Metagenes and Molecular Pattern Discovery using Matrix Factorization

Supplemental Information

Jean-Philippe Brunet[1], Pablo Tamayo[1], Todd. R. Golub[1, 2,] Jill P. Mesirov[1]

[1] Center for Genome Research, Massachusetts Institute of Technology, 320 Charles Street, Cambridge MA 02141 USA.

[2] Dana-Farber Cancer Institute and Harvard Medical School, 44 Binney Street, Boston MA 02115 USA.

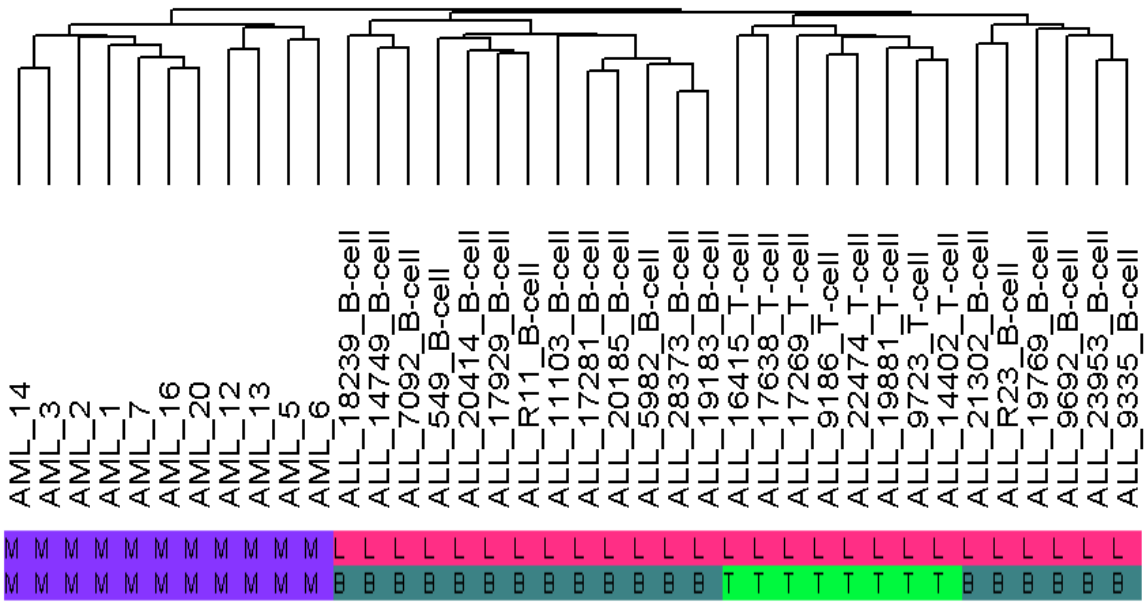

Figure S1 : Dendrogram showing agglomerative clustering of ALL and AML samples using a hierachical method with average linkage and N=3000 genes selected by their coefficient of variation (www.biostat.harvard.edu/complab/dchip/). Upper color bar indicates ALL (pink) and AML (purple). Lower color bar indicates further partitioning of ALL samples into B (teal) and T (green) cells.

