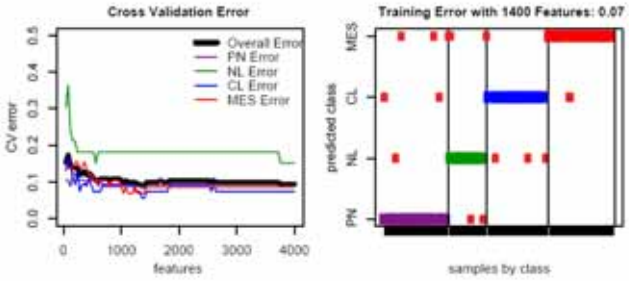
Supplemental Information

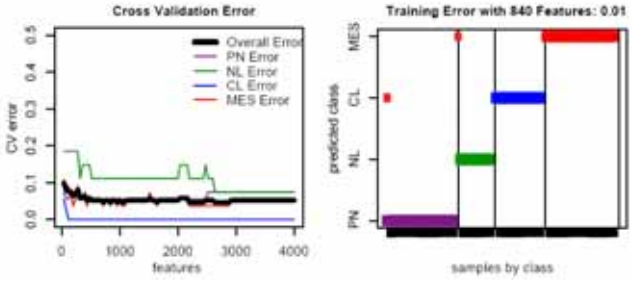
Inventory

Verhaak et al., Figure S1

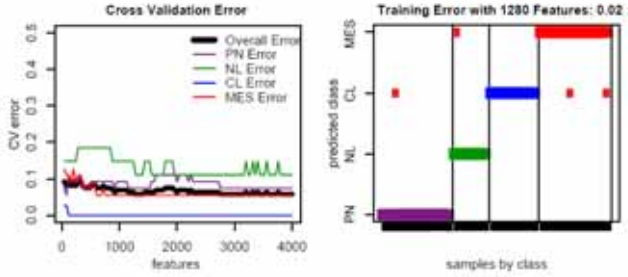
A Unified - all samples



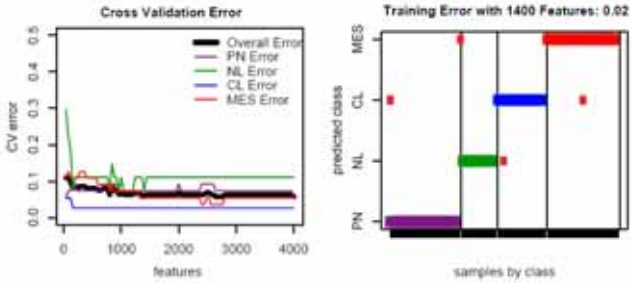
B Unified - core samples



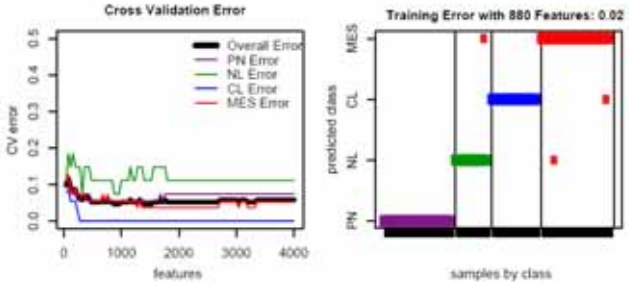
C Affymetrix U133A



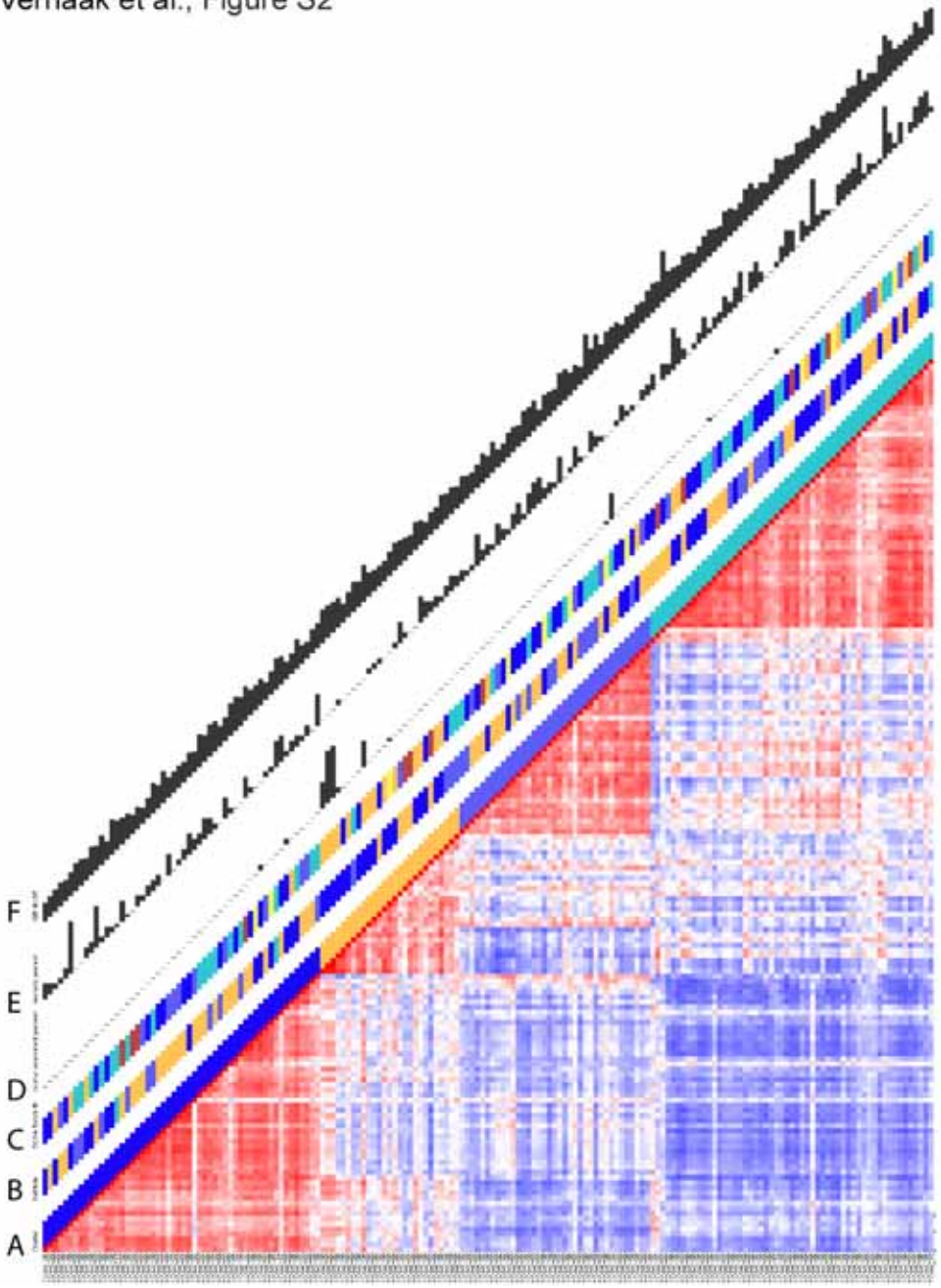
D Affymetrix Exon

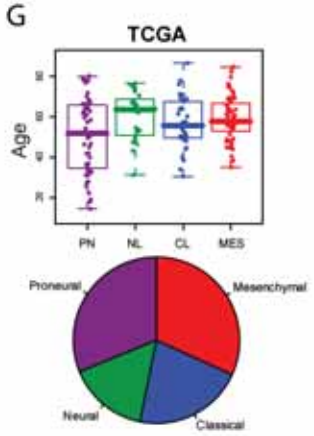


