Additional File 2

(Over-optimism in bioinformatics: an illustration)

1 Prediction error rates

In the following, we give an overview of the absolute error rates obtained for the methods rlda.TD, rlda.TF, rlda.TG including the ten variants rlda.TG$^{(1)}$, . . . , rlda.TG$^{(10)}$, DLDA, NSC and SVM. In addition to the error rate, both the sensitivity and the specificity are reported.

1.1 Golub data

Method	p	10 × five-fold CV error			Sensitivity			Specificity		
		t-test	Limma	Wilcoxon test	t-test	Limma	Wilcoxon test	t-test	Limma	Wilcoxon test
rlda.TD	100	0.028	0.029	0.103	0.960	0.960	0.956	0.979	0.977	0.867
rlda.TF	100	0.036	0.037	0.105	0.960	0.960	0.928	0.966	0.964	0.877
rlda.TG	100	0.029	0.031	0.090	0.956	0.956	0.952	0.979	0.977	0.887
dlda	100	0.044	0.046	0.118	0.880	0.872	0.992	0.998	0.998	0.824
rlda.TD	200	0.028	0.028	0.065	0.960	0.960	0.956	0.979	0.979	0.924
rlda.TF	200	0.028	0.028	0.089	0.960	0.960	0.916	0.979	0.979	0.909
rlda.TG	200	0.029	0.028	0.170	0.956	0.960	0.860	0.979	0.979	0.814
dlda	200	0.038	0.035	0.092	0.900	0.904	1.000	0.996	0.998	0.860
rlda.TD	500	0.032	0.031	0.046	0.948	0.952	0.960	0.979	0.979	0.951
rlda.TF	500	0.031	0.031	0.065	0.952	0.952	0.952	0.979	0.979	0.926
rlda.TG	500	0.032	0.039	0.168	0.944	0.936	0.824	0.981	0.975	0.836
dlda	500	0.031	0.036	0.112	0.916	0.916	0.992	0.998	0.990	0.833
rlda.TD	1000	0.031	0.032	0.057	0.952	0.948	0.940	0.979	0.979	0.944
rlda.TF	1000	0.028	0.029	0.064	0.956	0.952	0.952	0.981	0.981	0.928
rlda.TG	1000	0.049	0.060	0.124	0.928	0.912	0.856	0.964	0.955	0.887
dlda	1000	0.029	0.025	0.133	0.920	0.928	0.964	0.998	1.000	0.816
nsc	7 129	0.021	0.021	0.021	0.940	0.940	0.940	1.000	1.000	1.000
svm	7 129	0.054	0.054	0.054	0.952	0.952	0.952	0.942	0.942	0.942

Table 1: Overview of the 10 × five-fold CV errors obtained for the different variants of RLDA (rlda.TD, rlda.TF, rlda.TG) as well as for the methods DLDA, NSC and SVM. For each method except NSC and SVM the top 100, 200, 500 and 1000 genes of the two-class data **Golub** (n=72) are employed. The results of three variable selection methods are compared. Correspondingly, both the sensitivity and the specificity are obtained.

