SUPPLEMENTARY MATERIAL

Application of Clustering Analyses to the Diagnosis of Huntington Disease in Mice and Other Diseases with Well-Defined Group Boundaries.

Jason B. Nikas, DPT and Walter C. Low, PhD

SUPPLEMENTARY TABLES (S1-S11)

&

SUPPLEMENTARY FIGURES (S1, S2)

*Corresponding Author: Jason B. Nikas, DPT
Department of Neurosurgery, Medical School
University of Minnesota
4-218 McGuire Translational Research Facility
2001 Sixth St., SE
Minneapolis, MN 55455, USA
T: 612-625-2868 / F: 612-626-9201
E-mail: nikas001@umn.edu

K-Means Cluster Analysis Report

Distance Section for Cluster 1

Row	Cluster	Dist1	Dist2
1	1	3.0484	6.2439
2	1	3.3598	5.1079
3	1	2.4599	4.7514
4	1	2.2589	4.9319
5	1	2.8342	5.8760
6	1	2.6353	5.0683
7	1	2.3079	5.4675
8	1	2.8974	5.9031
9	1	1.2467	5.0675
10	1	3.0832	5.8090
11	1	2.4314	5.3043
12	1	1.8541	5.2802
13	1	2.1405	5.0057
14	1	2.0506	5.1658
15	1	2.0683	5.3467
16	1	2.5516	5.1310
17	1	1.9951	5.9111

Count = 17

Distance Section for Cluster 2

Row	Cluster	Dist1	Dist2
18	2	4.3590	2.4987
19	2	4.8567	3.6213
20	2	4.9546	3.7667
21	2	6.2601	4.4283
22	2	6.2500	4.4591
23	2	5.2714	3.1049
24	2	4.7910	4.1525
25	2	6.3389	3.2945
26	2	6.6912	3.5422
27	2	7.7911	5.1342
28	2	6.0151	2.8845
29	2	6.4565	3.5405
30	2	9.9016	7.0177

Count = 13

Minimum Iteration Section

Iteration No.	No. of Clusters	Percent of Variation	Bar Chart of Percent
2	2	65.76	|||||||||||||||||||||

Table S1: Classification of the 30 original mice (17 WT & 13 R6/2) into two clusters according to the unsupervised KMCA. Rows #1-17 are the WT mice, whereas rows #18-30 are the R6/2 mice. Dist 1 & Dist 2 are the distances of each subject from the center of Cluster 1 and Cluster 2 respectively. The unsupervised KMCA correctly identified all of the 30 mice. All of the WT mice were classified into Cluster 1, whereas all of the R6/2 into Cluster 2. All 17 IVs were used in this setting, and the Percent of Variation was 65.76.

K-Means Cluster Analysis Report

Distance Section for Cluster 1

Row	Cluster	Dist 1	Dist 2
1	1	0.7891	3.9294
2	1	1.2586	2.3867
3	1	0.6668	2.9516
4	1	0.7301	3.5212
5	1	0.8012	4.0580
6	1	0.5311	3.9329
7	1	0.3766	3.7267
8	1	0.3705	3.5628
9	1	0.5965	3.8337
10	1	1.0336	3.8094
11	1	0.4837	3.4846
12	1	0.3604	3.3653
13	1	0.7213	2.9754
14	1	0.8561	3.0042
15	1	1.0921	4.1624
16	1	0.5316	3.9053
17	1	0.4197	3.7268

Count = 17

Distance Section for Cluster 2

Row	Cluster	Dist 1	Dist 2
18	2	2.4043	1.1988
19	2	3.8064	0.9200
20	2	3.1157	1.1089
21	2	3.9263	1.0913
22	2	3.5241	0.7401
23	2	3.3865	0.9139
24	2	2.8613	0.7893
25	2	4.2573	0.9148
26	2	3.7675	1.3327
27	2	4.5809	1.6957
28	2	3.3639	1.1376
29	2	3.9215	0.4911
30	2	4.3145	1.6575

Count = 13

Minimum Iteration Section

Iteration No.	No. of Clusters	Percent of Variation	Bar Chart of Percent
1	2	22.04	|||||||

Cluster Means

Variables	Cluster1	Cluster2
Cr+PCr	8.943529	11.52708
Gln	3.448176	5.975923
Cr	5.472059	7.150462
NAA	7.653588	6.222077
Count	17	13

Table S2: Classification of the 30 original mice (17 WT & 13 R6/2) into two clusters according to ROC-supervised KMCA. Rows #1-17 are the WT mice, whereas rows #18-30 are the R6/2 mice. Dist 1 & Dist 2 are the distances of each subject from the center of Cluster 1 and Cluster 2 respectively. The ROC-supervised KMCA correctly identified all of the 30 mice. All of the WT mice were classified into Cluster 1, whereas all of the R6/2 into Cluster 2. The top four most significant IVs (AUC > 98%) according to ROC curve analysis were used in this setting, and the Percent of Variation was 22.04.