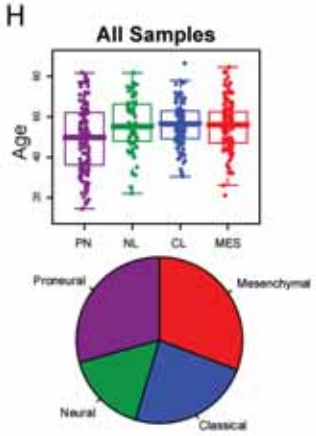
E Agilent 244K

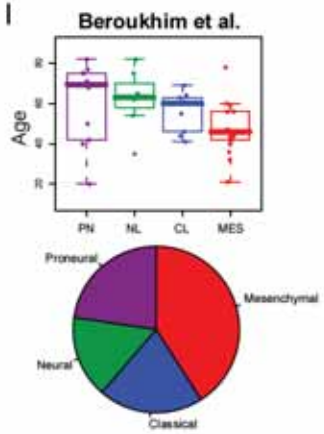


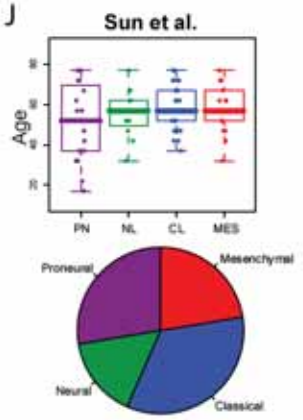
Verhaak et al., Figure S2

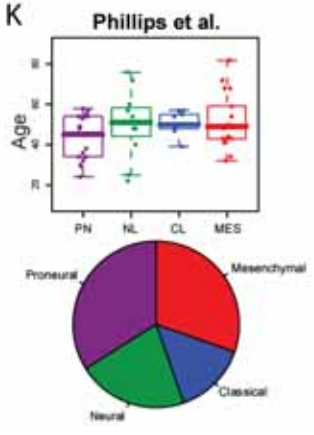


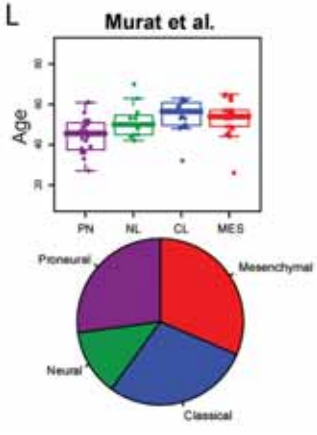












Verhaak et al., Figure S3

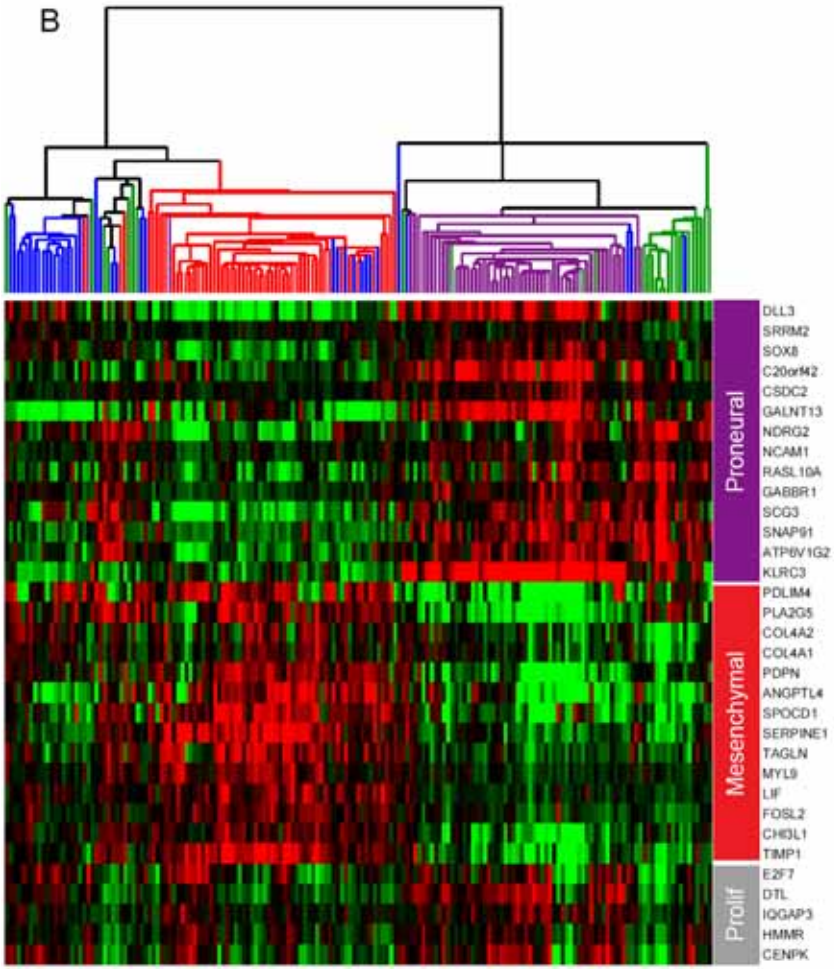
A

		Phillips et al. Subtype		
		PN	MES	PROLIF
TCGA Subtypes	PN	35	1	18
	NL	20	5	2
	CL	6	20	11
	MES	1	50	4