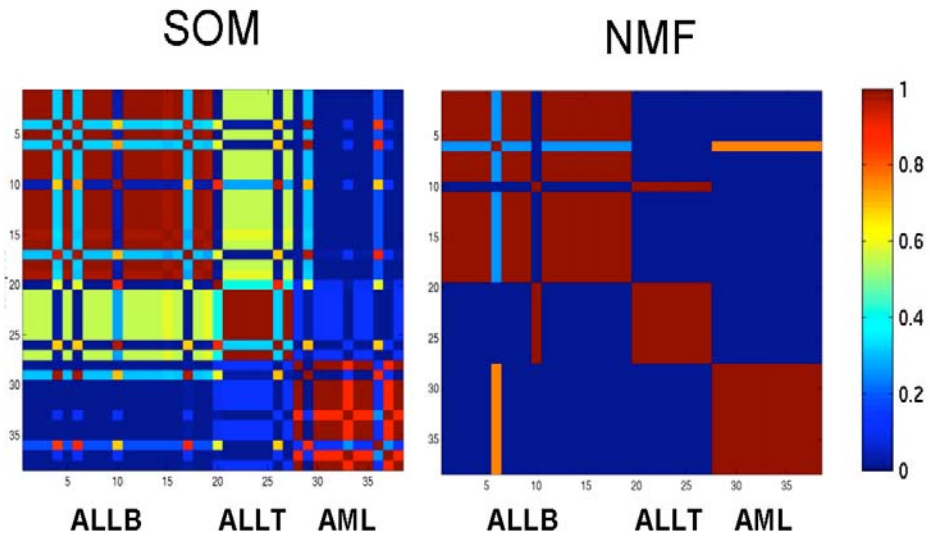

Figure S2: SOM and NMF consensus clustering matrices for the leukemia dataset with 5000 genes at k=3.

Sample	Subtype	k=5	k=3	k=2
Brain_MD_7	Classic	2	0	0
Brain_MD_59	Classic	3	2	1
Brain_MD_20	Classic	1	1	1
Brain_MD_21	Classic	4	1	1
Brain_MD_50	Classic	1	1	1
Brain_MD_49	Classic	0	0	0
Brain_MD_45	Classic	1	1	1
Brain_MD_43	Classic	1	1	1
Brain_MD_8	Classic	1	1	1
Brain_MD_42	Classic	2	0	0
Brain_MD_1	Classic	4	0	0
Brain_MD_4	Classic	2	0	0
Brain_MD_55	Classic	2	0	0
Brain_MD_41	Classic	1	1	1
Brain_MD_37	Classic	3	2	1
Brain_MD_3	Classic	2	0	0
Brain_MD_34	Classic	4	0	0
Brain_MD_29	Classic	1	1	1
Brain_MD_13	Classic	2	0	0
Brain_MD_24	Classic	2	0	0
Brain_MD_65	Classic	1	1	1
Brain_MD_5	Classic	3	2	1
Brain_MD_66	Classic	3	2	1
Brain_MD_67	Classic	3	2	1
Brain_MD_58	Classic	2	0	0
Brain_MD_53	Desmoplastic	0	0	0
Brain_MD_56	Desmoplastic	0	0	0
Brain_MD_16	Desmoplastic	0	0	0
Brain_MD_40	Desmoplastic	0	2	1
Brain_MD_35	Desmoplastic	0	0	0
Brain_MD_30	Desmoplastic	0	0	0
Brain_MD_23	Desmoplastic	0	0	0
Brain_MD_28	Desmoplastic	4	1	0
Brain_MD_60	Desmoplastic	3	2	1

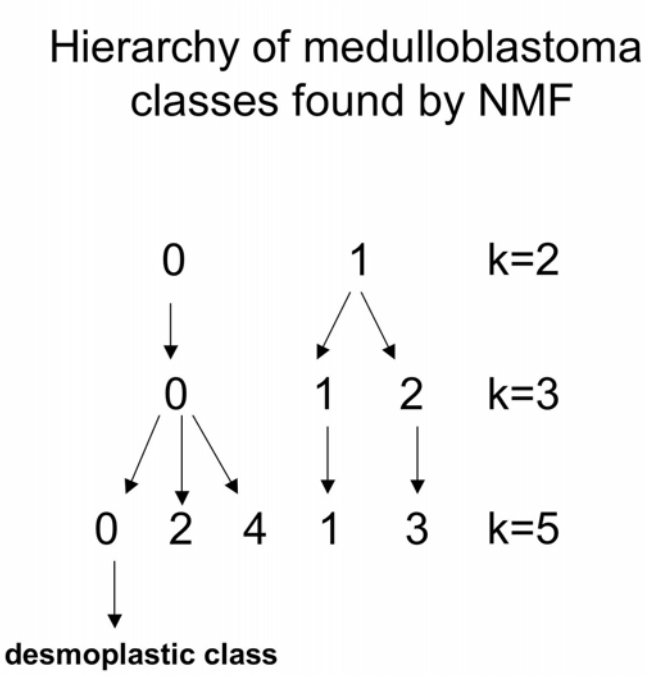

Figure S3: Approximate nesting pattern observed for medulloblastoma classes discovered by NMF. The rank-4 factorization, which model selection does not predict to be robust, is omitted.