Method	p	10 × five-fold CV error			Sensitivity			Specificity		
		t-test	Limma	Wilcoxon test	t-test	Limma	Wilcoxon test	t-test	Limma	Wilcoxon test
rlda.TG$^{(1)}$	100	0.030	0.029	0.252	0.956	0.956	0.792	0.977	0.979	0.726
rlda.TG$^{(2)}$	100	0.032	0.033	0.100	0.944	0.940	0.924	0.981	0.981	0.887
rlda.TG$^{(3)}$	100	0.028	0.029	0.104	0.960	0.960	0.948	0.979	0.977	0.869
rlda.TG$^{(4)}$	100	0.028	0.029	0.110	0.960	0.960	0.952	0.979	0.977	0.858
rlda.TG$^{(5)}$	100	0.030	0.032	0.084	0.928	0.924	0.936	0.992	0.992	0.905
rlda.TG$^{(6)}$	100	0.030	0.029	0.077	0.956	0.956	0.968	0.977	0.979	0.899
rlda.TG$^{(7)}$	100	0.032	0.029	0.079	0.956	0.960	0.968	0.975	0.977	0.897
rlda.TG$^{(8)}$	100	0.033	0.033	0.283	0.944	0.948	0.744	0.979	0.977	0.702
rlda.TG$^{(9)}$	100	0.033	0.033	0.276	0.956	0.956	0.740	0.972	0.972	0.716
rlda.TG$^{(10)}$	100	0.028	0.029	0.093	0.960	0.960	0.956	0.979	0.977	0.880
rlda.TG$^{(1)}$	200	0.029	0.030	0.258	0.956	0.956	0.768	0.979	0.977	0.727
rlda.TG$^{(2)}$	200	0.029	0.028	0.069	0.956	0.960	0.944	0.979	0.979	0.924
rlda.TG$^{(3)}$	200	0.029	0.028	0.061	0.956	0.960	0.956	0.979	0.979	0.930
rlda.TG$^{(4)}$	200	0.028	0.028	0.072	0.960	0.960	0.948	0.979	0.979	0.917
rlda.TG$^{(5)}$	200	0.021	0.025	0.079	0.944	0.936	0.940	0.998	0.996	0.911
rlda.TG$^{(6)}$	200	0.029	0.032	0.072	0.956	0.948	0.976	0.979	0.979	0.903
rlda.TG$^{(7)}$	200	0.031	0.032	0.079	0.952	0.948	0.968	0.979	0.979	0.897
rlda.TG$^{(8)}$	200	0.096	0.117	0.180	0.868	0.844	0.896	0.925	0.906	0.780
rlda.TG$^{(9)}$	200	0.059	0.062	0.165	0.900	0.896	0.864	0.962	0.960	0.819
rlda.TG$^{(10)}$	200	0.029	0.029	0.121	0.956	0.956	0.916	0.979	0.979	0.860
rlda.TG$^{(1)}$	500	0.059	0.068	0.298	0.916	0.880	0.660	0.955	0.960	0.725
rlda.TG$^{(2)}$	500	0.026	0.026	0.058	0.944	0.944	0.932	0.990	0.990	0.947
rlda.TG$^{(3)}$	500	0.032	0.031	0.042	0.948	0.952	0.964	0.979	0.979	0.955
rlda.TG$^{(4)}$	500	0.032	0.031	0.047	0.948	0.952	0.956	0.979	0.979	0.951
rlda.TG$^{(5)}$	500	0.018	0.019	0.077	0.952	0.952	0.892	0.998	0.996	0.939
rlda.TG$^{(6)}$	500	0.031	0.031	0.058	0.952	0.952	0.960	0.979	0.979	0.933
rlda.TG$^{(7)}$	500	0.031	0.031	0.066	0.952	0.952	0.968	0.979	0.979	0.916
rlda.TG$^{(8)}$	500	0.220	0.173	0.317	0.740	0.772	0.712	0.803	0.857	0.668
rlda.TG$^{(9)}$	500	0.113	0.078	0.215	0.852	0.880	0.796	0.906	0.944	0.780
rlda.TG$^{(10)}$	500	0.032	0.031	0.203	0.948	0.952	0.812	0.979	0.979	0.788
rlda.TG$^{(1)}$	1000	0.096	0.097	0.248	0.892	0.872	0.744	0.912	0.920	0.757
rlda.TG$^{(2)}$	1000	0.024	0.025	0.056	0.952	0.948	0.928	0.990	0.990	0.953
rlda.TG$^{(3)}$	1000	0.031	0.032	0.049	0.952	0.948	0.948	0.979	0.979	0.953
rlda.TG$^{(4)}$	1000	0.031	0.032	0.054	0.952	0.948	0.944	0.979	0.979	0.947
rlda.TG$^{(5)}$	1000	0.026	0.026	0.062	0.944	0.944	0.896	0.990	0.990	0.960
rlda.TG$^{(6)}$	1000	0.025	0.025	0.050	0.964	0.964	0.964	0.981	0.981	0.943
rlda.TG$^{(7)}$	1000	0.025	0.024	0.065	0.964	0.968	0.960	0.981	0.981	0.922
rlda.TG$^{(8)}$	1000	0.219	0.190	0.301	0.748	0.772	0.692	0.799	0.831	0.705
rlda.TG$^{(9)}$	1000	0.158	0.113	0.242	0.820	0.844	0.740	0.854	0.910	0.768
rlda.TG$^{(10)}$	1000	0.032	0.031	0.176	0.948	0.952	0.824	0.979	0.979	0.824

Table 2: Overview of the 10 × five-fold CV errors obtained for the different variants of rlda.TG. For each method the top 100, 200, 500 and 1000 genes of the two-class data **Golub** (n=72) are employed. The results of three variable selection methods are compared. Correspondingly, both the sensitivity and the specificity are obtained.