K-Means Cluster Analysis Report

Distance Section for Cluster 1

Row	Cluster	Dist 1	Dist 2
1	1	0.7902	3.8639
2	1	1.2696	2.2961
3	1	0.6710	2.8675
4	1	0.7381	3.4535
5	1	0.8140	3.9904
6	1	0.5394	3.8603
7	1	0.3805	3.6557
8	1	0.3724	3.4848
9	1	0.5998	3.7614
10	1	1.0408	3.7326
11	1	0.4892	3.4100
12	1	0.3653	3.2895
13	1	0.7329	2.8942
14	1	0.8642	2.9192
15	1	1.0957	4.0956
16	1	0.5395	3.8313
17	1	0.4253	3.6509

Count = 17

Distance Section for Cluster 2

Row	Cluster	Dist 1	Dist 2
18	2	2.4386	1.1149
19	2	3.8438	0.9030
20	2	3.1619	1.0975
21	2	3.9816	1.1657
22	2	3.5657	0.7220
23	2	3.4384	0.9571
24	2	2.8977	0.6920
25	2	4.3046	0.9839
26	2	3.7964	1.2956
27	2	4.6501	1.8347
28	2	3.3970	1.0941
29	2	3.9734	0.6220
30	2	4.3701	1.7312
31	2	2.1247	1.8607

Count = 14

Minimum Iteration Section

Iteration No.	No. of Clusters	Percent of Variation	Bar Chart of Percent
1	2	24.86	|||||||||

Table S3: Classification of the 30 original mice (17 WT & 13 R6/2) and one unknown mouse into two clusters according to ROC-supervised KMCA. Rows #1-17 are the WT mice, rows #18-30 are the R6/2 mice, and row #31 is the unknown mouse. Dist 1 & Dist 2 are the distances of each subject from the center of Cluster 1 and Cluster 2 respectively. All of the WT mice are classified into Cluster 1, whereas all of the R6/2 into Cluster 2. The unknown mouse (#31) was correctly identified as an R6/2 mouse by the ROC-supervised KMCA. The top four most significant IVs (AUC > 98%) according to ROC curve analysis were used in this setting.

K-Means Cluster Analysis Report

Distance Section for Cluster 1

Row	Cluster	Dist1	Dist2
1	1	1.0827	1.4372
2	1	0.3848	2.1191
3	1	0.1526	2.3371
4	1	0.6097	2.1726
5	1	0.2713	2.4490
6	1	0.4182	2.0032
7	1	1.0664	3.2842

Count = 7

Distance Section for Cluster 2

Row	Cluster	Dist1	Dist2
8	2	1.3609	0.8573
9	2	2.3820	0.3837
10	2	2.0063	0.2201
11	2	1.5916	0.8114
12	2	2.0422	0.2545
13	2	4.0331	1.8310

Count = 6

Minimum Iteration Section

Iteration No.	No. of Clusters	Percent of Variation	Bar Chart of Percent
1	2	33.78	||||||||||||

Cluster Means

Variables	Cluster1	Cluster2
R62_TTau	0.4010287	0.354751
R62_GPCPC	2.043571	3.182
Count	7	6

Table S4: Classification of the 13 original R6/2 mice [(7) 8-wk old & (6) 12-wk old] into two clusters according to the R6/2-ROC-supervised KMCA. Rows #1-7 are the 8-wk old R6/2, whereas rows #8-13 are the 12-wk old R6/2 mice. Dist 1 & Dist 2 are the distances of each subject from the center of Cluster 1 and Cluster 2 respectively. Our KMCA DBM correctly identified all of the R6/2 mice. All of the 8-wk old R6/2 mice were classified into Cluster 1, whereas all of the 12-wk old R6/2 mice into Cluster 2. TTau and GPC+PC, the top two most significant IVs (AUC>95%) according to R6/2 ROC curve analysis, were used in this setting, and the Percent of Variation was 33.78.