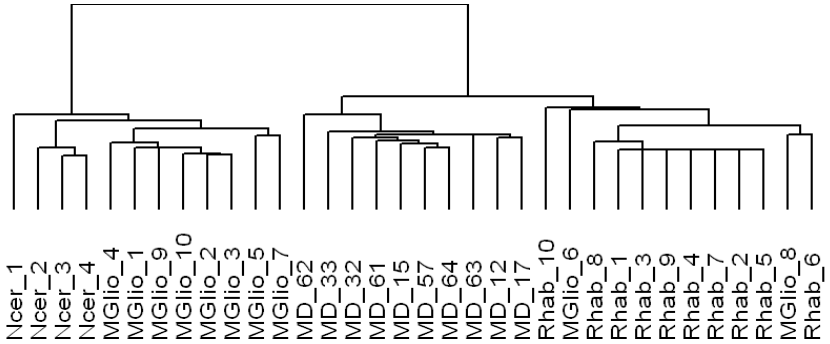

Figure S4: NMF analysis of central nervous system embryonal tumors. The data is dataset A from Pomeroy et al. *Gene expression-based classification and outcome prediction of central nervous system embryonal tumor*', Nature 415, 2002, 436-442, and is available at http://www.broad.mit.edu/cancer. We extracted a set of 34 samples including 10 classic medulloblastomas (MD), 10 malignant gliomas (Mglio), 10 Rhabdoids (Rhab), and 4 normals (Ncer). Dendrogram from hierarchical clustering in Figure 7a of main paper with sample labels included.