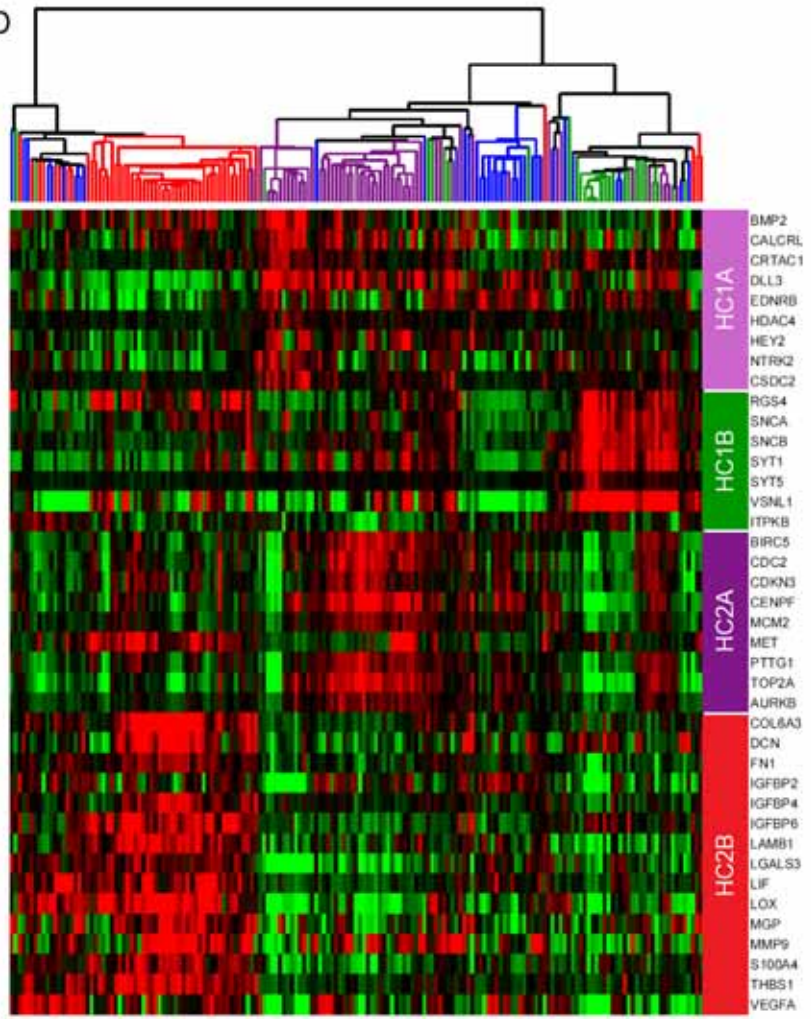
C

		Freije et al. Subtype			
		HC1A	HC1B	HC2A	HC2B
TCGA Subtypes	PN	20	1	32	1
	NL	12	11	3	1
	CL	7	5	10	15
	MES	0	3	0	52