Fuzzy Clustering Report

Cluster Medoids Section

Variable	Cluster1	Cluster2
Cr+PCr	8.553	12.117
Gln	3.513	6.169
Cr	5.54	7.33
NAA	7.85	6.189
Row	7	29

Membership Matrix Section

Row	Cluster	Prob in 1	Prob in 2
1	1	0.8594	0.1406
2	1	0.6576	0.3424
3	1	0.8299	0.1701
4	1	0.8490	0.1510
5	1	0.8618	0.1382
6	1	0.9018	0.0982
7	1	0.9111	0.0889
8	1	0.9072	0.0928
9	1	0.8873	0.1127
10	1	0.8074	0.1926
11	1	0.8874	0.1126
12	1	0.9003	0.0997
13	1	0.8185	0.1815
14	1	0.7923	0.2077
15	1	0.8184	0.1816
16	1	0.8965	0.1035
17	1	0.9097	0.0903
18	2	0.2907	0.7093
19	2	0.1482	0.8518
20	2	0.1972	0.8028
21	2	0.1624	0.8376
22	2	0.1262	0.8738
23	2	0.1694	0.8306
24	2	0.1709	0.8291
25	2	0.1418	0.8582
26	2	0.2234	0.7766
27	2	0.2422	0.7578
28	2	0.2214	0.7786
29	2	0.1174	0.8826
30	2	0.2482	0.7518

Summary Section

Number Clusters	Average Distance	Average Silhouette	F(U)	Fc(U)	D(U)	Dc(U)
2	7.091622	0.654794	0.7313	0.4626	0.0622	0.1244

Table S5: Classification of our 30 mice (17 WT & 13 R6/2) into two clusters according to the ROC-supervised FCA. Rows #1-17 are the WT mice, whereas rows #18-30 are the R6/2 mice. Prob in 1 & Prob in 2 are the membership probabilities of each subject in Cluster 1 and Cluster 2 respectively. The ROC-supervised FCA correctly identified all of the 30 mice. All of the WT mice were classified into Cluster 1, whereas all of the R6/2 into Cluster 2. The top four most significant IVs (AUC>98%) were used in this setting, as can be seen from the Cluster Medoids Section. Subjects # 7 and # 29 were the medoids (closest to the center) of Cluster 1 and Cluster 2 respectively for this setting. As can be seen from the values of the Average Silhouette, Fc(U), and Dc(U) in the Summary Section, as well as from the membership probabilities, this was the best setting and was used in the development of our FCA DBM. The Average Distance denotes the value of the average dissimilarity for this setting. [AS=0.655].

Fuzzy Clustering Report

Cluster Medoids Section

Variable	Cluster1	Cluster2
Cr+PCr	8.553	12.117
Gln	3.513	6.169
Cr	5.54	7.33
NAA	7.85	6.189
Row	7	29

Membership Matrix Section

Row	Cluster	Prob in 1	Prob in 2
1	1	0.8653	0.1347
2	1	0.6581	0.3419
3	1	0.8275	0.1725
4	1	0.8501	0.1499
5	1	0.8557	0.1443
6	1	0.8964	0.1036
7	1	0.9107	0.0893
8	1	0.9018	0.0982
9	1	0.8902	0.1098
10	1	0.8029	0.1971
11	1	0.8911	0.1089
12	1	0.8969	0.1031
13	1	0.8197	0.1803
14	1	0.7915	0.2085
15	1	0.8249	0.1751
16	1	0.8928	0.1072
17	1	0.9032	0.0968
18	2	0.2880	0.7120
19	2	0.1480	0.8520
20	2	0.1954	0.8046
21	2	0.1643	0.8357
22	2	0.1271	0.8729
23	2	0.1666	0.8334
24	2	0.1706	0.8294
25	2	0.1423	0.8577
26	2	0.2241	0.7759
27	2	0.2435	0.7565
28	2	0.2241	0.7759
29	2	0.1192	0.8808
30	2	0.2511	0.7489
31	1	0.7633	0.2367

Summary Section

Number Clusters	Average Distance	Average Silhouette	F(U)	Fc(U)	D(U)	Dc(U)
2	7.432293	0.651339	0.7272	0.4544	0.0641	0.1282

Table S6: Classification of our 30 mice (17 WT & 13 R6/2) and one unknown mouse into two clusters according to the ROC-supervised FCA. Rows #1-17 are the WT mice, rows #18-30 are the R6/2 mice, and row #31 is the unknown mouse. Prob in 1 & Prob in 2 are the membership probabilities of each subject in Cluster 1 and Cluster 2 respectively. All of the WT mice were classified into Cluster 1, whereas all of the R6/2 into Cluster 2. The unknown mouse (#31) was correctly identified as a WT mouse and placed in Cluster 1. The top four most significant IVs (AUC>98%) were used in this setting, as can be seen from the Cluster Medoids Section. Subjects # 7 and # 29 were the medoids (closest to the center) of Cluster 1 and Cluster 2 respectively for this group of 31 mice.

