

Mathematically universal and biologically consistent astrocytoma genotype encodes for transformation and predicts survival phenotype

Methods, Figs. S1–S18, and Tables S1–S3

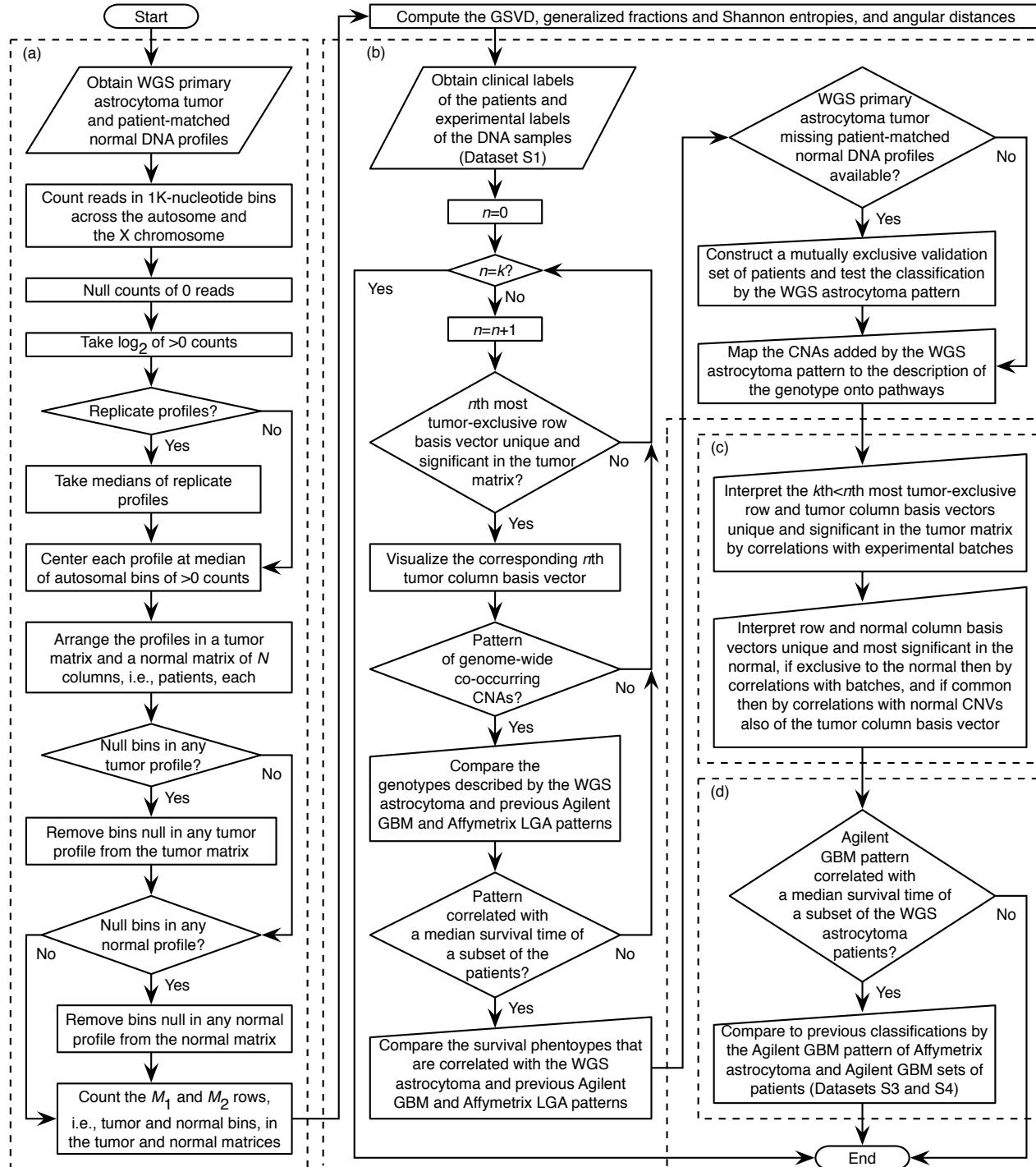


Fig. S1. The GSVD invariably identifies the same genotype and phenotype as significant in and exclusive to the WGS astrocytoma tumor relative to the patient-matched normal profiles, here like in the previous GSVDs of Agilent GBM and, separately, Affymetrix LGA tumor and normal profiles. (a) Construction of the WGS astrocytoma tumor and patient-matched normal datasets. (b) Identification of the WGS astrocytoma tumor-exclusive genotype and phenotype. (c) Blind separation from normal and experimental sources of copy-number variation. (d) Technology- and grade-independent prediction of astrocytoma survival.

Construction of the WGS astrocytoma tumor and patient-matched normal datasets. We obtained WGS binary alignment map (BAM) files of primary adult astrocytoma tumor and patient-matched normal DNA from a set of 52 GBM and 33 LGA patients from TCGA at the Genomic Data Commons (GDC) together with the clinical labels of the patients and experimental labels of the corresponding DNA samples^{4,5} (Dataset S1).

The total size of the TCGA raw level 1 BAM files is ≈ 23 terabytes or 0.02 petabytes. In each BAM file, we counted the number of Illumina HiSeq 2000- or Genome Analyzer II-measured sequence reads that map to each nonoverlapping 1K-nucleotide bin across the autosome and the X chromosome of the reference human genome hg19 by using copy-number estimation by a mixture of Poissons (CN.MOPS).^{10,11} Each profile lists the \log_2 of the positive read counts, centered at the median of the bins that map to the autosome and are with positive counts in the corresponding BAM file, across the tumor, and, separately, normal bins with positive counts in all tumor or normal BAM files, respectively.

We used the same computational workflow to construct the WGS astrocytoma set of patients as we previously used to construct the Agilent GBM and Affymetrix LGA discovery and validation sets^{6,7} (Fig. S1). The resulting tumor and normal datasets have the structure of two matrices of $N = 85$ matched columns, i.e., patients, and $M_1 = 2,827,037$ and $M_2 = 2,828,152$ rows, i.e., tumor and normal bins. Of the 85 patients, 24, i.e., $\approx 28\%$, complement the previous discovery sets of 251 GBM and 59 LGA patients. Of the 24 patients, 14 complement the previous validation sets of 184 GBM and 74 LGA patients. The ≈ 2.8 M bins, across the autosome and the X chromosome, include almost all of the 213K CGH and 934K SNP probes. In addition, the bins fill in gaps in the genome that are not covered by either set of probes. Note that bin sizes below 1K nucleotides would have increased the sparsity of the 2.5X- to 75X-coverage WGS profiles, whereas bin sizes above 1K nucleotides would have reduced the resolution of the WGS profiles relative to the previous Agilent and Affymetrix microarray profiles as well as the ≈ 25 K-nucleotide median size of a human protein-encoding gene.

We find that, like the GSVDs of the microarray profiles, the GSVD of the WGS profiles is robust to, i.e., quantitatively similar and qualitatively the same under, perturbations to the datasets, e.g., due to changes in the preprocessing of the BAM files, including changing the bin size in the range of 100–2.5K nucleotides. Note that the robustness of the GSVD, the unique and significant row and column basis vectors, and the interpretations of the vectors, implies robustness of the single coherent model that the GSVD creates from the WGS profiles.

Formulation of the GSVD as a comparative spectral decomposition. That the GSVD (Fig. 1) can simultaneously identify the similar and dissimilar between two column-matched but row-independent matrices, and, therefore, create a single coherent model from two datasets recording different aspects of interrelated phenomena,^{22,23} is possible because the GSVD is exact, exists, and has uniqueness properties that directly generalize those of the SVD.^{28,29}

Theorem S1 (Uniqueness properties of the GSVD). *The GSVD of two column-matched but row-independent real matrices $D_i \in \mathbb{R}^{M_i \times N}$, each with full column rank $N \leq M_i$, is unique up to phase factors of ± 1 of each triplet of corresponding column and row basis vectors, i.e., $u_{i,n}$ and v_n , except in degenerate subspaces defined by subsets of pairs of generalized singular values of equal ratios, i.e., $\sigma_{1,n}/\sigma_{2,n}$.*

First proof. Consider the GSVD as it is computed by using the eigenvalue decomposition of the balanced arithmetic mean of all pairwise quotients of $A_1 = D_1^T D_1$ and $A_2 = D_2^T D_2$, i.e., $S = \frac{1}{2}[A_1 A_2^{-1} + (A_1 A_2^{-1})^{-1}]$. We proved that the eigenvectors of S can be used to compute the normalized row basis vectors v_n , such that the eigenvalue decomposition gives $SV = V\Lambda$, where $\Lambda = \text{diag}(\lambda_n)$.^{24,25} We also proved that for normalized v_n , the eigenvalues satisfy $\lambda_n = \frac{1}{2}[(\sigma_{1,n}/\sigma_{2,n})^2 + (\sigma_{1,n}/\sigma_{2,n})^{-2}] \geq 1$, where $(\sigma_{1,n}/\sigma_{2,n})^2$ and $(\sigma_{1,n}/\sigma_{2,n})^{-2}$ are the eigenvalues of $A_1 A_2^{-1}$ and $(A_1 A_2^{-1})^{-1}$, respectively. The corresponding column basis vectors $u_{1,n}$ and $u_{2,n}$ and generalized singular values $\sigma_{1,n}$ and $\sigma_{2,n}$ can, therefore, be computed by normalizing the columns of $D_1 V^{-T}$ and $D_2 V^{-T}$, and arranged, together with v_n , in U_1 and U_2 , Σ_1 and Σ_2 , and V^T , respectively, in a decreasing order of $\sigma_{1,n}/\sigma_{2,n}$.