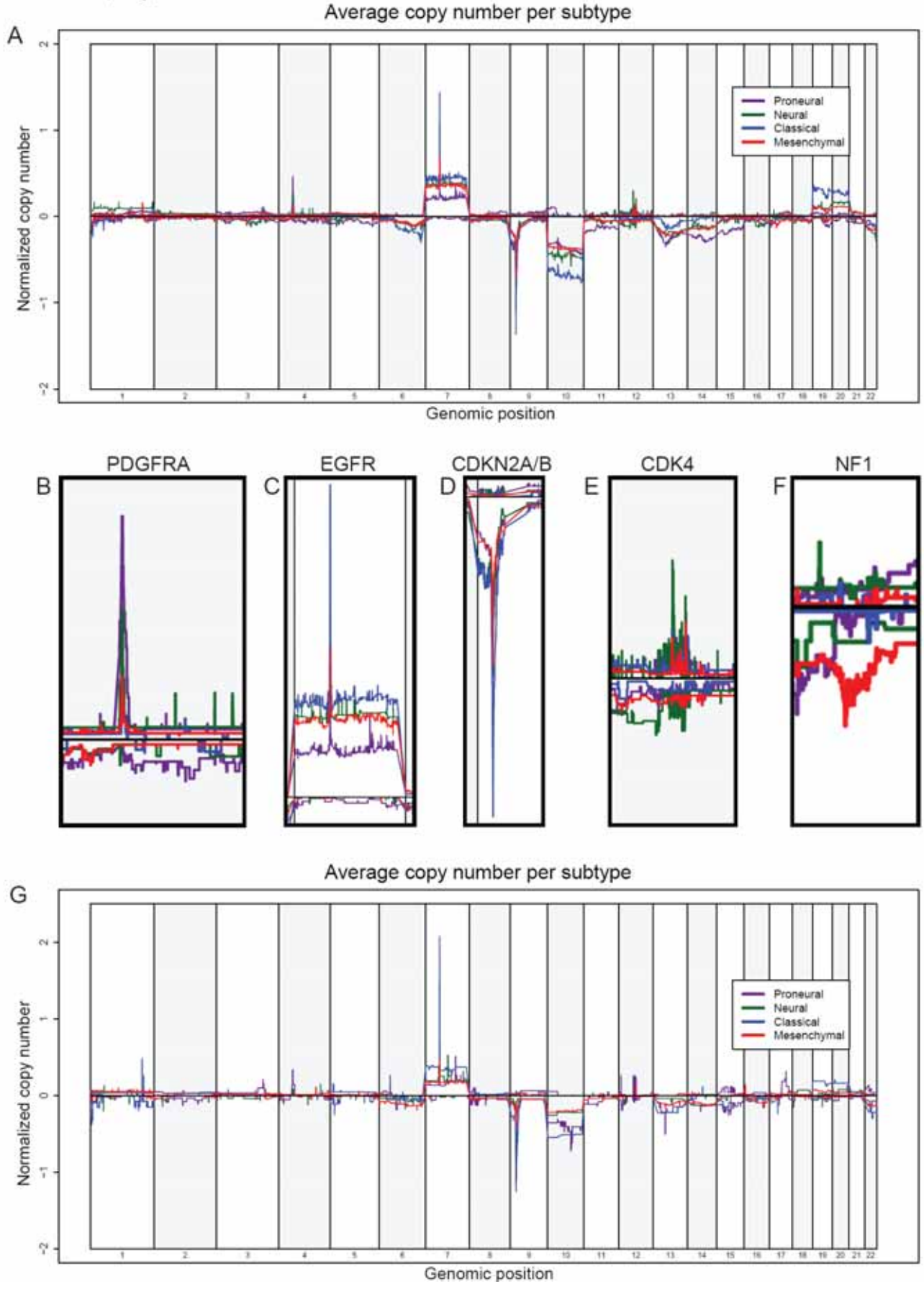
B



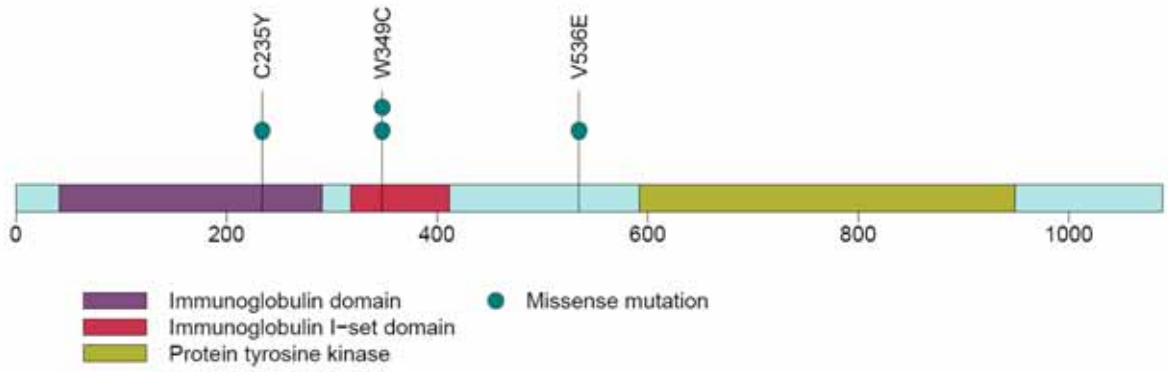
D



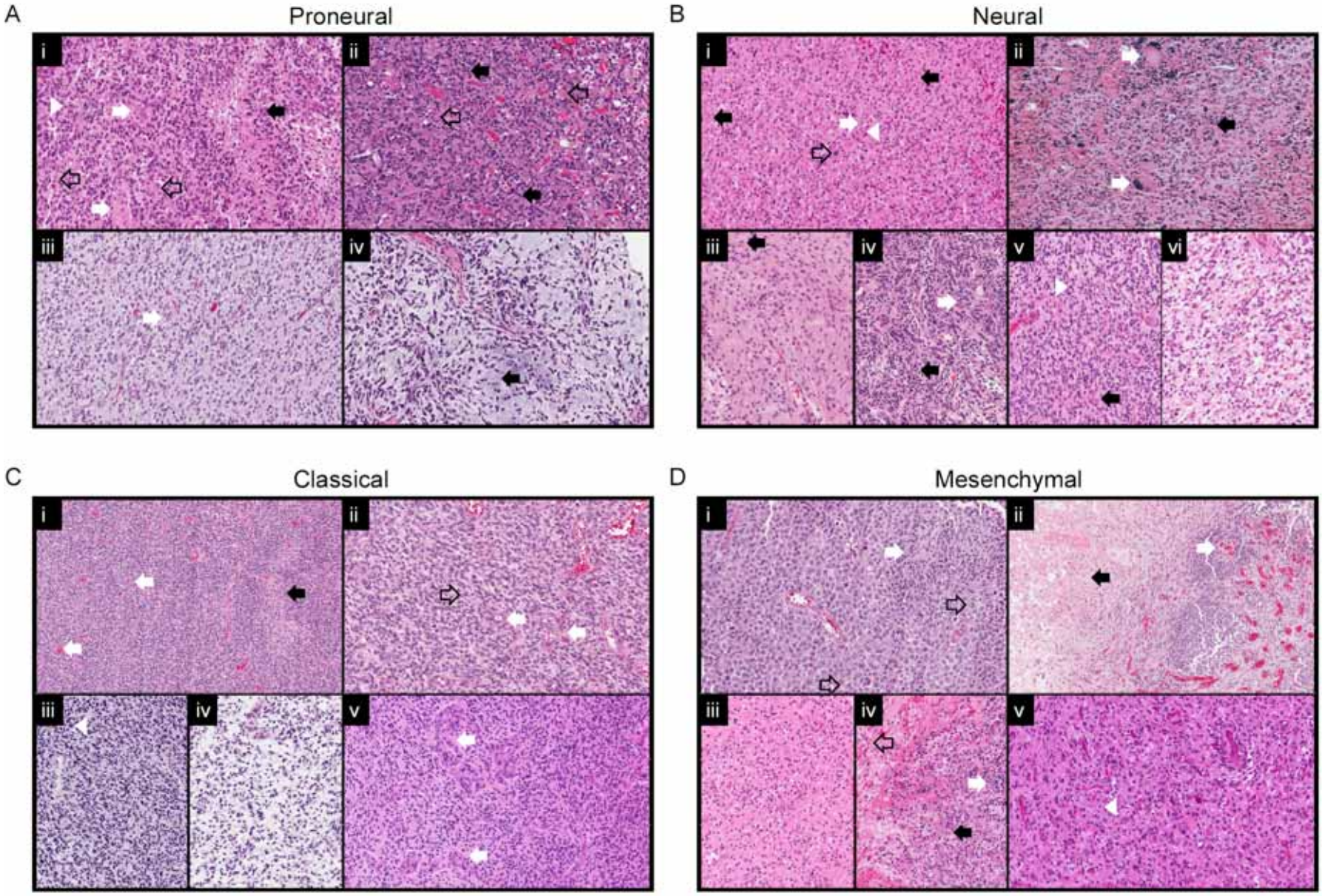
Verhaak et al., Figure S4



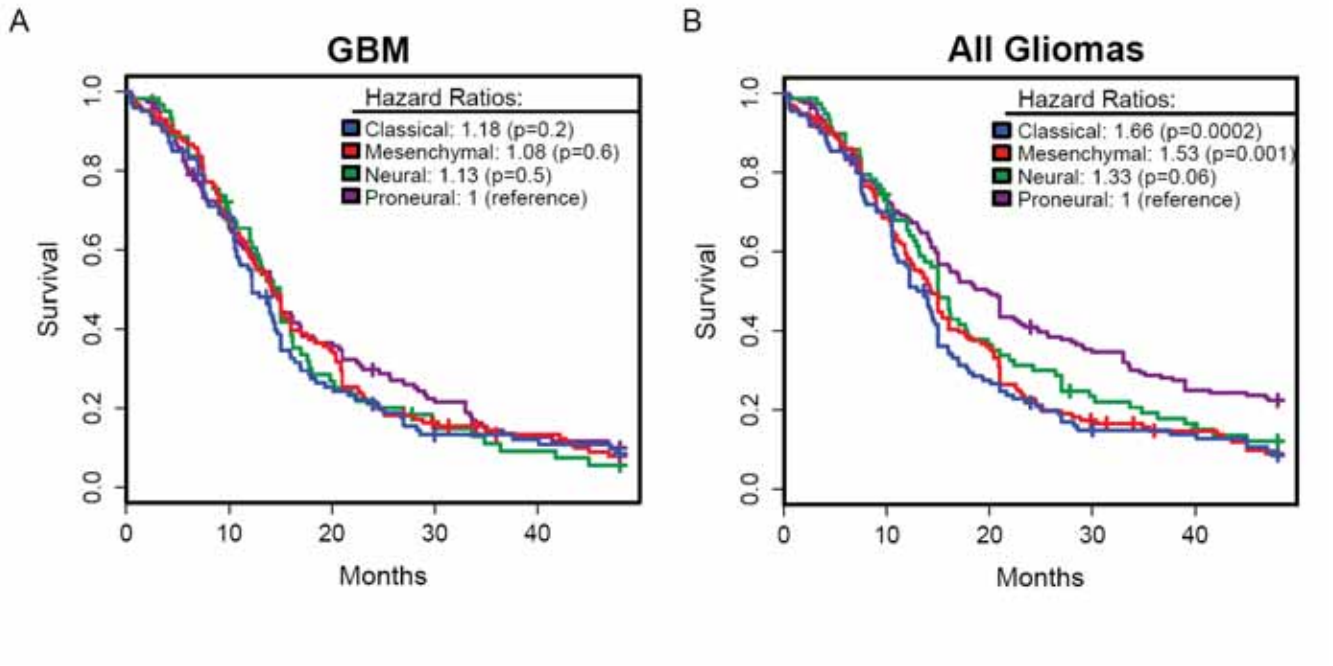
Verhaak et al., Figure S5



Verhaak et al., Figure S6



Verhaak et al., Figure S7



Supplemental Figure Legends

Figure S1. Improved cross validation and training error rates using unified core dataset versus each individual dataset. Cross validation and training error rates for (A) Entire unified dataset (n=202), (C) Core unified dataset (n=173), (C) Core Affymetrix U133A, (D) Core Affymetrix-HuEx, and (E) Core Agilent 244K Custom.