Fuzzy Clustering Report

Cluster Medoids Section

Variable	Cluster1	Cluster2
R62_TTau	0.4054208	0.3580374
R62_GPCPC	2.036	3.042
Row	3	10

Membership Matrix Section

Row	Cluster	Prob in 1	Prob in 2
1	1	0.5124	0.4876
2	1	0.8758	0.1242
3	1	0.9199	0.0801
4	1	0.8221	0.1779
5	1	0.8938	0.1062
6	1	0.8174	0.1826
7	1	0.7822	0.2178
8	2	0.2785	0.7215
9	2	0.1637	0.8363
10	2	0.0841	0.9159
11	2	0.2212	0.7788
12	2	0.0892	0.9108
13	2	0.3095	0.6905

Summary Section

Number Clusters	Average Distance	Average Silhouette	F(U)	Fc(U)	D(U)	Dc(U)
2	3.016061	0.525337	0.7115	0.4230	0.0996	0.1992

Table S7: Classification of the 13 original R6/2 mice [(7) 8-wk old & (6) 12-wk old] into two clusters according to the R6/2-ROC-supervised FCA. Rows #1-7 are the 8-wk old R6/2, whereas rows #8-13 are the 12-wk old R6/2 mice. Prob in 1 & Prob in 2 are the membership probabilities of each subject in Cluster 1 and Cluster 2 respectively. The R6/2-ROC-supervised FCA correctly identified all of the R6/2 mice. All of the 8-wk old R6/2 mice were classified into Cluster 1, whereas all of the 12-wk old R6/2 mice into Cluster 2. The top two most significant IVs (TTau & GPC+PC) (AUC>95%) according to the R6/2 ROC curve analysis were used in this setting. Subjects # 3 and # 10 were the medoids (closest to the center) of Cluster 1 and Cluster 2 respectively for this group of 13 mice. [AS=0.525].

Clustering by Medoid Partitioning Report

Variables   R62_Cr+PCr to R62_PCr/Cr
Method: Kaufman - Rousseeuw, Objective Function: Silhouette
Distance Type: Euclidean, Scale Type: Standard Deviation

Row Detail Section

Row	Cluster	Nearest Neighbor	Average Distance Within	Average Distance Neighbor	Silhouette Value	Silhouette Bar
3	1	2	40.45	71.04	0.4306	|IIIIIIIIIIIIIIIIIIIII
5	1	2	42.96	72.13	0.4044	|IIIIIIIIIIIIIIIIIIII
6	1	2	42.80	62.92	0.3197	|IIIIIIIIIIIIIIII
7	1	2	47.74	69.64	0.3144	|IIIIIIIIIIIIIIII
4	1	2	55.33	75.66	0.2687	|IIIIIIIIIIIII
2	1	2	50.26	64.40	0.2195	|IIIIIIIIIII
1	1	2	50.07	57.66	0.1316	|IIIIIII
Cluster Average	1	(7)	47.09	67.64	0.2984	
13	2	1	70.02	91.24	0.2325	|IIIIIIIIIIII
9	2	1	52.43	64.88	0.1920	|IIIIIIIIII
10	2	1	59.81	73.85	0.1901	|IIIIIIIIII
12	2	1	49.42	60.55	0.1838	|IIIIIIIII
11	2	1	50.15	59.09	0.1513	|IIIIIIII
8	2	1	55.52	56.21	0.0122	|I
Cluster Average	2	(6)	56.23	67.64	0.1603	
Overall Average		(13)	51.31	67.64	0.2347	= SC

Table S8: Classification of the 13 original R6/2 mice [(7) 8-wk old & (6) 12-wk old] into two clusters according to the unsupervised MPCA. Rows #1-7 are the 8-wk old R6/2, whereas rows #8-13 are the 12-wk old R6/2 mice. The unsupervised MPCA correctly identified all of the R6/2 mice. All of the 8-wk old R6/2 mice were classified into Cluster 1, whereas all of the 12-wk old R6/2 mice into Cluster 2. All 17 IVs were used in this setting. The Average Silhouette (AS) value was 0.235. "Average Distance Within" is the average distance of a subject with respect to all other members of the same cluster, whereas "Average Distance Neighbor" is the average distance of a subject with respect to all members of the other cluster.

Clustering by Medoid Partitioning Report

Variables   Cr+PCr, Gln, Cr, NAA
Method: Kaufman - Rousseeuw, Objective Function: Silhouette
Distance Type: Euclidean, Scale Type: Standard Deviation