1.2 CLL data

Method	p	10 × five-fold CV error			Sensitivity			Specificity		
		t-test	Limma	Wilcoxon test	t-test	Limma	Wilcoxon test	t-test	Limma	Wilcoxon test
rlda.TD	100	0.255	0.251	0.200	0.450	0.450	0.620	0.907	0.913	0.900
rlda.TF	100	0.254	0.246	0.237	0.450	0.450	0.580	0.907	0.920	0.870
rlda.TG	100	0.234	0.237	0.192	0.510	0.490	0.630	0.907	0.913	0.907
dlda	100	0.235	0.228	0.223	0.480	0.500	0.590	0.920	0.920	0.883
rlda.TD	200	0.266	0.256	0.146	0.420	0.420	0.700	0.907	0.920	0.940
rlda.TF	200	0.250	0.248	0.212	0.440	0.440	0.630	0.920	0.920	0.883
rlda.TG	200	0.269	0.274	0.159	0.430	0.430	0.690	0.897	0.887	0.930
dlda	200	0.231	0.236	0.184	0.480	0.480	0.660	0.927	0.920	0.903
rlda.TD	500	0.228	0.216	0.146	0.470	0.470	0.660	0.940	0.960	0.960
rlda.TF	500	0.220	0.212	0.192	0.480	0.480	0.630	0.947	0.960	0.913
rlda.TG	500	0.260	0.233	0.185	0.450	0.460	0.610	0.897	0.940	0.923
dlda	500	0.220	0.250	0.223	0.490	0.450	0.610	0.940	0.910	0.863
rlda.TD	1000	0.216	0.211	0.154	0.480	0.480	0.630	0.953	0.960	0.967
rlda.TF	1000	0.199	0.203	0.170	0.510	0.490	0.660	0.960	0.967	0.927
rlda.TG	1000	0.222	0.225	0.221	0.500	0.510	0.620	0.933	0.927	0.867
dlda	1000	0.280	0.243	0.228	0.460	0.470	0.610	0.863	0.910	0.857
nsc	12 625	0.333	0.333	0.333	0.380	0.380	0.380	0.833	0.833	0.833
svm	12 625	0.398	0.398	0.398	0.050	0.050	0.050	0.900	0.900	0.900

Table 3: Overview of the 10 × five-fold CV errors obtained for the different variants of RLDA (rlda.TD, rlda.TF, rlda.TG) as well as for the methods DLDA, NSC and SVM. For each method except NSC and SVM the top 100, 200, 500 and 1000 genes of the two-class data **CLL** (n=22) are employed. The results of three variable selection methods are compared. Correspondingly, both the sensitivity and the specificity are obtained.