The uniqueness properties of the GSVD follow from the uniqueness properties of the eigenvalue decomposition of S . \square

Second proof. Consider the GSVD as it is computed by using the QR decomposition of the appended D_1 and D_2 , followed by the SVD of the block of the column-wise orthonormal Q that corresponds to D_1 , i.e., Q_1 ,

$$\begin{bmatrix} D_1 \\ D_2 \end{bmatrix} = QR = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} R = \begin{bmatrix} U_{Q_1} \Sigma_{Q_1} V_{Q_1}^T \\ Q_2 \end{bmatrix} R = \begin{bmatrix} U_{Q_1} \Sigma_{Q_1} \\ U_{Q_2} \Sigma_{Q_2} \end{bmatrix} V_{Q_1}^T R, \quad (S1)$$

where R is upper triangular.^{19,20} Since D_1 and D_2 are with full column rank, then Q_1 and Q_2 are also with full column rank, $V_{Q_1}^T$ is orthonormal, and Σ_{Q_1} is positive diagonal. It follows from Eq. (S1) that the diagonal $\Sigma_{Q_2} = (I - \Sigma_{Q_1}^2)^{\frac{1}{2}}$ is also positive, and that

$$\begin{aligned}
U_{Q_2} &= Q_2 V_{Q_1} (I - \Sigma_{Q_1}^2)^{-\frac{1}{2}} \text{ is column-wise orthonormal,} \\
I &= Q^T Q = Q_1^T Q_1 + Q_2^T Q_2 \\
&= V_{Q_1} \Sigma_{Q_1}^2 V_{Q_1}^T + Q_2^T Q_2, \\
\Sigma_{Q_2}^2 &= I - \Sigma_{Q_1}^2 = (Q_2 V_{Q_1})^T (Q_2 V_{Q_1}) > 0, \\
U_{Q_2}^T U_{Q_2} &= I \\
&= [Q_2 V_{Q_1} (I - \Sigma_{Q_1}^2)^{-\frac{1}{2}}]^T [Q_2 V_{Q_1} (I - \Sigma_{Q_1}^2)^{-\frac{1}{2}}]. \tag{S2}
\end{aligned}$$

That is, the SVD of Q_1 also defines an SVD of Q_2 , where the singular values are arranged in Σ_{Q_2} in an increasing order, because the singular values of Q_1 are arranged in Σ_{Q_1} in a decreasing order.

It follows from Eq. (S2) then that the SVD of Q_1 factorizes D_1 and D_2 into the GSVD,

$$\begin{aligned}
U_1 &= U_{Q_1}, \\
U_2 &= U_{Q_2}, \\
\Sigma_1 &= \Sigma_{Q_1} \{\text{diag}[(V_{Q_1}^T R)(V_{Q_1}^T R)^T]\}^{\frac{1}{2}}, \\
\Sigma_2 &= (I - \Sigma_{Q_1}^2)^{\frac{1}{2}} \{\text{diag}[(V_{Q_1}^T R)(V_{Q_1}^T R)^T]\}^{\frac{1}{2}}, \\
V^T &= \{\text{diag}[(V_{Q_1}^T R)(V_{Q_1}^T R)^T]\}^{-\frac{1}{2}} V_{Q_1}^T R, \tag{S3}
\end{aligned}$$

where U_1 and U_2 are column-wise orthonormal, Σ_1 and Σ_2 are positive diagonal, and V^T , identical in both factorizations, has normalized rows. The positive generalized singular values are arranged in $\Sigma_1 \Sigma_2^{-1} = \Sigma_{Q_1} (I - \Sigma_{Q_1}^2)^{-\frac{1}{2}}$ in a decreasing order.

The QR decomposition is unique and, from Eq. (S3), the uniqueness properties of the GSVD follow from the uniqueness properties of the SVDs of Q_1 and Q_2 . \square

We defined the significance of the row and corresponding column basis vector v_n and $u_{i,n}$ in the corresponding matrix D_i to be the generalized fraction, which is interpreted to represent the fraction of information captured by v_n and $u_{i,n}$ in the corresponding dataset (Fig. S2). The generalized fraction, of the Frobenius norm of the outer product $\sigma_{i,n} u_{i,n} \otimes v_n^T$ in the norm of D_i , is proportional to the corresponding generalized singular value $\sigma_{i,n}$,

$$p_{i,n} = \frac{\|\sigma_{i,n} u_{i,n} \otimes v_n^T\|^2}{\|D_i\|^2} = \frac{\sigma_{i,n}^2}{\sum_{n=1}^N \sigma_{i,n}^2} > 0. \tag{S4}$$

We defined the complexity of D_i , i.e., the generalized normalized Shannon entropy, which is interpreted to represent the distribution of information among the row and column basis vectors v_n^T and $u_{i,n}$, to be proportional to the arithmetic mean of $p_{i,n} \log p_{i,n}$,

$$0 < d_i = -(\log N)^{-1} \sum_{n=1}^N p_{i,n} \log p_{i,n} \leq 1. \tag{S5}$$

At its lower bound, an entropy of $d_i \rightarrow 0$ corresponds to an ordered and redundant dataset, where, as in Eq. (S6), all the information is captured by one row basis vector and the corresponding column basis vector, i.e., v_1^T and $u_{i,1}$,

$$p_{i,n} \rightarrow \begin{cases} 1, & n = 1, \\ 0, & n \neq 1. \end{cases} \tag{S6}$$

An entropy of $d_i = 1$ corresponds to a disordered and random dataset, in which all v_n^T and $u_{i,n}$ are of equal significance and capture equal fractions of the information, i.e., $p_{i,n} = 1/n$ for all n .

Here we find that the two most tumor-exclusive, i.e., the first and second row basis vectors, with the angular distances $\theta_1, \theta_2 > \pi/6$, are also the first and second most significant in the tumor dataset, with the generalized fractions $p_{1,1}, p_{1,2} \gtrsim 0.08$. The most normal-exclusive, i.e., the 85th row basis vector, with $\theta_{85} < -\pi/6$, is also the most significant in the normal dataset, with $p_{2,85} > 0.23$. The second most significant row basis vector in the normal dataset, i.e., the 82nd row basis vector, is also the fifth most significant in the tumor dataset, with $p_{1,82}, p_{2,82} > 0.02$. With $|\theta_{82}| < \pi/10$, the 82nd row basis vector is common to both the normal and tumor datasets.

Segmentation of the WGS astrocytoma pattern. To compare the WGS astrocytoma pattern to the Agilent GBM and Affymetrix LGA patterns, we mapped the hg18 genomic start and end coordinates of the 130 segments previously identified in the Agilent GBM pattern to the reference human genome hg19, and classified the 111 genomic segments of at least five Agilent probes in length as amplified, unaltered, or deleted in the WGS astrocytoma pattern (Dataset S2). We then compared the classifications to the previously computed classifications of the same segments in the Affymetrix LGA and Agilent GBM patterns.

To expand upon the description of the tumor-exclusive genotype by the microarray patterns, we segmented the WGS pattern by using CBS,³¹ and classified the segments as amplified, unaltered, or deleted in the WGS pattern (Fig. 2). We then mapped the segments in relation to gaps in the genome that are not covered by either the Agilent or the Affymetrix probes, where the DNA copy number was not measured by the microarrays.

A segment is classified as amplified or deleted if the difference between the relative copy-number mean of the segment and the autosome is greater than twice the standard deviation of the autosome, or if the difference between the mean of the segment and the chromosome it maps to is greater than the standard deviation of the chromosome, and is consistent with the difference between the segment and the autosome. The mean and standard deviation of the autosome are computed excluding the astrocytoma outlying chromosomes 7 and 10 and chromosome arm 9p.

Estimation of the consistency between the DNA CNAs and mRNA expression. We obtained TCGA level 3 Illumina HiSeq 2000 RNA sequencing profiles, which were available for the primary tumors of 62 of the 85 WGS astrocytoma patients. There are 29 genes highlighted in the tumor-exclusive genotype that corresponds to the one-year survival phenotype. We assessed the mRNA expression of each of these genes in the subset of patients that have high weights of the WGS astrocytoma pattern in their primary tumor DNA copy-number profiles, relative to the complementary subset of patients that have low weights, by using box-plots and computing the corresponding MWW *P*-values (Figs. S4–S7).

Visualization and interpretation of the significant row and column basis vectors. To visualize the first tumor and 85th normal column basis vectors, we segmented the vectors by using CBS (Figs. S8 and S9). To interpret the vectors, we computed the correlations of the first tumor and 85th normal column basis vectors with the vectors that list the \log_2 of the fractional GC content across the tumor and normal bins, respectively. The fractional GC content was computed for the tumor and normal bins by counting the numbers of the A, C, G, and T nucleotides in the nonoverlapping 1K-nucleotide sequences in the reference human genome hg19 that correspond to the bins. We also assessed the distribution of the relative copy numbers listed in each vector between bins of $\leq 50\%$ and $> 50\%$ GC content by using box-plots and computing the corresponding MWW *P*-values (Fig. S10).