Row Detail Section

Row	Cluster	Nearest Neighbor	Average Distance Within	Average Distance Neighbor	Silhouette Value	Silhouette Bar
7	1	2	15.74	75.45	0.7913	|IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
17	1	2	15.86	75.48	0.7899	|IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
6	1	2	16.94	79.39	0.7866	|IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
8	1	2	15.64	72.35	0.7839	|IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
16	1	2	17.21	78.80	0.7816	|IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
12	1	2	15.77	68.66	0.7703	|IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
9	1	2	17.81	77.42	0.7700	|IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
11	1	2	16.89	70.77	0.7613	|IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
5	1	2	20.73	81.83	0.7467	|IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
1	1	2	20.50	79.32	0.7415	|IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
4	1	2	20.13	71.64	0.7190	|IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
15	1	2	25.25	83.68	0.6982	|IIIIIIIIIIIIIIIIIIIIIIIIIIIIII
3	1	2	19.13	60.90	0.6859	|IIIIIIIIIIIIIIIIIIIIIIIIIIIIII
10	1	2	24.67	77.07	0.6799	|IIIIIIIIIIIIIIIIIIIIIIIIIIIIII
13	1	2	19.93	61.15	0.6741	|IIIIIIIIIIIIIIIIIIIIIIIIIIIII
14	1	2	21.81	61.70	0.6464	|IIIIIIIIIIIIIIIIIIIIIIIIIIII
2	1	2	28.32	50.48	0.4390	|IIIIIIIIIIIIIIIIIII
Cluster Average	1	(17)	19.55	72.12	0.7215	
29	2	1	24.49	77.70	0.6848	|IIIIIIIIIIIIIIIIIIIIIIIIIIIIII
25	2	1	28.15	84.15	0.6655	|IIIIIIIIIIIIIIIIIIIIIIIIIIIII
22	2	1	25.02	69.91	0.6421	|IIIIIIIIIIIIIIIIIIIIIIIIIIII
19	2	1	27.68	75.35	0.6326	|IIIIIIIIIIIIIIIIIIIIIIIIIIII
21	2	1	29.56	77.75	0.6198	|IIIIIIIIIIIIIIIIIIIIIIIIIII
23	2	1	27.78	67.37	0.5877	|IIIIIIIIIIIIIIIIIIIIIIIIII
27	2	1	40.96	90.59	0.5479	|IIIIIIIIIIIIIIIIIIIIIIII
24	2	1	26.12	57.18	0.5431	|IIIIIIIIIIIIIIIIIIIIIIII
26	2	1	34.40	74.62	0.5391	|IIIIIIIIIIIIIIIIIIIIIIII
30	2	1	39.94	85.43	0.5325	|IIIIIIIIIIIIIIIIIIIIIIII
28	2	1	31.73	66.90	0.5256	|IIIIIIIIIIIIIIIIIIIIIII
20	2	1	29.50	62.08	0.5248	|IIIIIIIIIIIIIIIIIIIIIII
18	2	1	32.42	48.59	0.3328	|IIIIIIIIIIIIIII
Cluster Average	2	(13)	30.60	72.12	0.5676	
Overall Average		(30)	24.34	72.12	0.6548	= SC

Table S9: Classification of the 30 original mice (17 WT & 13 R6/2) into two clusters according to the ROC-supervised MPCA. Rows #1-17 are the WT mice, whereas rows #18-30 are the R6/2 mice. The ROC-supervised MPCA correctly identified all of the 30 mice. All of the WT mice were classified into Cluster 1, whereas all of the R6/2 mice into Cluster 2. The top four most significant IVs (AUC > 98%) were used in this setting, and the Average Silhouette (AS) value was 0.655.

Clustering by Medoid Partitioning Report

Variables   Cr+PCr, Gln, Cr, NAA
Method: Kaufman - Rousseeuw, Objective Function: Silhouette
Distance Type: Euclidean, Scale Type: Standard Deviation