Method	p	10 × five-fold CV error			Sensitivity			Specificity		
		t-test	Limma	Wilcoxon test	t-test	Limma	Wilcoxon test	t-test	Limma	Wilcoxon test
rlda.TG$^{(1)}$	100	0.286	0.290	0.246	0.390	0.380	0.630	0.890	0.890	0.830
rlda.TG$^{(2)}$	100	0.252	0.243	0.146	0.460	0.460	0.720	0.907	0.920	0.933
rlda.TG$^{(3)}$	100	0.249	0.251	0.199	0.470	0.450	0.640	0.907	0.913	0.893
rlda.TG$^{(4)}$	100	0.238	0.247	0.196	0.500	0.460	0.630	0.907	0.913	0.900
rlda.TG$^{(5)}$	100	0.195	0.184	0.151	0.530	0.550	0.720	0.960	0.967	0.923
rlda.TG$^{(6)}$	100	0.268	0.279	0.199	0.390	0.380	0.620	0.920	0.910	0.903
rlda.TG$^{(7)}$	100	0.258	0.251	0.221	0.400	0.420	0.620	0.933	0.933	0.873
rlda.TG$^{(8)}$	100	0.270	0.274	0.227	0.460	0.460	0.600	0.877	0.870	0.870
rlda.TG$^{(9)}$	100	0.246	0.262	0.199	0.480	0.450	0.620	0.907	0.900	0.900
rlda.TG$^{(10)}$	100	0.263	0.255	0.196	0.440	0.440	0.630	0.900	0.913	0.900
rlda.TG$^{(1)}$	200	0.247	0.243	0.258	0.500	0.510	0.580	0.900	0.900	0.840
rlda.TG$^{(2)}$	200	0.241	0.236	0.133	0.440	0.450	0.690	0.933	0.933	0.960
rlda.TG$^{(3)}$	200	0.254	0.253	0.141	0.450	0.440	0.700	0.907	0.913	0.947
rlda.TG$^{(4)}$	200	0.262	0.253	0.132	0.430	0.440	0.730	0.907	0.913	0.947
rlda.TG$^{(5)}$	200	0.184	0.180	0.129	0.530	0.540	0.750	0.980	0.980	0.943
rlda.TG$^{(6)}$	200	0.221	0.209	0.172	0.490	0.510	0.700	0.947	0.953	0.897
rlda.TG$^{(7)}$	200	0.222	0.214	0.187	0.500	0.510	0.650	0.940	0.947	0.907
rlda.TG$^{(8)}$	200	0.297	0.260	0.181	0.490	0.540	0.690	0.820	0.853	0.883
rlda.TG$^{(9)}$	200	0.233	0.235	0.179	0.520	0.510	0.670	0.897	0.900	0.903
rlda.TG$^{(10)}$	200	0.261	0.247	0.152	0.460	0.470	0.680	0.893	0.907	0.947
rlda.TG$^{(1)}$	500	0.202	0.195	0.400	0.580	0.590	0.560	0.923	0.927	0.623
rlda.TG$^{(2)}$	500	0.222	0.222	0.134	0.500	0.480	0.670	0.933	0.947	0.973
rlda.TG$^{(3)}$	500	0.214	0.218	0.146	0.500	0.480	0.670	0.947	0.953	0.953
rlda.TG$^{(4)}$	500	0.220	0.210	0.146	0.470	0.490	0.660	0.953	0.960	0.960
rlda.TG$^{(5)}$	500	0.185	0.188	0.181	0.530	0.500	0.640	0.973	0.980	0.917
rlda.TG$^{(6)}$	500	0.194	0.188	0.166	0.530	0.550	0.700	0.960	0.960	0.913
rlda.TG$^{(7)}$	500	0.182	0.182	0.180	0.580	0.580	0.630	0.953	0.953	0.927
rlda.TG$^{(8)}$	500	0.298	0.299	0.253	0.540	0.590	0.620	0.793	0.773	0.810
rlda.TG$^{(9)}$	500	0.280	0.342	0.270	0.550	0.470	0.590	0.817	0.760	0.803
rlda.TG$^{(10)}$	500	0.236	0.211	0.167	0.470	0.500	0.620	0.927	0.953	0.947
rlda.TG$^{(1)}$	1000	0.241	0.214	0.333	0.480	0.550	0.540	0.917	0.917	0.733
rlda.TG$^{(2)}$	1000	0.215	0.211	0.175	0.500	0.520	0.570	0.940	0.940	0.967
rlda.TG$^{(3)}$	1000	0.216	0.201	0.158	0.480	0.500	0.620	0.953	0.967	0.967
rlda.TG$^{(4)}$	1000	0.216	0.206	0.169	0.480	0.500	0.600	0.953	0.960	0.960
rlda.TG$^{(5)}$	1000	0.213	0.209	0.210	0.460	0.480	0.530	0.967	0.960	0.930
rlda.TG$^{(6)}$	1000	0.214	0.194	0.190	0.510	0.530	0.650	0.940	0.960	0.900
rlda.TG$^{(7)}$	1000	0.194	0.194	0.204	0.540	0.540	0.600	0.953	0.953	0.907
rlda.TG$^{(8)}$	1000	0.335	0.338	0.294	0.490	0.460	0.650	0.750	0.763	0.737
rlda.TG$^{(9)}$	1000	0.283	0.291	0.239	0.460	0.520	0.630	0.853	0.807	0.837
rlda.TG$^{(10)}$	1000	0.216	0.211	0.157	0.480	0.480	0.620	0.953	0.960	0.967

Table 4: Overview of the 10 × five-fold CV errors obtained for the different variants of rlds.TG. For each method the top 100, 200, 500 and 1000 genes of the two-class data **CLL** (n=22) are employed. The results of three variable selection methods are compared. Correspondingly, both the sensitivity and the specificity are obtained.