Figure S2. (A-F) Sample-sample heatmap showing pair-wise correlation between all samples. The colors of the cells relate to Pearson's correlation coefficient values, with deeper colors indicating higher positive (red) or negative (blue) correlations. Additional information is shown alongside the diagonal. (A) Sample subtype; Proneural in blue, Neural in orange, Classical in purple, Mesenchymal in cyan. (B) Sample providing institute; Henry Ford Hospital in blue, MD Anderson Cancer Center in orange, University of California San Francisco in purple, Duke University in cyan. (C) Batch; batch 1 in purple, batch 2 in brown, batch 3 in yellow, batch 4 in purple, batch 5 in blue, batch 6 in turquoise. (D) The percentage normal tissue in the sample, scaled to a maximum of 100%. (E)Percentage necrosis, scaled to a maximum of 50%. (F) IQR NUSE, a metric that indicates the amount of random noise in an expression profile, scaled to a maximum of 0.10. (G-L) Subtype proportions and association with age stratified by dataset. Overall, and within the TCGA samples, age was correlated with the Proneural subtype. The proportions of the subtypes are shown for each dataset and were approximately equal. PN, Proneural; NL, Neural; CL, Classical; MES, Mesenchymal.

Figure S3. Comparison of TCGA with previously published glioma subtypes. (A) Matrix of subtype calls for TCGA core samples (n=173) as predicted by TCGA or Phillips et al. centroids (B) TCGA Core samples clustered with Phillips et al. 35 gene signature. To obtain the best coverage of the Phillips gene signature, gene expression data for 33 of the 35 gene signatures were mapped to the Agilent data platform. (C) Matrix of subtype calls for TCGA core samples (n=173) as predicted by TCGA or Freije et al. gene lists. (D) TCGA Core samples clustered with Freije et al. 44 gene signature. Gene expression data for 40 of the 44 gene signatures were mapped to the Unified dataset. Genes were ordered according to Freije et al. and samples were hierarchically clustered. Samples were color coded based on the ClaNC subtype predictions (Proneural, purple; Neural, green; Classical, blue; Mesenchymal, red).

Figure S4. Genome wide view of copy number alterations in the four major subtypes of glioblastoma. Average copy number was determined for positive and negative copy number estimates separately using 244,000 probes and 150 tumor samples derived from the largest single copy number dataset collected as part of TCGA project (MSKCC 244K Agilent aCGH platform). Normalized copy number was the $\log_2$ ratio to copy number 2. (A) Average copy number calculated for each the four major subtypes of GBM. Genome regions with alterations seen significantly more frequent in one of the subtypes are enlarged. (B) *PDGFRA*, (C) *EGFR*, (D) *CDKN2A/B*, (E) *CDK4*, (F) *NF1*. (G) Average copy number calculated for each of the four major subtypes of GBM as predicted in the Beroukhim et al. dataset (Beroukhim et al., 2007), one of the four datasets in the Validation dataset.

Figure S5, related to Table 3. Gene structure of *PDGFRA* indicating missense mutations found in the TCGA sample set.

Figure S6. Variable histopathological features of each GBM subtype (A) \Proneural. One tumor (i) contained neoplastic fibrillary and gemistocytic (open arrows) astrocytes, numerous mitoses (arrowhead), widespread endothelial hyperplasia (white arrows), and pseudopalisading necrosis (black arrow). A second tumor (ii) contained predominantly neoplastic gemistocytic astrocytes (open arrows), a lymphocytic infiltrate (black arrows), focal endothelial hyperplasia (not shown), and geographic, infarct-like necrosis (not shown). A third tumor (iii, iv) contained neoplastic fibrillary astrocytes that diffusely infiltrated the neocortex (iii) and featured perineuronal satellitosis (iii, white arrow), focal endothelial hyperplasia (not shown), and no necrosis. Distinct areas of this tumor contained a myxoid background (iv, black arrow). (B) Neural. One tumor (i, ii) contained neoplastic fibrillary (i, open arrow) and multinucleated (giant cell) astrocytes (ii, white arrows), minigemistocytes (i, black arrows), and oligodendroglia (i, white arrows), the latter of which were characterized by round, regular nuclei with crisp nuclear membranes and a prominent nucleolus. Scattered mitoses (i, arrowhead), perivascular lymphocytes (ii, black arrow), focal endothelial hyperplasia (not shown), and no necrosis were evident in this tumor. A second tumor (iii, iv) contained neoplastic fibrillary (iii) and gemistocytic (iii, black arrow) astrocytes, a prominent lymphocytic infiltrate (iv, black arrow), widespread endothelial hyperplasia (iv, white arrow), and focal geographic, infarct-like necrosis (not shown). A third tumor (v, vi) contained densely (v) and loosely (i) packed neoplastic fibrillary astrocytes and featured scattered mitoses (v, arrowhead), florid endothelial hyperplasia (v, black arrows), and focal pseudopalisading necrosis (not shown). (C) Classical. One tumor (i, ii) contained neoplastic fibrillary astrocytes (i) and oligodendroglia (ii, open arrows), florid endothelial hyperplasia (i, ii, white arrows), and pseudopalisading necrosis (i, black arrow). A second tumor (iii, iv) contained densely (iii) and loosely (iv) packed neoplastic fibrillary astrocytes and featured abundant mitoses (iii, arrowhead) and both geographic and pseudopalisading necrosis (not shown). A third tumor (v) contained neoplastic fibrillary astrocytes and florid endothelial hyperplasia (white arrows). (D) Mesenchymal. One tumor (i, ii) contained predominantly neoplastic gemistocytic astrocytes (i, open arrow) and featured focal endothelial hyperplasia (ii, white arrow), geographic, infarct-like necrosis (ii, black arrow), and focal pseudopalisading necrosis (i, white arrow). A second tumor (iii, iv) contained neoplastic fibrillary astrocytes (iii), a prominent lymphocytic infiltrate (iv, white arrow), widespread endothelial hyperplasia (iv, black arrow), and focal geographic, infarct-like necrosis (not shown). A third tumor (v) contained neoplastic fibrillary and gemistocytic astrocytes, scattered mitoses (v, arrowhead), focal endothelial hyperplasia (not shown), and no necrosis.

Figure S7. Survival analysis according to GBM subtype. Data for the TCGA dataset and the validation dataset were combined for survival analysis. (A) Glioblastoma samples (B) Glioblastoma and lower grade gliomas.

Table S1. Frequencies and associated p-values from a Fisher Exact two sided tests of copy number alterations and GBM subtype and Chi-squared Tests of Independence. (see excel file).

Table S2. Frequencies and associated Fisher Exact two sided test p-values of mutations and GBM subtype. Only mutations found in at least four samples were tested.