To interpret the corresponding first and 85th row basis vectors, we assessed the subsets of patients that are of either high or low copy numbers in each vector for enrichment in any one of the experimental labels of the tumor and normal DNA samples, e.g., the TCGA GCCs or TSSs. The *P*-value of each enrichment was computed assuming a hypergeometric probability distribution of the K labels among the N patients, and of the $k \subseteq K$ labels among the n patients of either high or low copy numbers, i.e.,

$$P(k; n, N, K) = \binom{N}{n}^{-1} \sum_{i=k}^n \binom{K}{i} \binom{N-K}{n-i}. \quad (\text{S7})$$

In each row basis vector, we also assessed the distribution of the copy numbers between the subset of patients corresponding to each of these labels and the complementary subset by using boxplots and computing the corresponding MWW *P*-values (Fig. S11).

Similarly, to interpret the 82nd row basis vector, we assessed the subsets of patients that are of high or low

copy numbers in the vector for enrichment in gender, i.e., females or males. We also assessed the distribution of the copy numbers between the female and male patients (Fig. S12). To interpret the 82nd tumor and normal column basis vectors, we assessed the distribution of copy numbers between bins that map to the autosome and the X chromosome (Figs. S13–S15).

Classification of the WGS astrocytoma, Affymetrix astrocytoma, and Agilent GBM tumor profiles by correlation with the Agilent GBM pattern. Of the 212,696 probes of the Agilent Human Genome CGH 244A microarray platform that constitute the GBM pattern, 212,619 were mapped by Agilent onto the reference human genome hg19. To classify the WGS astrocytoma tumor profiles by correlation with the Agilent GBM pattern, we mapped each CGH probe to the WGS bin that contains the hg19 genomic start coordinate of the probe. When more than one probe mapped to one bin, the probe closest to the hg19 genomic start coordinate of the bin was selected, resulting in a one-to-one mapping of 211,096 probes of unique hg19 coordinates onto 211,096 nonoverlapping bins.

To compare the KM analyses and Cox models of the WGS astrocytoma patients (Figs. 4 and S3) to those of the Affymetrix astrocytoma, i.e., GBM and LGA, patients, we used the previously computed correlations of the Agilent GBM pattern with the minimally preprocessed TCGA raw level 2 tumor profiles of the discovery and validation sets of 59 and 74 LGA patients as well as 364 of the discovery and validation sets of 251 and 184 GBM patients, measured by the Affymetrix Genome-Wide Human SNP Array 6.0 microarray platform (Fig. S16, Table S2, and Dataset S3). To compare to the KM analyses and Cox models of the Agilent GBM patients, we used the previously computed correlations of the Agilent GBM pattern with the minimally preprocessed TCGA raw level 1 tumor profiles of 364 of the discovery and validation sets of 251 and 184 GBM patients, measured by the Agilent Human Genome CGH 244A microarray platform (Fig. S17, Table S3, and Dataset S4). We used the correlation cutoff of 0.15 as was previously established for the Agilent GBM discovery set of patients and validated for the Agilent GBM validation, and Affymetrix LGA discovery and validation sets of patients.

To estimate the *MGMT* promoter methylation status of a tumor, we used the TCGA raw level 1 of the Illumina Infinium Human Methylation 27 or 450 BeadChip-measured DNA methylation levels.⁵⁰ The *IDH1* mutation status of the LGA and GBM tumors is from TCGA.⁵¹

Fig. S2. The most significant row basis vectors uncovered by the GSVD of the WGS astrocytoma tumor and normal datasets. (a) The ten largest generalized fractions of Eq. (S4) in the WGS astrocytoma tumor dataset are depicted in a bar chart, showing that the two most tumor-exclusive row basis vectors, i.e., the first and second, are also the first and second most significant in the tumor dataset and capture $\approx 29\%$ and 8% of the information, respectively. The corresponding generalized normalized Shannon entropy of Eq. (S5) is 0.78. (b) The ten largest generalized fractions in the normal dataset are depicted in a bar chart, showing that the most normal-exclusive row basis vector, i.e., the 85th, is also the most significant in the normal dataset and captures $\approx 23\%$ of the information. The 82nd row basis vector, which is approximately common to both datasets, is the second and fifth most significant and captures $\approx 14\%$ and 2% of the information in the normal and tumor datasets, respectively.

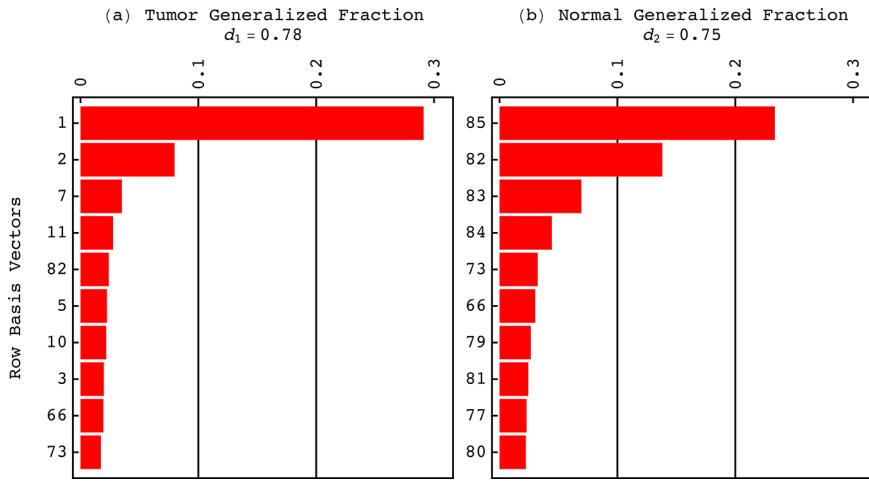
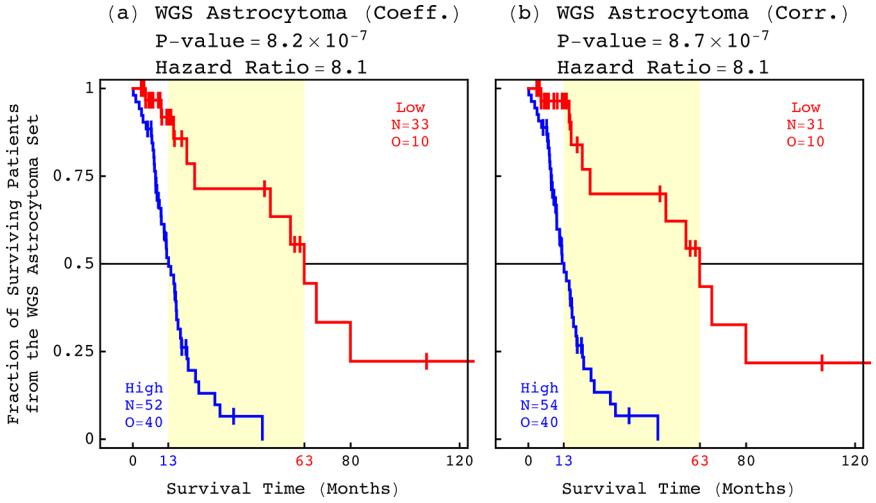


Fig. S3. Survival analyses of the WGS astrocytoma patients based upon the GSVD of the WGS datasets. (a) The classification of the 85 patients into low (red) or high (blue) superposition coefficients based upon the second most tumor-exclusive row basis vector is depicted in KM curves showing a 50-month median survival time difference (yellow). The corresponding log-rank test P -value is $<10^{-6}$. The univariate Cox proportional hazard ratio is ≈ 8 . (b) The classification of the 85 patients based upon the correlations of their tumor profiles with the second tumor column basis vector.