1.3 Wang data

Method	p	10 × five-fold CV error			Sensitivity			Specificity		
		t-test	Limma	Wilcoxon test	t-test	Limma	Wilcoxon test	t-test	Limma	Wilcoxon test
rlda.TD	100	0.380	0.381	0.387	0.427	0.426	0.443	0.735	0.735	0.714
rlda.TF	100	0.378	0.377	0.388	0.421	0.425	0.408	0.742	0.742	0.733
rlda.TG	100	0.382	0.383	0.384	0.421	0.419	0.445	0.736	0.736	0.719
dlda	100	0.381	0.381	0.385	0.512	0.511	0.529	0.683	0.684	0.666
rlda.TD	200	0.368	0.368	0.371	0.454	0.452	0.483	0.739	0.740	0.716
rlda.TF	200	0.368	0.369	0.382	0.458	0.455	0.430	0.736	0.736	0.730
rlda.TG	200	0.375	0.375	0.379	0.457	0.453	0.485	0.725	0.727	0.702
dlda	200	0.375	0.376	0.390	0.566	0.566	0.584	0.660	0.659	0.626
rlda.TD	500	0.354	0.353	0.353	0.473	0.474	0.507	0.749	0.750	0.731
rlda.TF	500	0.356	0.353	0.374	0.470	0.474	0.475	0.747	0.750	0.717
rlda.TG	500	0.383	0.384	0.409	0.462	0.462	0.505	0.710	0.708	0.642
dlda	500	0.369	0.369	0.402	0.594	0.592	0.581	0.653	0.654	0.608
rlda.TD	1000	0.347	0.348	0.336	0.474	0.474	0.497	0.760	0.758	0.764
rlda.TF	1000	0.348	0.348	0.358	0.472	0.476	0.491	0.759	0.757	0.732
rlda.TG	1000	0.380	0.376	0.402	0.474	0.481	0.516	0.708	0.710	0.648
dlda	1000	0.365	0.363	0.410	0.604	0.607	0.578	0.654	0.655	0.596
nsc	22 283	0.376	0.376	0.376	0.359	0.359	0.359	0.783	0.783	0.783
svm	22 283	0.341	0.341	0.341	0.272	0.272	0.272	0.890	0.890	0.890

Table 5: Overview of the 10 × five-fold CV errors obtained for the different variants of RLDA (rlda.TD, rlda.TF, rlda.TG) as well as for the methods DLDA, NSC and SVM. For each method except NSC and SVM the top 100, 200, 500 and 1000 genes of the two-class data **Wang** (n=286) are employed. The results of three variable selection methods are compared. Correspondingly, both the sensitivity and the specificity are obtained.