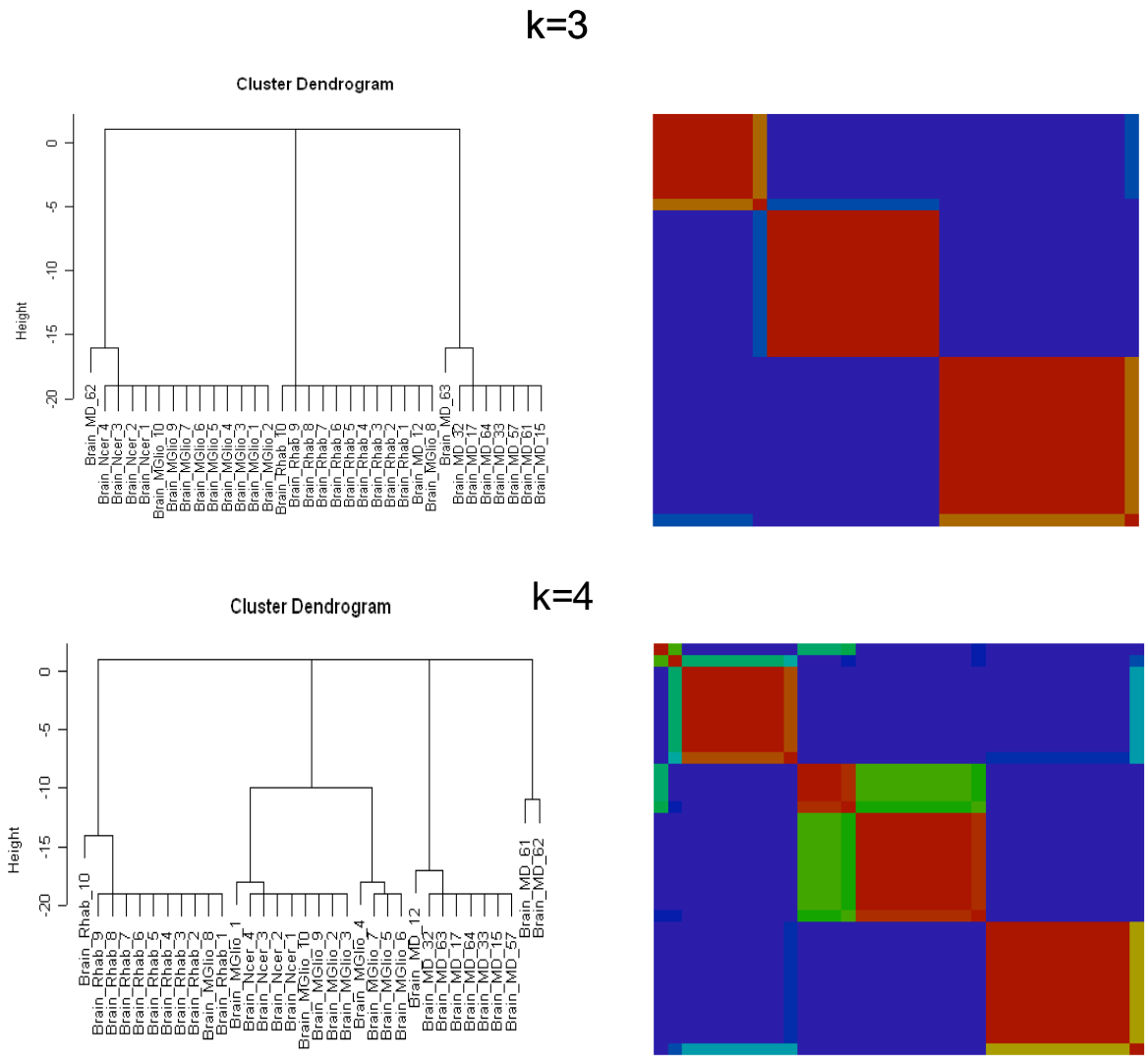

Figure S5: This picture shows more details for the k=3 (3-centroid) and k = 4 (4-centroid) SOM results shown in Figure 7c in the main paper: clustering consensus clustering matrix (right) and associated dendrogram (left).

Class Assignment of Central Nervous System Tumors

samples	type	class
Brain_MD_12	medulloblastoma	3
Brain_MD_61	medulloblastoma	3
Brain_MD_15	medulloblastoma	3
Brain_MD_57	medulloblastoma	3
Brain_MD_33	medulloblastoma	3
Brain_MD_64	medulloblastoma	3
Brain_MD_17	medulloblastoma	3
Brain_MD_62	medulloblastoma	3
Brain_MD_63	medulloblastoma	3
Brain_MD_32	medulloblastoma	3
Brain_MGlio_1	malignant glioma	4
Brain_MGlio_2	malignant glioma	4
Brain_MGlio_3	malignant glioma	4
Brain_MGlio_4	malignant glioma	4
Brain_MGlio_5	malignant glioma	4
Brain_MGlio_6	malignant glioma	4
Brain_MGlio_7	malignant glioma	4
Brain_MGlio_8	malignant glioma	1
Brain_MGlio_9	malignant glioma	4
Brain_MGlio_10	malignant glioma	4
Brain_Rhab_1	rhabdoid	1
Brain_Rhab_2	rhabdoid	1
Brain_Rhab_3	rhabdoid	1
Brain_Rhab_4	rhabdoid	1
Brain_Rhab_5	rhabdoid	1
Brain_Rhab_6	rhabdoid	1
Brain_Rhab_7	rhabdoid	1
Brain_Rhab_8	rhabdoid	1
Brain_Rhab_9	rhabdoid	1
Brain_Rhab_10	rhabdoid	2
Brain_Ncer_1	normal	2
Brain_Ncer_2	normal	2
Brain_Ncer_3	normal	2
Brain_Ncer_4	normal	2

Figure S6: Class assignments for rank k=4 NMF clustering shown in Figure 7b of main paper.