Row Detail Section

Row	Cluster	Nearest Neighbor	Average Distance Within	Average Distance Neighbor	Silhouette Value	Silhouette Bar
17	1	2	17.37	75.50	0.7699	|||||||||||||||||||||||||||||||
7	1	2	17.53	75.36	0.7674	|||||||||||||||||||||||||||||||
8	1	2	16.96	72.38	0.7657	|||||||||||||||||||||||||||||||
6	1	2	18.71	79.36	0.7642	|||||||||||||||||||||||||||||||
16	1	2	18.92	78.76	0.7598	||||||||||||||||||||||||||||||
12	1	2	17.12	68.63	0.7506	||||||||||||||||||||||||||||||
9	1	2	19.58	77.25	0.7466	|||||||||||||||||||||||||||||
11	1	2	18.43	70.62	0.7391	|||||||||||||||||||||||||||||
5	1	2	22.50	81.81	0.7250	||||||||||||||||||||||||||||
1	1	2	22.35	79.09	0.7174	||||||||||||||||||||||||||||
4	1	2	21.68	71.48	0.6967	||||||||||||||||||||||||||
3	1	2	19.69	60.92	0.6768	|||||||||||||||||||||||||
15	1	2	27.00	83.38	0.6762	|||||||||||||||||||||||||
10	1	2	25.44	77.22	0.6706	||||||||||||||||||||||||
13	1	2	20.76	61.06	0.6600	|||||||||||||||||||||||
14	1	2	22.23	61.72	0.6399	||||||||||||||||||||||
2	1	2	27.85	50.52	0.4486	||||||||||||||
31	1	2	42.55	43.23	0.0157	||
Cluster Average	1	(18)	22.04	70.46	0.6661	
29	2	1	24.47	75.99	0.6780	||||||||||||||||||||||||
25	2	1	28.03	82.00	0.6581	|||||||||||||||||||||||
22	2	1	24.97	67.99	0.6328	|||||||||||||||||||||
19	2	1	27.56	73.07	0.6228	|||||||||||||||||||||
21	2	1	29.55	76.27	0.6125	||||||||||||||||||||
23	2	1	27.65	66.29	0.5830	|||||||||||||||||||
27	2	1	40.95	89.49	0.5424	|||||||||||||||||
24	2	1	26.08	55.53	0.5304	|||||||||||||||||
26	2	1	34.25	72.20	0.5255	|||||||||||||||||
30	2	1	39.98	83.84	0.5232	|||||||||||||||||
20	2	1	29.42	60.94	0.5173	|||||||||||||||||
28	2	1	31.72	64.88	0.5111	|||||||||||||||||
18	2	1	32.31	47.48	0.3194	|||||||||||
Cluster Average	2	(13)	30.53	70.46	0.5582	
Overall Average		(31)	25.60	70.46	0.6209	= SC

Table S10: Classification of the 30 original mice (17 WT & 13 R6/2) and one unknown mouse into two clusters according to the ROC-supervised MPCA. Rows #1-17 are the WT mice, rows #18-30 are the R6/2 mice, and row #31 is the unknown mouse. All of the WT mice are classified into Cluster 1, whereas all of the R6/2 into Cluster 2. The unknown mouse (#31) – an R6/2 by genotyping – was incorrectly identified as a WT mouse and classified in Cluster 1, along with the WT mice. The Silhouette Value of 0.0157 of mouse #31 (the unknown), by being so close to 0, indicates that the decision on its classification was hardly better than the chance decision (50-50). The top four most significant IVs (AUC > 98%) were used in this setting.

Clustering by Medoid Partitioning Report

Variables   R62_TTau, R62_GPC+PC
Method: Kaufman - Rousseeuw, Objective Function: Silhouette
Distance Type: Euclidean, Scale Type: Standard Deviation

Row Detail Section

Row	Cluster	Nearest Neighbor	Average Distance Within	Average Distance Neighbor	Silhouette Value	Silhouette Bar
3	1	2	12.46	45.89	0.7285	|IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
5	1	2	13.62	48.06	0.7166	|IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
2	1	2	14.00	41.86	0.6656	|IIIIIIIIIIIIIIIIIIIIIIIIIIIIII
6	1	2	14.97	39.66	0.6226	|IIIIIIIIIIIIIIIIIIIIIIIIIIII
7	1	2	24.40	63.40	0.6151	|IIIIIIIIIIIIIIIIIIIIIIIIIII
4	1	2	16.87	42.87	0.6064	|IIIIIIIIIIIIIIIIIIIIIIIIIII
1	1	2	24.81	28.85	0.1400	|IIIIIIII
Cluster Average	1	(7)	17.30	44.37	0.5850	
10	2	1	15.85	39.63	0.6002	|IIIIIIIIIIIIIIIIIIIIIIIIII
12	2	1	16.18	40.32	0.5987	|IIIIIIIIIIIIIIIIIIIIIIIIII
9	2	1	20.07	46.99	0.5729	|IIIIIIIIIIIIIIIIIIIIIIIII
13	2	1	44.64	80.42	0.4449	|IIIIIIIIIIIIIIIIIIII
11	2	1	21.26	31.39	0.3229	|IIIIIIIIIIIIIII
8	2	1	22.12	27.48	0.1949	|IIIIIIIIII
Cluster Average	2	(6)	23.35	44.37	0.4558	
Overall Average		(13)	20.10	44.37	0.5253	= SC