Method	p	10 × five-fold CV error			Sensitivity			Specificity		
		t-test	Limma	Wilcoxon test	t-test	Limma	Wilcoxon test	t-test	Limma	Wilcoxon test
rlda.TG$^{(1)}$	100	0.352	0.352	0.360	0.467	0.468	0.498	0.756	0.756	0.726
rlda.TG$^{(2)}$	100	0.378	0.377	0.389	0.445	0.446	0.442	0.728	0.729	0.712
rlda.TG$^{(3)}$	100	0.384	0.385	0.387	0.426	0.425	0.441	0.730	0.729	0.716
rlda.TG$^{(4)}$	100	0.380	0.380	0.386	0.425	0.424	0.446	0.736	0.737	0.714
rlda.TG$^{(5)}$	100	0.372	0.373	0.350	0.441	0.441	0.488	0.739	0.738	0.747
rlda.TG$^{(6)}$	100	0.343	0.344	0.351	0.470	0.470	0.480	0.768	0.768	0.750
rlda.TG$^{(7)}$	100	0.343	0.345	0.349	0.473	0.472	0.481	0.766	0.765	0.752
rlda.TG$^{(8)}$	100	0.385	0.384	0.388	0.424	0.425	0.442	0.730	0.730	0.714
rlda.TG$^{(9)}$	100	0.383	0.384	0.383	0.421	0.419	0.446	0.735	0.735	0.719
rlda.TG$^{(10)}$	100	0.382	0.383	0.384	0.425	0.422	0.446	0.733	0.733	0.717
rlda.TG$^{(1)}$	200	0.376	0.382	0.408	0.480	0.474	0.508	0.710	0.704	0.641
rlda.TG$^{(2)}$	200	0.359	0.359	0.361	0.483	0.485	0.494	0.734	0.734	0.726
rlda.TG$^{(3)}$	200	0.371	0.372	0.373	0.448	0.443	0.482	0.737	0.739	0.714
rlda.TG$^{(4)}$	200	0.373	0.373	0.374	0.449	0.446	0.483	0.733	0.735	0.712
rlda.TG$^{(5)}$	200	0.346	0.345	0.363	0.482	0.483	0.484	0.757	0.758	0.728
rlda.TG$^{(6)}$	200	0.342	0.344	0.344	0.477	0.484	0.504	0.766	0.760	0.747
rlda.TG$^{(7)}$	200	0.343	0.343	0.340	0.475	0.473	0.510	0.765	0.767	0.750
rlda.TG$^{(8)}$	200	0.381	0.381	0.375	0.444	0.445	0.488	0.723	0.723	0.707
rlda.TG$^{(9)}$	200	0.389	0.388	0.391	0.440	0.442	0.472	0.713	0.714	0.690
rlda.TG$^{(10)}$	200	0.375	0.376	0.377	0.446	0.441	0.481	0.732	0.733	0.709
rlda.TG$^{(1)}$	500	0.385	0.391	0.426	0.491	0.479	0.497	0.689	0.687	0.621
rlda.TG$^{(2)}$	500	0.365	0.362	0.345	0.456	0.458	0.511	0.742	0.745	0.740
rlda.TG$^{(3)}$	500	0.357	0.357	0.352	0.469	0.470	0.509	0.747	0.747	0.732
rlda.TG$^{(4)}$	500	0.351	0.352	0.354	0.477	0.476	0.503	0.752	0.750	0.732
rlda.TG$^{(5)}$	500	0.359	0.359	0.347	0.460	0.456	0.507	0.740	0.751	0.740
rlda.TG$^{(6)}$	500	0.343	0.343	0.334	0.474	0.476	0.511	0.766	0.766	0.759
rlda.TG$^{(7)}$	500	0.340	0.339	0.325	0.475	0.479	0.514	0.770	0.770	0.771
rlda.TG$^{(8)}$	500	0.402	0.399	0.408	0.476	0.462	0.503	0.672	0.684	0.645
rlda.TG$^{(9)}$	500	0.384	0.384	0.407	0.480	0.486	0.496	0.697	0.693	0.651
rlda.TG$^{(10)}$	500	0.360	0.359	0.398	0.481	0.480	0.497	0.735	0.736	0.665
rlda.TG$^{(1)}$	1000	0.407	0.408	0.426	0.470	0.476	0.532	0.667	0.660	0.600
rlda.TG$^{(2)}$	1000	0.361	0.360	0.335	0.464	0.465	0.517	0.744	0.744	0.754
rlda.TG$^{(3)}$	1000	0.348	0.348	0.340	0.470	0.471	0.499	0.761	0.760	0.756
rlda.TG$^{(4)}$	1000	0.407	0.345	0.338	0.470	0.477	0.497	0.667	0.762	0.761
rlda.TG$^{(5)}$	1000	0.339	0.339	0.343	0.465	0.465	0.505	0.779	0.778	0.747
rlda.TG$^{(6)}$	1000	0.345	0.344	0.328	0.463	0.467	0.519	0.771	0.769	0.764
rlda.TG$^{(7)}$	1000	0.338	0.343	0.319	0.465	0.463	0.534	0.780	0.774	0.769
rlda.TG$^{(8)}$	1000	0.436	0.449	0.432	0.468	0.440	0.528	0.621	0.617	0.592
rlda.TG$^{(9)}$	1000	0.432	0.433	0.406	0.491	0.490	0.513	0.614	0.614	0.642
rlda.TG$^{(10)}$	1000	0.390	0.394	0.399	0.485	0.468	0.512	0.684	0.688	0.654

Table 6: Overview of the 10 × five-fold CV errors obtained for the different variants of rlda.TG. For each method the top 100, 200, 500 and 1000 genes of the two-class data **Wang** (n=286) are employed. The results of three variable selection methods are compared. Correspondingly, both the sensitivity and the specificity are obtained.