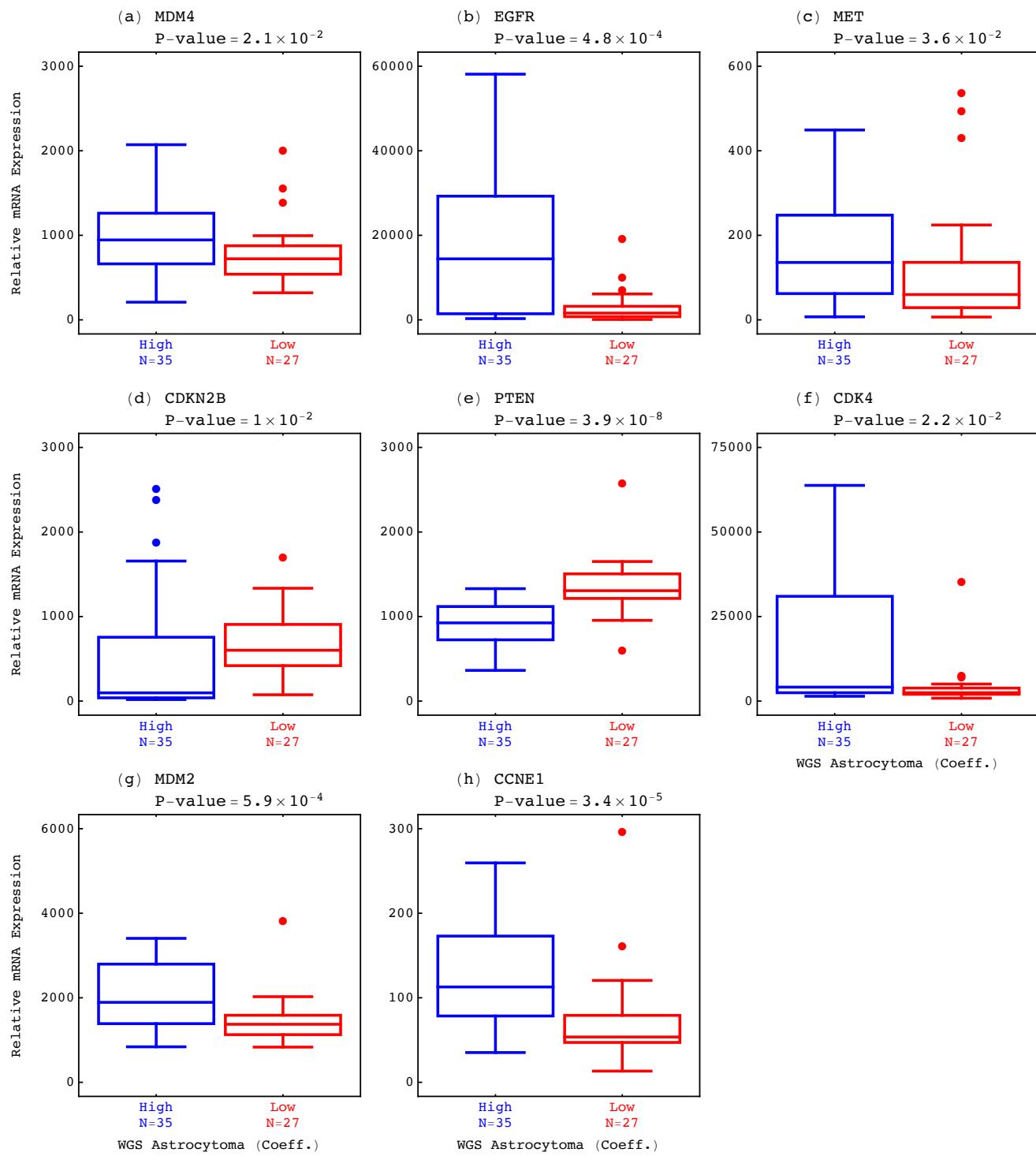


Fig. S4. Differential mRNA expression in the Ras pathway consistent with the DNA CNAs. The differential mRNA expression of genes highlighted in the Ras pathway in the subset of patients that have high weights of the WGS astrocytoma pattern in their primary tumor DNA copy-number profiles, i.e., the patients that have the approximately one-year survival phenotype, is depicted in boxplots with the corresponding MWU P-values. These genes consistently map to amplifications or deletions in the tumor-exclusive genotype (Fig. 3).

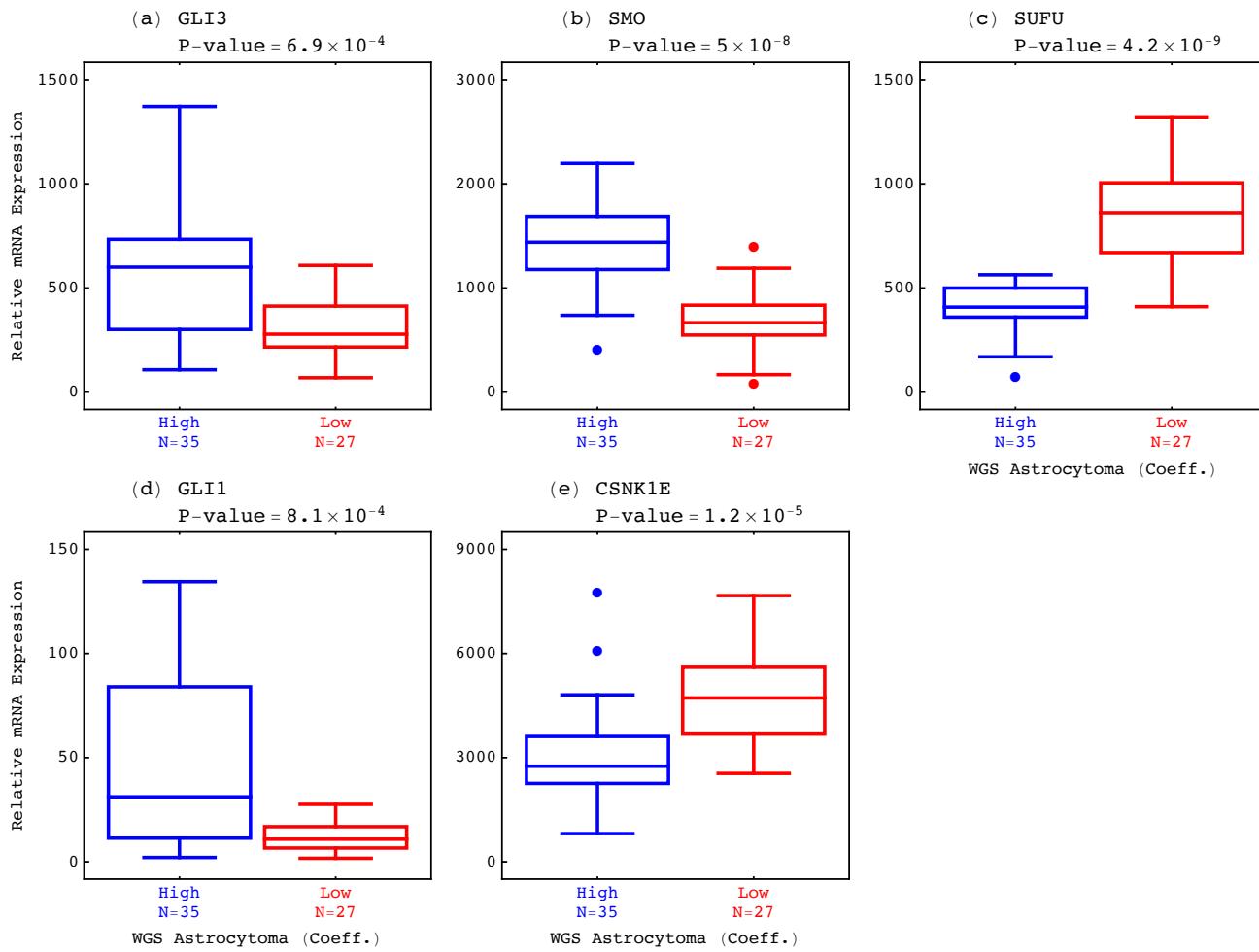


Fig. S5. Differential mRNA expression in the Shh pathway consistent with the DNA CNAs.

Fig. S6. Differential mRNA expression in the Notch pathway consistent with the DNA CNAs.

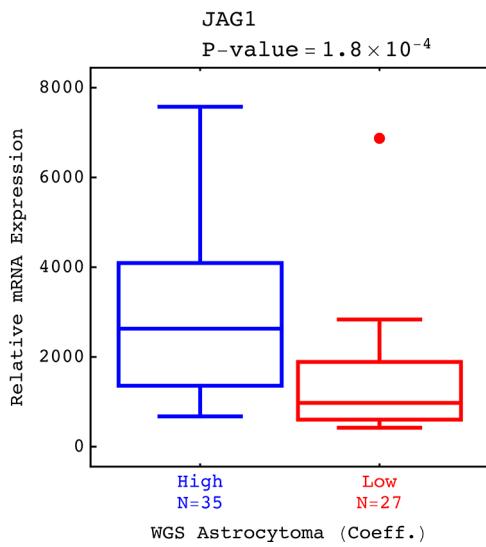


Fig. S7. Differential mRNA expression outside the Ras, Shh, and Notch pathways consistent with the DNA CNAs.

