.038
FCPC			0.015	0.020	0.030	0.031	0.031	0.034	0.042	0.045	0.041	0.044
BH95		0.75	0.025	0.025	0.025	0.026	0.028	0.027	0.030	0.040	0.029	0.030
FCPC			0.017	0.019	0.025	0.028	0.031	0.033	0.036	0.049	0.036	0.048
BH95		0.9	0.023	0.023	0.022	0.022	0.023	0.025	0.025	0.032	0.028	0.030
FCPC			0.020	0.026	0.025	0.032	0.029	0.035	0.039	0.045	0.044	0.046
BH95	20	0.25	0.033	0.033	0.033	0.032	0.035	0.036	0.034	0.032	0.040	0.043
FCPC			0.016	0.021	0.026	0.032	0.037	0.046	0.043	0.044	0.054	0.057
BH95		0.5	0.030	0.030	0.030	0.027	0.032	0.040	0.032	0.031	0.036	0.045
FCPC			0.017	0.023	0.024	0.032	0.035	0.041	0.034	0.041	0.046	0.053
BH95		0.75	0.026	0.025	0.027	0.025	0.034	0.032	0.026	0.025	0.034	0.035
FCPC			0.019	0.021	0.029	0.029	0.035	0.038	0.036	0.037	0.039	0.043
BH95		0.9	0.021	0.023	0.023	0.020	0.027	0.028	0.017	0.023	0.026	0.035
FCPC			0.026	0.028	0.031	0.027	0.040	0.041	0.036	0.040	0.040	0.050
BH95	50	0.25	0.031	0.030	0.036	0.036	0.032	0.033	0.035	0.044	0.035	0.034
FCPC			0.019	0.020	0.035	0.037	0.037	0.040	0.045	0.062	0.051	0.051
BH95		0.5	0.029	0.025	0.032	0.035	0.029	0.032	0.035	0.048	0.034	0.029
FCPC			0.019	0.017	0.031	0.036	0.034	0.038	0.038	0.051	0.040	0.043
BH95		0.75	0.023	0.020	0.028	0.030	0.024	0.025	0.030	0.044	0.029	0.024
FCPC			0.023	0.018	0.049	0.034	0.036	0.034	0.035	0.051	0.035	0.030
BH95		0.9	0.019	0.018	0.024	0.023	0.020	0.021	0.030	0.041	0.025	0.019
FCPC			0.025	0.024	0.035	0.037	0.037	0.034	0.046	0.054	0.037	0.031
BH95	100	0.25	0.032	0.030	0.036	0.033	0.033	0.032	0.037	0.037	0.040	0.033
FCPC			0.018	0.022	0.032	0.035	0.034	0.035	0.044	0.058	0.049	0.046
BH95		0.5	0.030	0.028	0.035	0.026	0.031	0.028	0.036	0.038	0.038	0.029
FCPC			0.020	0.023	0.032	0.028	0.033	0.031	0.043	0.051	0.048	0.043
BH95		0.75	0.027	0.027	0.030	0.024	0.023	0.025	0.034	0.035	0.032	0.028
FCPC			0.022	0.025	0.029	0.026	0.032	0.027	0.039	0.042	0.047	0.041
BH95		0.9	0.022	0.024	0.023	0.019	0.017	0.023	0.026	0.033	0.030	0.027
FCPC			0.029	0.034	0.034	0.030	0.029	0.031	0.043	0.047	0.052	0.044