1.4 Singh data

Method	p	10 × five-fold CV error			Sensitivity			Specificity		
		t-test	Limma	Wilcoxon test	t-test	Limma	Wilcoxon test	t-test	Limma	Wilcoxon test
rlda.TD	100	0.083	0.086	0.126	0.897	0.890	0.869	0.938	0.940	0.880
rlda.TF	100	0.080	0.081	0.127	0.897	0.896	0.907	0.944	0.944	0.840
rlda.TG	100	0.081	0.081	0.135	0.892	0.892	0.872	0.948	0.948	0.858
dlda	100	0.126	0.126	0.189	0.870	0.872	0.960	0.878	0.876	0.656
rlda.TD	200	0.084	0.088	0.129	0.899	0.897	0.866	0.934	0.928	0.878
rlda.TF	200	0.081	0.083	0.121	0.903	0.903	0.915	0.936	0.932	0.844
rlda.TG	200	0.133	0.125	0.178	0.863	0.871	0.831	0.872	0.880	0.814
dlda	200	0.149	0.152	0.209	0.855	0.851	0.920	0.848	0.846	0.656
rlda.TD	500	0.087	0.084	0.112	0.917	0.915	0.882	0.910	0.918	0.896
rlda.TF	500	0.086	0.083	0.105	0.917	0.917	0.944	0.912	0.918	0.846
rlda.TG	500	0.166	0.182	0.158	0.853	0.833	0.844	0.816	0.804	0.840
dlda	500	0.181	0.183	0.207	0.830	0.828	0.888	0.808	0.806	0.694
rlda.TD	1000	0.090	0.089	0.095	0.915	0.917	0.899	0.906	0.906	0.912
rlda.TF	1000	0.084	0.081	0.103	0.917	0.919	0.936	0.916	0.920	0.858
rlda.TG	1000	0.211	0.224	0.197	0.789	0.791	0.826	0.790	0.762	0.780
dlda	1000	0.224	0.224	0.221	0.798	0.798	0.860	0.754	0.754	0.696
nsc	12 625	0.130	0.130	0.130	0.859	0.859	0.859	0.882	0.882	0.882
svm	12 625	0.100	0.100	0.100	0.881	0.881	0.881	0.920	0.920	0.920

Table 7: Overview of the 10 × five-fold CV errors obtained for the different variants of RLDA (rlda.TD, rlda.TF, rlda.TG) as well as for the methods DLDA, NSC and SVM. For each method except NSC and SVM the top 100, 200, 500 and 1000 genes of the two-class data **Singh** (n=102) are employed. The results of three variable selection methods are compared. Correspondingly, both the sensitivity and the specificity are obtained.