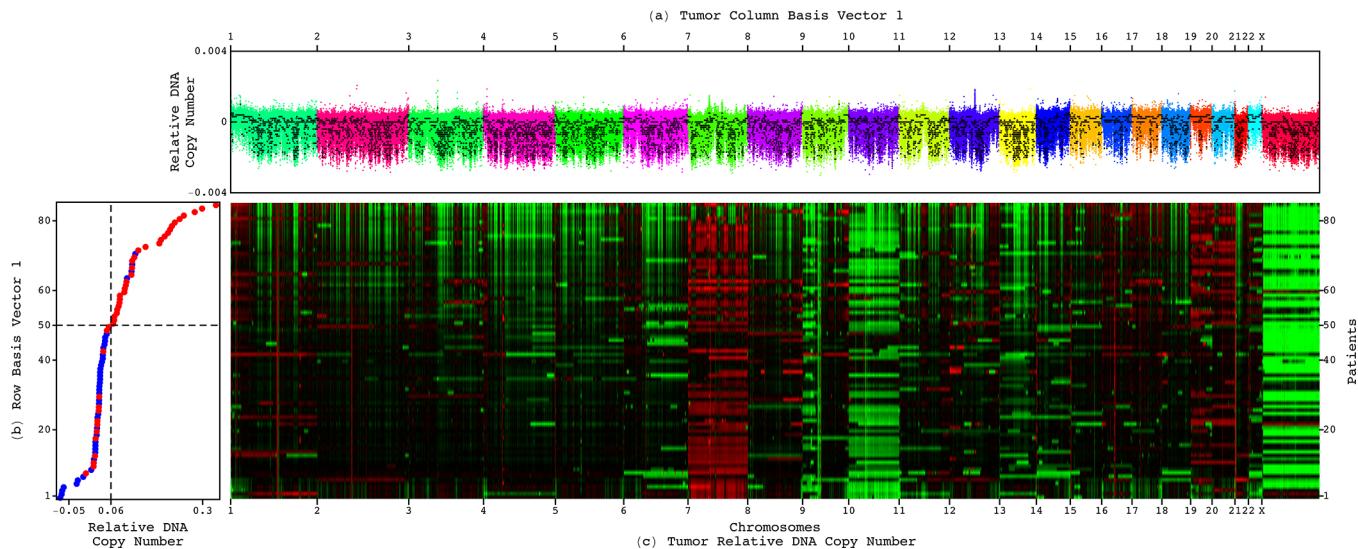
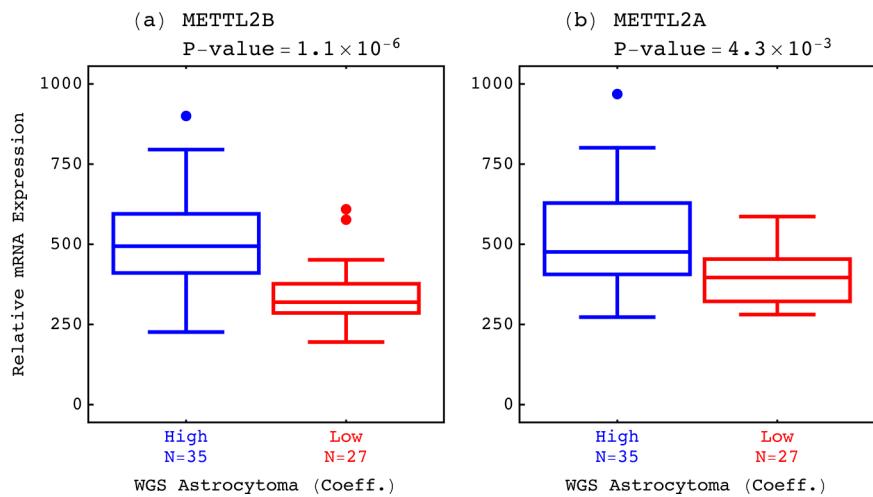


Fig. S8. The first, most tumor-exclusive row basis vector and corresponding tumor column basis vector. (a) The first tumor column basis vector is depicted in a plot of relative copy numbers, ordered and colored by their genomic coordinates and segmented by CBS (black lines), roughly describing frequent spikes of reduced copy numbers superimposed on an invariant baseline. The correlation of the vector with the fractional GC content across the tumor bins is 0.78. (b) The corresponding first, most tumor-exclusive row basis vector is depicted in a plot showing an enrichment of the BI (08) GCC (red) relative to the other centers (blue) among the 35 patients with high superposition coefficients of the first tumor column basis vector in their tumor profiles. The corresponding hypergeometric P -value of Eq. (S7) is $<10^{-7}$. (c) The WGS astrocytoma tumor dataset is depicted in a raster, with relative WGS read-count, i.e., DNA copy-number amplification (red), no change (black), and deletion (green), showing the GC content variation and its correlation with the experimental batches.

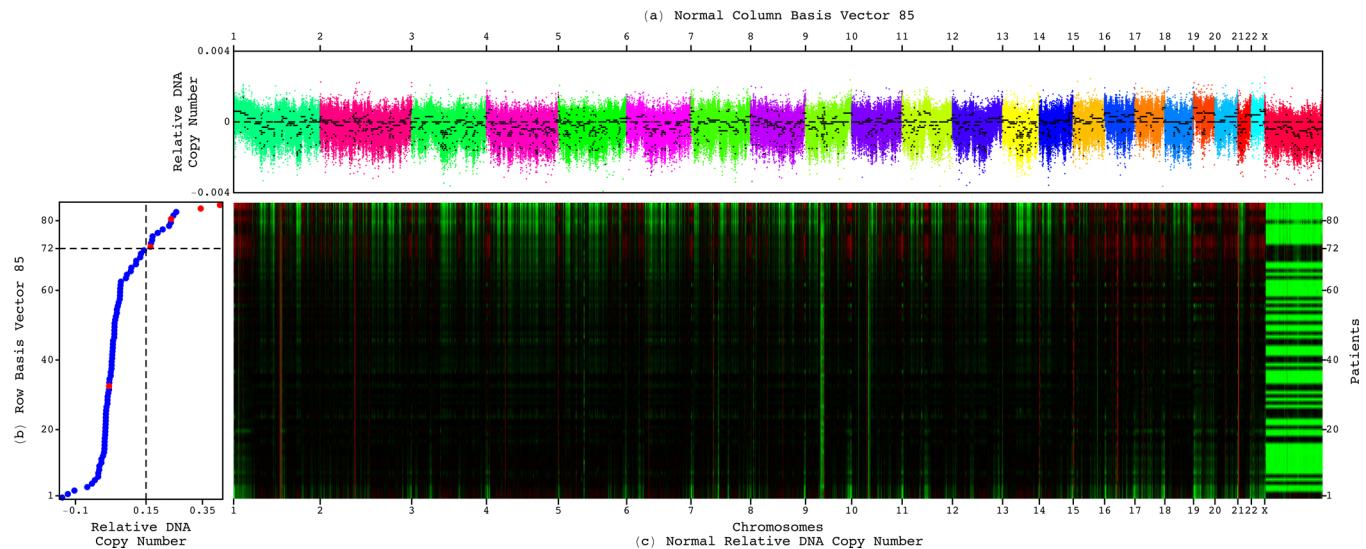


Fig. S9. The 85th, most normal-exclusive row basis vector and corresponding tumor column basis vector. (a) The 85th normal column basis vector is depicted in a plot of copy numbers, ordered and colored by their genomic coordinates and segmented by CBS (black lines), roughly describing frequent spikes of reduced copy numbers superimposed on an invariant baseline. The correlation of the vector with the fractional GC content across the normal bins is 0.91. (b) The corresponding 85th, most normal-exclusive row basis vector is depicted in a plot showing an enrichment of the TJU (CS) TSS (red) relative to the other sites (blue) among the 13 patients with high superposition coefficients of the 85th column basis vector in their normal profiles. The corresponding hypergeometric P -value is $<10^{-2}$. (c) The WGS normal dataset is depicted in a raster showing the GC content variation and its correlation with the experimental batches.

Fig. S10. The first tumor and 85th normal column basis vectors are correlated with the fractional GC content across the tumor and normal genomes. The distributions of the copy numbers listed in the (a) first tumor and (b) 85th normal column basis vectors between tumor and normal bins, respectively, of $>50\%$ and $\leq 50\%$ GC content are depicted in boxplots with the corresponding MWU P -values.

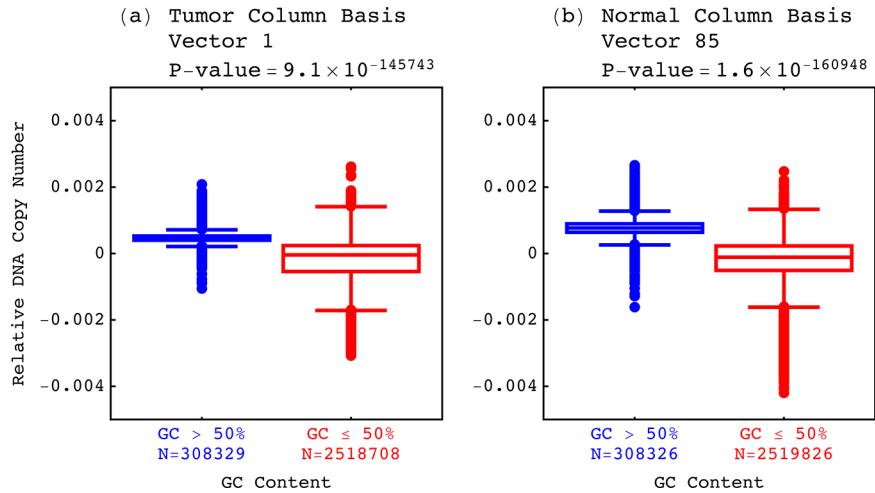


Fig. S11. The first and 85th row basis vectors are correlated with experimental batches. The distributions of the copy numbers listed in the (a) first and (b) 85th row basis vectors between GCCs and TSSs, respectively, are depicted in boxplots with the corresponding MWW *P*-values.