Table S11: Classification of the 13 original R6/2 mice [(7) 8-wk old & (6) 12-wk old] into two clusters according to the R6/2-ROC-supervised MPCA. Rows #1-7 are the 8-wk old R6/2, whereas rows #8-13 are the 12-wk old R6/2 mice. The R6/2-ROC-supervised MPCA correctly identified all of the R6/2 mice. All of the 8-wk old R6/2 mice were classified into Cluster 1, whereas all of the 12-wk old R6/2 mice into Cluster 2. The top two most significant IVs (TTau & GPC+PC) (AUC>95%) according to the R6/2 ROC curve analysis were used in this setting. The Average Silhouette value (AS) was 0.525.
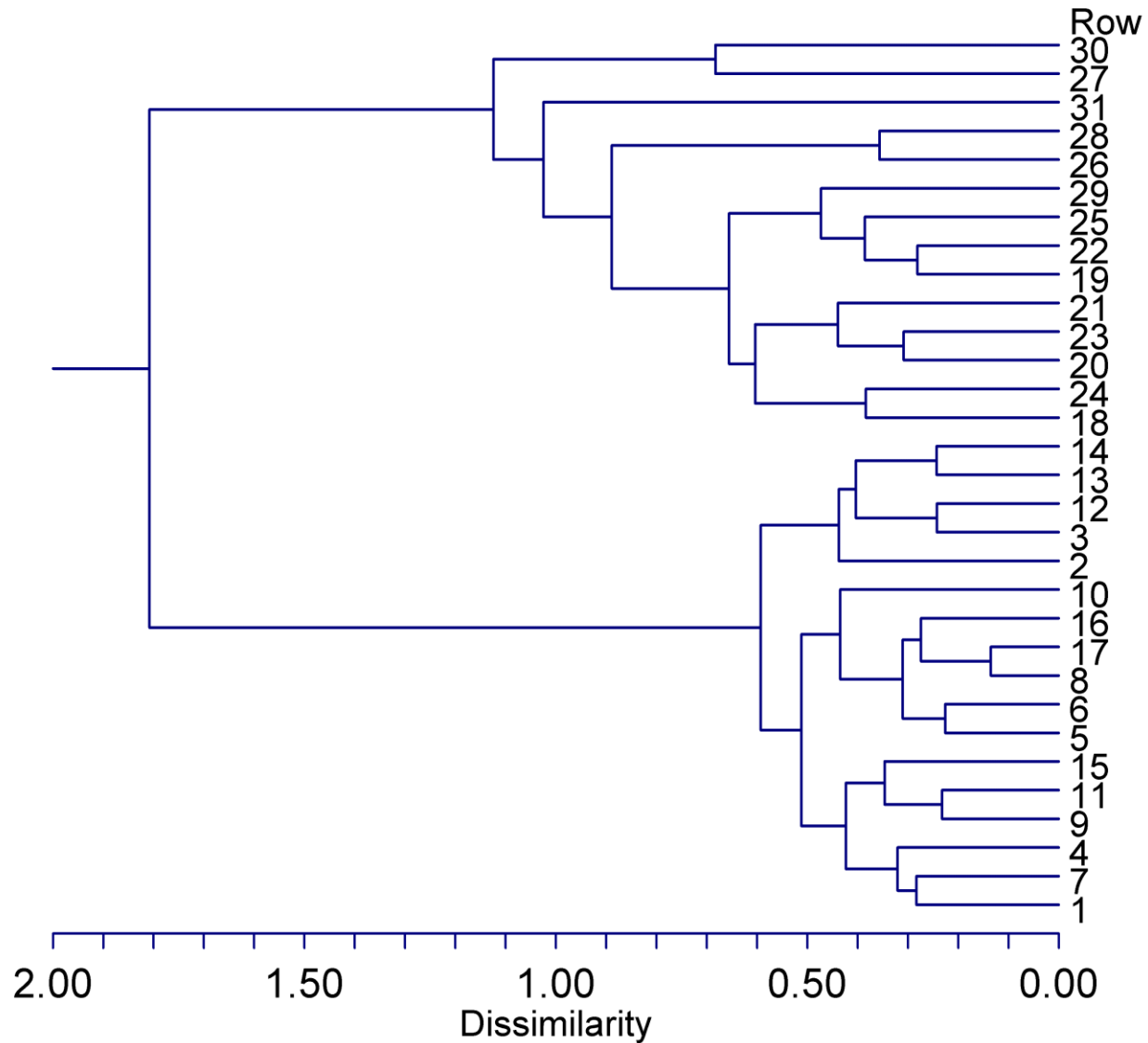
Figure S1. Dendrogram classification of the 30 original mice and one unknown mouse by the ROC-supervised HCA. Rows #1-17 are the WT mice, rows #18-30 are the R6/2 mice, and row #31 is the unknown mouse. The ROC-supervised HCA correctly identified that unknown mouse as an R6/2 (confirmed via genotyping). More specifically, looking in that area of the dendrogram where there are only two clusters (Dissimilarity range of ~ 1.2-1.7), one can see that subjects # 1-17 (all of the WT mice) were classified into the lower cluster, whereas subjects #18-30 (all of the R6/2 mice) and # 31 (the unknown mouse) were classified together into the upper cluster. The top four most significant IVs (AUC > 98%) according to ROC curve analysis were used in this setting.
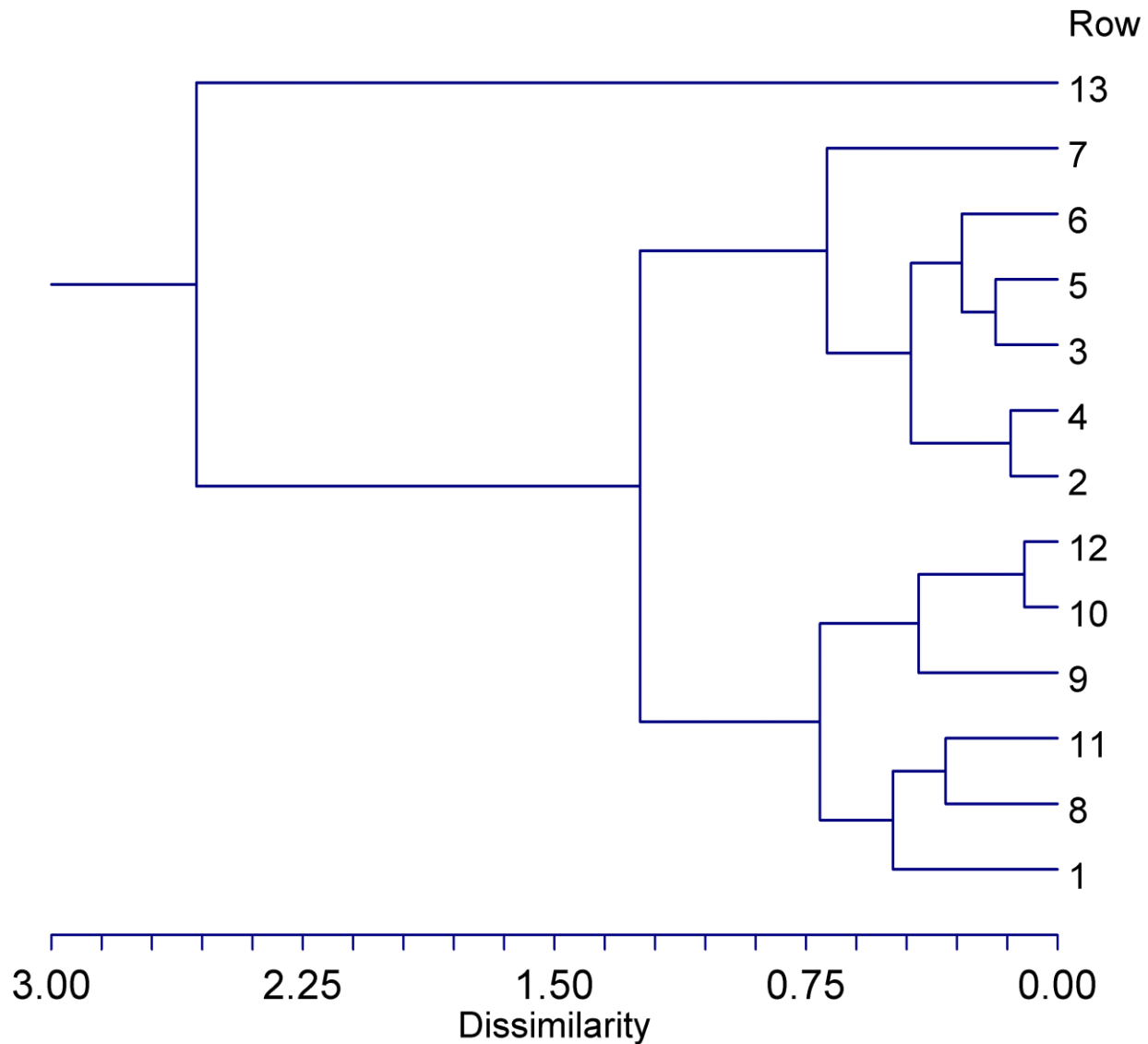
Figure S2. Dendrogram classification of the 13 original R6/2 mice by the R6/2-ROC-supervised HCA. Rows #1-7 are the 8-wk old R6/2 mice, whereas rows #8-13 are the 12-wk old R6/2 mice. The R6/2-ROC-supervised HCA failed to identify correctly the R6/2 mice. Looking in the area where there are only two clusters (Dissimilarity range of ~ 1.4-2.3), one can see that subject #13 was classified all by itself into the upper cluster, whereas all of the remaining subjects were classified into the lower cluster. This classification makes little sense. The top two most significant IVs (TTau & GPC+PC) (AUC>95%) according to the R6/2 ROC curve analysis were used in this setting.