Method	p	10 × five-fold CV error			Sensitivity			Specificity		
		t-test	Limma	Wilcoxon test	t-test	Limma	Wilcoxon test	t-test	Limma	Wilcoxon test
rlda.TG$^{(1)}$	100	0.126	0.129	0.175	0.869	0.870	0.857	0.880	0.872	0.792
rlda.TG$^{(2)}$	100	0.087	0.088	0.124	0.896	0.896	0.857	0.932	0.930	0.898
rlda.TG$^{(3)}$	100	0.081	0.082	0.127	0.895	0.896	0.871	0.944	0.942	0.876
rlda.TG$^{(4)}$	100	0.081	0.082	0.127	0.894	0.892	0.875	0.946	0.946	0.872
rlda.TG$^{(5)}$	100	0.151	0.148	0.151	0.862	0.863	0.892	0.838	0.842	0.806
rlda.TG$^{(6)}$	100	0.118	0.117	0.131	0.868	0.869	0.887	0.898	0.898	0.852
rlda.TG$^{(7)}$	100	0.123	0.116	0.126	0.871	0.879	0.893	0.886	0.892	0.856
rlda.TG$^{(8)}$	100	0.079	0.078	0.137	0.896	0.895	0.863	0.948	0.950	0.864
rlda.TG$^{(9)}$	100	0.086	0.085	0.138	0.888	0.888	0.872	0.942	0.944	0.852
rlda.TG$^{(10)}$	100	0.081	0.080	0.136	0.890	0.892	0.873	0.950	0.950	0.856
rlda.TG$^{(1)}$	200	0.231	0.212	0.205	0.787	0.794	0.815	0.752	0.782	0.774
rlda.TG$^{(2)}$	200	0.095	0.093	0.143	0.902	0.902	0.845	0.910	0.914	0.870
rlda.TG$^{(3)}$	200	0.090	0.092	0.134	0.894	0.892	0.866	0.928	0.926	0.868
rlda.TG$^{(4)}$	200	0.085	0.087	0.134	0.903	0.899	0.863	0.928	0.928	0.870
rlda.TG$^{(5)}$	200	0.151	0.152	0.141	0.861	0.865	0.894	0.830	0.832	0.824
rlda.TG$^{(6)}$	200	0.115	0.115	0.117	0.886	0.886	0.899	0.886	0.884	0.868
rlda.TG$^{(7)}$	200	0.130	0.129	0.116	0.878	0.878	0.906	0.862	0.864	0.862
rlda.TG$^{(8)}$	200	0.257	0.263	0.241	0.750	0.737	0.772	0.736	0.738	0.746
rlda.TG$^{(9)}$	200	0.166	0.165	0.268	0.837	0.840	0.773	0.832	0.830	0.690
rlda.TG$^{(10)}$	200	0.092	0.099	0.188	0.897	0.894	0.807	0.920	0.910	0.818
rlda.TG$^{(1)}$	500	0.311	0.306	0.282	0.699	0.730	0.741	0.680	0.658	0.694
rlda.TG$^{(2)}$	500	0.101	0.098	0.119	0.905	0.909	0.886	0.894	0.896	0.878
rlda.TG$^{(3)}$	500	0.089	0.087	0.113	0.913	0.913	0.884	0.910	0.914	0.892
rlda.TG$^{(4)}$	500	0.091	0.088	0.122	0.911	0.909	0.884	0.908	0.916	0.874
rlda.TG$^{(5)}$	500	0.118	0.115	0.113	0.894	0.892	0.895	0.870	0.878	0.880
rlda.TG$^{(6)}$	500	0.099	0.100	0.105	0.907	0.907	0.907	0.896	0.894	0.884
rlda.TG$^{(7)}$	500	0.097	0.096	0.093	0.907	0.907	0.921	0.900	0.902	0.894
rlda.TG$^{(8)}$	500	0.331	0.260	0.280	0.643	0.735	0.744	0.696	0.744	0.696
rlda.TG$^{(9)}$	500	0.330	0.279	0.267	0.677	0.727	0.754	0.664	0.716	0.712
rlda.TG$^{(10)}$	500	0.149	0.186	0.163	0.872	0.819	0.867	0.830	0.810	0.806
rlda.TG$^{(1)}$	1000	0.282	0.277	0.348	0.727	0.742	0.685	0.710	0.704	0.620
rlda.TG$^{(2)}$	1000	0.094	0.094	0.111	0.911	0.911	0.898	0.902	0.902	0.882
rlda.TG$^{(3)}$	1000	0.092	0.088	0.098	0.915	0.915	0.899	0.902	0.910	0.906
rlda.TG$^{(4)}$	1000	0.093	0.095	0.097	0.913	0.911	0.899	0.902	0.900	0.908
rlda.TG$^{(5)}$	1000	0.101	0.098	0.104	0.908	0.911	0.913	0.890	0.892	0.880
rlda.TG$^{(6)}$	1000	0.080	0.080	0.097	0.911	0.911	0.926	0.930	0.930	0.880
rlda.TG$^{(7)}$	1000	0.083	0.084	0.085	0.909	0.911	0.930	0.926	0.922	0.900
rlda.TG$^{(8)}$	1000	0.352	0.370	0.360	0.656	0.633	0.650	0.640	0.626	0.628
rlda.TG$^{(9)}$	1000	0.327	0.297	0.272	0.680	0.687	0.780	0.666	0.720	0.674
rlda.TG$^{(10)}$	1000	0.237	0.239	0.198	0.770	0.780	0.831	0.756	0.742	0.772

Table 8: Overview of the 10 × five-fold CV errors obtained for the different variants of rlda.TG. For each method the top 100, 200, 500 and 1000 genes of the two-class data **Singh** (n=102) are employed. The results of three variable selection methods are compared. Correspondingly, both the sensitivity and the specificity are obtained.