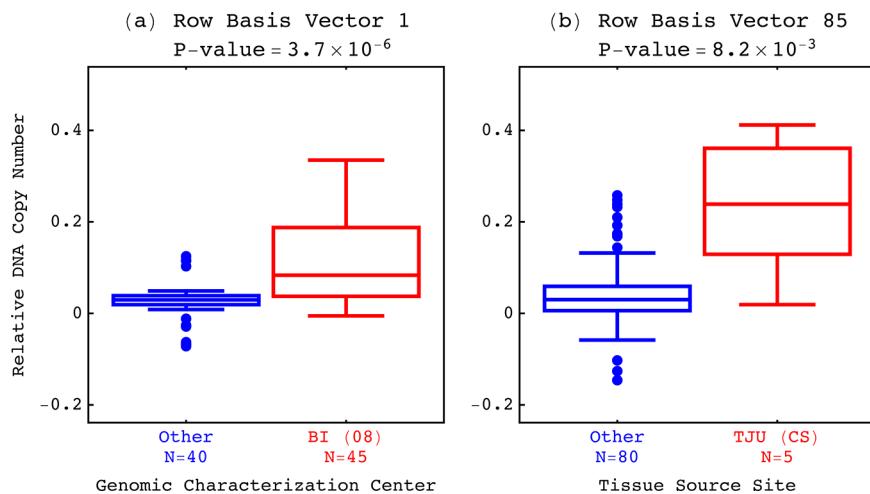
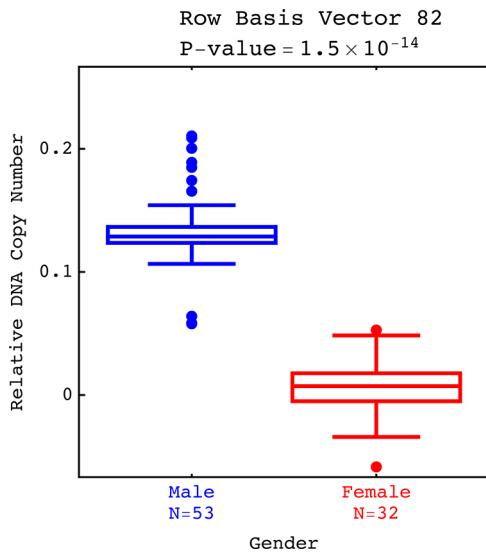


Fig. S12. The 82th row basis vector is correlated with gender. The distribution of the copy numbers listed in the 82nd row basis vectors between females and males is depicted in a boxplot with the corresponding MWW *P*-value.



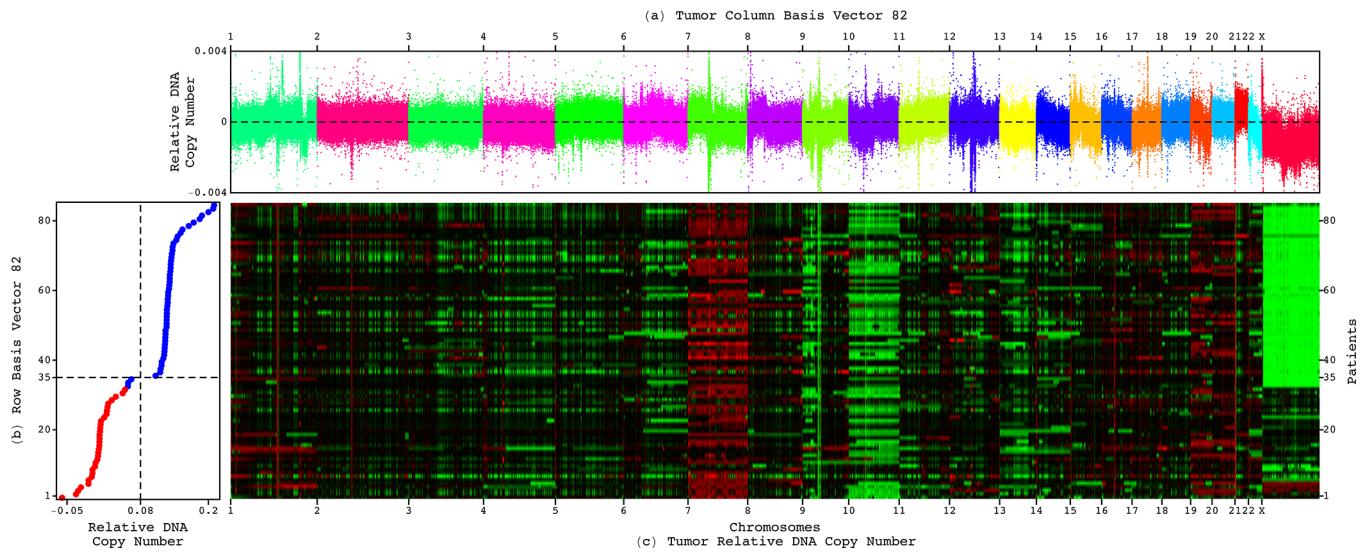


Fig. S13. The 82nd row basis vector and corresponding tumor column basis vector. (a) The 82nd tumor column basis vector is depicted in a plot of copy numbers describing a deletion of the X chromosome relative to the autosome across the tumor bins. (b) The corresponding 82nd row basis vector is depicted in a plot showing an enrichment of the males (blue) relative to the females (red) among the 50 patients with high superposition coefficients of the 82nd tumor column basis vector in their tumor profiles. The corresponding hypergeometric P -value is $<10^{-19}$. (c) The WGS astrocytoma tumor dataset is depicted in a raster showing the normal male-specific X chromosome deletion conserved in the tumors.

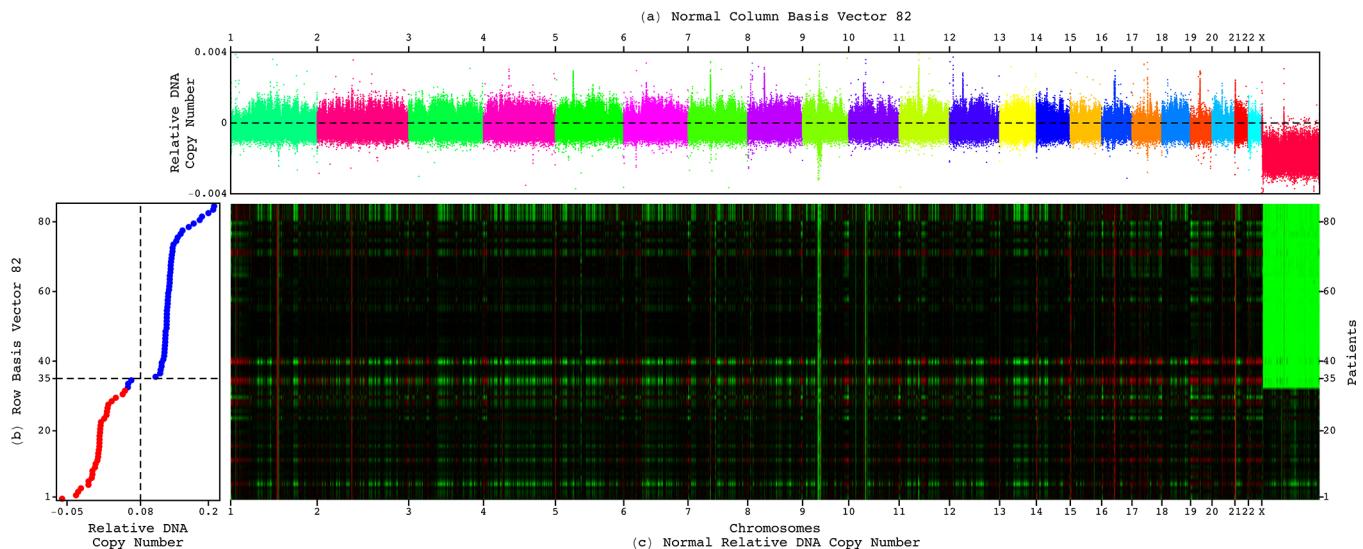


Fig. S14. The 82nd row basis vector and corresponding normal column basis vector. (a) The 82nd normal column basis vector is depicted in a plot of copy numbers describing a deletion of the X chromosome across the normal bins. (b) The corresponding 82nd row basis vector is depicted in a plot. (c) The WGS normal dataset is depicted in a raster showing the normal male-specific X chromosome deletion.

Fig. S15. The 82nd tumor and normal column basis vectors are correlated with a deletion of the X chromosome relative to the autosome across the tumor and normal genomes. The distributions of the copy numbers listed in the 82nd (a) tumor and (b) normal column basis vectors between tumor and normal bins, respectively, which map to the autosome and the X chromosome, are depicted in boxplots with the corresponding MWU P -values.