Gene	Proneural (n=37)	Neural (n=19)	Classical (n=22)	Mesenchymal (n=38)	Total # Mut	Chi-square Test of Indep.
TP53	20 (0.10)	4 (>0.20)	**0 (0.04)**	12 (>0.20)	36	<0.01
PTEN	6 (>0.20)	4 (>0.20)	5 (>0.20)	12 (>0.20)	27	>0.20
NF1	2 (>0.20)	3 (>0.20)	1 (>0.20)	**14 (0.07)**	20	0.05
EGFR	6 (>0.20)	5 (>0.20)	7 (>0.20)	2 (>0.20)	20	>0.20
IDH1	**11 (<0.01)**	1 (>0.20)	0 (>0.20)	0 (>0.20)	12	<0.01
PIK3R1	7 (>0.20)	2 (>0.20)	1 (0.28)	0 (>0.20)	10	>0.20
DST	3 (>0.20)	0 (>0.20)	5 (>0.20)	1 (>0.20)	9	>0.20
RB1	1 (>0.20)	1 (>0.20)	0 (>0.20)	5 (>0.20)	7	>0.20
ERBB2	2 (>0.20)	3 (>0.20)	1 (>0.20)	1 (>0.20)	7	>0.20
EGFRvIII	1 (>0.20)	0 (>0.20)	5 (>0.20)	1 (>0.20)	7	>0.20
PIK3CA	3 (>0.20)	1 (>0.20)	1 (>0.20)	1 (>0.20)	6	>0.20
FKBP9	3 (>0.20)	1 (>0.20)	2 (>0.20)	0 (>0.20)	6	>0.20
SYNE1	3 (>0.20)	0 (>0.20)	1 (>0.20)	1 (>0.20)	5	>0.20
PDGFRA	4 (>0.20)	0 (>0.20)	0 (>0.20)	0 (>0.20)	4	>0.20

Shaded rows correspond to regions with significant chi-squared tests after adjusting for multiple testing. Numbers between parentheses represent adjusted p-values of the probability that the number of events in a specific subtype is different than that of the remaining samples combined (two sided Fisher's exact test). Bolded entries show significant differences.

Table S3. Gene ontology analysis for genes highly expressed in GBM Subtypes from the ClaNC 840 list. (see excel file)

Table S4. Distribution of co-mutations across GBM subtypes.

Gene	Proneural n=37	Neural n=19	Classical n=22	Mesenchymal n=38	Total # Mut
PTEN_TP53	3 (8%)	1 (5%)	0 (0%)	7 (18%)	11
IDH1_TP53	9 (24%)	1 (5%)	0 (0%)	0 (0%)	10
EGFR_TP53	4 (11%)	2 (11%)	0 (0%)	1 (3%)	7
NF1_PTEN	0 (0%)	1 (5%)	0 (0%)	6 (16%)	7
PIK3R1_TP53	5 (14%)	1 (5%)	0 (0%)	0 (0%)	6
NF1_TP53	1 (3%)	0 (0%)	0 (0%)	4 (11%)	5
PTEN_RB1	1 (3%)	1 (5%)	0 (0%)	3 (8%)	5
EGFR_PTEN	2 (5%)	0 (0%)	2 (9%)	0 (0%)	4
RB1_TP53	0 (0%)	1 (5%)	0 (0%)	3 (8%)	4

Table S5. Clinical Analysis statistical tests and p-values

Variable	Test Performed	p-value
Age*	2-way, unbalanced ANOVA (controlling for collection center)	0.04 (TCGA) 0.06 (TCGA—no Duke) 0.07 (Validation)
	1-way ANOVA	0.01 (TCGA) 0.05 (Validation)
	Multinomial generalized linear model	<0.001 (TCGA) 0.01 (Validation)
Karnofsky Score	2-way, unbalanced ANOVA	0.23 (TCGA)
	1-way ANOVA	0.04 (TCGA)
	Multinomial generalized linear model	<0.001 (TCGA)
Sex	Chi-squared test of independence	0.13 (TCGA) 0.50 (Validation)
Survival	Mantel-Haenszel test	0.90 (TCGA) 0.40 (Validation)
MGMT Methylation	Chi-squared test of independence	0.54 (TCGA)
Secondary or Recurrent	Chi-squared test of independence	0.33 (TCGA)
TCGA Batch	Chi-squared test of Independence	0.24 (TCGA)
Collection Center*	Chi-squared test of Independence	0.01 (TCGA)
% Necrosis*	2-way ANOVA, after logit transformation, controlling for collection center	0.01 (TCGA)
% Tumor Nuclei	2-way ANOVA, after logit transformation, controlling for collection center	0.68 (TCGA)

Table S6. Number of bases sequenced per sample per gene. (see excel sheet)

Table S7. Patient Characteristics. (See excel sheet)

Experimental Procedures

Factor Analysis We assume, *per gene*, a standard factor analysis model for the relationship of the standardized gene measurement (*y*) to the true underlying gene expression (*x*),

$$y_{ai} = \beta_a x_i + \varepsilon_{ai},$$

$$y_{ei} = \beta_e x_i + \varepsilon_{ei},$$

$$y_{ui} = \beta_u x_i + \varepsilon_{ui},$$

where *i* indexes a sample, *a, e,* and *u* index the Agilent, Exon and U133 platforms, respectively, and ε is the platform specific noise. The x_i are assumed to have a normal distribution with mean zero and standard deviation one; the vector of error terms ε_i is assumed to have a multivariate normal distribution with zero mean and a diagonal covariance matrix, **Ψ**. The parameters **β** and **Ψ** are unknown and are estimated using an EM algorithm within the package `factanal`. With these estimates and a given vector ***y*** of observed measurements, estimates of *x* are a linear combination of the observed *y* measurements given by,

$$\hat{x} = \hat{\beta}^T(\hat{\beta}\hat{\beta}^T + \hat{\Psi})^{-1}y.$$

Because the inputs *y* from each platform are standardized values, the estimates $\hat{\beta}$ are exactly equivalent to the correlation across all samples of the input *y* and the estimate $\hat{x}$.

Rescaling of Unified Expression Measures Each expression platform has a different estimate of its variation between samples. To estimate a single estimate of variation, *s*, we focus on using information from only those platforms whose gene measurements were well correlated with our unified estimate $\hat{x}$. When possible, we estimated the rescaling factor *s* as the average MAD value of the platforms that had $\hat{\beta} \geq 0.70$ (possibly only a single platform). This gave an estimate of *s* for all but 631 genes (and all of

the non-filtered genes used in the estimates of clusters). To obtain an estimate for all 11,861 genes with unified estimates (used in other parts of the analysis) we lowered the cutoff to $\hat{\beta} \geq 0.50$. This gave an estimate of *s* for all but 158 genes. For these remaining genes, we just took the average MAD across all three platforms.

Gene Filtering. As explained in the main text, three filters were applied to the genes. The first filter kept only genes in which at least two of the platforms had a high correlation with the unified gene expression estimates ($\hat{\beta} \geq 0.70$ for at least two platforms). 9,255 genes passed this criterion. The goal of this filter was not to reduce the analysis to a small number of genes, but rather to eliminate poorly-behaving genes. The second filter eliminated genes that did not have large variability across patients; this was the main filter that was used to reduce the number of genes to a reasonable size for clustering. We filtered based on the individual MAD values per platform. Specifically, we required that the MAD of each platform be greater than 0.5, but only in those platforms that had a high correlation with the unified expression value ($\hat{\beta} \geq 0.50$) and thus had a reasonably high contribution to the unified gene expression estimate. We chose a slightly smaller $\hat{\beta}$ value here because even for genes with $\hat{\beta} \approx 0.50$, the contribution of the platform to unified gene expression value can be as much as 20% (though for genes that pass the first filter, this is not generally a problem). This resulted in 2,120 that satisfied this requirement. Combined with the first filter based on the $\hat{\beta}$ values this gave 1,903 genes. The last filter removed genes with extremely variable estimates of standard deviation. Specifically, we excluded genes if the ratio of the individual MAD and the averaged MAD (or its inverse) was greater than 1.5 (again restricted to platforms with $\hat{\beta} \geq 0.50$). 8,795 genes passed this criteria and in combination with the previous two filters, this resulted in 1,740 genes.