**Table S2.** Powers of BH95 and FCPC under random dependency

Method	Scenario		Number of genes (in thousand)									
	$m$	$\rho$	1	2	3	4	5	6	7	8	9	10
BH95	1	0.25	0.829	0.682	0.593	0.482	0.430	0.407	0.356	0.306	0.281	0.259
FCPC			0.859	0.829	0.799	0.783	0.783	0.766	0.757	0.740	0.750	0.731
BH95		0.5	0.825	0.679	0.596	0.487	0.436	0.427	0.381	0.330	0.304	0.280
FCPC			0.869	0.839	0.812	0.785	0.787	0.771	0.757	0.734	0.746	0.724
BH95		0.75	0.821	0.679	0.606	0.494	0.449	0.444	0.410	0.360	0.336	0.315
FCPC			0.888	0.863	0.844	0.820	0.814	0.803	0.798	0.785	0.780	0.769
BH95		0.9	0.822	0.686	0.617	0.513	0.467	0.470	0.435	0.387	0.358	0.356
FCPC			0.921	0.911	0.897	0.891	0.884	0.871	0.876	0.872	0.858	0.846
BH95	20	0.25	0.830	0.695	0.590	0.503	0.413	0.407	0.347	0.316	0.272	0.247
FCPC			0.862	0.823	0.806	0.775	0.760	0.773	0.749	0.759	0.734	0.730
BH95		0.5	0.828	0.694	0.592	0.508	0.424	0.426	0.369	0.334	0.294	0.270
FCPC			0.867	0.835	0.813	0.785	0.767	0.769	0.759	0.753	0.729	0.735
BH95		0.75	0.826	0.695	0.596	0.523	0.439	0.449	0.393	0.359	0.324	0.313
FCPC			0.892	0.866	0.840	0.820	0.811	0.803	0.798	0.797	0.777	0.777
BH95		0.9	0.826	0.700	0.609	0.541	0.461	0.471	0.418	0.391	0.361	0.343
FCPC			0.929	0.908	0.900	0.883	0.878	0.874	0.875	0.871	0.873	0.853
BH95	50	0.25	0.824	0.684	0.574	0.493	0.442	0.399	0.333	0.309	0.286	0.258
FCPC			0.868	0.836	0.811	0.795	0.777	0.767	0.754	0.768	0.749	0.710
BH95		0.5	0.820	0.681	0.579	0.498	0.455	0.414	0.352	0.330	0.311	0.282
FCPC			0.878	0.846	0.819	0.791	0.784	0.774	0.749	0.761	0.754	0.718
BH95		0.75	0.815	0.679	0.597	0.512	0.471	0.436	0.379	0.359	0.334	0.310
FCPC			0.902	0.875	0.853	0.828	0.827	0.816	0.795	0.810	0.797	0.769
BH95		0.9	0.811	0.687	0.599	0.532	0.491	0.457	0.404	0.386	0.368	0.340
FCPC			0.935	0.927	0.904	0.895	0.888	0.882	0.881	0.877	0.881	0.867
BH95	100	0.25	0.825	0.675	0.584	0.506	0.437	0.381	0.338	0.308	0.280	0.238
FCPC			0.870	0.826	0.799	0.789	0.786	0.767	0.749	0.735	0.736	0.720
BH95		0.5	0.822	0.673	0.585	0.511	0.450	0.398	0.357	0.333	0.302	0.266
FCPC			0.875	0.836	0.801	0.777	0.787	0.767	0.748	0.720	0.748	0.673
BH95		0.75	0.822	0.677	0.594	0.526	0.467	0.420	0.381	0.365	0.333	0.299
FCPC			0.900	0.864	0.842	0.819	0.823	0.813	0.789	0.768	0.772	0.734
BH95		0.9	0.819	0.691	0.606	0.545	0.488	0.442	0.405	0.404	0.370	0.335
FCPC			0.939	0.916	0.902	0.892	0.894	0.892	0.876	0.877	0.879	0.861



**Fig. S1.** Horizontal axis:  $\gamma$ , the proportion of total variance explained by the first PC. Vertical axis: The count (in thousands) of  $\gamma$ -values. This histogram is based on the  $\gamma$ -values of 100221 tight clusters of level  $\geq 0.8$  in Table 1, which are produced by 1000 simulated replications of the scenario where  $\underline{m}=1$ ,  $\underline{\rho}=0.5$ , and  $M_t=5000$  in Section 3.1. In addition to the characteristics listed in Table 1, 98197 (97.98%) of the 100221  $\gamma$ -values are larger than 0.90. In other words, the first PC of each of the 98197 clusters can explain more than 90% of its total variance. Together with Table 1, the first PC of a tight cluster of level 0.8 can explain typically more than 90% of its total variance.



**Fig. S2.** Each curve is based on a gene pair expressed in 20 controls and 40 treatments. For a control, the expression of the pair is sampled from the bivariate normal distribution with mean zero, variance 1, and correlation  $\rho$ , and for a treatment, the expression of the pair is sampled from the bivariate normal distribution with mean  $(\mu, \nu)'$ , variance 1, and correlation  $\rho$ . Horizontal axis:  $\rho_0$ . Vertical axis:  $\Pr(\hat{\rho} > \rho_0)$ , where  $\hat{\rho}$  is gene-to-gene correlation.