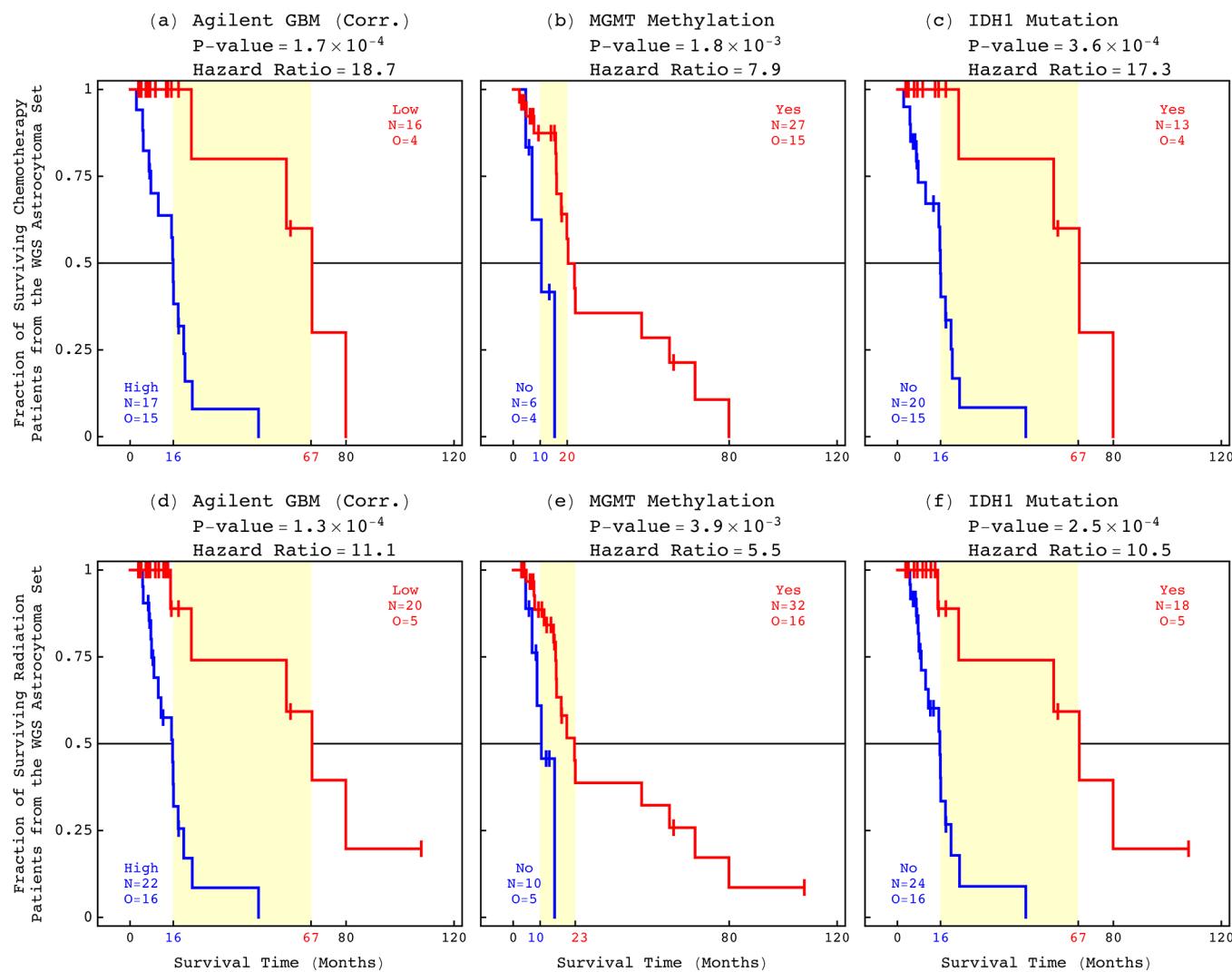
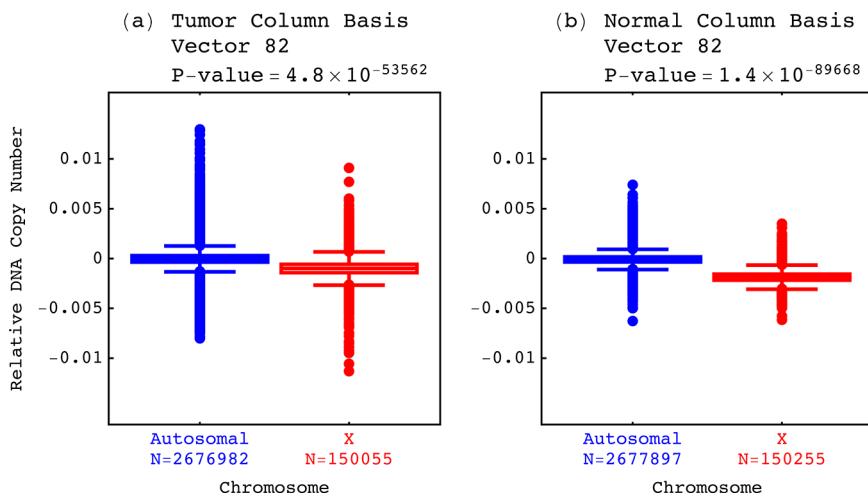


Fig. S16. Survival analyses of the chemotherapy- and radiation-treated WGS astrocytoma patients with *MGMT* promoter methylation and *IDH1* mutation test results. The classifications of the chemotherapy-treated patients based upon (a) the Agilent GBM pattern, (b) *MGMT* promoter methylation, or (c) *IDH1* mutation, and the classifications of the radiation-treated patients (d)–(f), are depicted in KM curves highlighting median survival time differences (yellow) with the corresponding log-rank P -values and Cox hazard ratios.

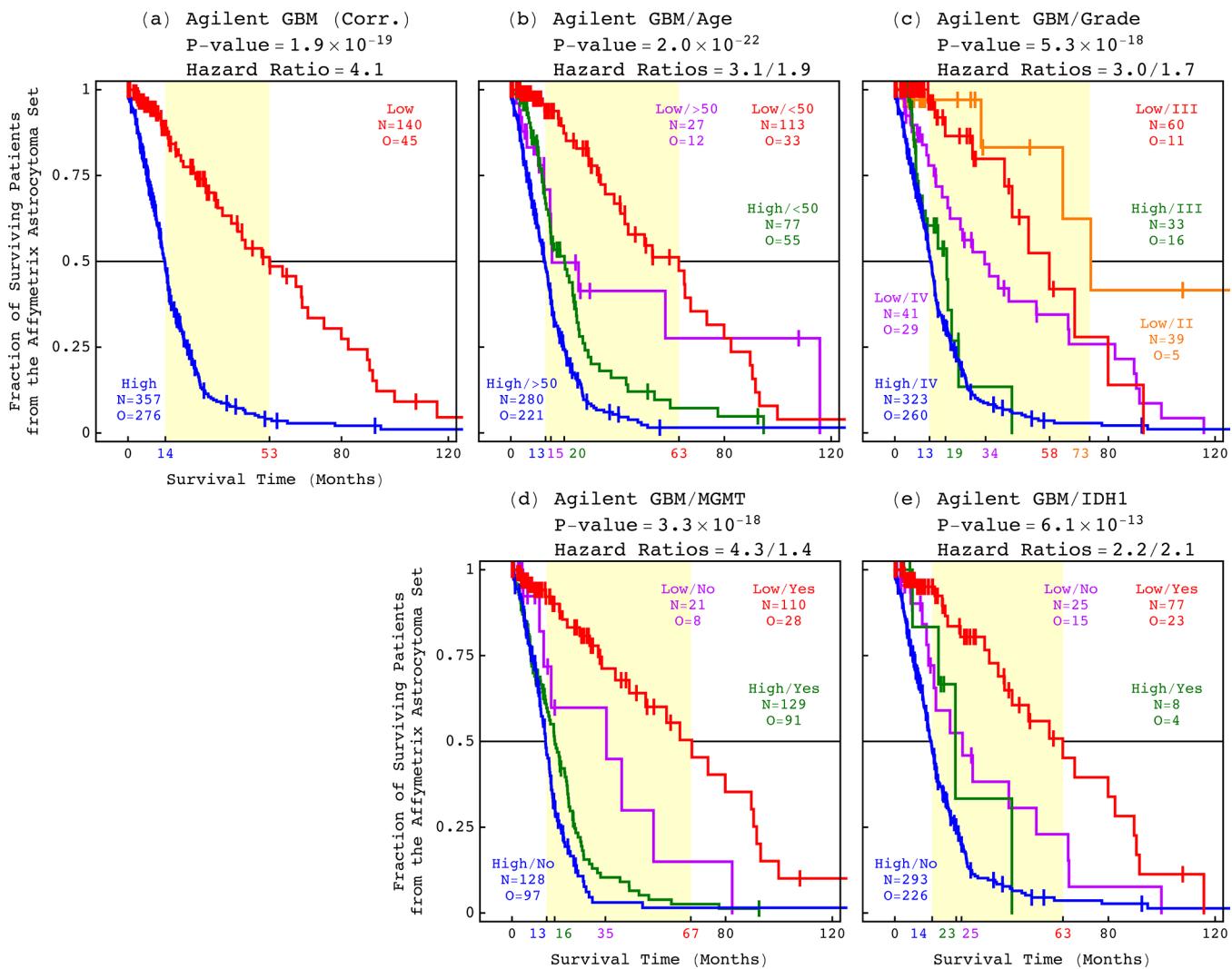


Fig. S17. Survival analyses of the Affymetrix astrocytoma patients. The classifications of the 497 patients based upon (a) the Agilent GBM pattern and, in addition, (b) age, (c) grade, (d) *MGMT* promoter methylation, or (e) *IDH1* mutation are depicted in KM curves highlighting median survival time differences (yellow) with the corresponding log-rank P-values and Cox hazard ratios.

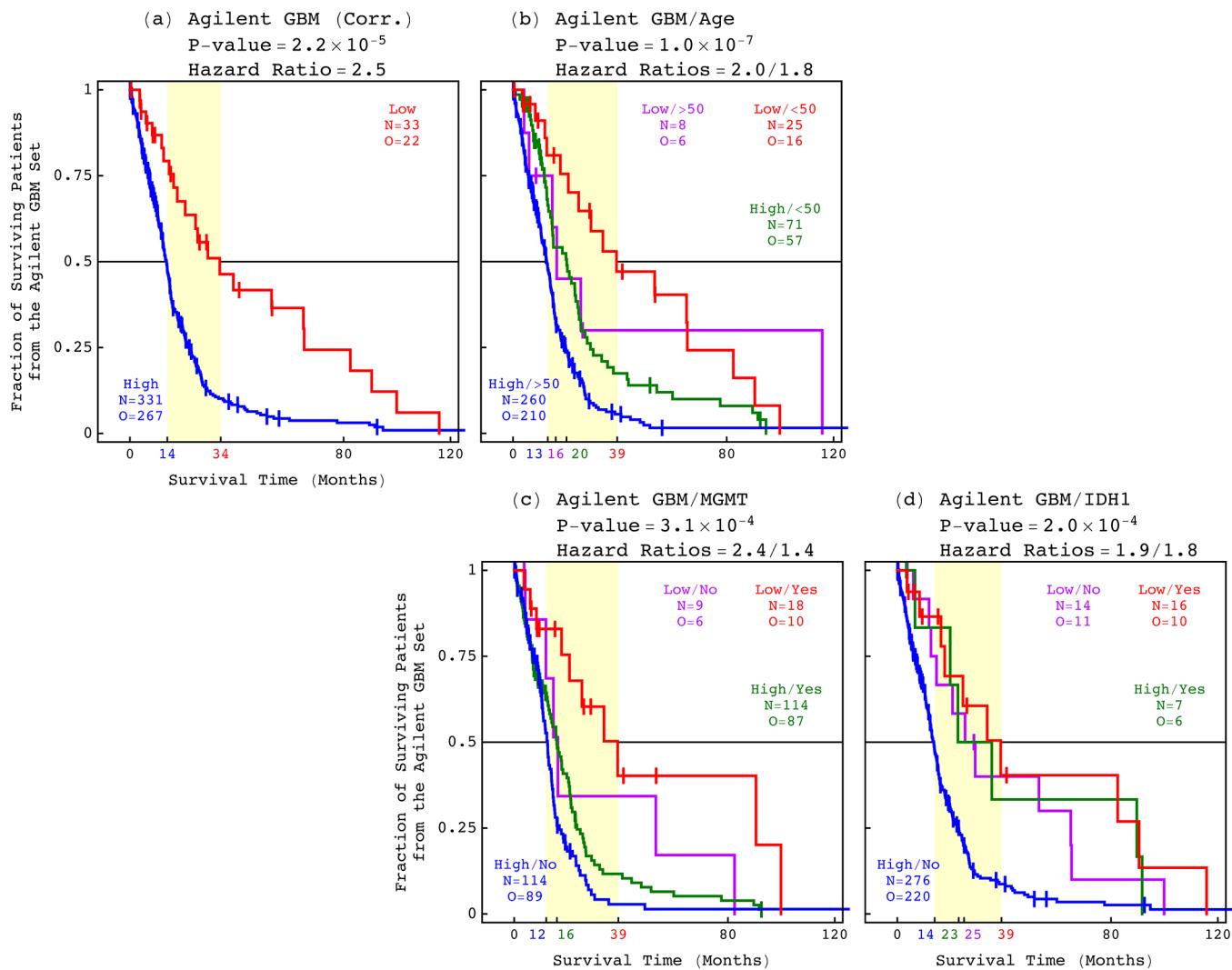


Fig. S18. Survival analyses of the Agilent GBM patients. The classifications of the 364 patients based upon (a) the Agilent GBM pattern and, in addition, (b) age, (c) *MGMT* promoter methylation, or (d) *IDH1* mutation are depicted in KM curves highlighting median survival time differences (yellow) with the corresponding log-rank P-values and Cox hazard ratios.

Cox Proportional Hazards Model	Number of Patients	Predictor	Hazard Ratio	95% Confidence Interval	P-value	Concordance Index
Univariate	85	Agilent GBM (Corr.)	5.3	2.4– 11.9	4.3×10^{-5}	0.95
		WGS Astrocytoma (Coeff.)	8.1	3.1– 21.3	2.0×10^{-5}	0.87
		WGS Astrocytoma (Corr.)	8.1	3.1– 21.3	2.1×10^{-5}	0.89
		Age	4.2	1.9– 9.2	3.6×10^{-4}	0.83
		Grade	3.5	1.8– 6.8	2.1×10^{-4}	0.75
	57	Agilent GBM (Corr.)	14.9	3.4– 64.3	3.1×10^{-4}	0.96
		MGMT Methylation	3.6	1.6– 8.0	2.2×10^{-3}	0.79
	75	Agilent GBM (Corr.)	10.8	3.2– 36.1	1.2×10^{-4}	0.95
		IDH1 Mutation	8.9	3.0– 26.1	6.7×10^{-5}	0.95
Bivariate	85	Agilent GBM (Corr.)	3.9	1.6– 9.4	2.3×10^{-3}	0.86
		Age	2.7	1.1– 6.4	3.1×10^{-2}	
		Agilent GBM (Corr.)	3.1	1.3– 7.6	1.3×10^{-2}	0.80
		Grade	2.3	1.1– 4.8	2.8×10^{-2}	

Table S1. Cox proportional hazards models of the WGS astrocytoma patients. The classifications of the 85 patients based upon the Agilent GBM pattern, age, and grade, of the subset of 57 patients with *MGMT* promoter methylation test results based upon the Agilent GBM pattern or the *MGMT* promoter methylation status, and of the subset of 75 patients with *IDH1* mutation test results based upon the Agilent GBM pattern or the *IDH1* mutation status.

Cox Proportional Hazards Model	Number of Patients	Predictor	Hazard Ratio	95% Confidence Interval	P-value	Concordance Index
Univariate	497	Agilent GBM (Corr.)	4.1	3.0– 5.8	2.9×10^{-17}	0.85
		Age	2.8	2.2– 3.6	1.4×10^{-15}	0.80
		Grade	2.8	2.0– 3.7	1.4×10^{-10}	0.82
	388	Agilent GBM (Corr.)	4.8	3.3– 7.0	3.5×10^{-16}	0.86
		MGMT Methylation	2.1	1.6– 2.7	8.6×10^{-8}	0.66
	403	Agilent GBM (Corr.)	3.6	2.5– 5.1	1.9×10^{-12}	0.84
		IDH1 Mutation	3.9	2.6– 5.8	5.2×10^{-11}	0.87
	497	Agilent GBM (Corr.)	3.1	2.2– 4.4	5.7×10^{-10}	0.79
		Age	1.9	1.4– 2.5	7.5×10^{-6}	
		Agilent GBM (Corr.)	3.0	2.1– 4.3	4.2×10^{-9}	0.80
		Grade	1.7	1.2– 2.4	1.5×10^{-3}	
Bivariate	388	Agilent GBM (Corr.)	4.3	2.9– 6.3	2.4×10^{-13}	0.73
		MGMT Methylation	1.4	1.1– 2.0	1.0×10^{-2}	
	403	Agilent GBM (Corr.)	2.2	1.3– 3.5	2.0×10^{-3}	0.83
		IDH1 Mutation	2.1	1.2– 3.7	8.0×10^{-3}	

Table S2. Cox proportional hazards models of the Affymetrix astrocytoma patients.

Cox Proportional Hazards Model	Number of Patients	Predictor	Hazard Ratio	95% Confidence Interval	P-value	Concordance Index
Univariate	364	Agilent GBM (Corr.)	2.5	1.6– 4.0	4.1×10^{-5}	0.76
		Age	2.1	1.6– 2.7	3.4×10^{-7}	0.72
	255	Agilent GBM (Corr.)	2.5	1.5– 4.2	5.0×10^{-4}	0.73
		<i>MGMT</i> Methylation	1.5	1.1– 2.0	5.0×10^{-3}	0.58
	313	Agilent GBM (Corr.)	2.4	1.5– 3.8	1.7×10^{-4}	0.77
		<i>IDH1</i> Mutation	2.4	1.5– 4.1	7.3×10^{-4}	0.77
	Bivariate	Agilent GBM (Corr.)	2.0	1.3– 3.2	3.2×10^{-3}	0.71
		Age	1.8	1.3– 2.4	6.5×10^{-5}	
		Agilent GBM (Corr.)	2.4	1.4– 4.0	1.1×10^{-3}	0.61
		<i>MGMT</i> Methylation	1.4	1.1– 1.9	1.7×10^{-2}	
	313	Agilent GBM (Corr.)	1.9	1.2– 3.1	1.1×10^{-2}	0.76
		<i>IDH1</i> Mutation	1.8	1.0– 3.1	4.8×10^{-2}	

Table S3. Cox proportional hazards models of the Agilent GBM patients.