SAM/ROC

For the SAM statistics table, SAM (package siggenes version 1.12.0) is used to identify genes that are differentially expressed in bi-classifications of one subtype vs. all the others. The SAM *d* statistic is shown in the table for each gene and each subtype.

For the ROC table, the ROC (Receiver Operating Characteristic) curves are drawn as follows. The expression values for each gene are used to order the samples. If a sample belongs to a particular subtype, it is considered true, and false otherwise. The AUC (Area Under the Curve) is then calculated for each gene and each subtype. R package 'ROC' version 1.12.0 is used for the calculations.

Comparison of Subtypes with Previous High-Grade Glioma Classifications. Prior profiling studies have identified subgroups of high grade gliomas (Freije et al., 2004; Phillips et al., 2006). Using the signatures from Phillips et al and Freije et al, subtypes were predicted on the TCGA core samples (n=173). For the Phillips et al subtype prediction, we used the Agilent platform as many of the genes that represented the Proliferation subgroup were not present on the unified data set. We used a single platform, Agilent, to compare the 33/35 gene signature with our four subgroups. One Proneural gene and Mesenchymal gene were not mapped to the Agilent platform. For the Frieje et al subtype prediction, 338 genes from the 595 probe sets were identified and used in the unified data set. Subtype predictions on the TCGA data was performed using single sample predictor (Hu et al., 2006). TCGA core samples were clustered with Phillips et al. 35 gene signature or Freije et al. reduced 44 gene signature and color coded according to TCGA four GBM subtypes. High similarity was observed for clusters that identify the Proneural and Mesenchymal subtypes in both the Phillips and Freije gene signatures (Figure S3). Freije et al. HC1B gene set was similar to the Neural subtype. However, neither signature contained a set of genes with high expression limited to the Classical subtype.

References

Freije, W. A., Castro-Vargas, F. E., Fang, Z., Horvath, S., Cloughesy, T., Liau, L. M., Mischel, P. S., and Nelson, S. F. (2004). Gene expression profiling of gliomas strongly predicts survival. Cancer Res *64*, 6503-6510.

Hu, Z., Fan, C., Oh, D. S., Marron, J. S., He, X., Qaqish, B. F., Livasy, C., Carey, L. A., Reynolds, E., Dressler, L.*, et al.* (2006). The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics *7*, 96.

Phillips, H. S., Kharbanda, S., Chen, R., Forrest, W. F., Soriano, R. H., Wu, T. D., Misra, A., Nigro, J. M., Colman, H., Soroceanu, L.*, et al.* (2006). Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. Cancer Cell *9*, 157-173.

The Cancer Genome Atlas Research Network

Tissue source sites: Duke University Medical School
Roger McLendon[1], Allan Friedman[2] & Darrell Bigner[1]

Emory University
Erwin G. Van Meir[3,4,5], Daniel J. Brat[5,6], Gena M. Mastrogianakis[3] & Jeffrey J. Olson[3,4,5]

Henry Ford Hospital
Tom Mikkelsen[7] & Norman Lehman[8]

MD Anderson Cancer Center
Ken Aldape[9], W. K. Alfred Yung[10] & Oliver Bogler[11]

University of California San Francisco
Scott VandenBerg[12], Mitchel Berger[13] & Michael Prados[13]

Genome sequencing centers: Baylor College of Medicine
Donna Muzny[14], Margaret Morgan[14], Steve Scherer[14], Aniko Sabo[14], Lynn Nazareth[14], Lora Lewis[14], Otis Hall[14], Yiming Zhu[14], Yanru Ren[14], Omar Alvi[14], Jiqiang Yao[14], Alicia Hawes[14], Shalini Jhangiani[14], Gerald Fowler[14], Anthony San Lucas[14], Christie Kovar[14], Andrew Cree[14], Huyen Dinh[14], Jireh Santibanez[14], Vandita Joshi[14], Manuel L. Gonzalez-Garay[14], Christopher A. Miller[14,15], Aleksandar Milosavljevic[14,15,16], Larry Donehower[17], David A. Wheeler[14] & Richard A. Gibbs[14]

Broad Institute of MIT and Harvard
Kristian Cibulskis[18], Carrie Sougnez[18], Tim Fennell[18], Scott Mahan[18], Jane Wilkinson[18], Liuda Ziaugra[18], Robert Onofrio[18], Toby Bloom[18], Rob Nicol[18], Kristin Ardlie[18], Jennifer Baldwin[18], Stacey Gabriel[18] & Eric S. Lander[18,19,20]

Washington University in St Louis
Li Ding[21], Robert S. Fulton[21], Michael D. McLellan[21], John Wallis[21], David E. Larson[21], Xiaoqi Shi[21], Rachel Abbott[21], Lucinda Fulton[21], Ken Chen[21], Daniel C. Koboldt[21], Michael C. Wendl[21], Rick Meyer[21], Yuzhu Tang[21], Ling Lin[21], John R. Osborne[21], Brian H. Dunford-Shore[21], Tracie L. Miner[21], Kim Delehaunty[21], Chris Markovic[21], Gary Swift[21], William Courtney[21], Craig Pohl[21], Scott Abbott[21], Amy Hawkins[21], Shin Leong[21], Carrie Haipek[21], Heather Schmidt[21], Maddy Wiechert[21], Tammi Vickery[21], Sacha Scott[21], David J. Dooling[21], Asif Chinwalla[21], George M. Weinstock[21], Elaine R. Mardis[21] & Richard K. Wilson[21]

Cancer genome characterization centres: Broad Institute/Dana-Farber Cancer Institute
Gad Getz[18], Wendy Winckler[18,22,23], Roel G. W. Verhaak[18,22,23], Michael S. Lawrence[18], Michael O'Kelly[18], Jim Robinson[18], Gabriele Alexe[18], Rameen Beroukhim[18,22,23], Scott Carter[18], Derek Chiang[18,22], Josh Gould[18], Supriya Gupta[18], Josh Korn[18], Craig Mermel[18,22], Jill Mesirov[18], Stefano Monti[18], Huy Nguyen[18], Melissa Parkin[18], Michael Reich[18], Nicolas Stransky[18], Barbara A. Weir[18,22,23], Levi Garraway[18,22,23], Todd Golub[18,22,23] & Matthew Meyerson[18,22,23]

Harvard Medical School/Dana-Farber Cancer Institute
Lynda Chin[22,24,25], Alexei Protopopov[24], Jianhua Zhang[24], Ilana Perna[24], Sandy Aronson[26], Narayan Sathiamoorthy[26], Georgia Ren[24], Sachet Shukla[24], W. Ruprecht Wiedemeyer$_{22}$, Hyunsoo Kim[26], Sek Won Kong[27,28], Yonghong Xiao[24], Isaac S. Kohane[26,27,29], Jon Seidman[30], Peter J. Park[26,27,29] & Raju Kucherlapati[26]

Johns Hopkins/University of Southern California
Peter W. Laird[31], Leslie Cope[32], James G. Herman[33], Daniel J. Weisenberger[31], Fei Pan[31], David Van Den Berg[31], Leander Van Neste[34], Joo Mi Yi[33], Kornel E. Schuebel[33] & Stephen B. Baylin[33]

HudsonAlpha Institute/Stanford University
Devin M. Absher[35], Jun Z. Li[36], Audrey Southwick[37], Shannon Brady[37], Amita Aggarwal[37], Tisha Chung[37], Gavin Sherlock[37], James D. Brooks[38] & Richard M. Myers[35]

Lawrence Berkeley National Laboratory
Paul T. Spellman[39], Elizabeth Purdom[40], Lakshmi R. Jakkula[39], Anna V. Lapuk[39], Henry Marr[39], Shannon Dorton[39], Yoon Gi Choi[41], Ju Han[39], Amrita Ray[39], Victoria Wang[40], Steffen Durinck[39], Mark Robinson[42], Nicholas J. Wang[39], Karen Vranizan[41], Vivian Peng[41], Eric Van Name[41], Gerald V. Fontenay[39], John Ngai[41], John G. Conboy[39], Bahram Parvin[39], Heidi S. Feiler[39], Terence P. Speed[40,42] & Joe W. Gray[39]

Memorial Sloan-Kettering Cancer Center
Cameron Brennan[43], Nicholas D. Socci[44], Adam Olshen[45], Barry S. Taylor[44,46], Alex Lash[44], Nikolaus Schultz[44], Boris Reva[44], Yevgeniy Antipin[44], Alexey Stukalov[44], Benjamin Gross[44], Ethan Cerami[44], Wei Qing Wang[44], Li-Xuan Qin[45], Venkatraman E. Seshan[45], Liliana Villafania[47], Magali Cavatore[47], Laetitia Borsu[48], Agnes Viale[47], William Gerald[48], Chris Sander[44] & Marc Ladanyi[48]

University of North Carolina, Chapel Hill
Charles M. Perou[49,50], D. Neil Hayes[51], Michael D. Topal[50,52], Katherine A. Hoadley[49], Yuan Qi[51], Sai Balu[52], Yan Shi[52] & Junyuan Wu[52]

Biospecimen Core Resource
Robert Penny[53], Michael Bittner[54], Troy Shelton[53], Elizabeth Lenkiewicz[53], Scott Morris[53], Debbie Beasley[53] & Sheri Sanders[53]

Data Coordinating Center
Ari Kahn[55], Robert Sfeir[55], Jessica Chen[55], David Nassau[55], Larry Feng[55], Erin Hickey[55], Jinghui Zhang[56] & John N. Weinstein[57]

Project teams: National Cancer Institute
Anna Barker[58], Daniela S. Gerhard[58], Joseph Vockley[58], Carolyn Compton[58], Jim Vaught[58], Peter Fielding[58], Martin L. Ferguson[59], Carl Schaefer[56], Subhashree Madhavan[56] & Kenneth H. Buetow[56]

National Human Genome Research Institute
Francis Collins[60], Peter Good[60], Mark Guyer[60], Brad Ozenberger[60], Jane Peterson[60] & Elizabeth Thomson[60]

[1]Department of Pathology,
[2]Department of Surgery, Duke University Medical Center, Durham, North Carolina 27710, USA.
[3]Department of Neurosurgery,
[4]Department of Hematology and Medical Oncology,
[5]Winship Cancer Institute,
[6]Department of Pathology and Laboratory Medicine, Emory University School of Medicine, Atlanta, Georgia 30322, USA.
[7]Department of Neurological Surgery,
[8]Department of Pathology, Henry Ford Hospital, Detroit, Michigan 48202, USA.
[9]Department of Pathology,
[10]Department of Neuro-Oncology,
[11]Department of Neurosurgery, University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030, USA.

[12]Department of Pathology,
[13]Department of Neurosurgery, University of California San Francisco, San Francisco, California 94143, USA.
[14]Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, USA.
[15]Graduate Program in Structural and Computational Biology and Molecular Biophysics,
[16]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA.
[17]Department of Molecular Virology and Microbiology, Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, USA.
[18]The Eli and Edythe L. Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts 02142, USA.
[19]Department of Biology, Institute of Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA.
[20]Department of Systems Biology, Harvard University, Boston, Massachusetts 02115, USA.
[21]The Genome Center at Washington University, Department of Genetics, Washington University School of Medicine, St Louis, Missouri 63108, USA.
[22]Department of Medical Oncology,
[23]Center for Cancer Genome Discovery,
[24]Belfer Institute for Applied Cancer Science, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA.
[25]Department of Dermatology, Harvard Medical School, Boston, Massachusetts 02115, USA.
[26]Harvard Medical School-Partners HealthCare Center for Genetics and Genomics, Boston, Massachusetts 02115, USA.
[27]Informatics Program,
[28]Department of Cardiology, Children's Hospital, Boston, Massachusetts 02115, USA.
[29]Center for Biomedical Informatics,
[30]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA.
[31]USC Epigenome Center, University of Southern California, Los Angeles, California 90089, USA.
[32]Biometry and Clinical Trials Division,
[33]Cancer Biology Division, The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University, Baltimore, Maryland 21231, USA.
[34]Department of Molecular Biotechnology, Faculty of Bioscience and Engineering, Ghent University, Ghent B-9000, Belgium.
[35]HudsonAlpha Institute for Biotechnology, Huntsville, Alabama 35806, USA.
[36]Department of Human Genetics, University of Michigan, Ann Arbor, Michigan 48109, USA.
[37]Department of Genetics,
[38]Department of Urology, Stanford University School of Medicine, Stanford, California 94305, USA.
[39]Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.
[40]Department of Statistics,
[41]Department of Molecular and Cellular Biology, University of California at Berkeley, Berkeley, California 95720, USA.
[42]Walter and Eliza Hall Institute, Parkville, Victoria 3052, Australia.
[43]Department of Neurosurgery,
[44]Computational Biology Center, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA.
[45]Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA.
[46]Department of Physiology and Biophysics, Weill Cornell Graduate School of Medical Sciences, New York, New York 10065, USA.
[47]Genomics Core Laboratory, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA.
[48]Department of Pathology, Human Oncology and Pathogenesis Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA.
[49]Department of Genetics,
[50]Department of Pathology and Laboratory Medicine,
[51]Department of Internal Medicine, Division of Medical Oncology, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA.
[52]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA.

[53]International Genomics Consortium, Phoenix, Arizona 85004, USA.
[54]Computational Biology Division, Translational Genomics Research Institute, Phoenix, Arizona 85004, USA.
[55]SRA International, Fairfax, Virginia 22033, USA.
[56]Center For Biomedical Informatics and Information Technology, National Cancer Institute, Rockville, Maryland 20852, USA.
[57]Department of Bioinformatics and Computational Biology, M.D. Anderson Cancer Center, Houston, Texas 77030, USA.
[58]National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA.
[59]MLF Consulting, Arlington, Massachusetts 02474, USA.
[60]National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA.